Wao Terero Lexical Suffixes

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By Noah Diewald BA, MA

Graduate Program in Linguistics

The Ohio State University

2025

Dissertation Committee:

Professor Carl Jesse Pollard, Co-advisor

Professor Andrea Dorothy Sims, Co-advisor

Professor Robert Daigon Levine

Professor Simeon Floyd

Abstract

The topic of this thesis is the Wao Terero lexical suffix system within the context of morphological theory. Cross-linguistically, lexical affixes are closed-class, bound elements with lexical meanings. A lexical meaning is a concrete meaning, like ‘boat’ or ‘hand’, rather than a more abstract or functional meaning, like plurality or causativity. Unlike lexical elements that occur in incorporation or compounding, lexical suffixes are not derivable from open class items, such as nouns.

In addition to a description of the lexical suffix system and my field methods, the thesis provides a formal framework and model of the form-meaning relationships of the Wao Terero system. An important aspect of my model concerns the functional-lexical divide. In the literature the functional and lexical are often considered distinct categories with additional properties, beyond a particular meaning type. For instance, functional items are closed class, while lexical items are open class. Notably, lexical suffixes explicitly defy this divide by being both closed class and concrete. This is problematic in current theories. My model unifies the treatment of lexical and functional morphological content in a word-based realizational system. Other theories represent realization as a functional relation between morphosyntactic feature structures and morphological forms. Concrete meanings are poorly represented as feature structures so they explicitly fall outside the realizational system. I model the realizational calculation as a formal proof that relates word-forms directly to semantic expressions and lexical entries, in a manner that is neutral with respect to meaning type.

The framework is formalized in a dependent type theory, the calculus of inductive constructions, which allows for a computational implementation that can be verified for correctness. The morphological system interfaces with a categorial grammar, Linear Categorial Grammar, which previously lacked a morphological component.

Vita

Master of Arts in Linguistics, Ohio State University, 2020

Bachelor of Arts in Linguistics, University of Wisconsin – Madison, 2007

2021 Jacobs Research Funds Individual Grant

2021 Lewis and Clark Fund for Exploration and Field Research, American Philosophical Society

2021 Summer Foreign Language and Area Studies Fellowship for Lowland Ecuadorian Kichwa, University of Wisconsin – Madison

2020–2021 Dissertation Foreign Language and Area Studies Fellowship for Wao Tededo, Ohio State University

2020 International Research and Scholarship Grant, Office of International Affairs, Ohio State University

2020 Summer Foreign Language and Area Studies Fellowship for Wao Tededo, Ohio State University

2019–2020 Academic Year Foreign Language and Area Studies Fellowship for Ecuadorian Kichwa, Ohio State University

2019 Summer Foreign Language and Area Studies Fellowship for Wao Tededo, Ohio State University

2018 Summer Foreign Language and Area Studies Fellowship for Lowland Ecuadorian Kichwa, Florida International University

2014–2015 University Fellowship, Ohio State University

2015–2019, 2021–2023 Graduate Teaching Associate, The Ohio State University Department of Linguistics,

Publications

"A Word-and-Paradigm Workflow for Fieldwork Annotation" (2022) Maria Copot, Sara Court, Noah Diewald, Stephanie Antetomaso and Micha Elsner, Proceedings of the 5th Workshop on Computational Methods for Endangered Languages (ComputEL-5), Dublin, Ireland

"Modeling morphological learning, typology, and change: What can the neural sequence-to-sequence framework contribute?" (2019) with Micha Elsner, Andrea Sims, et al. *Journal of Language Modelling* 7.1 (2019): 53–98.

Field of Study

Major Field: Linguistics

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The topic of this thesis is the Wao Terero[1] (Saint and K. L. Pike, 1962; Peeke, 1968; Peeke, 1979; Fiddler, 2011; Dickinson et al., 2013; Fawcett, 2012; Fawcett, 2018; Fawcett, 2023)[2] lexical suffix system within the context of morphological theory. Cross-linguistically, lexical affixes are closed-class, bound elements associated with lexical meanings (Sapir, 1911; Kinkade, 1963). A lexical meaning may be a concrete meaning, like 'boat' or 'hand', rather than a more abstract or functional meaning, like plurality or causativity. Unlike lexical elements that occur in incorporation or compounding, lexical suffixes are not derivable from open class items, such as nouns.

1.1 Basic Attributes of the Wao Terero Lexical Suffixes

In the Wao Terero system, lexical suffixes contribute to noun-noun compound-like meanings in nominal constructions, as can be seen in (1).[3] There one sees a free nominal *daa*, 'thorn'. The addition of -*ka*, which may be associated with fruit, heads, stones, and other meanings, results in 'thorny fruit'. With the addition of -*wẽ*, which has meaning associations with plants, poles and trees, among other meanings, the result is 'thorny tree'.

[1]In line with previous work by Fawcett (2018) and Fawcett (2023) and others, I use the spelling "Wao Terero" when referring to the language in English or Spanish. Some Wao consultants are firmly against the use of 'r' in their orthography.

[2]The ISO 639-3 code (based on Ethnologue (Simons and Fennig, 2017)) is often provided to identify languages. The ISO code for Wao Terero references a slur. This is not uncommon in language coding systems. Out of respect for speaker communities, I do not use ISO 639-3, Glottocode or other systems.

[3]See Appendix E for information on glossing conventions.

(1) a. *daa*

thorn

‘thorn’

b. *daa-ka*

thorn-LS.stone

‘thorny fruit’

c. *daa-wẽ*

thorn-LS.plant

‘thorny tree’

Compound meanings may exist for other parts of speech, but commonly lexical suffixes are used as classifier-like elements with non-nominals. A verbal usage can be seen in (2). There one sees an example of *doubling*, which demonstrates that the verbal classifier does not *saturate* the argument position of the verb, and allows an overt argument *a · wẽ*. This is considered an important diagnostic of classifier behavior (Sara T. Rosen, 1998).

(2) *õ · be* *a · wẽ* *pa-wẽ-ta-bo-pa*

∅ · LS.territory ∅ · LS.plant cut-CLF.plant-PST-1-DECL

‘In the garden, I cut a pole.’

Example (2) may look like agreement, at first glance. It isn’t uncommon that such alignments take place, but concord is semantic. Felicitous classifier-like usage is determined by salient characteristics of the discourse context, and is not grammatically obligatory. The adjectival examples in (3) are equally felicitous, given the context of the image.

(3) 

a. *giitã-ka* *eibe* *ĩ-pa*

small-CLF.stone above COP-DECL

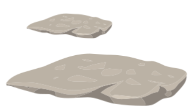
'The small one is above.'

b. *giitã* *eibe* *ĩ-pa*

small above COP-DECL

'The small one is above.'

The *-ka* affix is appropriate for rocks, but it also has a 'round' interpretation in classifier-like usage. A flat stone may still be referenced with *-ka*, but it may also be referenced by a classifier specific to the flat shape, such as *-pa*. This can be seen in (4).

(4) 

a. *giitã-pa* *eibe* *ĩ-pa*

small-CLF.board above COP-DECL

'The small one is above.'

b. *giitã-ka* *eibe* *ĩ-pa*

small-CLF.stone above COP-DECL

'The small one is above.'

c. *giitã* *eibe* *ĩ-pa*

small above COP-DECL

'The small one is above.'

For more information on the typological significance of the examples above, and a more detailed explanation of terms such as lexical suffix and classifier, see Chapter 2.

1.2 Thesis Goals

In addition to a description of the lexical suffix system and the field methods that support my findings, the primary contribution of this thesis is a formal theoretical framework that models the form-meaning relationships of the Wao Terero lexical suffix system. An important aspect of this concerns the notion of the functional-lexical divide. In the literature there is a distinction between functional (similar to *grammatical*) and lexical that goes beyond an intuitive description of how concrete a meaning is. The functional and lexical are often considered distinct categories. According to this perspective, functional and lexical items have additional properties beyond a particular meaning type. For instance, functional items are not only more abstract but are also closed class, while lexical items are open class in addition to being more concrete. Notably, lexical suffixes explicitly contradict this divide by being both closed class and having concrete meanings. This is problematic in current morphological theories for reasons that will be elaborated in this introduction, and is the central topic of §4.1. An important aspect of my model is that it unifies the treatment of so-called lexical and functional morphological content in a realizational (Zwicky, 1985b; Anderson, 1992; Halle and Marantz, 1993; Aronoff, 1994; Stump, 2001), Word and Paradigm (WP) (Robins, 1959; Singh and Starosta, 2003; Blevins, 2006) system.

In this thesis I address a broad audience which includes theoretical morphologists, formal theorists (in the sense of those who use logic or set theory in their models), Amazonianists, and semanticists who are interested in the morphology-semantics interface. For that reason, it is worthwhile to provide a preliminary explanation of realizationalism and WP early in the introduction. Even morphologists who are familiar with these terms will find this useful because my take on the two concepts may be unfamiliar.

1.3 Realizationalism

There are a number of definitions of realizationalism. The two most popular realizational theories, Paradigm Function Morphology (PFM) (Stump, 2001) and Distributed Morphology (DM) (Halle and Marantz, 1993), define realization as a relationship between morphological forms and morphosyntactic features, such as {NUMBER : plural} or [+plural]. Such features are claimed to exist at syntactic nodes. I generalize realization to include a broader notion of *context*, which can include syntax, compositional semantics, discourse and other information. Realization is then a relationship between morphological forms and such contexts.[4] The shared insight between the less general conception of realization and my own is that the interpretation of a morphological form depends on extrinsic information. Although a morphological form, such as the word-form *run* or the affix *-s* in English may be associated with a conventional range of meanings that can be thought of as intrinsic, their particular interpretation in normal usage is dependent on context. As an example of the difference, consider that when an English speaker is asked what *run* means in isolation, they will likely have a response. This will be informed by a range of conventional uses that the speaker is familiar with. The familiarity with convention and previously encountered patterns is what is often referred to as the "intrinsic" meaning of a morphological form. Intrinsic knowledge mediates but does not determine the interpretation of *run* in context. In normal speech it will be context that determines whether the understood meaning should be 'operate', 'manage', 'move rapidly', or something else. In the sentence *I run a company.* the 'manage' interpretation is fixed and not usurped by the knowledge that *run* can also mean 'move rapidly'. Neither is there any ambiguity. Likewise, the *-s* suffix will be interpreted differently in the following sentences: *He runs*, *His runs are rapid*, and *His run's time was slow.* In these sentences the syntax alone goes a long way in providing the necessary context for interpretation.

1.3.1 Relevance of Realizationalism to Lexical-Functional Dichotomy

One reason realization is relevant to the lexical-functional dichotomy is that despite a speaker's knowledge of conventional patterns, the morphological form is just a form. The interpretation is an external factor. Since lexical and functional meanings of a word-form or affix are matters of the interpretation of a form in context, a word-form or affix cannot be said to be intrinsically of only one meaning category or another, just as *-s* cannot be said to be only one of either plural or third-person singular. It is the case that some forms may simply happen to only associate with

[4] This does not imply the entirety of one's encyclopedic knowledge of the world is needed to construct a context. Lexical entries, and dynamic and lexical semantics are sufficient to model these contexts.

lexical or functional meanings, but this is not enforced by some synchronic mechanism. So-called grammaticality clines are not uncommon in languages (Muysken, 2008). An example from Wao Terero is the lexical suffix *-koo* that may be associated with woven things and clothing (5b) but also serves as a kind of plural for clusters or collections of things (5a). This may be diachronically motivated by weaving involving a binding together of many similar things, but firm evidence is currently lacking. It should be noted that Wao Terero has no generic plural, a point I explain in greater detail in §3.2.2.

(5) a. *a · wẽ-de* *deye-koo* *bõ-pa*
∅ · LS.plant-LOC spider.monkey-CLF.group sleep-DECL
'Within the tree, spider monkeys sleep.'

b. *awẽde okiye wepẽĩ we · koo ẽya-pa*
leader woman red ∅ · LS.group wear-DECL
'The boss lady is dressed in red.'

This is not to say that previous realizational theories have solved the issue of the lexical-functional dichotomy. They have likely done more to deepen the categorical divide, and complicate matters. Popular realizational theories do not treat lexical meanings realizationally. Despite my presentation of the form-context relationship of *run* as analogous to that of *-s*, in the DM and PFM models, these are considered two distinct and unrelated phenomena. For instance, in PFM, RUN is treated as an abstract word called a lexeme. A lexeme is fixed to a specific meaning and stem-form. To the extent that the lexeme's form and meaning may vary, the mechanisms for modeling such variation are distinct from those used for affixes with functional meanings.

Given this stance in popular realizational theory, one of the first steps in unifying treatments of lexical and functional content in a realizational system involves extending realization to lexical content. Even if one does not have a theoretical motivation, such as parsimony, elegance, or uniformity of approach, extending realization to lexical meaning contexts remains necessary for Wao Terero. Lexical suffixes in the language exhibit many aspects of functional item behavior since they are bound, closed class and as classifiers have a semi-grammatical role. Yet, their meanings are indisputably concrete and lexical. In PFM, lexical suffixes cannot be lexemes because lexemes are associated with only stems and not affixes. Affixes in PFM are associated with morphosyntactic features, which are specific to non-concrete, functional meanings. There simply is no place for lexical suffixes in the theory. DM has a looser correspondence between abstract morphological objects and form types. The issues that theory faces are more subtle since a $\sqrt{}$root may be associated with bound, closed-class content, according to some authors. I discuss the problems both theories have given their notion of the lexical-functional divide in detain in §4.1.

1.4 Word and Paradigm

WP encompasses diverse models and concepts. In this work the notion of WP fundamentally is a recognition of two points that have far-reaching consequences. The first point addresses the notion of W(ord) in WP. I equate W with word-form, rather than a notion like *lexeme*. Free word-forms logically precede any more abstract analysis of morphological patterning, which has consequences for a model of a morphological system. The second point concerns the P(aradigm) in WP. Patterns in the covariation of form and meaning seen across groups of stem-related word-forms constitute morphological categories with causal consequence.

1.4.1 Relations Between Word-Forms

Imagine a linguist encountering an undocumented language called English for the first time. Before the linguist may posit that there is an affix *-s* in the language and that that affix is associated with plurality, they need to make comparisons of free forms. Such an observation requires minimal pairs like *dog* and *dogs*, which demonstrably contrast in only plurality. Such comparisons are essential in morphological discovery procedures. Their necessity is such that it is better to say that the linguist who observes the *dog*-*dogs* contrast has not discovered the suffix *-s*, as some independent unit, but something subtly different, a relation between word-forms. To make this relation a little more concrete, I'll use the common morphological notation for such relations, $X_N \Rightarrow X_N s$. In this expression X is some base, N is some category that constrains the relation to a class of stems. The process of affixation is represented through implicit concatenation. The N can be thought of as a mnemonic for nominal, but it isn't a syntactic category. One can read the relation notation as, "When there is an X of category N there is an X_N affixed by s." The relation is declarative, not a transformation. Note that the inverse relation must also hold to accurately capture what was measured by the discovery procedure. If one demonstrates that there is a specific $X_N s$, such as *houses* – and that *s* is a signaling -*s*, and not simply a stem ending in *s* coincidentally – then the un--*s*-affixed X_N must also exist. In short, the W in WP indicates that the focus is on relations between word-forms.

Imagine a theory that individuates items like affixes as distinct theoretical objects that serve as inputs to morphological rules, essentially what is called Item and Arrangement (IA) (Bloomfield, 1934; Z. S. Harris, 1942; Hockett, 1954). There is no explicit statement or formal mechanism in IA to express that affixation is a relation between two (classes of) word-forms. An affix, like -*s* is simply a lexical entry, like *dog* or some other *morpheme*. There will be restrictions on how a morpheme may combine with other morphemes in IA, but there is no notion that some morpheme's status as an affix depends on minimal pairs that are related by the morpheme's presence or absence. This fact is left as an informal assumption or assumed meta-theory. Perhaps it is made explicit in an external theory of discovery procedures. It is not enforced in the formal IA theory, itself. The implication is that there could be some affix-stem combination where the following are true:

- The stem does not otherwise exist freely.
- The stem does not otherwise exist bound with other affixes or processes.
- The affix does not appear with any other stem.

I don't know how one would demonstrate that such an affix was an affix in synchronic linguistics. It is not clear what it would even mean for such a thing to be an affix outside a formalism that allowed for its definition. A word-form-orientation captures something necessary about affixation and other morphological processes that may be absent in other theories, which is that processes are relationships between word-forms that share a stem.

1.4.2 WP and Non-productive Patterns

The word-based perspective is also helpful in finding a middle ground between overstating productivity and positing covert forms on the one hand and ignoring salient patterns on the other. In Wao Terero, as in many languages, there are bound stems. In my interlinear glosses, I indicate affixation to such stems using ‘ · ’ rather than ‘-’ since conceptually different relationships are involved. As an example, there are two words that share a bound stem *te · pẽ*, ‘chicha (fermented drink)’, and *te · wẽ*, ‘chonta palm’. Both *-pe* and *-wẽ* are lexical suffixes associated with liquids and plants, respectively. There is no free *te* form that seems reasonable to associate with the pattern. Historically, the stems may have shared a meaning relationship since chicha can be made from chonta fruit. There is currently a different name for chicha made from chonta fruit *dagẽ-ka-pẽ*, ?-LS.stone-LS.liquid. Other lexical suffixes do not attach to the base. The affix *-yabo*, LS.leaf$_1$, cannot be used to form **te-yabo*. One could choose to posit that a bound item *te* exists as an independent morphological entity, but it will be difficult to make a convincing case. There is stronger evidence for two distinct lexicalized word-forms *te · pẽ* and *te · wẽ*. The most important reason for this is because of a lack of productivity. Additionally, there is no obviously shared meaning, which is relevant to morpheme-based theories, where a morpheme is defined as a pair of form and meaning. Yet, by positing two unrelated forms, the salience to speakers of the *-pẽ* and *-wẽ* endings is ignored. Word-form relations allow for a middle ground. One can state a very specific relation X_{te}pẽ $\Rightarrow$ X_{te}wẽ, where ‘te’, here, is serving as the name of the category of *te-*. This relation expresses something different from the kind of affixation that involves a free stem but given some supporting theoretical machinery, which will be introduced in due course, the observation that these forms are related to others where *-pẽ* and *-wẽ* are present can be properly indicated. It may seem like I am doing nothing here but adding some additional rules for special cases. Yet, there are more opportunities for parsimony than may be apparent, at first. The *te*-relation is not unique to only one pair, despite its category. There are other free items in Wao Terero that can take the two affixes in addition to other affixes. Let us call the category of such items β. Given that for every member of β it is both true that $X_\beta \Rightarrow X_\beta$wẽ and $X_\beta \Rightarrow X_\beta$pẽ, β is a subcategory of category te because X_βpẽ $\Rightarrow$ X_βwẽ holds. This places specific patterns within the context of more general

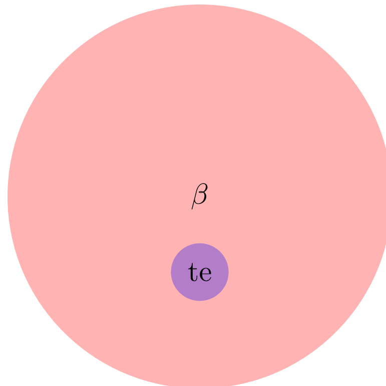


Figure 1.1: The relationship of te category relations to β category relations.

patterns while not exaggerating productivity in specific cases. The relationship between te and β is represented in Figure 1.1. This way of modeling relationships provides an informative explanation. It expresses that though the te pattern is extremely limited, it is not coincidental relative to the morphological system as a whole.

Once it is stated that β exhibits the te pattern, providing additional, β specific, rules $X_\beta \Rightarrow X_\beta$wẽ and $X_\beta \Rightarrow X_\beta$pẽ is redundant. If one states that β subsumes the te pattern all one needs in order to infer $X_\beta \Rightarrow X_\beta$wẽ and $X_\beta \Rightarrow X_\beta$pẽ is some relation that declares the existence of X_β, the un-affixed, free form. For instance, we could specify that it is the case for all free forms that $X \Rightarrow X$ holds, for any category. Or, some other category that is subsumed by β may state that the relation $X_\beta \Rightarrow X_\beta$ka holds – an affixation rule for another lexical suffix which requires the existence of an unaffixed form. Since the relationships are declarative, not transformations, all one needs to be able to do is prove, for a particular stem, that the form described on the left of the arrow exists and the form on the right exists for the relation to hold.

In some cases, it may be that more productive categories like β are definable as the summation of less productive categories, as is represented in Figure 1.2, where κ_n are various categories with more limited productivity. The circle indicating β is not filled in in comparison to Figure 1.1 to informally indicate that β has no relations defined independently of the categories it subsumes. This largely fits the pattern of nominal uses of lexical suffixes in comparison to the more productive demonstrative and adjectival classifier uses in Wao Terero. For this reason, stating relationships

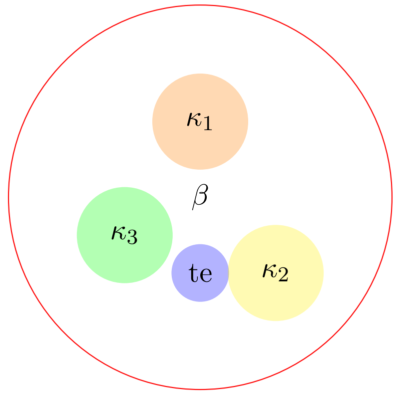


Figure 1.2: The relationship of less general categories to a more general category β. The less general categories are named te and κ_n.

for small, specific categories is not redundant with respect to more general categories that exhibit the same pattern. In fact, the model implied by Figure 1.2 shows that in such a system, the general pattern may logically depend on less general patterns. For instance, if β were specified as productive classifier uses, and te and the various κ_ns all happened to be nominal uses, it would imply that classifier patterns logically depend on nominal usage. This appears to roughly be the case. I discuss possible implications later.

1.4.3 Paradigms as Linguistic Categories

This discussion of fine-grained patterns leads to the second point about WP, which concerns the P(aradigm) in WP. Patterns in the covariation of form and meaning seen across groups of stem-related word-forms constitute morphological categories with causal consequence. The category β was an example of such a category. Another well known case of categories of morphological relations is Spanish conjugations. Pedagogical grammars recognize three conjugations and name them for the ending of the infinitive for each conjugation: AR, *hablar*, 'to speak'; ER, *comer*, 'to eat'; IR, *escribir*, 'to write'. As is true in other languages with conjugation classes, the inflections for person, tense and other language specific inflections follow contrasting patterns within these categories, as can be seen in Table 1.1.

The reason that these three categories exist is diachronic. There isn't any deep synchronic reason for the conjugations to exist and membership in them cannot be explained through appeal

	AR	ER	IR
1	hablo	como	escribo
2	hablas	comes	escribes
1 plural	hablamos	comemos	escribimos

Table 1.1: Example forms for the AR, ER and IR Spanish conjugations in the present indicative using *hablar*, ‘to speak’; *comer*, ‘to eat’; and *escribir*, ‘to write’.

to some other aspect of grammar. For instance, there is a rare instance where a word exists in both ER and IR, with infinitives *converger* and *convergir*. Coincidentally, the word means converge. If the phonology of the stem or meaning of the word was decisive in determining class, this should not only be rare but impossible.

There may be certain qualities, such as phonological form, meaning patterns or default status that may make membership in an AR, ER or IR class more or less likely. Yet, such qualities do not *determine* membership. There may also be ways that speakers leverage the classes in ways that are beneficial to communication. Yet, these, if they exist, would not explain the existence of these three particular classes, even if such a finding might help explain why such classes exist at all and persist through generations. In short, the three conjugations are not the result of some more basic, underlying causal phenomenon. Since membership in the category determines how a lexeme inflects, the category has a consequence.

Wao Terero does not have Spanish-style conjugation classes. I used the Spanish inflectional patterns simply to make the point that morphological systems contain categories that cannot be explained in terms of other grammatical systems. Spanish is more familiar to a general audience than Wao Terero. Despite the lack of conjugation classes, as discussed above, Wao Terero has other paradigmatic patterns that are relevant to lexical suffixes. Wao Terero lexical suffixes have a broad distribution and are found in most parts of speech. Yet, the distribution of the affixes is not uniform. Not all affixes attach to all bases and there is not always a phonological or semantic explanation. This requires some notion of morphological category. More is said about this in §3.3 and Chapter 6.

1.4.4 A Generative Morphology

As an additional note concerning the form-to-form relationships and form-class categories I have introduced is that their definition in this theory results in a generative morphological system. I will be replacing the relatively informal notation for form-to-form relations used above and adding additional information to rules but the basic notion holds. Once one knows one form of something in category β, they can generate a larger set of forms that must exist as a consequence. This is unlike PFM or DM. In those theories there is no independent, generative morphological system.

1.5 Types and Categories of WP Relationships

The import of WP to the goals of this work can be seen in the formalism. The theory I propose is formalized in a modern type theory (Martin-Löf, 1984), called the Calculus of Inductive Constructions (CiC) (Coquand and Huet, 1988; Coquand and Paulin-Mohring, 1990; Luo, 1990). Modern type theory is helpful in defining WP relationships and taxonomies because modern type theories provide an expressive means of defining fine grained types.

1.5.1 Introducing the Concept of Types with Numbers

To introduce the basic notion of types, it is helpful to think of something simpler (in the sense of being more predictable) than conjugation classes, like numbers. A type is something like $\mathbb{N}$ or $\mathbb{Z}$ for the natural numbers or integers, respectively. The number referred to by the symbol 7 may have a representation that is of type $\mathbb{N}$ or $\mathbb{Z}$. The symbol -5 corresponds to a representation of type $\mathbb{Z}$. It cannot be represented as $\mathbb{N}$, which includes only positive whole numbers and 0, whereas $\mathbb{Z}$ includes negative numbers. I say "represented" because whether in set theory or using binary computer code, numbers of these types will require distinct structures. Types are often defined in a manner which makes the properties of their terms, such as 7 or -5, clear. Types can be used to both constrain the applicability of functions and to enable one to reason about functions. For instance, natural number addition as a function takes two terms of type $\mathbb{N}$ as arguments and returns a term of type $\mathbb{N}$. The type for natural number addition is written $\mathbb{N} \rightarrow \mathbb{N} \rightarrow \mathbb{N}$ in type notation.

The integer version is $\mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}$. The first two items connected by arrows are the types of the arguments to addition. The last is the type of what one gets as a result of addition. One could also define a function with a type like $\mathbb{N} \to \mathbb{N} \to \mathbb{Z}$ which converts the result to an integer. For an addition-like function, $\mathbb{Z} \to \mathbb{Z} \to \mathbb{N}$ is a type that would need to coincide to something like the addition of absolute values of the input integers because numbers of type $\mathbb{N}$ cannot be negative. The types alone can tell one quite a bit about what actual functions can serve as terms. The type constraints make it easier to reason about the consequences of addition. One can rely on the fact that natural number addition will always represent a number equal to or greater than either input number. This would not be the case for integers. Improving the ability to reason about a formal system is the major motivation for using types.

1.5.2 Why Use Type Theory?

This point about improving one's ability to reason about a formal system is born out in computer programming, where types are used to ensure that large classes of errors in reasoning simply cannot occur. This is useful for linguistic theories as well. One could express all of my formalism in set theory, without types (or only one type), but it would be much harder. The likelihood of serious errors would be greater and useful abstractions and patterns may not be as obvious. To make an analogy, one can write the equivalent of Python or R code by flipping a switch on and off, since it all reduces to binary, but not only will the programmer likely lose count, they'll be so focused on switch flipping that they may miss out on the bigger picture that abstractions provide.

Other linguistic formalisms such as those used by unification grammars like Head Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994), use abstractions over set theory for similar reasons. There are meaningful differences between my theory and a unification grammar that go beyond formal implementation but one reason to prefer type theory as an abstraction is that it is used far more broadly in the scientific community. By not using a specialist system just for linguists, linguistic science becomes more interdisciplinary and one can benefit from the work of logicians, mathematicians, computer scientists and philosophers who are engaged in research with type theory. This is born out in formal semantics, where the use of logic and types are the norm.

In fact, semantics is the area where modern type theories have found the greatest use so far in linguistics (Asher, 2011; Chatzikyriakidis and Luo, 2013; Luo, 2012a; Sundholm, 1989; Luo, 2012b; Ranta, 1994; Bekki, 2012; Kubota, Mineshima, et al., 2019).

As an example benefit of speaking an interdisciplinary language, I am able to verify my formalism using theorem provers, also called proof assistants. Three popular ones that can be used to work in CiC are Coq (The Coq Development Team, 2019), Agda (van der Walt and Swierstra, 2013) and Lean (Moura and Ullrich, 2021). These are specialized computer programming languages for working in type theory. They were not developed with linguistics in mind but serve my purposes without requiring large and complicated specialist libraries. Using them I can ensure that I have not made any errors, develop my ideas with greater confidence and scale my model to a broader fragment of a language with greater ease. Just as a corpus linguist can benefit from general statistical libraries, I do not need to reinvent the wheel, so to speak. Likewise, such stat libraries, like computational implementations of type theory, are not just about convenient computer programming, as is the case with general purpose programming languages like Prolog or Python. They are "real math" so to speak, which is important for general understanding by others. Since these systems are designed by experts in the mathematical systems involved, it allows one to have confidence in the implementation and design of the systems.

1.5.3 More Advanced Types

All of what I have said so far about types is consistent with *basic* type theory. A modern type theory allows for greater expressiveness and more specific constraints. The goal of ensuring that one can better reason about a formal system is maintained, but the system also allows one to say more interesting things. For instance, one can quantify over types, or have types that depend on the value of a term. Returning to the example of the very specific te category above, there are more general categories in the language that exhibit the te pattern in addition to allowing affixation by a wider range of suffixes. We called that category β above. Since β has all the properties of te we can designate their relationship as $\beta \leq$ te. The idea is that anything less than or equal to te has all the properties of te. With dependent types we could provide a type for the X_{te}pẽ $\Rightarrow$ X_{te}wẽ relation analogous to $\mathbb{N} \to \mathbb{N} \to \mathbb{N}$ for natural number addition. I don't want to introduce the whole type for the te relation here since it will look complicated if one isn't used to modern type theory. I'll leave a more thorough introduction to the aspects of type theory needed to understand my formalism for §5.1. For now, I only want to introduce the part of the type that ensures that the relation only holds for things that exhibit the te pattern. This part of the type would look like $\forall a : \kappa . a \leq \text{te}$. Here, a is a variable. κ is the type of categories like β or te. The type says first $\forall a : \kappa$, for all categories a. Importantly, whatever kind of thing the relation applies to, it is something that has a morphological category. The next part is the more fine-grained condition, which filters out non-conforming classes that the relation should not apply to, $a \leq \text{te}$, read a less than or equal to te. Remember that this is how the relationship between β and te was defined. The type, all together, can be read, for all a such that a is a morphological category and is less than or equal to te. So if something with the category β were to be provided as an argument, the type would match. Likewise, te would match. Those familiar with logic but not modern type theory will note that the type, $\forall a : \kappa . a \leq \text{te}$, looks like an expression in first-order logic. This is kind of the idea. There is a notion that types are propositions and their terms are a proof of the proposition. This should give one some insight into the expressiveness of modern type theory. I further elaborate on using type theory to express WP taxonomies in Chapter 6.

1.6 Realization and Proof

The ability to specify taxonomies utilizing types is very useful, but the system must also provide a means of expressing the realizational relationship between morphological forms and context. Another property of using CiC, which is a particular kind of modern type theory called a *constructive* type theory, is that demonstrating that morphological forms correspond to particular meanings is done by constructive proof. Such proofs can be thought of as analogous to the notion of *derivation* used in transformational theories of syntax, like Minimalism (Chomsky, 1995). There are a number of important differences between proofs and such derivations. A prominent difference is that the proof is not a transformation from covert structures to surface structures. Instead, it is a declarative relation. Despite the differences, the "mechanics" of a proof and a Minimalism-style derivation are sometimes superficially similar. This is similar to the way that $X \Rightarrow Xy$ style relations may look like transformational rules, even though the meaning is declarative and should also imply $Xy \Rightarrow X$.

To prove something constructively, one constructs an example object. As an example, above I noted that a type is a proposition and the term its proof. So for the type $\mathbb{N}$ a constructive proof would be an example of the type, like 0, 43 or 1,000,001. For realization (we'll ignore the types of realizations for now), that object will be a morphological form that is compatible with a context. One could imagine that there is some "almost" sentence *The cat ____.* that must have a corresponding meaning $past(jump(\iota(cat)))$, where ι is some definiteness function/operator and the rest expresses that this definite cat made some motion away from the ground in the past. There is a meaning and a syntactic context that some form must be compatible with in order to fill the blank. The form one puts in the blank must be grammatical and felicitous. The question is, is there such a form? As stated above, the expectation is that a taxonomy is defined as expressions in CiC. The taxonomy describes the morphological system. The expressions that define the taxonomy are axioms of the theory. Axioms are what one assumes to be true. In a scientific theory the axioms largely correspond to hypotheses. Some of these are strongly supported by evidence, such as rules of English plural formation. Others may be exploratory and proposed with the goal of comparing empirical tests with the predictive consequences of the axiom within the model. For instance, the axioms may predict that a stem X takes suffix y. The field linguist then formulates a sentence with that construction to test whether a speaker will accept the morphological construction. For Wao Terero, which lacks the level of documentation seen for English, many axioms would necessarily be of the latter type.

Returning to the "almost" sentence and its corresponding meaning, given a reasonable set of axioms for English, it should be possible to prove that any one of the morphological forms *jumped*, *leaped* or *leapt* exist and are compatible with the given context. There may be others, but any one of the three could fill the blank. In the theory's formalism each such form would be proven independently. Different pieces of evidence would allow for each of the three proofs. One neither proves them as a collection all at once, nor does one specify that one proof depends on another proof of a form being true or false. One proof of a form cannot negate or block another. This is unique among overtly realizational theories. PFM and DM, for instance, make use of concepts such as blocking and competition in their realizational models. Neither the early Stump (2001) PFM nor current DM would allow both *leaped* and *leapt*. The result would need to be exclusively one or the other. For some inflectional category such as the past tense, realization in these theories involves competition such that only one potential form can "win", blocking other potential forms from also winning.

Later versions of PFM do accommodate both *leaped* and *leapt* but for reasons other than competition or blocking, PFM would still not allow both *jumped* and *leaped* to be realized. In DM and PFM one also needs to incorporate information into the realizational computation about which word one is concerned with. Conceptually, the almost sentences for these theories would have to look like: *The cat (*JUMP*).* or *The cat (*LEAP*).*, with some additional information about which word is allowed to fill the blank. As noted above, the theories assume that syntax has abstract units called lexemes or √roots in its structure, which will usually determine the form of the stem in a non-realizational manner.

1.6.1 Morphology without Reference to Lexemes and √Roots

The preceding subsection should give the reader a strong clue as to how functional and lexical morphology are unified in my theory. I am modeling realization in terms of compatibility with a context. In a fieldwork test, if I supplied the sentences *The cat leaped.*, *The cat leapt.* and *The cat jumped.*, and provided an image of a cat jumping and asked speakers for the acceptability of the sentences given the stimulus, acceptance of all three forms is most of the evidence I would need to claim that my model is making the correct predictions. This does not require covert lexemes and √roots. As mentioned above, in both DM and PFM it is the presence of those covert elements, and other covert elements called features, that encode the lexical-functional dichotomy. Lexemes and √roots have non-realizational relationships to meaning while other covert elements do have realizational relationships to meaning. By erasing both these representations and focusing on more measurable constructs, I simplify the system and allow for the concept of realization to be generalized to a broader range of morphological phenomena. This point is described in greater detail in §4.1.

1.6.1.1 Covert Lexical Elements and Generative Morphology

One reason I can get away without lexemes and √roots is because, unlike PFM or DM, I propose a generative morphology. What this means is that given the axioms of the system, one can conceptually prove all the licit word-forms in a language. Those forms simply have no meaning associated with them until they are related to a context at the interface. PFM and DM make use of lexemes or √roots as objects that pick out stems and then have a function that takes the stem and morphosyntactic features as input to calculate a word-form. My system generates word forms directly from stems. There are abstract categories that aid this, like te and β, but they are not meaning oriented. Lexemes and √roots are concepts that exist in the theory but have no formal instantiation since they are irrelevant to the computation of form. For instance, a lexeme is a category of a particular instantiation of an inflectional paradigm, *walk*, *walks*, *walked* and *walking* would all be forms under the WALK lexeme category, but *walker* would not be. WALKER is a distinct category that includes *walker* and *walkers*. A √root has a derivational domain so all of WALK (verb) is included in the category √walk as well as WALKER. I believe both category types are useful. I would personally describe inflectional suppletion, for instance, as relying on the notion of lexeme, distinct from the notion of stems, which is inline with more recent PFM (Stump, 2016). The conceptual usefulness of the concept does not require that lexemes should be treated as explicit objects on syntactic nodes. I raise but leave aside the question of whether the concept of lexeme is an epiphenomenon of the nature of syntactic category and the interface with a stem-oriented morphology. The question is beyond the scope of this thesis.

In PFM the lexeme is not very relevant to form building. It is used to establish a stem but afterward may not be referred to in realizational rules. This is not to say that lexemes are limited to serving as pointers to stems. They also play another non-syntactic role. In PFM and DM the lexemes and √roots are not only used for picking out a base but are also used for assigning lexical meanings. Another reason I do not need them is due to a different take on the interface(s) to semantics.

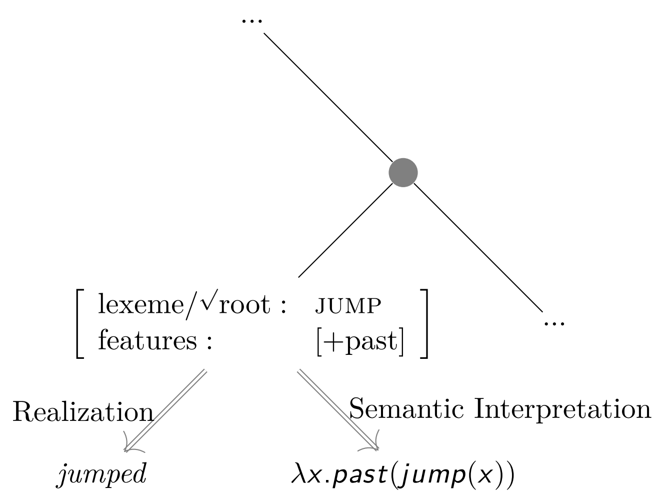


Figure 1.3: In PFM and DM the meaning of a morphological form is an interpretation of a syntactic structure that includes some combination of lexical and functional representations. Realization is parallel to this and cannot access the semantics.

1.7 The Interface to Syntax and Semantics

The interface to syntax and non-lexical semantics in my framework uses a categorial grammar (CG) (Ajdukiewicz, 1935; Bar-Hillel, 1953; Lambek, 1958), called Linear Categorial Grammar (LCG) (Mihaliček, 2012; Mihaliček and Pollard, 2012; Worth, 2014; Worth, 2016; Needle, 2022). CGs are known for their emphasis on the syntax semantics interface. LCG exemplifies this. It has been used in research on the interfaces to compositional and discourse-level semantic theories (Martin, 2013; Martin and Pollard, 2014; Pollard and Worth, 2015; Yasavul, 2017). This is important given that realization in my theory is relative to a notion of context that includes these meaning domains. LCG and the semantic theories it encompasses have also been formalized in Needle (2022) and Needle (2024) using CiC. This makes it easier to integrate the theories. The LCG syntax-semantics interface is quite different than what is assumed in PFM and DM.

Both PFM and DM assume that meaning is interpreted from morphosyntactic content. This means that meaning representation follows syntactic derivation or unification. This is schematized in Figure 1.3 where there is some substructure of a syntactic tree that contains lexical information and functional information. The substructure receives an interpretation as a secondary step to provide an actual semantic expression. This interface appears to have been conceived by Lewis (1970) and popularized in the undergraduate textbook of Heim and Kratzer (1998). Notably, the semantic interpretation is parallel to the realizational calculation. That means that to whatever extent the realizational calculation relies on semantic content, that content must be pushed up into the syntax, so to speak.

This is not the case in LCG. In LCG, the meaning and the form of a phrase are composed in tandem. Consider that I posed the problem of providing a proof of a morphological form using a pair of a sentence with a blank and a meaning: $\langle$ *The cat* ____., $\mathit{past}(\mathit{jump}(\iota(\mathit{cat})))\rangle$. I do not literally use a pair like this in the formalism. I posed a question that realization answers in this manner as a simplified analogy for the more technical system that I introduce later, but the idea of pairing is important. LCG assumes something very similar. Syntactic information is always paired with a semantic expression. An individuated noun is paired with a meaning $\langle \mathsf{N}, \mathit{cat}\rangle$. A syntactic category that takes an argument according to its subcategorization requirements is paired with a meaning that takes an argument. An example of this is the pair $\langle \mathsf{NP} \multimap \mathsf{S}, \lambda x.\mathit{past}(\mathit{jump}(x))\rangle$, where $\mathsf{NP} \multimap \mathsf{S}$ is the LCG category for an intransitive verb, a category that when provided a noun phrase, NP results in a complete sentence, S. The symbol '$\multimap$', lollipop, is linear implication. It expresses that in order to achieve the category on the right, one must provide something of the category on the left. The details of linear logic are not important in this thesis but see Wadler (1991) for a quick introduction.

One may note that in the pair $\langle \mathsf{N}, \mathit{cat}\rangle$, there is no form, just the meaning and a syntactic category. A full LCG *sign* is needed for syntactic combination. The sign will include a form in addition to the syntactic category and meaning. This notion of sign is similar to the notion of Saussurean sign (de Saussure, 2011) with the addition of a syntactic component. It is very near the notion of sign in HPSG because LCG originated as a categorial reinterpretation of HPSG. The form part of the sign is provided by a realizational proof. The result of realization is a sign triple, such as $\langle \text{cat}, \mathsf{N}, \mathit{cat}\rangle$. So if one considers $\langle \mathsf{NP} \multimap \mathsf{S}, \lambda x.\mathit{past}(\mathit{jump}(x))\rangle$, there is a actually a "blank" that the realizational proof fills.

To give one a more concrete notion of how this looks, here are the basic steps to prove *The cat* *jumped*:

1. Realizational proof of $\langle$cat, N, $\mathit{cat}\rangle$.

2. Realizational proof of $\langle\lambda$s.s • jumped, $\mathsf{NP} \multimap \mathsf{S}$, $\lambda x.\mathit{past}(\mathit{jump}(x))\rangle$.

3. Realizational proof of $\langle\lambda$s.the • s, $\mathsf{N} \multimap \mathsf{NP}$, $\lambda x.\iota(x)\rangle$.

4. Syntactic proof that given 1 and 3, $\langle$the • cat, NP, $\iota(\mathit{cat})\rangle$

5. Syntactic proof that given 2 and 4, $\langle$the • cat • jumped, S, $\mathit{past}(\mathit{jump}(\iota(x)))\rangle$

The steps could be slightly different but if we assume that we're proving the initial signs of words first, also known as the lexical entries of syntax, there will be 3 initial realizational proofs to show those signs exist. Afterward one proves various syntactic combinations until a sentence is produced. I assume that the reader has seen some lambda calculus in the past but if not, it is simply a notation for function where the λ symbol introduces variables. People are accustomed to seeing these in semantics but here we're seeing them on the very left of the triples, which is where the surface strings are specified. For now, it is enough to know that this controls relative word order, which is not handled by the abstract syntax. The '•' is a concatenation operator. The end result of the combination is something with a syntactic category S, the string *the cat jumped* and the correct meaning.

Full signs are required for syntactic combination. This is because for any combination of syntactic category there must always be a corresponding semantic and string combination. Such signs are only available though a realizational calculation. This may imply that a realizational proof is logically prior to syntactic combination. Yet, the concept of before or after does not really apply. The proof as a whole, including syntactic combination and realization is declarative. There is no sense of one level strictly feeding another or of some kind of transformation of covert information and structure to a surface representation. It is analogous to being given a collection of puzzle pieces and getting the edges to fit together to make a coherent picture.

Due to the fact that the morphology interfaces with such a system, the realizational proof has direct information from compositional semantics available to it. Metaphorically, the realizational proof has a window into multiple domains of grammar. This is not the case in PFM and DM, with their post-syntactic meaning interface. There is no window into meaning in their realizational calculations. For this reason a messenger is needed, so to speak, and the theories require lexemes and √roots in the syntax as meaning-bearing items. Having a more direct relationship to meaning paired with a generative morphological system are important ingredients that help make the use of objects like lexemes and √roots superfluous and remove barriers to providing a realizational treatment of lexical meaning content.

As a kind of side note, it is interesting that a theory, like PFM, where a common assumption is that it interfaces with a more HPSG-like syntax, does not have direct access to meaning representations in the same way that my own system does. This is because HPSG also is a system of sign combination so there is a window into semantics but PFM is not interested in the view. Even in Information-Based Morphology (IbM) (Crysmann and Bonami, 2016), a PFM-like theory that is explicitly and tightly integrated with HPSG, there is a parallel system of information specific to morphological realization that makes no reference to the available semantic representation. Instead, in IbM, a similar notion of lexeme is used for lexical meaning.

There are two other important ingredients relevant to the interface to semantics that I leave for later. One is the interface to lexical semantics, where a stem may be associated with a variety of conventional meanings. The other is discourse, which ensures felicity. It is possible to ignore felicity to prove *Colorless green ideas sleep furiously.* but if we're concerned about whether the use of a classifier is correct in a context, it is necessary to consider discourse.

1.8 Structure of this Work

Following this introductory chapter there are six additional chapters.

Chapter 2, A Description of Wao Terero and its Lexical Suffixes, provides an overview of typological categories that are relevant to Wao Terero. §2.1, Lexical affixes Cross-linguistically, describes lexical affix systems in other languages and how they relate to Wao Terero. §2.2, (Bound) Classifier Typologies, relates Wao Terero to themes in the classifier literature, including the classifier-gender dichotomy, issues with verbal classifiers, and word-formation in multi-classifier systems.

Chapter 3, A Description of Wao Terero and its Lexical Suffixes, provides a descriptive overview of relevant areas of Wao Terero grammar. It is not intended to be comprehensive. The chapter provides sufficient information to understand examples and make comparisions with previous literature on the language. It also expands on the description of lexical suffixes provided in the introduction and clarifies the content that an adequite analysis must cover. There are four sections.

§3.1, Phonology and Orthography, provides a basic description of the phonology and orthography of Wao Terero. Wao Terero data in this thesis is provided in a phonemic transcription. The section allows the reader to understand the evidence that provides validity to this treatment and how the transcription relates to various orthographies. It will also provide some diagnostics for free versus bound content.

§3.2, Inflection and Morphosyntax, focuses on basic inflectional morphology. The section focuses primarily on person and number morphology with some description of tense-aspect-mood related affixes. Morphosyntactic description is provided for the parts of speech relevant to lexical suffixes, including verbs, nouns, adjectives, demonstratives and the numeral/quantifier *adoke*, 'one'.

§3.3, Description of the Lexical Suffixes, provides a descriptive account of the lexical suffixes. A description for each of the parts of speech where the suffixes occur is provided. There is also a description of a selection of affixes of particular interest to this study.

§3.3.6, Lexical suffixes and anaphoric properties, provides a discussion of anaphoric properties of lexical suffixes. Particularly in the literature on classifiers there is claim that lexical suffix-like phenomena have anaphoric properties. Peeke (1968) and Peeke (1991) claims that Wao Terero lexical suffixes, when used as classifiers have these properties. I provide some elicitation-based data that provides greater insight on this claim from the point of view of formal pragmatics.

Chapter 4, Issues for a Formal Analysis, presents the issues that previous approaches to morphology have with lexical suffix data. This is laid out in detail in §4.1, The Functional-Lexical Divide, where there is a particular focus on issues for feature-based realizational theories, such as PFM and DM.

The chapter concludes with §4.3, Concepts of the Formalism, which provides a high level overview of the formal architecture that is used in my analysis.

Chapter 5, Formal Preliminaries, provides the formal preliminaries needed to understand the analysis. This includes a deeper dive into type theory in §5.1 as well as a more detailed look at the syntactic (§5.2) and semantic (§5.3) theories that I use.

Chapter 6, WP and Lexical Realization, provides the formal details of the theoretical framework and an analysis, essentially an outline of a grammatical fragment, which focuses on the lexical suffix system and basic verbal inflection. §6.1, What is a Formal Grammatical Fragment?, discusses the nature of a grammatical fragment. §6.3, Framework Architecture, provides an overview of the framework architecture and compares it similar theories. Following this more comparative section, §6.4, Some Example Wao Tededo Paradigm Instances, provides explicit examples of Wao Terero data as it is represented in the framework. Once the data representation is established §6.5, Some Example Wao Tededo Paradigm Instances, introduces how information is related within form paradigm instances and between form paradigm instances and sign paradigm instances, which are the two types of paradigms the framework uses to express the relationship between word-forms and their syntactic distributions. §6.6, What the analysis covers, describes the data that is covered by the analysis. The analysis begins in §6.7, Morphotactic analysis, which covers morphotactics and properties of form paradigms. The analysis continues in §6.8, The Wao Terero sign paradigm, which demonstrates realizational proofs, how the forms generated by morphotactics relate to syntax and semantics. The final section, §6.9, WP Properties, introduces how the system defines biconditional rules within and between paradigm instances.

Chapter 7 is the conclusion. A summary will be provided and remarks on the utility of the theoretical perspective.

Following the conclusion there is a bibliography and appendices.

Appendix A contains the formal theory in Coq, the theorem prover. In discuss how the syntax of the computer code relates to the notation used in the analysis.

Appendix B contains a brief analysis of intransitive English *walk* and two Spanish verbs. I use the word *walk* as a running example, and the analysis makes the relationship to the framework architecture explicit. The Spanish example provides a taste of allomorphy, which is not otherwise emphasized in the Wao Terero data.

Appendix C contains some clarification on the definitions and roles of paradigms and paradigm objects.

Appendix D provides the listings of lexical suffixes provided by Peeke that are found in previous works. I note where my own research findings reinforce or differ from her account. There are references back to examples where the lexical suffixes in the listings occur in the thesis.

Appendix E discusses glossing conventions. There is a master list of all non-Leipzig glossing conventions used. Each gloss contains a brief description, references to other literature and page number references to examples where it occurs in the thesis.

Chapter 2

Typological Context

I provided some basic examples which demonstrate important characteristics of the Wao Terero lexical suffix system in §1.1. That brief explanation is contextualized here from a cross-linguistic perspective. This is to aid the reader who would like to know more about such systems, or who may be unfamiliar with the literature that I borrow terminology from. Although I primarily supply a mildly opinionated review of the literature in this chapter, there is a unique contribution to discussions of comparative categories. There are a number of Amazonian languages that are said to have multi-classifier systems (Aikhenvald, 2000; Rose and Van linden, 2022; Krasnoukhova, 2012), where the same affixes are used on a number of parts of speech. Wao Terero has been classified as one of these (Aikhenvald, 2000). There is good evidence that there are classifier uses of the affixes in question in all such systems. The issue is that multi-classifier systems are also multi-function (Krasnoukhova, 2012). Some functions are derivational and compound-like (Aikhenvald, 2000; Rose and Van linden, 2022; Krasnoukhova, 2012; D. L. Payne, 1987; Derbyshire and D. L. Payne, 1990; Seifart and D. L. Payne, 2007). There are popular diagnostics for classifier status, which are specifically intended to exclude such behaviors (Sara T. Rosen, 1998; Passer, 2016). Therefore, referring to the multi-functional affix system as simply a type of classifier system is misleading.

There is also a very well-established North American literature on lexical affixes (Sapir, 1911; Kinkade, 1963; Mithun, 1997), which has since been popularized beyond the continent. Lexical suffix systems are also multi-function, which includes classifier-like behavior (Wiltschko, 2009; Kinkade, 1998). Lexical suffixes must be closed class affixes with lexical meanings, but otherwise may play a wide variety of roles. By applying the term lexical suffix as an alternative designation to Wao Terero, I contribute a merger of two traditions, which is clarifying. One can point to diagnostics that contradict a pure classifier designation of the Wao Terero's system, but one cannot say that classifier usage within the system is contradictory to multi-function lexical suffixes. It is also useful to look at multi-function systems in a broader context. Aikhenvald (2000), Rose and Van linden (2022), and Krasnoukhova (2012) point out that multi-classifier systems are atypical, and largely specific to the Amazonian region. There is a near certainty that this uniqueness is based on the choice of looking at such multi-functional systems as primarily classifiers. The behaviors described look very much like phenomena in the North American languages I discuss, as well as other languages of the world with lexical suffixes.

Wao Terero is spoken in the Amazon at the base of the Ecuadorian Andes. In addition to Amazonian languages, it comes into contact with languages in the Quechuan family (Camacho Rios, Floyd, and Julca Guerrero, 2024), as well as other language families that have been described as Andean, such as Chicham (Adelaar and Muysken, 2004). Among modern speakers I have observed substantial contact and bilingualism with Spanish and Lowland Kichwa. Therefore, the spheres of potential influence on the language have been diverse. Areal classifications have placed Wao Terero among the Amazonian languages (Aikhenvald, 2000; Derbyshire and D. L. Payne, 1990; D. L. Payne, 1987), which are known for their verbal classifier systems (D. L. Payne, 1987; Aikhenvald, 1994; Barnes, 1990; Gomez-Imbert, 2007; Grinevald, 2000; Grinevald and Seifart, 2004; Key, 1967; D. L. Payne, 1986; D. L. Payne, 1987; Seifart, 2002; Seifart, 2004; Seifart and D. L. Payne, 2007; Seifart, 2009; Valenzuela, 2016). Due to the traditional categorization, a typological contextualization requires some commentary on Amazonian languages. In my work in the Pastaza and Napo provinces of Ecuador I had no contact with speakers of the languages that are described in comparative works such as Derbyshire and D. L. Payne (1990) and D. L. Payne (1987), except Wao Terero. For that reason, areal comparison will be limited to basic commentary on the literature. Wao Terero cannot be compared to other languages of the same family because it is an isolate.

I am sparing with examples from other languages that I have not personally worked with. In secondary works, there is often a folklore of examples that are passed from one author to the next, layering interpretations without addressing speaker verification. Some of the examples that I do provide from typological works are three citations deep. Knowing that my own description of Wao Terero is a collection of reasonable hypotheses, rather than the result of verification from multiple researchers, this is worrying. Consider the debates surrounding a well-studied language like English. Making claims without scientifically engaging with a language is questionable. In the field I have often found it difficult to reproduce Wao Terero examples in Peeke (1968). This could be for a variety of reasons such as missing context, or language change. There is always the possibility that it may have been an error in the initial documentation. I strongly believe that most often, the issue is one of working with different populations at a different time, and that Peeke's documentation is excellent. Yet, the issues of reproducibility raise questions concerning the scientific foundation of any work using her examples uncritically, especially when adding new interpretations and claims.

In the following section I begin with a discussion of lexical suffixes. Lexical suffixes converge and contrast with incorporation and compounding in the typological literature due to the fact that meanings and some syntactic characteristics are similar, despite formal dissimilarities. I then discuss proposed contrasts between classifier systems and gender systems. These are sometimes described as being on a grammaticality cline with classifiers (Grinevald and Seifart, 2004), or as having significant overlapping properties (Fedden and Corbett, 2017). Having at least some qualities associated with gender are important in establishing classifier-like uses of a lexical suffix, as it demonstrates a contrast with compound-like uses.

The invocation of comparative categories in this work is only to aid in description. Considering only the category of classifier, debates concerning how it *should* be defined have diverse and contradictory conclusions (Dixon, 1986; Fedden and Corbett, 2017; Allan, 1977; Aikhenvald, 2000; Seifart, 2009). I am ambivalent to debates concerning comparative categories. I do not believe that concepts like classifier or lexical suffix are natural, only descriptive. My preferences for such terminology are based on being consistent and avoiding misleading characterizations.

2.1 Lexical affixes Cross-linguistically

Before getting into the details of specific systems, I would like to highlight the core characteristics of a lexical affix.

- The item must be a bound affix.
- The item must belong to a closed class.
- The item must not be systematically derivable from a free or open class.
- The item may be associated with multiple meanings, but some must be concrete, lexical meanings.

Cross-linguistically, hosts may be any part of speech. In Wao Terero lexical suffixes are used on most parts of speech. In Salishian, discussed below, they are used on the two major lexical categories of noun and verb. They are equally lexical suffixes if they occur only with a single part of speech. They tend to be multi-function, but this is not required.

The term lexical suffix was first introduced to describe a phenomenon of Salishian languages (Kinkade, 1963). Prior to the coining of this term, Sapir (1911) described such affixes as *substantives*. There have been other terms used for similar phenomena in the literature (Kinkade, 1998). There are also many systems that are known under other names which may be classified as lexical suffixes. I believe that most Algonquian medials would qualify (Jones, 1904; Denny, 1983; Bischoff, 2011; Macaulay and Salmons, 2017). I also believe that many Amazonian classifier systems would qualify, if they are known to have compounding or derivational uses, due to the fact that they violate fundamental tests for classifier status when used in this way (Passer, 2016; Sara T. Rosen, 1998; Aikhenvald, 2000). I save the discussion of Amazonian classifiers for later subsections.

When Sapir (1911) brought linguists' attention to lexical suffixes, he was primarily concerned with the topic of noun incorporation. At the time there were debates concerning the existence of incorporation, whether lexical or syntactic. Sapir, while arguing that incorporation existed, used Salishian lexical suffixes as an example of something that should not count as incorporation. The Salishian suffixes are closed class and are not derived from free nominals. Some example pairs

-as 'face, round object'	sʔaθəs 'face'
-cəs 'hand, finger'	celəš 'hand'
-šən 'foot, leg'	sx̌əńə 'foot'
-wił 'rig, vessel'	ləwəx̌ 'rib'

Table 2.1: A subset of a listing taken from Gerdts (2003, p. 346). On the left Halkomelem lexical suffixes are listed. On the right, free nouns with similar meanings are listed. There is no systematic phonological similarity.

of affixes and nouns with similar meanings from Halkomelem listed in Gerdts (2003, p. 346) are provided in Table 2.1. The lack of any consistent form similarity argues against any systematic process of noun incorporation when the suffixes are used with verbs. Neither do they qualify as affixoids, affixes with a clear correspondence to free word-forms (Ascoop and Leuschner, 2006), when used with nouns. I discuss the issue of whether Wao Terero affixes may be derived from free forms in §3.3.5.1.2. Despite having the formal characteristics of normal affixes, the meanings of the affixes are often noun-like, concrete, and therefore lexical.[1]

2.1.1 Lexical Affix Meanings in Salishian and Other Languages

Gerdts (2003) states that most Salishian languages have around one hundred lexical suffixes. Kinkade (1998) provides more specific numbers and citations, for those looking for details. The meanings include body parts; environmental concepts, such as 'fire' or 'rock'; cultural items, such as 'canoe'; and human relations, such as 'child'. Notably, all but the last category are represented in the Wao Terero system.

In Salishian languages, especially the body-part affixes may have meanings that extend to shape, locative, and relational meanings (Gerdts, 2003; Hinkson, 1999). In Wao Terero, most lexical suffixes have body-part interpretations, but I have not noted that lack of a body-part interpretation made an affix less likely to be associated with shapes. Body part meanings are common in lexical suffix systems, and exhibit a variety of characteristics cross-linguistically.

Gerdts (2003) states that the suffixes may have derivational uses. This is questionable in Wao

[1]As a warning, it is common in the Distributed Morphology (DM) (Halle and Marantz, 1993) incorporation literature to analyze lexical suffixes as a form of incorporation (Johns, 2017). This is a theoretical point of view I discuss in §4.2.

Terero. Derivation in such cases would not be the class changing derivation discussed in the gender and classifier literature, which I will describe later, since classes are not an issue. Whether lexical suffixes are used for derivation such as deverbalization in Wao Terero is discussed in §3.3.5.2.4.

Meanings associated with particular affixes are often described in terms of pluralities, rather than prototypes. Mithun (1997), speaking from a cross-linguistic perspective, describes lexical suffix meaning associations as being network-like. Gerdts (2003) also describes Salishian lexical suffixes as having multiple meanings. Kinkade (1998) claims that though such pluralities (or extensions) do not always have obvious internal relationships, there is *always* some cultural explanation. An association with a plurality of meanings is consistent with Wao Terero lexical suffixes. Multiple meanings of Wao Terero suffixes do not always have synchronic relationships. There is evidence of recent phonological mergers of distinct affixes (§3.1), which may have increased meaning inventories without any need for lexical semantic explanation. I believe that meaning inventories may also have increased due to phonological analogy, a hypothesis I briefly describe in §3.3.5.1.1.

2.1.2 Verbal Lexical Suffixes, Incorporation, and Syntactic Effects

There have been debates in the Salishian literature concerning the role of lexical suffixes within verbs. The debates are relevant to all languages that allow lexical affixes (or classifiers) within verbs. An important diagnostic in the debate is borrowed from the incorporation literature. The origin of the diagnostics are Mithun (1984) and Mithun (1986), where the author describes a diachronic process which begins with verb-noun compounds and results in verbal classifiers. Free nominal incorporation, where the nouns have relatively fixed concrete meanings, gradually becomes a system of closed class, bound affixes, with relatively general, vague, or abstract meanings. In the initial stage, the incorporated nominal is expected to serve as a verbal argument. In later stages, the classifier-like element is expected to be related to a verbal argument, but does not serve as a verbal argument, itself. Mithun's proposal may be a reasonable diachronic explanation for some systems. For synchronic studies, a positive quality of the contrast Mithun describes is that it provides a clear divide between two roles that nominal information can play within a verbal unit. One can find both patterns in incorporation systems. These are called compound noun incorporation (NI), and classifier NI, and are described below. Later, it will be clear that these are not the only two options.

Compound NI: The incorporated nominal serves as a verbal argument.

Classifier NI: The incorporated nominal meaning is related to the verbal argument, but does not serve as the verbal argument.

Since lexical suffixes are not incorporation, I use the terms *compound-like* and *classifier-like* in order to avoid the term incorporation. The primary diagnostics for differentiating the two types is the *doubling* test (Sara T. Rosen, 1998). This involves demonstrating that the incorporated element does not *saturate* the argument position of a verb. A lack of saturation allows a lexical noun to serve as a syntactic argument. The test is considered equally relevant in lexical suffix usage. I already provided an example of doubling for Wao Terero in (2), repeated here as (6). The test seems simple, but Salishian cases demonstrate that argumentation around doubling tests may require a great deal of supporting evidence. I raise additional issues in §2.2.2.1.

(6)	*õ · be*	*a · wẽ*	*pa-wẽ-ta-bo-pa*
	∅ · LS.territory	∅ · LS.plant	cut-CLF.plant-PST-1-DECL

‘In the garden, I cut a pole.’

The literature on the syntactic and semantic role of lexical suffixes in Salishian languages contains significant variation, and debate. Gerdts (2003) describes the affixes as altering argument structure, whereas others have described the affixes as having no effect on verbal argument structure (Anderson, 1992; Mithun, 1997; Wiltschko, 2009). Mithun (1997) emphasizes that the suffixes neither serve as grammatical arguments, nor refer to grammatical arguments. This presents a third option between compound and classifier-like behavior, which is found in some South American systems (§2.2.2.1). The widely cited definition of verbal classifiers provided by Aikhenvald (2000) requires that a classifier must relate to an argument of its host. The characterization of authors like Mithun (1997) may explain why some authors argue against an *adverb*-like role of the Salishian suffixes, rather than a classifier-like role Gerdts (2003). Note that I borrow this usage, adverb-like, for cases of lexical affixes and (so-called) classifiers, which modify the predicate or otherwise do not relate to an argument. Gerdts (2003, p. 355) states, “Lexical suffixation functions like compounding noun incorporation”, with no qualifications. This contrasts with Wiltschko (2009) who describes verbal lexical suffixes as being classifier-like. She reports doubling. For instance, in (7) the highlighted lexical suffix is glossed as having a similar meaning as the lexical noun that occurs in the sentence.

(7) *th'éxw*-**wíl**-*t-es te ló:thel*

wash-dish-TRANS-3.SG

‘He washed the dish.’ (Wiltschko, 2009, pp.208, 211)

The similar meanings attributed to the affix and noun strongly suggest a relation between the lexical suffix and the verbal argument. The existence of such a relation would entail that, at least in some cases, the observation of Mithun (1997) does not hold. There are relations between affixes and arguments. For Wiltschko (2009), given the generality of her analysis, there also is no room for the data of Gerdts (2003). All lexical suffixes with all stems are expected to be classifier-like. For more details on the analysis, see §4.2.

Notably, Kinkade (1963) and Kinkade (1998) provide examples of lexical suffixes filling a number of thematic roles within a verb, implying compound-like uses, but the author also documents classifier-like lexical suffixes. Despite clear documentation of a multi-functional nature, theoretical works explore highly generalized syntactic analyses. These generalizations come to contradictory conclusions.

Contrary to Wiltschko (2009), Bischoff (2011)[2] follows Gerdts (2003) in describing Salishian lexical suffixes as being compound-like within verbs. In order to accommodate doubling, Bischoff (2011) invokes a non-configurational syntactic analysis (Hale, 1983; Jelinek, 1984), where lexical nouns are verbal adjuncts, rather than arguments. I do not know enough about the languages to assess this claim, but since I draw parallels with Wiltschko (2009) and Wao Terero in §4.2, I will raise some questions concerning the argumentation.

[2]Bischoff (2011) focuses on a different Salishian language, Coeur d'Alene. Wiltschko (2009) and Gerdts (2003) focus on Halkomelem.

Gerdts (2003) provides convincing examples that demonstrate valence effects. Her evidence that lexical suffixes are verbal arguments requires that speakers reject additional lexical noun arguments during elicitation. A non-configurational analysis should allow lexical nouns, even when verbal argument positions are filled by lexical suffixes, a point that Bischoff (2011) relies on. This would mean that between the two works, Gerdts (2003) provides configurational-style arguments for valence changing, where lexical noun arguments are blocked by lexical suffixes. Bischoff (2011) assumes compound-like lexical semantics based on works such as Gerdts (2003), but must assume a non-configurational syntax, which would not provide the evidence needed in Gerdts (2003). There are languages that have been claimed to have mixed configurational and non-configurational systems. For Blackfoot, Ritter and Sara Thomas Rosen (2010) invoke a contrast between DP and non-DP noun phrases, where an overt determiner signals argument, rather than adjunct, status. This distinction was not made in Bischoff (2011), nor would his examples support such a claim. Both determiner containing noun phrases and non-determiner containing noun phrases were claimed to serve as adjuncts. An example with a determiner can be seen in (8).

(8) *čn-šɛpəpí-ɬxʷ* *xʷɛ* *hn-cɛtxʷ*

1.NOM-finish-LS.house DET 1.GEN-house

'I finished building my house.' (Bischoff, 2011, p. 13)

Gerdts (2003) does not address doubling data such as (7) or (8) in her examples. Given this, questions of the logic in Bischoff (2011), and the documentation of multi-functionalism in Kinkade (1998), there is no compelling reason to reject the data of Wiltschko (2009). Despite theoretical debate, documented examples support the conclusion that there are *both* compound and classifier-like uses of lexical suffixes in Halkomelem.

In Wao Terero, I have so far found compound-like verbal uses of lexical suffixes to be less common, and more idiomatic, but they do exist. I would not be surprised to find some productive patterns in the future.

Despite the variety of Salishian analyses, above, none of the characterizations are incompatible with lexical suffix status. There is no restriction on the semantic or syntactic role of a lexical affix, whether derivational, compound-like, classifier-like, or otherwise.

2.1.3 Lexical Suffixes in Non-Salishian Languages

Mithun (1997) appears to be one of the first to extend the term lexical affix outside the Salishian family. She described similar affixes in Eskimoan, Tsimshian, and Wakashan language families. It is now common to use the designation for similar phenomena in a wide variety of languages. At a workshop on lexical suffixes I attended in 2024 at the *57th Annual Meeting of the Societas Linguistica Europaea*, the language families of Murrinhpatha, Chukotko-Kamchatkan, Inuit-Yupik-Unangan, Tacanan, and Sino-Tibetan were represented, as well as Wao Terero. The only published Amazonian work that I know of that uses the term explicitly is Fleck (2006) for Matses, a Panoan language. The author describes a system of mostly body-part prefixes that occur on verbs, adjectives, and nouns. Similar Panoan systems were previously described as incorporation by D. L. Payne (1987). Much of Fleck (2006) is dedicated to providing evidence that the prefixes are not synchronic incorporation.

Some systems that are not entirely lexical suffix systems may contain lexical suffixes. I believe this to be the case in some Algonquian medial systems. Medials are a class of affixes that tend to have nominal meanings that follow the root of a verb in Algonquian languages. They also occur on nouns. Medials are often divided between two types. The first is called an "incorporating" or "common" medial. The second type is called a "classificatory" medial (Denny, 1981; Drapeau and Lambert-Brétière, 2011). The incorporating medial is not assumed to be the incorporation of a free nominal, though there may be some phonological similarity. For instance, in the Plains Cree example in (9), there is a medial -*ākon*-.[3] There is no free nominal with phonological identity. The noun *kōna*, 'snow' (Wolvengrey, 2001, p. 537), does have some similarity with the medial, though across the system, like Wao Terero, this similarity is insufficient to make a case for productive incorporation based on systematic synchronic morphophonological relationships. Medial affixes are generally closed class, bound content.

[3]The AI stands for animate intransitive, a verb class of Algonquian languages.

(9) *mow-ākon-ē*

eat-snow-AI

'S/he eats snow.' (Biedny et al., 2021, p. 4)

There is a salient meaning distinction between the two medial classes, where classifier medials are expected to have *superordinate* meanings such as substance or shape (Biedny et al., 2021), rather than more concrete meanings, like 'snow'. Likewise, there is an expectation that an incorporating medial serves a compound role with the verbal root, resulting in verbal meanings like (9). The classifier medials describe or constrain some quality of an argument. "Incoporating" and "classificatory" medial categories are not complementary classes (Biedny et al., 2021). There is some overlap in properties. The pattern is common across Algonquian languages (Biedny et al., 2021).

Algonquian languages do not usually show productive noun incorporation in the medial position of a verb. According to Slavin (2012), an exception is Ojicree, where the author has documented productive noun incorporation. The incorporated nominal inhabits the same position as a bound medial. This leads to an interesting question of whether the bound medials should be considered part of an open class system in the language, or whether there are two systems serving similar functions.

The utility of the term *lexical affix* is that it describes specific qualities, while allowing for a variety of sub-phenomena. In every work I have seen that discusses lexical suffixes, there in an emphasis on multiple uses. Even attempts to provide broad scope syntactic analyses of verbal uses of Salishian suffixes acknowledged nominal uses, where the lexical suffix serves as neither an argument nor a classifier. Referring to the affix systems described above as exclusively one of classifiers, derivation, or bound element compounds would be misleading, since it obscures systematic diversity in function, which deserves explanation.

2.2 (Bound) Classifier Typologies

Classifiers are a broad category. The issues of Burmese classifiers (Burling, 1965) are quite different than the Wao Terero lexical suffix system. In the former case there are free items in an open class that only modify numerals, and are grammatically obligatory. In contrast, it appears possible to speak grammatical Wao Terero while using classifier-like lexical suffixes rarely, if at all – though lexical suffix avoidance may not be the most natural, direct way to express oneself. The classifier-like uses of lexical suffixes involve closed class, bound forms. There are specific typological issues for languages with bound, closed class classifiers. One is their relation to gender systems, where some authors have claimed that classifiers are free, and gender is bound (Dixon, 1986). The other is that many systems of bound classifiers have compound-like or derivational qualities, where they have additional, non-classificatory uses (Aikhenvald, 1994; Krasnoukhova, 2012; Rose and Van linden, 2022). Passer (2016) also provides examples of verbal classifiers being used to do things like introduce arguments without an explicit nominal, which is inconsistent with classifier status. I consider these issues in relation to Wao Terero, and briefly review influential typologies, below. First, I consider the question of agreement. Then, I consider the questions concerning behavior inconsistent with classifier status, such as derivation, compounding, adverb-like uses, or discourse referent introduction.

2.2.1 Gender and Agreement

In Wao Terero, lexical suffixes are used in classifier-like ways with verbs, adjectives, demonstratives, and numerals. I am less sure about the status of interrogatives. A classifier-like usage with non-verbal parts of speech is closer to a prototypical classifier ideal than verbal use, since the parts of speech are within the noun phrase, which is the norm for classifiers (Aikhenvald, 2000; Fedden and Corbett, 2017; Passer, 2016; Dixon, 1986). I evaluate these non-verbal parts of speech similarly to verbs due to the fact that lexical suffixes may act in non-classifier-like ways across the system. The doubling diagnostic passes for non-verbal cases. Yet, especially given their bound status, more evidence is needed to demonstrate their use should be described as classifier-like, rather than gender. Gender agreement also fails to block lexical arguments. This is not a coincidence of the doubling test. The test is designed to align incorporation, and lexical suffixes with nominal classification, rather than compounding. The test therefore seeks to find an intersection with gender-like qualities.

Gender and classifier systems are often described in terms of overlap. Authors describe a cline between the two types, which may be a diachronic relationship in some cases (Grinevald and Seifart, 2004). Authors also describe mixtures of prototypical properties found within systems, arguing for an alternative to broad comparative categories (Seifart, 2009). Fedden and Corbett (2017) collapse the distinction between gender and classifiers because, according to the authors, gender has clear characteristics, but classifiers systems do not appear to form a coherent separate category. Yet, within typologies of classifier systems the argument is consistently made that despite similar qualities, classifiers and gender are usefully considered distinct (Dixon, 1986; Grinevald, 2000; Aikhenvald, 2000; Allan, 1977; Passer, 2016). I am ambivalent on this issue. Broad comparative categories may not be the best way to move forward with an understanding of all the systems involved. Despite this, gender would be a poor description of what I am calling classifier-like uses of Wao Terero lexical suffixes. This is due to the fact that gender systems entail qualities such as syntactic agreement (Corbett, 1991).

As a starting point to discussing the literature on gender-classifier divisions, an influential contrast is provided by Dixon (1986). Aikhenvald (1994) considered the division out of date, especially from her focus on bound classifiers. Fedden and Corbett (2017) also provide a range of criticisms. Yet, similar assumptions to Dixon (1986) are evident in Passer (2016), who questions the existence of verbal classifiers. Krasnoukhova (2012) also finds it to be a relevant guide to prototypical gender and classifier qualities. For that reason, the division Dixon (1986) provides is still relevant.

The typology of Dixon (1986) encodes the idea that free numeral classifiers in isolating languages of East Asia are in typological opposition to bound gender in European languages, though both may serve similar functions. Fedden and Corbett (2017, p. 5) provide a tabular summary of the contrasts proposed by Dixon (1986), which I've reproduced in Table 2.2.

Table 2.2 is notable in defining classifiers as free lexemes. Unlike vague language such as "around 20" and "Fair number", which can be seen as guideline-like, gender is "Always a closed grammatical

	Gender	Classifiers
Size	All nouns classified	Some nouns not classified, almost always
	Small number of classes (2 to around 20)	Fair number, at least a score, with 100+ being common
	Noun-to-gender relation is one-to-one	Noun-to-classifier relation is one-to-many
Realization	Always a closed grammatical system	Always a free form
Scope	Never entirely within the noun word	Never any reference outside the NP
	Little variation between speakers	Classifier use often indicates style/mode differences
Semantics	Affix has fairly fixed meaning	Classifier is a lexeme, with greater possibilities, context of use is important

Table 2.2: Criteria for distinguishing gender and classifiers (Dixon, 1986)

system" and classifiers are "Always a free form". Amazonianists, who have a history of describing verbal systems as classifiers, which precedes Dixon (1986), such as Peeke (1968), working on Wao Terero, would obviously take issue with such a typology. The *scope* definition is also contrary to verbal classifiers, which are outside the NP of their arguments.

Fedden and Corbett (2017) present an important counter-perspective that they frame using Table 2.2. I do not dive into the specifics of the argumentation used by Fedden and Corbett (2017), since many of the details are not relevant to central questions of this chapter. Their arguments for collapsing gender and classifiers distinctions are listed below. Items reference sections of Table 2.2, when relevant.

- (Size) The total number of classes is irrelevant according to their perspective on typological categories.
- (Size) In gender systems, it is not always the case that every noun is lexically specified for a particular gender.
- (Size) In classifier systems, it is not always the case that classification is variable.
- (Realization) Realization as free forms is not in contrast to being a closed class.

- (Scope) Verbal classifiers demonstrate that classifiers may be outside the NP.
- There are languages that combine gender and classifiers, where the two systems are not orthogonal (Aikhenvald, 1994).
- There are types that are intermediate between gender and classifiers (Seifart, 2009; Seifart and D. L. Payne, 2007).

Fedden and Corbett (2017) state that some traits of gender provided in Table 2.2 are sound. They point out some necessary traits of gender systems.

- All nouns are classified.
- There is a closed class of noun classes in opposition.
- There is grammatical agreement.

Given the assumptions of Canonical Morphology (CaM) (Corbett, 2005), which the Fedden and Corbett (2017) practice, definitions are stated in terms of canonical traits. The canonical traits are used as a point of comparison with actual systems, which are expected to deviate from the canonical ideal. Even though not all classifier systems have the necessary traits of gender, they may still be subsumed under the gender category as non-canonical gender. According to the authors, gender is simply the more canonical system of noun classification.

> We therefore order the theoretical space of nominal classification by using the notion of canonical gender, following Corbett and Fedden (2016), and use "classifier" as a label of convenience for a wide range of phenomena which have been named as such by various scholars in the literature, but which are impossible to unify under a single definition. Classifier systems can then be treated as various systems of nominal classification which are typically further away in the theoretical space from canonical gender. (Fedden and Corbett, 2017, p. 6)

This approach may be compared to Seifart (2009), who describes classification in Miraña, which has a system of classification with qualities that have been associated with both gender and classifiers. He makes similar points to Fedden and Corbett (2017) about the lack of maintainability of the two large comparative categories *gender* and *classifier* existing in contrast, though he does so through a detailed description of a single language, rather than meta-analyses of written grammars. Rather than setting up one category as the canonical ideal, he proposes that large comparative categories may be counterproductive in providing meaningful typologies of classifying systems. He suggests looking at noun classification systems as collections of properties.

I find the perspective potentially powerful. As an extrapolation of the proposal of Seifart (2009), field studies may be performed using standardized specifications for the creation of language specific diagnostic instruments, which measure properties that have been considered relevant to noun classification. Rather than forcing a split between two large categories, either gender or classifier, the goal would be to place noun classification systems within a multidimensional space. The utility may be that one could encode the measurements as mathematical objects that would allow for greater comparative insight.

2.2.1.1 Wao Terero in a Typology of Gender

Ignoring the canonical perspective, Fedden and Corbett (2017) laid out three important characteristics of a gender system, which are simple and general. These characteristics do not apply in a systematic manner to Wao Terero. They are repeated below.

- All nouns are classified.
- There is a closed class of noun classes in opposition.
- There is grammatical agreement.

It is tricky saying whether all nouns are classified or not in Wao Terero, because the system is semantic and descriptive. There are certainly words like *awoto*, ‘automobile’, that have no obvious default lexical suffix correspondence. Yet, just as nominals or adjectives can be used in a descriptive manner, it is likely that there are lexical suffixes that could describe characteristics of some vehicle. I do not believe this possibility is equivalent to the grammatical classification of gender systems. See section §3.3.5.3, in particular, for examples of adjectival classifier-like usage that is often descriptive.

There is a closed class of lexical suffixes in Wao Terero, but there is no reason to characterize them as a closed class of classes. For instance, there is overlap in meanings when lexical suffixes are used as classifiers. The “classes”, in terms of meaning, are only in opposition to the extent that morphotactics enforces the use of a single affix at a time on most parts of speech. The overlap in lexical suffix meanings are described throughout Chapter 3, with a detailed example in §3.3.5.3.2.

The most important issue is grammatical agreement. Corbett and N. M. Fraser (2000, p. 293) state “Gender systems have agreement as their defining characteristic.”

In order to demonstrate syntactic agreement, one must show not only that some forms co-occur sometimes, but that there is an obligatory, *syntactically determined* relationship. It is difficult to claim that co-occurrence is determined by syntax if one can choose to either use, or not use a classifier in any syntactic context. That classifier use is not obligatory in Wao Terero can be seen in (3), elaborated with lexical nouns in (10). Each minimally paired sentence remains equally acceptable when a lexical suffix is used as a classifier on *giitã*, and when it is not used. Though arguably, the lack of a lexical suffix and lexical noun could be considered less descriptive than the other options.

(10) 

a. *giitã-ka* *eibe* *ĩ-pa*
small-CLF.stone above COP-DECL

‘The small one is above.’

b. *giitã* *eibe* *ĩ-pa*
small above COP-DECL

‘The small one is above.’

c. *giitã-ka* *di · ka* *eibe* *ĩ-pa*

small-CLF.stone ∅ · CLF.stone above COP-DECL

'The small stone is above.'

d. *giitã* *di · ka* *eibe* *ĩ-pa*

small ∅ · CLF.stone above COP-DECL

'The small stone is above.'

Lexical suffix usage is sparse. I found it difficult to elicit examples in translation tasks, or by asking for descriptions of images. In order to perform focused studies of the lexical suffixes, I often had to construct examples to ensure the affixes were present. (D. L. Payne, 1987; Kalashnikova, Oliveri, and Mattock, 2018) both point out that similar optional occurrence in Desano (Miller, 1999), another Amazonian language that is considered to have a multi-classifier system, argue against agreement.

One may ask, even if agreement is sparse, whether it may exhibit other important qualities associated with agreement, such as redundancy (Corbett and Fedden, 2016; Corbett, 2003). In agreement systems, the nominal controller should be sufficient on its own, and a morphological "echo" of its class on various targets adds no additional information. I cannot state that lexical suffixes are never used redundantly in Wao Terero. Certainly, for the purposes of elicitation, as in (10), I constructed redundant examples, which were accepted. Yet, since classifier-like lexical suffix usage is optional in Wao Terero, and there is no known context where it is required, it is reasonable to suppose that when a lexical suffix occurs, it is for specific, non-redundant reasons.

(11) *õ · be* *a · wẽ* *pa-wẽ-ta-bo-pa*

∅ · LS.territory ∅ · LS.plant cut-CLF.plant-PST-1-DECL

'In the garden, I cut a pole.'

Considering (2), repeated here as (11), the use of *-wẽ* in the verb looks redundant. There are a number of reasons to doubt this. One is the extent to which the lexical noun argument is necessary. Example (11) demonstrates classifier-like qualities in terms of argument saturation, but it does not tell one how the verb is interpreted in the absence of a lexical noun. In Wao Terero, grammatical verbal lexical suffix usage does not require a lexical noun to serve as controller (§3.3.6). The lack of a contextually salient property or entity may result in infelicity when a verbal lexical suffix occurs, but these do not require a specific previous mention in all cases. For instance, all examples in (12) were judged to be felicitous in contexts where no previous phrase explicitly mentioned a head. In the first case there is a lexical suffix and a lexical noun. In the second, there is no lexical suffix, but there is a lexical noun. In the third case, there is no lexical noun, but there is a lexical suffix. In these examples, the lexical suffix is both optional, and sufficient to express the notion that it is a head aching. There are pragmatic limitations. In (12) a head is contextually entailed. Even though *-ka* has other interpretations, which makes it vague, the context of pain entails a body-part interpretation. The paradigm in (12) does not hold for all lexical suffixes and all verbs, but such cases do exist within the diverse system.

(12) a. *dãta-ka-bo-pa* *bo-to* *okabõ*
hurt-CLF.stone-1-DECL 1-PRO head
'My head hurts.'

b. *dãta-bo-pa* *bo-to* *okabõ*
hurt-1-DECL 1-PRO head
'My head hurts.'

c. *dãta-ka-bo-pa*
hurt-CLF.stone-1-DECL
'My head hurts.'

In (12a), the repetition of 'head' is perhaps redundant, though neither the lexical noun nor the lexical suffix is clearly more redundant. There are many examples in Chapter 3 where both the lexical suffix and lexical noun make semantic contributions that are non-redundant. For instance, in (13) the lexical suffix *-bẽ*, often associated with ropes and cords, combines with the adjective for 'big', and expresses the dimension of largeness. Notably, the lexical suffixes in these instance do not match the lexical nouns, even though *-bo* and *-ka* can be used in a classifier-like manner with these nouns. The examples in (13) were constructed, but they were accepted and translated to Spanish by a native speaker.

(13) a. *yẽdẽ-bẽ* *eke · bo* *ĩ-pa*
big-CLF.cord ∅ · LS.egg COP-DECL
'It's a long egg.'

b. *yẽdẽ-bẽ* *di · ka* *ĩ-pa*
big-CLF.cord ∅ · LS.stone COP-DECL
'It's a long stone.'

The examples above are intended to argue for non-redundancy. A point of clarification is needed. I do not have evidence that lexical suffixes are involved in the conventional introduction of entities into discourse. I have no reason to believe that they are anything other than neutral with regard to definite or indefinite status of their host's arguments. I hypothesize that classifiers are not signals of discourse entity introduction or reference. When there is no overt nominal, the lexical suffixes serve as the descriptive content of what has variously been called *null-headed anaphora* (Nerbonne, Lida, and Ladusaw, 1989; Nerbonne and Mullen, 2000), or noun ellipsis (Gũnther, 2011), which is closely related to one anaphora in English (Gardiner, 2003; Khullar, Bhattacharya, and Shrivastava, 2020). See §3.3.6 for more information.

Related to the notion of non-redundancy, Passer (2016) argues that, cross-linguistically, verbal classifiers, and perhaps other bound classifiers should not be considered classifiers. Though lexical suffixes may not be blocking lexical arguments, they may have a degree of independence from their nominal controller that is inconsistent with popular notions of classification. They may even be said to signal pronoun-like qualities.

In the next section, in addition to the verbal classifier critique of Passer (2016), I consider whether it is coherent to call multi-classifier systems classifier systems. Given well-established diagnostics in the literature, they appear to be mislabeled, due to issues raised by Passer, and clear derivational, and compound uses.

2.2.2 Bound Classifiers Behaving Badly

According to the discussion so far, bound classifiers face a categorization dilemma. On the one hand, they may be too gender-like. On the other, they may be too compound-like. Yet, there may be additional issues that affect the categorization of bound classifiers. Although the doubling diagnostic provides a simple measurement, I have already discussed some complexities that may surround it. There are many ways that (supposed) bound classifiers may behave in non-classifier-like ways, even if doubling superficially succeeds. Recall that Mithun (1997) described verb-hosted lexical suffix usage where the lexical suffix meaning was unrelated to the nominal argument. Fleck (2006) described similar non-relatedness in Matses. There is also the question raised by Bischoff (2011) concerning non-configurational languages. Additionally, doubling tests only demonstrate behavior when a lexical noun is present. It does not measure behavior of the verb-classifier combination when there is no lexical noun in the clause. Further, verbs are complex, and bound forms are subject to more lexical irregularities and small patterns than is the norm in phrasal syntax. It is not certain that all classifier-verb combinations behave identically in the same language, or even that any single classifier-verb combination behaves the same in all contexts. There are reasons to question whether a verbal classifier is only a classifier, or whether it has multiple uses. Below I summarize some issues that Passer (2016) has noted in his review of the verbal classifier literature.

Related to the above, in some languages, similar to Wao Terero, the classifiers used with verbs are also used on other parts of speech. In these systems, it is common to see the same affixes used for derivation or compounding on nouns. One of the major reasons I have claimed Wao Terero's suffixes are lexical suffixes with classifier-like uses, rather than simply classifiers, is because the affixes contribute to compound-like meanings on nominals – and sometimes elsewhere. I briefly review some multi-classifier literature on the issue. I am not aware of any developed argumentation as to why derivation and compound-like behavior is considered consistent with multi-classifiers and nowhere else. Authors who describe it and do not find it inconsistent state that it is not prototypical of classifier systems (Aikhenvald, 2000; Rose and Van linden, 2022; Krasnoukhova, 2012).

2.2.2.1 Verbal Classifier Cross-linguistic Issues

Passer (2016) raises a number of questions concerning verbal classifiers. The issues he raises center on three issues. First, he argues that some systems are not primarily verbal classifier systems. Second, he argues that the affixes of some systems are adverb-like – modifying or classifying the verbal predicate, rather than the nominal arguments. Third, he argues that the classifiers of many systems, rather than being related to a nominal argument, are used for introducing arguments, and other functions. His focus is on verbal classifiers, which includes multi-classifier systems.

The argument against verbal classifiers existing in multi-classifier systems hinges on the extent to which the verbal system is primary. For the language Mundurukú, a Tupi language, he emphasizes that Aikhenvald (2000) describes the verbal system to be less productive than other host types, much like Wao Terero. He claims this implies that there is no verbal classifier system, and states that verbal uses are likely diachronically derived from more standard NP-internal systems. The author does not explain how historical status of the affixes would invalidate synchronic, classifier-like behavior within verbs, even if the behavior were more limited than is found with other hosts.

Passer (2016) also considers a number of verbal classifier languages where the classifiers appear to play an adverbial role. These include Manam, with prefixes that are consistently action oriented (according to glosses), such as *tata*-, 'throwing and breaking action'. The language Diegueño appears to have some nominal-like glosses for classifiers, such as *c*-, 'an indefinitely large number of small/soft/pliable/liquid objects', but others that reference causativity, or other predicate modifiers. Great Andamanese also has glosses that are worded in a manner that suggests verbal modification, rather than classification, such as *a*-, 'mouth-related activity, origin'. Yet, as the last gloss demonstrates, some of these might have been presented in a more noun-like manner. For instance, the last many have been glossed as 'mouth, origin', alone. Actions performed by the mouth, or to the mouth are mouth related, so it isn't clear that the gloss isn't simply verbose. Great Andamanese is also listed with glosses such as 'resultative state', which seem less likely to be related to a concrete nominal.

Relying on glosses, alone, can be misleading. In addition to issues I've already raised, there is the question of whether some affixes in the three languages may have had more than one meaning

– some adverbial, and some classifying. Despite my doubts concerning the specifics of the data, Passer (2016) makes his point that some verbal classifier systems plausibly contain affixes that serve adverbial roles, and that such uses are inconsistent with classifier definitions. The issue with Passer's broader argument is that some listed affixes also had classifier-like nominal meanings. There was a mix of different meaning types represented. Some affixes might pass diagnostics for classifier-like behavior taken individually. One cannot invalidate positive categorizations based on the existence of some negatives. In a more fine-grained approach, if a group of affixes forms a system based on shared roles, morphotactics, or other characteristics, yet there is clear diversity of function within the system, the system would be designated in a manner that neither amplified one function, such as calling all members classifiers, nor erased a function, by denying classifier status.

The bulk of the argumentation provided by Passer (2016) concerns what he calls *argument marking systems*, which differ from classifiers in a number of ways. An argument marking system may introduce an argument into discourse in the absence of a lexical noun. They may also have definite qualities. These are related to my discussion of "non-redundancy" above, though I have seen no evidence of lexical suffixes being signals of definite or indefinite status in Wao Terero. The classifier-argument marker distinction is also related to works such as Bresnan and Mchombo (1987), which contrasts the notion of an incorporated pronoun and agreement.

Classifiers are not supposed to be definite or indefinite signals. Even if classifiers are based on semantic, rather than syntactic concord, according to Passer (2016), they are expected to classify an element within the clause (Passer, 2016; Krasnoukhova, 2012). There is some room for anaphoric classifier usage, but the nominal-classifier relationship must have been established explicitly at a prior point in the discourse (Krasnoukhova, 2012). The requirement is not arbitrarily. It implies some syntactic constraints for verbal classifiers that are not put forward by authors such as Aikhenvald (2000) and Mithun (1986), but it does not only represent a preference for a more grammatical definition of classifiers. There is a consequence when one allows classifiers without overt arguments. Demonstrating that a classifier-verb combination passes a doubling test is insufficient to demonstrate that some incorporated element, or affix, cannot act in a compound-like way. It may be the case that the same classifier-verb combination could be used without a lexical noun argument in another context. Without explicit use of a nominal, it leaves open the possibility that the verbal construction is able to signal its own argument, which is compound-like.

Passer (2016) considers only one Amazonian language in the section on argument marking systems, Terêna, an Arawakan language. He presents one example that supports classifier-like usage in the language, which can be seen in (14). Passer (2016) questions whether *hi* is a classifier or noun incorporation. Although a lexical suffix must be a member of a bound, closed class that cannot be derived from another open class, that is not true of classifiers. Dixon (1986) states that classifiers should be lexemes. Passer (2016) is clearly seeking to disqualify bound, verbal classifiers without simply stating that bound or word internal status is forbidden. The idea is to demonstrate that bound classifiers violate definitions that are independent of bound status. For this reason I see no weight to the argument that incorporated elements cannot serve as classifiers.

(14) *moyó-hi-ti-raú* *úhi-ti*
dry-lxclf.grass-PROG-DEM grass-POSS
'Grass is dry.' (Butler and Ekdahl, 1979; Aikhenvald, 2000; Passer, 2016)

What a classifier shouldn't be able to do is introduce a discourse referent. The examples in (15) are intended to show indefinite usage. This is difficult to prove, since the context is absent. Yet, it raises the necessary question. In Wao Terero, contextually entailed entities, such as body-parts, do not appear to require either previous mention, or a lexical nominal – at least, not for all speakers. Some Wao Terero speakers appear comfortable with the use of lexical suffixes in verbs, adjectives, etc., without previous mention. Krasnoukhova (2012) provides a number of examples from a variety of languages in her discussion of multi-classifiers that demonstrate similar behavior. In addition to my claim that contextual entailment is important, Krasnoukhova (2012) points to the concrete meanings of some affixes. Their lack of vagueness may make an additional signal of descriptive content redundant, even if it is allowed. This seems plausible, and there may be other explanations for the behavior seen in Wao Terero, Terêna, and other languages. Given this, the assertion that the examples in (15) are indefinite is not unrealistic based on cross-linguistic data. At the very least, there is a question that is raised. If the arguments of the verbs are not embedded, where are they?

(15) a. *oye-pu'i-co-ti*

cook-CLF.round-THEME-PROG

'He is cooking (round things)'

(Butler and Ekdahl, 1979; Aikhenvald, 2000; Passer, 2016)

b. *movó-cava*

dry-CLF.branch

'It (branch) is dry' (Butler and Ekdahl, 1979; Aikhenvald, 2000; Passer, 2016)

An example of the classifier not relating to an explicit verbal argument can be seen the Terêna example (16). In that example, the affix describes the location where the food was found, not the food itself, at least according to the gloss. Another possibility is that the food is typified by its location, and is water-food. Again, the important effect of the example is more to raise the question. That some system may behave as described is completely plausible. Gerdts (2003) and Kinkade (1998) describe locative uses of lexical suffixes in Salishian languages.

(16) *neve-nó'e-co-ti* *nica*

select-CLF.liquid-PL-PROG food

'The are picking food (from the water).'

(Butler and Ekdahl, 1979; Derbyshire and D. L. Payne, 1990; Passer, 2016)

A consistent relationship between verbal argument and classifier is questionable in the Terêna data, only one of the languages considered in Passer's critique of the concept of verbal classifier. It is for the typologists to debate their terminology. I do not believe the terminology relates to natural categories. Yet, the existence of patterns like Terêna in the literature may argue for a broader application of terms such as lexical suffix, or new designations for systems that contain quasi-classifier behaviors, such as argument marking systems, so as not to be misleading. Yet, neither does it make sense to dismiss classifier-like qualities within a lexical suffix or argument marking system. If some aspect of the subsystem behaves like a classifier system, that is relevant, and is not erased by non-classifier behavior. The claim of Passer (2016) that specific examples invalidate the concept of verbal classifier is a faulty generalization. One cannot invalidate the notion of verbal classifiers by demonstrating that there is additional, non-classifier-like behavior in the system. One must comprehensively show that the positive evidence is flawed. Morpho-lexical systems are known to be subject to semi-productivity. It is entirely possible to have a set of affixes that are recognizable as a system due to distributional characteristics, which have multiple uses that may be uniform or varied within the set.

As an additional observation, even though the arguments of Passer (2016) focused on verbal classifiers, I see no reason that other types of classifiers might not exhibit similar, non-classifier-like behaviors. If a language allows classifiers with other parts of speech in the absence of a lexical noun, there are questions concerning the role the classifier is playing in those instances, as well. This likely extends beyond bound versus free status.

2.2.2.2 Derivation and Compounding in Multi-classifier Systems

In a multi-classifier system the forms of classifier elements on nominals and other parts of speech are so systematically similar in form and meaning that one must identify them as such, or provide some less parsimonious explanation for the coincidence. The issue is that despite concrete aspects of meaning being fairly consistent, within the norm of variation expected in a lexical system, there are clearly different functions for the same affixes, which often, but not always, depend on part of speech. These include nominal derivation and compound-like behavior. Krasnoukhova (2012, p. 209) states that nominal derivation plays a characteristic role in multi-classifier systems. For this reason the author prefers the term *multi-functional classifier* system, which acknowledges non-classifier roles. The term multi-classifier is only intended to emphasize multiple host types within a typology based on host or modified element, such as verbal, nominal, numeral, etc. I will not say that any language that I have not studied in detail has a lexical suffix system. Isolated examples from papers and grammars can be misleading. Yet, when affixes belong to a closed class, have concrete meanings, and are multi-function, they are very similar to lexical suffixes.

There are many issues to consider in the nominal derivation-compounding space of multi-classifiers, too many to review here. I pick two examples of word-formation and derivation below that can be compared with Wao Terero. Both examples demonstrate properties that are inconsistent with classifier usage. The first is a single example of verbal derivation in Chayahuita, which is relevant to claims that have been made about Wao Terero. The second demonstrates how Tariana mixes compound-like derivation and grammatical agreement. This is to compare to Wao Terero, which has very similar word-formation dynamics, but does not have agreement. Due to the fact that Tariana's agreement extends to utilizing compounds, it allows for a brief discussion of how I use terms such as compound, derivation, and compound-like, which are intended to emphasize slightly different properties.

I find verbal derivation to be especially contrary to the notion of classifier status. This is because deriving a noun from a verb by means of a suffix with a nominal meaning may embed an argument. Such derivation is explicitly contrary to the goals of doubling tests. D. L. Payne (1987) provides an example from Chayahuita. An example from the author, presented in (17), plausibly embeds an argument.

(17) *shipi-ro'*
get.wet-CLF.earth
'mud' (D. L. Payne, 1987, p. 33)

D. L. Payne (1987) stated that Wao Terero (referred to as Waurani) was exactly the same as Chayahuita in this regard, so much so that it was not necessary to provide separate Wao Terero examples. Peeke (1968) provided some example nominalizations in a table-like listing of (what I call) lexical suffixes. I have reproduced the listing in Appendix D. Despite this, I have see no evidence of lexical suffixes being used for productive deverbalization. I discuss the issue in §3.3.5.2.4. There is evidence of lexical suffix use with deverbal nouns, but there are other deverbal mechanisms at play in the language, which make it ambiguous whether the classifier is playing the role of deverbalizer in Peeke's examples. Lexical suffixes have likely served argument-like functions in deverbalized nouns in the past. I do not have evidence one way or the other for modern speakers.

When classifiers are used for nominal derivation, they may have class changing effects in agreement systems. In Tariana (Aikhenvald, 1994), every derivational affix may be used as a classifier and there is also classifier NI. This is a nice use of terminology in some sense. Rather than stating that all the derivational affixes are classifiers, it is stated that they may be used as classifiers. As far as I am able to determine, the majority of the affixes in question have plausibly lexical meanings, with some examples being 'swamp', or 'trunk'. They do not include certain nominalizing, diminutive, and other derivational affixes that exist in the system. I saw no mention of valence changing affixes.

The derivational suffixes in Tariana have some optional uses. In addition to adding descriptive information, Aikhenvald (1994, p. 437) states that when present they emphasize a "singular character". For instance an affix associated with roundness, *-da*, may be added to a stem for 'his eye', *dithi*. This emphasizes that it is a single eye 'his (one) eye', *dithi-da*. Except for this notion of singular character, this is similar to what Rose and Van linden (2022) call a qualifying binominal combination, where a classifier emphasizes some aspect of a nominal. Qualifying combinations describe some quality of the referent, which may or may not be typical. The roundness of the eye be essentially intrinsic, but a round stone may be descriptive. Either may be considered qualifying. Qualifying is contrasted with derivational combinations, which show more compound-like combinations used for word-formation. Tariana derivational affixes are also used for word-formation. It is unclear how transparent this may be. The examples appear in (18).

(18) a. *heku-da*
tree-CLF.round
'fruit'

b. *episi-da*
iron-CLF.round
'motor' (Aikhenvald, 1994, p. 437)

Notably, Wao Terero does not appear to allow the more abstract meaning associated with classifying lexical suffixes in nominal constructions. Therefore, according to the dichotomy described by Rose and Van linden (2022), there are derivational, but not qualifying combinations.

An interesting aspect of the Tariana system is how it interacts with agreement. Derivational affixes are often used as agreement markers. Aikhenvald is not entirely clear on what *often* means in this context, but provides examples that demonstrate agreement, which is obligatory for adjectives. One of these is provided in (19).[4]

(19) *panisi-way* *hanu-way*
house-AFF.side big-AFF.side
'a big wall of the house' (Aikhenvald, 1994, p. 438)

Nominal uses of derivational affixes are not limited to a single affix. When there is more than one, "the word-final one is more likely to be used as a classifier" (Aikhenvald, 1994, p. 438).

Though Wao Terero does not exhibit grammatical agreement, there are similar instances where there is more than one lexical suffix on a nominal argument to a verb. Verbs, when they allow lexical suffixes, do not allow such aggregates, as is the case for Tariana adjectives. The verbal lexical suffix that is allowed, in such cases may be the final affix, but it may not be. For instance, in (20) it is the *-ka* from *o · ka-gi*, 'hair', that is used in the verb, rather than *-gi*. This is not necessarily the norm. It likely has less to do with order and more to do with verbs allowing only specific body-part affixes. *-Ka* is associated with heads.

(20) *kĩdãte* *o · ka-gi* *to-ka-bi*
why ∅ · LS.stone-LS.string cut-CLF.stone-2
'Why did you cut your hair?'

The combinations used in derivational word formation in Tariana have compound-like meanings in, at least, some cases. Notably, true noun-noun compounds behave similarly to the derivational combinations. The second element of the compound is used on the adjectival agreement target, as seen in (21).

[4] In (19), the gloss AFF was used instead of a classifier gloss, but the author described it as a derivational affix.

(21) *kuphe-yeri hanu-yeri*

fish-basket big-basket

'a big basket of fish' (Aikhenvald, 1994, p. 438)

The derivational affixes in Tariana are not forming true compounds. Affix-like elements do not form compounds. Yet, it is interesting to see converging uses of bound affixes with lexical meanings and true nominals. It is reminiscent of the medial system of Ojicree, mentioned above, were medials share a verbal position with incorporated nouns.

Though I would not argue against the use of the term derivation for the Tariana system, I tend to use compound-like for similar phenomena. I use compound-like due to the doubling diagnostic and the lexical semantics of nominal word-formation with lexical suffixes. I think that it is fair, in a typology where affixation is either inflection or derivation, to call the Tariana system derivation. In this work, I tend to reserve derivation for cases such a nominalization, valence changing, etc., since nominal-like lexical meaning combination is qualitatively different from these more functional operations. I do not claim that this reflects a coherent typology, but in terms of formal theory, dealing with causativity is quite different from dealing with lexical suffix-based word-formation. For that reason, I tend to describe lexical suffix combinations that have compound-like lexical semantics as compound-like. One is free to call it derivation that results in compound-like meanings.

2.3 Conclusion

The core reason for referring to the Wao Terero system as a lexical suffix system is that it is completely consistent with the term. In fact, Wao Terero is not only consistent, but fairly typical. The general functions of the affixes are similar to other lexical suffix systems. In contrast, multi-classifier systems are atypical within their categorization. There are clearly authors who would say that such systems are not classifier systems at all (Passer, 2016). For that reason, even if Wao Terero is typical of multi-classifier systems, multi-classifiers are not typical classifiers. Perhaps if there was some reason for saying that the definition of classifier should be shaken up in the way that multi-classifiers shake things up, it would be good to push in that direction. Instead, the concept of multi-classifiers appears to dismiss foundational diagnostics. As I said above, my primary concern is to not be misleading. The choice of categorization is clear.

At the end of this section, the reader should be familiar with cross-linguistic attributes of lexical suffixes, and questions that arise in such systems. I also provided information relevant to the gender-classifier dichotomy. There are interesting systems within that dichotomy, but many of their attributes are not immediately relevant to Wao Terero, which does not exhibit agreement within its lexical suffix system. I explored some additional questions concerning the lexical semantic and discourse roles of classifier systems. There may be reason to believe that even many of the superficially classifier-like behaviors in the Wao Terero system do not exhibit a grammatical relationship with a nominal argument. There is also reason to believe that many verbal classifier systems are similar. Yet, there is also no reason to assume that just because some constructions are not classifier-like that none are. Finally, I provided some word-formation and derivation examples from other systems where affixes with lexical meanings have classifier, even agreement usage, but also function in ways that are inconsistent with the term classifier.

This will hopefully have helped situate the reader in the typological context of this study. The next chapter goes deeper into the data and provides a basic description of Wao Terero phonology, inflection and the lexical suffixes.

Chapter 3

A Description of Wao Terero and its Lexical Suffixes

I introduced features of the Wao Terero lexical suffix system in the last chapter. The purpose of this chapter is to go in depth. It serves as a reference, especially when used with the index of glosses in Appendix E. The chapter begins with a description of Wao Terero phonology and orthography. Following the discussion of the sound system and writing, I provide an overview of some topics in Wao Terero morphosyntax. I then describe the lexical suffix system, which includes information on most affixes in Peeke (1968), as well as some that have not been listed in previous sources. There is a description of how the lexical suffixes are used with each part of speech, and a section on potential anaphoric properties.

3.1 Phonology and Orthography

Wao Terero data in this thesis is provided using a semi-phonemic transcription. This section provides some insight into the evidence for these phonemic categories and their relation to actual pronunciation and various orthographies. The section also provides diagnostics for free versus bound content. It should be noted that there is no standardized orthography for Wao Terero.

In claiming that the transcription is phonemic, I am not claiming to have executed an adequate phonological analysis of the language. The truth is, there are many mysteries in the phonology of Wao Terero and there have been no focused phonological studies since Saint and K. L. Pike (1962).[1] I favor my transcriptions for their morphological utility. With my transcriptions, one can predict how a Wao Terero person would write and pronounce a word while preserving an orthographic identity for morphs. For instance, [epẽ], 'water', and [mãnĩmẽ], 'this (liquid)', both end in /pẽ/, which is a lexical suffix associated with liquids.[2] These are transcribed as *e · pẽ* and *bãdĩ-pẽ* in my examples so that the morphological identity is recognizable.

Despite the goal of adding transparency to the morphological description through the use of my particular transcription conventions, various unknowns lurk beneath the surface. For instance, I made a choice in my transcriptions to accept that two vowels that had previously been documented as /e/ and /æ/, with nasal variants, are effectively merged. If I am wrong about this, my transcriptions may effectively be hiding meaningful phonological contrasts.

I use notions of "underlying" and "surface" form in this section only for descriptive utility. I am agnostic concerning phonological theories and only wish to make the patterns evident to the reader.

3.1.1 Prosody and Meter

There is little written on prosody and meter in Wao Terero. K. L. Pike (1964) wrote an influential paper on Wao Terero stress patterns that dealt with so-called *bidirectional stress trains.* The claim was that there are two patterns of stress alternations in Wao Terero. The first initiates near the end of the word and progresses toward the beginning, such that the final syllable is unstressed, the penultimate stressed, etc., in a pattern similar to ...CaCáCa#. The second progresses from the beginning of the word, with stress on the initial syllable, ex. #CáCaCá.... There are various subrules and exceptions but the result is that the two patterns may or may not be "in phase". If they are in phase, the final syllable would not be predicted to be stressed due to the forward stress

[1]There are two later works with some phonological content, Yeti Caiga (2012) and Yeti Caiga and Tocari Ahua (2012). I do not use data from those sources in this section. They should be referenced with care due to some important issues. For instance, Wao Terero clearly has nasal vowels, but the authors of the works failed to recognize them.

[2]I discuss this affix later in light of a possible vowel merger. The proper transcription prior to the merger was /pæ̃/.

train, nor would the initial syllable be predicted to be unstressed due to the backward stress train. If they are not in phase, various rules of juncture govern stress patterns. Stress trains have been of interest to phonological theorists. The Wao Terero examples from Pike have proven popular in works that seek to demonstrate the viability of phonological theories and formalisms. Fitzgerald (2000) is one example and contains citations to others.[3]

The details of bidirectional stress trains are not particularly relevant here. Neither do I know whether K. L. Pike (1964) is a good description of modern Wao Terero.

Beyond K. L. Pike (1964), little has been written about prosody and meter. Saint and K. L. Pike (1962), the only focused work on Wao Terero phonemics, refers to a version of K. L. Pike (1964) that had not yet been published and offers no further description. Fawcett (2018) contains some basic information on the topic, but only as it relates to ideophones in the language. Peeke (1979) also contains some information intended to aid language learners in correct pronunciation.

Wao Terero follows a CV pattern, which allows vowel sequences. I have observed no onset consonant clusters, and none have been documented in the past. I have observed unambiguous codas only with ideophones, sound effect-like items, consistent with Fawcett (2018). Despite this, there is some evidence that codas may exist in the language. Modern speakers will often write a final 'n' on some words. Rather than indicating a coda, this is thought to indicate that the final vowel in a word is nasal (Connie Dickinson p.c.). This is reasonable, and very likely correct. There is no direct orthographic distinction between oral and nasal vowels when using the most popular writing conventions. For instance, both oral and nasal unrounded low back vowels are written 'a'. Word initially, the preceding consonant is often a cue to the nasality of the first vowel. Word internally, the orthographic conventions for the consonant following a vowel almost always makes it clear where nasal vowels occur. Word finally, there may not be any orthographic cue. This results in ambiguity, which Wao Terero speakers may attempt to resolve orthographically. For instance, the distal demonstrative *ĩ* is nasal but according to common writing conventions appears as *i*. This may motivate some speakers to write it as *in* at times. This is one hypothesis. Another is that the final 'n' indicates that codas have existed, or are currently entering the phonology. According to

[3]The popularity of the example in the literature is obscured by the fact that authors citing K. L. Pike (1964) refer to Wao Terero using the slur *auca*.

this hypothesis, native speakers are using the final 'n' as they would when writing in Spanish. That is to say that the symbol maps to a conceptual coda in both cases. The two hypotheses concerning what a final 'n' indicates are not necessarily mutually exclusive if there is some change in progress. Neither are they incompatible if some items have long had final codas while other others do not. The final 'n' could have a dual use.

Pre-nasalization, which is discussed below, may be a case where word internal syllable codas are developing. Orthographically, pre-nasalization is represented by an 'n' or 'm' that follows a nasal vowel and precedes a voiceless stop. Given the bilingualism of Wao Terero speakers and their knowledge of Spanish conventions, one might question whether they might draw parallels between the writing systems. The analogies made between writing systems could affect Wao Terero speech. Therefore, it is an open question whether a nasal consonant preceding a voiceless stop in writing represents a consonant coda for Spanish influenced modern speakers. In short, I do not believe that it can be said with certainty that Wao Terero syllables have no codas.

I have nothing further to say about prosodic, metrical patterns. It is an area that requires some attention.

3.1.2 Vowels

Each vowel category, based on height, advancement and rounding, in Wao Terero has a phonemic oral and nasal version. I am aware of only one possible phonological rule that alters the nasal status of a vowel, which can be seen in vowel sequences of ṼV. I will discuss this in §3.1.2.2. Otherwise, nasal vowels drive much of the allophony seen in consonants.

I do not believe that there has been enough modern work on the vowels of Wao Terero to perform a fine-grained characterization of their height and advancement. The descriptions of earlier Summer Institute of Linguistics (SIL) linguists do not completely match my experience. I think it is highly likely that younger speakers all have a Spanish accent, so to speak. This is reflected in how I transcribe the vowels in the IPA.

Table 3.1 lists the vowels of the language. The first column represents how I write the vowels in my morphological transcriptions. These match well with popular orthographic conventions.

Morph. Trans.	Modern IPA	SIL	Description
i	i	i	high front unrounded
o	o	ɨ o	mid back rounded
e	e	ɪ ɩ	mid front unrounded
a	ɑ	ɑ a	low back unrounded
e	e (æ)	æ	mid (low) front unrounded

Table 3.1: Wao Terero vowel categories and how they have been transcribed. The first column is the convention I use in my morphological transcriptions. The second column is how I transcribe the vowels of modern speakers in the IPA. The third column contains transcriptions used by SIL linguists. The last column is a description of the vowel.

The second column is how these correspond to IPA transcriptions. I place 'æ' in parentheses in that column because it appears to have merged with /e/. The column labeled SIL provides some common transcription conventions used by SIL linguists. The use of 'a' was apparently a typographic convenience. The use of 'ɩ' follows Pike's Americanist conventions. The description in the last column matches what I have in the IPA column. I place the word 'low' in parentheses to match the 'æ'.

The /i/ and /ɑ/ vowels need little commentary. They are described in a very similar manner through all works that offer any description of the vowels. They should match the reader's expectations given normal IPA usage cross-linguistically.

The e was transcribed as /ɩ/ in Pike's Americanist influenced notation (Peeke, 1979; Kelley, 1988). This should correspond roughly to IPA /ɪ/ and was transcribed as /I/ in Saint and K. L. Pike (1962). The description provided by those authors is of a higher vowel than I believe currently exists in the language. Saint and K. L. Pike (1962) says it was easily confused with /i/. In my own experience they are not difficult to distinguish. There are environments where the vowel raises to [i], but those are predictable. As the second element of a vowel sequence, I sometimes have difficulty telling the two vowels apart. The sequence /oe/ and /oi/ both sound similar to the English diphthong in *boy* to my foreign ear. This perceptual bias likely plays a role in my perception of the /e/ vowel, which, to me, sounds similar to Spanish /e/ or English /ɛ/ in the speech of my consultants. It is not surprising that my perceptions differ from what SIL described. I work with

Spanish-Wao Terero bilinguals between the ages of 19 and 50. SIL linguists worked with a much different population at the time of contact in the 1960s.

The /o/ was described as often unrounded, higher and more central by some SIL linguists. Saint and K. L. Pike (1962) gave the vowel as /o/. Later SIL linguists, while consistently maintaining an orthographic 'o', provided phonemic transcriptions using an equivalent of IPA /ɨ/ in some works (Peeke, 1979; Kelley, 1988). The vowel sounds much like a Spanish /o/ among my consultants.

I save the majority of the discussion of /æ/ for §3.1.2.1.

As noted, all vowels have oral and nasal versions. Saint and K. L. Pike (1962) noted that vowel nasalization is not as strong as in a language like French. This is consistent with my own observations. By ear, it can be difficult to tell if a vowel is nasal without listening to its effects on surrounding consonants.

3.1.2.1 Vowel Merger

There is a likely vowel merger between the vowels /e/ and /æ/. Though I believe the vowels have completely or partially merged, there are a number of reasons for doubt.

Words that were documented as having /æ/ by SIL linguists are spelled with 'e' in common orthographies by all modern speakers I have worked with. Speakers also use 'e' for what SIL linguists transcribed as /e/. All linguists who have worked on the language since Peeke have relied on native speakers to perform transcriptions of video and audio sources. For this reason, there is some uncertainty whether the orthography reflects a merger or not. It may simply be that the use of 'e', which has an obvious advantage for instruction in a Spanish-speaking country, is disguising the contrast.

Saint and K. L. Pike (1962) provided only one minimal pair for oral /e/ versus /æ/. It can be elicited from modern speakers. It is /pekã/, 'he calls', and /pækã/, 'he grows'. Even then, the 'call' word cannot easily be elicited in isolation but must be preceded by *aa* (meaning unknown) or possibly some other adverb-like item. I have not had any success eliciting the nasal pair the authors provided. All other pairs that I have managed to find in word lists involve a difference in nasalization, as well. They are not oral to oral or nasal to nasal. I have found no further potential pairs in my study.

The search for minimal pairs does not follow the line of reasoning one imagines in fieldwork. One would normally note that there were vowel sounds that differed, and then look at whether they were allophones of different phonemes or not. In the case of /e/ and /æ/, I have not met anyone who can clearly distinguish between the vowels, either researchers or native speakers. This does not mean the contrast does not exist. In English, it takes training to distinguish /ð/ and /θ/, so the fact that native speakers do not seem aware of a vowel distinction does not mean that it is not there. The number of linguists that have been in contact with the language and noted no clear difference between the vowels is at least 6, including myself. Given that the two vowels were at one time described to be as distinct as /ɪ/ and /æ/, or even /ɪ/ and /a/, it is notable that none of these researchers perceived a difference.

There is some ongoing work to make acoustic measurements. Preliminary measurements performed by Stephanie Antetomaso (p.c.) for her MA Thesis in progress showed no clearly distinct clusters based on F1 and F2 values that would distinguish the vowels, though her sample size was small, and vowel categories are often difficult to reliably measure. The most one can take from the measures is that the categories, if they exist, are not as distinct as /ɑ/ and /i/, which is no surprise.

I believe it is nearly certain that if a linguist were to step in from day one and have no knowledge of the SIL literature, they would not think to investigate a contrast. Despite this, there are reasons to suspect that the vowels have not completely merged. There are a number of morphophonological phenomena that could have purely phonological explanations if the vowel distinction exists. One of these will be described in §3.1.3.5. It involves differences in nasalization patterns for specific lexical suffixes that may be explained by contrasting vowel categories. Another piece of evidence concerns vowel raising in stems. As can be seen in Table 3.2, some stems raise their vowels when an affix contains /i/. As far as I know, the affix must have the /i/ in the first syllable. I have not seen vowel raising for the common suffix *-bõdi*, '1.PL', for instance, which has an intervening syllable with another vowel.

As can be seen in Table 3.2, the stems may contain oral or nasal vowels when raising occurs.

	1	2	
live	[kẽwẽmo]	[kĩwĩmi]	/ẽ/
eat	[kẽmo]	[kẽmi]	/æ̃/
do	[kebo]	[kebi]	/æ/
	3	2	
call	[pekã]	[pibi]	/e/
	3	INF	
grow	[pekã]	[pekĩ]	/æ/
drink	[bekã]	[bikĩ]	/e/

Table 3.2: Some verb stems raise their vowels when a CV affix contains /i/ or /ĩ/, as can be seen for 'live', 'call' and 'drink'. The last column contains the vowel that would be predicted to be in the stem based on SIL sources.

The first three rows use the first person and the second person as examples to demonstrate the contrast. These pairs didn't exist in the corpus for all forms. The minimal pairs for 'call' and 'grow' were not found in the corpus with the same inflections. I found the third-person form for both but needed to use the infinitive for 'grow'. I included 'drink' to demonstrate that the infinitive ending produces the vowel raising. The pattern corresponds to the difference in vowels that the SIL linguists documented for these items, which is demonstrated in the final column.

The table looks fairly neat, but speakers sometimes have uncertainty about these items. I have examples of *bekĩ*, 'drink'. I have examples such as *tedikĩ*, 'speak', which would be predicted as [tæɾækĩ] given SIL sources. These variants occur inconsistently, but are not, necessarily, corrected when presented to the speaker later. Plausibly, there are now post-merger form classes. Just as English speakers may exhibit variability for words such as LEAP in the past tense (*leaped* or *leapt*) competing patterns may now be at play for Wao Terero speakers that are not following a strictly phonological rule.

In terms of deciding whether to use as a working hypothesis *merged* or *non-merged* vowels, I made a practical choice. I have no reliable means of measuring whether a vowel is (was) /e/ or /æ/ in all environments. I cannot perceive the difference, myself. I know of no one I can ask to tell me which vowel is which in a word that hasn't been documented, yet. Therefore, there is an opaqueness and a categorical difference that only reveals itself through morphological diagnostics. For this reason, I favor a morphological analysis of the raising behavior, and treat the vowels as

	i	e	æ	ɑ	o	ĩ	ẽ	æ̃	ɑ̃	õ
i	ii			iɑ				iæ̃		
e		ee		eɑ			eẽ			eõ
æ	æi		ææ		æo			ææ̃		
ɑ	ɑi			ɑɑ	ɑo	ɑĩ			ɑɑ̃	ɑõ
o	oi				oo	oĩ				oõ
ĩ	ĩi					ĩĩ		ĩæ̃		ĩõ
ẽ							ẽẽ			ẽõ
æ̃	æ̃i		æ̃æ			æ̃ĩ		æ̃æ̃		
ɑ̃	ɑ̃i			ɑ̃ɑ		ɑ̃ĩ			ɑ̃ɑ̃	ɑ̃õ
õ	õi				õo	õĩ				õõ

Table 3.3: A chart of documented two vowel clusters in Wao Terero from Saint and K. L. Pike (1962, p. 22).

merged. Yet, I have no horse in the race. If it is eventually found that the vowels are not merged, it has no bearing on the central questions of my thesis.

3.1.2.2 Vowel Sequences

Vowel sequences are another tricky aspect of Wao Terero. Saint and K. L. Pike (1962, p. 22) provided a chart of documented two-vowel clusters within words. I am reproducing that chart here in Table 3.3. I believe there are some inaccuracies and missing information, but it should provide the reader with an initial reference point. What I would like the reader to get from the table is simply that there are many possible combinations.

It is likely that clusters are not restricted to two vowels. Saint and K. L. Pike (1962) claim that a cluster of up to four vowels may occur in a word. They provide /tõõõõba/, 'he blows straight', as an example. This is likely two words. There are adverb-like elements that often precede verbs. I have instances of *wo õkĩ*, 'to blow (a blowgun)', where plausibly the initial 'õ' of the second word is 'õõ'. Either way, *õ(õ)kĩ* is clearly the same verb as used in Saint and K. L. Pike (1962). The final *ba* on their example is likely the suffix *-pa*, 'DECL'. The *-kĩ*, 'INF', affix in the example I found in my corpus is used by Wao when translating from the Spanish infinitive, but has a variety of uses. Adverb-like elements such as *wo* and *tõõ* tend to provide information about how the action was performed. Sometimes the elements are fairly general. The adverb-like element *dao* can be used with more than one general motion verb to indicate that the motion was performed by walking.

In some cases the elements may be highly lexicalized and verb-specific. The verb *pekĩ*, ‘to call’, is usually preceded by *aa*, and even though the same element shows up in other contexts, I have failed to determine any independent meaning it provides, yet. The elements are not bound because they do not partake in word-internal phonological processes with the following verb. They are also sometimes the target of expressive morphology, such as repeating, which does not conventionally occur word internally. They share a syntactic super-category with conventionalized ideophones, and sometimes have sound-effect-like qualities (Fawcett, 2018). All of this is to say that this particular example of a four vowel cluster is likely an error.

There are reasons to believe three vowel clusters exist. There are suffixes that begin with or are entirely vowels, such as the imperative /ẽ/ or the plural-like affix /idi/. Lexical items are known to end in two vowel clusters. This means that word final clusters of three vowels may occur, but I have not looked carefully at the phonology of any particular instance.

Notably in Table 3.3, one can see that all vowels may pair with themselves. Additionally, there are clusters of identical vowel categories that differ only in nasality, such as /ɑɑ̃/.

One thing that is likely inconsistent with modern Wao Terero is that the /ẽ/ column of Table 3.3 should be full, due to the fact that the imperative *-ẽ* can occur on most, if not all, verbs, and verb stems end in vowels. The imperative is given in Peeke (1968) as /i/, which explains the nearly full /i/ column.

There is some evidence that a preceding nasal causes a following oral to behave like a nasal or nasalize. This can be seen with the affix *edẽ*, ‘DESI’. With *go*, ‘go’, the affix results in [goedẽ], ‘want to go’. With *põ*, ‘come’, the affix results in [põenẽ], ‘want to come’. The /d/ should not go to [n], unless preceded by a nasal vowel. Either the initial nasality of the cluster is determining the consonant nasalization, or the second vowel in the cluster has become nasalized by phonological rule.

Notably, the opposite is not the case. The word for ‘morning’, [bɑɑ̃neke], would be expected to have an initial [m], if the following vowel were a nasal.

3.1.3 Consonants

The consonant phonemes of Wao Terero can be seen in Table 3.4. There is no reason to propose phoneme categories for nasal consonants. The occurrence of a nasal can be predicted based on the nasal vowels in its environment.

	Bilabial	Alveolar	Palatal	Velar
Stop	p b	t d		k g
Approximant	w		j	w

Table 3.4: The consonant phonemes of Wao Terero

Wao Terero consonant phonemes have nasalization patterns when in the environment of nasal vowels. This may result in the consonant surfacing as a nasal stop or in pre-nasalization. Only /w/ shows no prominent nasalization pattern. It never surfaces as a nasal stop or receives notable pre-nasalization. There are pairs such as [mẽ^{m}po] and [wẽ^{m}po] that both mean ‘father’, but these are reported by some to contrast in whether one is speaking of their own father or the father of someone else. There is no consistent pattern such that any word that begins with /w/ followed by a nasal vowel has both [m] and [w] variants.

Past authors have proposed a phonemic /m/, including fairly recent works (Fawcett, 2023). I believe there is sufficient evidence to predict the distribution of [m], at this time.

3.1.3.1 Pre-Nasalization of Voiceless Stops after Nasal Vowels

Consonants in Wao Terero have nasalization patterns when the preceding vowel is nasal. The first example of this is expressed by the rule in (I).

$$\underset{\circ}{C} \rightarrow {}^{n}\underset{\circ}{C} / \tilde{V}\_\_ \qquad \text{(I)}$$

This states that a voiceless consonant receives pre-nasalization when preceded by a nasal vowel. This can be seen when the affixes *-po*, ‘LS.canoe’; *-ta*, ‘LS.shell’; or *-ka*, ‘LS.stone’ are affixed to a stem like *bãdĩ*, ‘this’, where the stem ends in a nasal. The surface form of /bãdĩ/ is [mãnĩ]. With the affixes it is [mãnĩ^{m}po], [mãnĩ^{n}ta], and [mãnĩ^{ŋ}ka]. In the common orthography of my consultants these are usually written *manimpo*, *maninta*, and *maninka*.

Peeke (1979) states that the voiceless consonant may also become voiced in these cases. Then one might optionally get [mãnĩᵐbo] from /bãdĩpo/. This seems reasonable. I wouldn't be surprise if measurements of voicing patterns eventually provide support for the claim. Though, there is a complication, which is that voiceless stops surface as voiced oral stops in some case when between nasal vowels. This is discussed in §3.1.3.6. In the case of the stop being between nasals, I believe that there is variation between it surfacing as a nasal or surfacing as a voiced oral stop, with the latter being more common. I do not know whether Peeke's observations took these differing environments into account. There is also the possibility that in some cases the voiceless stop between nasals does not surface as voiced or nasal, but remains voiceless with prenasalization. For that reason, I decided not to add an oral V following the consonant in (I).

3.1.3.2 Voiced Stops to Nasal Stops after Nasal Vowels

$$\begin{bmatrix} +\text{voice} \\ -\text{nasal} \\ +\text{oclusive} \end{bmatrix} \rightarrow \text{N}/\tilde{\text{V}}\_\_ \qquad \text{(II)}$$

The next example of how a preceding nasal affects the following vowel can be seen in rule (II). This is a rule for the voiced stops, which become nasals when preceded by nasal vowels. This can be seen with the first person affix *-bo*, which is [bo] on a non-nasal stem, as in [gobo], 'I go', and [mo] on a nasal stem, as in [põmo], 'I come'. It can be seen with the locative *-de*, which on a non-nasal stem, is [de], as in [ebode], 'in a plane', but on a nasal stem is [ne], [epẽne], 'at a river'. The word for tooth is [bɑgɑ] but the *ga* ending, when used as a lexical suffix on the proximal demonstrative is [ŋɑ], [mɑ̃nĩŋɑ], 'this (tooth)'.

3.1.3.3 Palatal Approximant to Palatal Nasal Stop after Nasal Vowels

The approximant /j/ becomes a palatal nasal [ɲ] when it is nasalized. This can be seen with *-yabo*, 'LS.leaf$_1$'. On the word for the leaf of *eugenia stipitata* (*arazá* in Ecuadorian Spanish), with the oral-ending stem /bĩgikɑ/, the affix has a surface form [yɑbo], [mĩŋikɑyɑbo], 'arazá leaf'. On the nasal-ending stem of the word for the manioc plant, /kẽwẽ/, it is [ɲɑbo], [kẽwẽɲɑbo], 'manioc leaf'.

$$j \rightarrow \text{ɲ}/\tilde{V}\_\_ \tag{III}$$

3.1.3.4 Non-velar Voiced Consonants Become Nasals Word Initially Before Nasal Vowels

Consonants may also nasalize in other environments. Word-initial non-velar voiced consonants go to their nasal variant when followed by a nasal vowel.

$$\begin{bmatrix} +\text{voice} \\ -\text{nasal} \\ -\text{velar} \end{bmatrix} \rightarrow N/\#\_\_\tilde{V} \tag{IV}$$

3.1.3.5 Voiced Bilabial Stop to Bilabial Nasal Before Some Nasal Vowels

The voiced bilabial stop becomes a voiced bilabial nasal when *preceding* non-low nasal vowels, as well. The rule is provided in (V) (Peeke, 1979).

$$b \rightarrow m/V\_\_ \begin{bmatrix} +\text{vowel} \\ +\text{nasal} \\ -\text{low} \end{bmatrix} \tag{V}$$

The low vowels were those that Saint and K. L. Pike (1962), and later SIL authors transcribed as ‘ɑ’ and ‘æ’. That the voiced bilabial stop surfaces as a nasal when preceding /ĩ/, and /õ/ is highly consistent. For example, it surfaces as [m] in the first and second person plural affixes [mõni] and [mĩni], respectively. That it surfaced as [m] before /ẽ/ can be seen in words like [bɑmẽ], ‘bone’, which is specifically for longer bones. There are two reasons that the /ɑ/ in /bɑbẽ/ cannot be nasal. First, there is another bone word for shorter bones that surfaces as [bɑye]. If the /ɑ/ were nasal, one would expect the approximant to nasalize. Second, one would also expect that the word-initial /b/ would nasalize if /ɑ/ was nasal due to the rule in (IV).

Further support for the accuracy of (V) comes from the low vowel. There are clearly cases where a non-word-initial /b/ does not nasalize before /ɑ̃/, such as /ebɑ̃do/, ‘how’ where the low vowel does not trigger a surface [m], [ebɑ̃no].

It is possible to find surface instances of [b] before nasal high vowels. I have found no instances where the underlying bilabial is not convincingly /p/. Though, I have at times needed to work to convince myself. This is relevant to the /e/-/æ/ merger. One would not predict the rule to apply for /bæ̃/ sequences. If instances of /b/, surfacing as [b], clearly precede what a Wao Terero speaker has written as *e*, and that *e* clearly corresponds to a nasal vowel, it results in two possibilities. First, the vowel merger may not be complete. Alternately, it may be an indication that the merger has resulted in a phonemic /m/, due to /b/ nasalization no longer being phonologically predictable. The latter would require a revision of (V), to reflect that fact. Therefore, much hinges on the correct analyses in cases where there is a surface [bẽ] (or [bæ̃]). Unfortunately, the facts are quite delicate. I have found only one example of a form that is transcribed with a [bẽ/æ̃] sequence, written *toobente*, [toõbẽ^{n}te], 'braking arm'.

For the 'braking arm' example, I tentatively favor an analysis where [bẽ] is underlyingly /pẽ/, which is a suffix for 'arm'. As described in §3.1.3.6, there is variability in whether /p/ surfaces as [m] or [b] between nasal vowels. In this case, I believe it is surfacing as [b].

An issue is that there is variation in the previous literature concerning the phonological form of the 'arm' affix. Consistent with my analysis, Fiddler (2011) has the arm suffix listed as /pæ̃/. Yet, the same suffix is given as /bæ̃/ in Peeke (1979). Both analyses were due to Peeke, since Fiddler borrowed her lists from a later work that was never published.

That Peeke reconsidered her initial analysis may be seen as providing weight to my stance. Despite this, I have never seen a surfacing [pæ̃] or [pẽ] for the arm suffix. Therefore, a phonemic /m/ analysis remains a possibility.

Though I lack key pieces of data at the time of writing, the issue is trivial to resolve. For instance, there is a verb *dãta*, 'to ache', which takes body-part affixes, and ends in a oral vowel. The arm affix should surface as [pẽ] when used with that verb, otherwise there is a phonemic /m/.

3.1.3.6 Voiceless Stop Nasalization Between Nasal Vowels

A voiceless consonant will sometimes become a nasal when between two nasal vowels. Sometimes the result is only voicing. There is evidence for instances where neither change takes place, which results in only pre-nasalization.

$$\text{C̥} \rightarrow \text{N}|\text{C̬}/\tilde{\text{V}}\_\_\tilde{\text{V}} \tag{VI}$$

The nasalization can be seen with *-pẽ* (LS.liquid), mentioned earlier, where the examples [epẽ], 'water', and [mɑ̃nĩmẽ], 'this (liquid)', were given. The sentient third person *-kɑ̃* also undergoes this change as can be seen in [gokɑ̃], 'He/She goes.', versus [põŋɑ̃], 'He/She comes'. The nasalization pattern is very common in my data.

I have already described an instance of voicing, alone, above. The suffix /pẽ/ is associated with arms. In the nominal *õdõpẽ*, 'arm', the surface form is [õnõmẽ]. When used with the verb *a toõkĩ*, 'to break', *toõpẽte*, 'breaking (arm)', the surface form is [toõbẽ^{n}te]. The verbal combination is less frequent than the nominal, which may influence the variation. That is the only explanation I can offer for the variability at this point. One will be aware, from the discussion in the previous subsection, that a lot rides on the arm-breaking example. That it is also an instance where variability is evident makes it still more vulnerable to an alternative analysis.

A case where neither voicing, nor complete nasalization occur is the word for 'house', /õkõ/. This example may not be a valid example. I am following previous analyses in proposing the initial nasal *o*. I doubt I would have proposed it otherwise based on the speech of my consultants. It may be that I am mishearing the nasalization. The nasality of the final vowel can be demonstrated by applying affixes. The voiceless [k] is very noticeable and difficult to mistake. If there is an initial nasal, the surface form is [õ$^{\text{ŋ}}$kõ].

I have doubts about the initial vowel of /õkõ/, but there are reasons to believe that it was accurately described in the SIL literature. Previous sources did not have a rule like (VI). Peeke (1979) stated that voiceless consonants would occasionally become voiced, but the norm was prenasalization. It is reasonable to suppose that some items still reflect previous patterns. Given that *õkõ* is used frequently in Wao Terero, frequency may be part of an explanation for the maintenance of the pattern.

There are alternatives to the notion that the pattern is largely lexical. It is possible that there are interactions between particular vowels and consonants that I do not understand. Alternately, there may be a sequence of vowels before or after the consonant. That would mean that /õkõ/ may be /õokõ/ or /õkoõ/.

Alexia Fawcett (p.c.) pointed out the complete nasalization pattern, which appears dominant among my consultants.

It is reasonable to ask if nasal consonants are becoming phonemic. For the time being, I feel that it is reasonable to continue under the assumption that they are not, since previously described phonological patterns largely hold.

3.1.3.7 Alveolar Voiced Stop to Alveolar Tap when Preceded by Oral Vowel

The reason that some authors write the name of the Wao language as Wao Terero and others write it Wao Tededo is that it is pronounced [wao#teɾeɾo] by most speakers. The /d/ goes to [ɾ] when following an oral vowel.

$$d \rightarrow ɾ/V\_\_ \tag{VII}$$

It is not uncommon to find the convention of using 'r' for the tap in non-technical orthographies. Some speakers are very adamant that there is no 'r' in Wao Terero and resent its usage in these orthographies. There is unlikely to be a resolution of this controversy in the near future. I respect the wishes and follow the practices of my consultants. It would be worth studying why Wao speakers find the Wao tap to be distinct from the tap represented in Spanish by 'r'. These should be very similar sounds. It may be as simple as the Wao speakers feeling that there is nothing like the trill, which is also represented by 'r' in Spanish.

3.1.4 Diagnostics for Free versus Bound Status of Morphological Units

None of the phonological rules described above are active across word boundaries. Demonstrating that an element is affected by the rules is sufficient to show that it is word internal.

When some element is not affected by word-internal phonological rules, there is usually additional evidence that it is not bound, such as the ability to occur in isolation, the ability to be a target of expressive morphology, such as repeating, or optionally occurring before or after some element that it modifies.

I have not seen the phonological rules above affecting compounds of free forms. For instance *awẽ daa*, 'tree thorn', is [ɑwẽ dɑɑ], not [ɑwẽ nɑɑ].

3.1.5 Orthographies

I will talk about two systems of writing. One was developed by SIL linguists and is reported on in Kelley (1988). The other is what is used most consistently by my younger consultants. There is often a mixture of the two among modern speakers though it is rare for a grapheme associated with a voiced stop or palatal approximant to be used for a nasal allophone.

The SIL system will be familiar to the reader as it is essentially the phonemic system that I use except that nasalization is represented with diaeresis instead of a tilde. The SIL authors would write *babë*, instead of my *babẽ* for 'bone'.

The mostly widely used modern orthographic convention does not consistently use annotations on a vowel to indicate nasality. The vowel characters 'a', 'e', 'i' and 'o' are used. The consonant characters can be seen in Table 3.5.

	Bilabial	Alveolar	Palatal	Velar
Stop	p b	t d		k g
Nasal	m	n	ñ	ng
Approximant	w		y	w

Table 3.5: The consonant characters that are commonly used in modern Wao Terero orthographies.

Pre-nasalization of voiceless stops is given by 'mp', 'nt', and 'nk'. One can use the phonological rules provided above to determine when to use a nasal rather than some non-nasal consonant. A comparison of my transcriptions with modern orthographic conventions can be seen in Table 3.6.

It is not uncommon to find a mixture of the modern writing system and the SIL system. Spanish con-

Morph. Trans.	Common
babẽ	bame
bãdĩ	mani
bãdĩka	maninka
bãdĩpo	manimpo
bãdĩta	maninta
gobopa	gobopa
põbopa	pomopa

Table 3.6: A comparison of my morphologically oriented transcriptions and modern orthographic conventions.

ventions may be used, such as writing 'gui' for [gi]. Some speakers use a final 'n' on words that end in nasal vowels. Sequences of identical vowels are inconsistently represented by Wao speakers.

The diaeresis may be used as a kind of stylistic element in some cases to accentuate the fact that the writing is Wao. On social media, for instance, I frequently see usage where diaereses do not correspond to actual nasal vowels. In my experience, speakers who consistently use the diaeresis correctly are less likely to represent pre-nasalization in their writing.

3.2 Inflection and Morphosyntax

I only cover basic information concerning general morphosyntax and inflection. It is a big topic, and much of it is not relevant to the central topic of the thesis. It is also the case that a great deal of work remains to be done. Peeke (1968), the only attempt to produce a semi-comprehensive morphosyntactic description, represents a major accomplishment but it has a number of flaws. First, it is extremely terse and deeply couched in tagmemic (K. L. Pike, 1958) conventions that are now obscure. A second issue is that Peeke's grammar is highly dated with respect to the modern language. The language appears to have changed very quickly after contact. Some older texts found in E. G. Pike and Saint (1988) are extremely difficult to read for modern speakers. Though not an unintended issue, Peeke's grammar is explicitly preliminary and incomplete, so even if none of the previously mentioned issues existed, a great deal of additional work was known to be necessary before the grammar could be considered comprehensive.

In this section, I begin with a discussion of pronouns and person, which includes some information on case. I mention some facts about where morphs for tense and aspect occur on verbs, but only to highlight their interaction with person marking on verbs. I also discuss plurality and negation.

3.2.1 Person and Number

In this subsection, I discuss grammaticalized person. Since much of this work deals with the lexical-functional divide, I should clarify what I mean by this. By grammaticalized, I intend person meanings and categories that correspond to particular affixes or pronouns in the system. For the most part, when there is an affix for person, there are at least two corresponding pronouns. I will start by talking about person marking with respect to the present tense verbal paradigm and will also provide a discussion of the past

	Singular	Plural	Dual
1	gobo	gobõdi	gobõda
1 Inclusive		gobõ	
2	gobi	gobĩdi	gobĩda
2 Maternal	gobĩ		
3 Feminine	godã		
3 Sentient	gokã	godãdi	goda
3 Non-sentient	go		

Table 3.7: Verbal person marking in Wao Terero using GO, 'go', as an example.

tense paradigm. Person marking does not exhibit allomorphy depending on tense, aspect or mood, except in the sense that positions of person morphs vary with respect to the past tense morph.

3.2.1.1 Person and Number Contrasts in the Verbal Domain

Table 3.7 displays all person contrasts using the verb *go*, 'go'. I largely follow the paradigm as drawn by Peeke (1979), with some tweaks that are explained throughout this section.

Within the first person there are contrasts for singular, dual, plural and inclusive. I do not know of any contrast within the inclusive between plural and dual. It is not clear whether the non-inclusive plural necessarily means "more than two" or "exclusive", explicitly, or if it is an underspecified first-person plural. It is not uncommon in the Wao system to allow for a less specific person to be used when a more specific person could have been used. A lack of strict agreement is not uncommon. This can be seen in (22), where the verbs have no person marking despite the use of the first person inclusive pronoun. The use of *-pa*, 'DECL', on the final verb indicates that it is the main, finite verb. The *-pa* marked verb contains the tense for the clause, and person marking for the subject – if they are marked at all.

(22) *bõ-tõ* *odẽ-de* *go-kĩ* *be te-kĩ-pa*
1.INCL-PRO heaven-LOC go-INF find-FUT-DECL

'We will go to find one another in heaven.'

Though I have labeled the affix-less form, *go*, in Table 3.7, as non-sentient and third person, it may make more sense to say that it is underspecified. I would not go so far as to say that one can *always* use some underspecified form grammatically and felicitously to agree with a specific referent, but there are contexts where less specific forms are used. Wao Terero allows nominal person and verbal person to mismatch in what is likely a constrained and predictable manner. A feminine subject need not have a feminine verbal marking. A plural verbal marking may not match a nominal plural marking, in all cases. The extent to which this is bound by grammar or semantics is currently unknown. In Chapter 6, I chose a grammatical agreement analysis. In the formal fragment, I must make a choice. The grammatical analysis allows me to demonstrate features of the framework design.

The second person contains the same three number contrasts as the first person, but also what I have labeled *Maternal*. The affixes on the second person maternal singular and feminine third-person singular were called honorifics by Peeke (1968) and Peeke (1979). According to Peeke, what I have as feminine and maternal were once used primarily for women of a particular social status, including mothers and other senior female members of the community. I have verified that the second person maternal is used by modern speakers for addressing one's mother. It is optional in that case. Speakers have stated that the general second person may be used when addressing one's mother, another example of the most specific usage not being obligatory. Whether the suffix continues to have broader cultural uses, such as addressing other senior female members of the community, is not known.

The same three number distinctions exist for the third person as the first and second persons, but there are also distinctions based on gender and sentience.

As stated above, Peeke considered what I have labeled feminine to be an honorific. Possibly influenced by Spanish bilingualism, there is a tendency for some speakers to use the third-person feminine more broadly than indicated by Peeke. It may now be used for female persons in general, and I see no reason to call it an honorific. Despite this, it is not syntactically obligatory when speaking of a feminine person. I frequently find that even speakers with Spanish as a native second language vary as to whether they would use the *-dã* suffix or *-kã* suffix, which I have labeled sentient, when speaking of a feminine subject. Example (23) provides two portions of sentences where *-kã* is used on verbs with a woman as the subject. An example using the third feminine with concord is provided in (24). I use the term concord in a manner distinct from agreement. Concord differs from agreement in that agreement describes systems that are strict and syntactic, while concord may occur due to semantic, pragmatic or other factors that result in optional usage. Here I do not mean to imply that the pattern cannot be grammatical, but describing the co-occurrence patterns as concord is neutral between options. I have not seen the feminine used in reference to a masculine person.

(23) a. *yẽdẽ okiye-kã põ-kã tõbẽ-kã põ-kã ate*

big woman-3.H come-3.H PRO-3.H come-3.H when

'When the fat woman came'

b. *ĩ-kã okiye-kã dõbe põ-kã tobã õdõ-pẽ-kã waponi yewai-kã*

PRO-3.H woman-3.H already come-3.H all ∅-LS.arm-LS.body good paint-3.H

'She, the woman, arriving already, her arms painted beautifully'

(24) *tobẽ-dã kowẽba owẽpo-dãbaĩ ĩ-dã-pa õdõ-ke wa a-dã-pa*

PRO-3.F always play-NEG COP-3.3.F-DECL ?-LIM observe-3.F-DECL

'She never plays. She only observes.'

It is true that the *põ* verbs in (23) are not the main verbs of the clause, which may affect concord characteristics. I used subclauses because the full sentences are quite long. I see the examples as supporting my point despite this status. Rather than a lack of affix, or a non-person associated affix, there is a *-kã* present on both verbal elements and pronouns. The femininity of the subject is quite clear due to the use of *okiye*, 'woman'. Therefore, these are valid examples of *-kã* usage in a feminine referencing context.

There are some further complications with the example that the reader should be aware of. Note that in (23b), I gloss *-kã* as both a person marker and as a lexical suffix. Peeke (1968) had claimed that *-kã* should be considered a classifier, and that the person marking usage was an extension of its classifier usage. As a lexical suffix (classifier) she glossed it as 'body' and I have verified that it is compatible with the concept of 'body', 'meat' or 'flesh'. Although I am certain that it has a more inflectional role in the verbal paradigm in Table 3.7 for modern speakers, I am not certain of how to analyze it in all cases. For that reason, I may be incorrect in some cases when glossing it as a lexical suffix or person marking. It may be that in (23b), that *yewai-kã* should be labeled as referencing a body, rather than a person. I find such an interpretation to be unlikely in this case, since *põ*, 'come', does not normally allow lexical suffixes. In general, I assume that for inanimate usage, *-kã* will have a lexical suffix status. For animate usage on nouns, it will have a person status. It is the only affix, other than *-koo* that appears to have something like a dual inflectional-lexical status. Lexical suffixes have a very strong tendency to describe inanimate properties. See §3.3.4.16 for more on the lexical suffix usage of *-kã*, as well as §3.2.2 and §3.3.4.20, for more information about *-koo*.

As a point of additional context, the women in (23), are supernatural beings, with animal natures. I have noted no pattern where such beings are more likely to be associated with the *-kã* suffix, but I do not feel certain why the affix is used on the noun *okiye*. Nominals do not obligatorily receive person marking. There may be some interesting subtleties here that I am not aware of.

The fact that the *-kã* suffix is compatible with feminine uses is documented back to the time of Peeke (1968). The examples in (23) are nice in being very explicit about the feminine nature of the subjects.

The sentient–non-sentient distinction in Wao Terero is similar to the contrast between *it*, on the one hand, and *she, he* and singular *they*, on the other, in English. Verbs often occur without person marking when they are non-finite. As previously described, lack of person marking on a finite verb does not require the subject to be non-sentient, nor does it always require a third person. In (22) the subject was not required to be third in the absence of an affix. Yet, a finite verb without person marking is clearly the convention when speaking of inanimate objects, (25a), and animals, (25b). The sentient usage, with *-kã*, is for people and optionally for animals. The plural, dual and feminine distinctions are largely reserved for sentient beings, as in (25c). The plural for an animal often involves no plural marking on the verb, as in (25d). There is flexibility in plural usage for animals, at least in translations from Spanish. This can be seen in (25e), where the verb has the third-person plural marking. Example (25e) was a translation from Spanish.

(25) a. *bãdĩ di·ka okoi ã-pa*

DEM ∅·LS.stone hot-DECL

'This stone is hot.'

b. *a·wẽ-de deye-koo bõ-pa*

∅·LS.plant spider.monkey-CLF.group sleep-DECL

'The spider monkeys sleep in the tree.'

c. *okiyẽ-dãdĩ* *tokĩ-dãdĩ*

woman-3.PL laugh-3.PL

'The women laugh.'

d. *ba·koo* *geye* *kewo-pa*

many·CLF.group fish swim-DECL

'Many fish swim.'

e. *ba·koo* *ẽyabõ-idi* *kewo-dãdĩ-pa*

many·CLF.group fly-COL fly-3.PL-DECL

'Many flies fly.'

One may note that in examples (25a-25e), there is a fair amount of variation in number marking on the nouns. This is discussed further below.

Before discussing pronouns, I would like to highlight some facts about person, number and the verbal paradigm when the past tense suffix *-ta* occurs. The pattern can be seen in 3.8. I have no example of the second person maternal in the past tense, but I have no reason to believe it would not pattern with the other second person. The pattern of note is that first persons have the *-ta* morph before person marking, while non-first persons have the *-ta* morph following person and number marking. The declarative morph, *-pa*, is included for all examples. This is because, for the speakers I work with, the *-ta* morph may not occur word finally. Given that questions are formed by omitting the *-pa* on a main verb, questions may be underspecified for tense. Other affixes that indicate tense or aspect do not exhibit the pattern. The *-kĩ* and *-ke* affixes, when associated with the future, precede the person and number marking for all persons. The *-ĩ* affix for completed actions, always follows person and number marking, and may be used for questions concerning the past for non-first persons, given the morphological restriction on word final *-ta*. The affixes *-ĩ* and *-ta* may be in competition. I have not seen them co-occur in any of my data.

	Singular	Plural	Dual
1	gotabopa	gotabõdipa	gotabõdapa
1 Inclusive		gotabõpa	
2	gobitapa	gobĩditapa	gobĩdatapa
2 Maternal	gobĩtapa		
3 Feminine	godãtapa		
3 Sentient	gokãtapa	godãditapa	godatapa
3 Non-sentient	gotapa		

Table 3.8: Verbal person marking for the past tense indicative in Wao Terero using GO, 'go', as an example.

3.2.1.2 Pronouns

The pronoun-like elements of Wao Terero are given in Table 3.9. The *ĩ* stem has also been described as a distal demonstrative (Peeke, 1968). There are short and long forms for all pronouns. The long forms appear to have an emphatic or topicalization-like function. The long form is often translated to Spanish as 'yo mismo', 'ellos mismos', 'ella misma', etc. ('myself', 'themselves' or 'herself'), which may carry the sense of having performed the task alone, or something like focus.

	Singular		Plural		Dual	
	Short	Long	Short	Long	Short	Long
1	botõ	tõbẽbo	bõditõ	tõbẽbõdi	bõdatõ	tõbẽda
1 Inclusive			bõtõ	tõbẽbõ		
2	bitõ	tõbẽbi	bĩditõ	tõbẽbĩdi	bĩdatõ	tõbẽbĩda
2 Maternal	bĩtõ	tõbẽbĩ				
3 Non-sentient	ĩ	tõbẽ				
3 Sentient	ĩkã	tõbẽkã	ĩdãdi	tõbẽdãdi	ĩda	tõbẽda
3 Feminine	ĩdã	tõbẽdã				

Table 3.9: Pronouns of Wao Terero

Verbal person-number distinctions are signaled with similar forms on pronoun-like elements, whether short or long, with some differences in patterning. A notable difference between third and non-third pronouns is that first and second person marking appears without any preceding stem on short-form pronouns. The final *tõ* on these forms is labeled 'PRO' in glosses.

All pronoun-like elements may be used as either possessive or personal pronouns.

There are some additional pronoun-like elements that do not appear in the table. I will not cover these due to my lack of experience with them. See Peeke (1968) for more information. I discuss the relative pronoun-like element *dãtõ* in connection to kinship terms in §3.3.4.12.

Case is not indicated directly on pronouns. The animate object of a verb is signaled using a collocation of a pronoun or nominal with what I label as the accusative. I label it as the accusative because it serves as a marker for the animate object of a transitive verb. It is a misidentification, in a strict sense, since the comparative concept of accusative does not imply animacy or sentience. The element takes person marking and what I label as the gerundial affix. There is some ambiguity between the element and other parts of speech. The stem is homophonous with the demonstrative, the copula, and the short third-person pronouns. The fact that the *-te*, GER, is a verbal affix, argues for a verb-like designation, yet it is the only element that takes both person marking and the *-te* suffix that I have seen. In the analysis in Chapter 6 I treat it as an accusative particle. Though it may not be historically distinct from the copula, the copula distributes similarly to verbs, with similar morphological characteristics. The accusative particle does not. Likewise, the syntactic distribution and morphological characteristics of the distal demonstrative and third-person pronouns are similarly distinct.

To signal that a nominal or pronoun is the object of a verb, the accusative directly follows the nominal or pronoun. I have not investigated whether it is an absolute grammatical requirement that the accusative follows directly. In a typical example, such as (26), the accusative is marked for person and the affix *-te*. With pronominal third-person objects, this has a reduplication-like quality.

(26) *ĩ-kã* *a-kã-pa* *ĩ-dã* *ĩ-dã-te*
PRO-3.H see-3.H-DECL PRO-3.F ACC-3.F-GER
'He sees her.'

It is very common to leave off the *-te* ending, as seen in (27a) and (27b). Due to the fact that the copula is homophonous with the stem of the short third-person pronouns, the seeming reduplication is increased in (27a).

(27) a. *ĩ-kã* *a-kã-pa* *ĩ-dã* *ĩ-dã*
PRO-3.H see-3.H-DECL PRO-3.F ACC-3.F
'He sees her.'

b. *kidi bo-to ĩ-bo wao waka-ta-pa*
cat 1-PRO ACC-1 scratch-PST-DECL
'The cat scratched me.'

The long pronominal form may be used in the construction, as can be seen in (28).

(28) *a-ta-bo-pa tõbẽ-dã ĩ-dã ekoda õkõ*
see-PST-1-DECL PRO-3.F ACC-3.F school
'I met her in school.'

Example (29) provides an example where a non-pronominal, nominal is used in the construction.

(29) *gĩta tawadiya ĩ-kã wẽ-dõ-pa*
dog chicken ACC-3.H die-CAUS-DECL
'The dog killed the chicken.'

Inanimate objects do not receive any indication of object-hood, as can be seen in (30a) and (30b). Example (30c) demonstrates that non-sentient animates may not have their object-hood indicated with the construction.

(30) a. *bo-to wao to-ta-bo-pa ado-ke õ-yabo*
1-PRO cut-PST-1-DECL one-LIM ∅-LS.leaf$_1$
'I cut a leaf.'

b. *awoto yẽ-pẽ-bo-pa*
car wash-CLF.liquid-1-DECL
'I'm washing a car.'

c. *õ-pa-ka õõ-ta-bo-pa gata*
∅-LS.board-INS hunt-PST-1-DECL monkey
'I killed the monkey with a dart.'

3.2.2 Plurals

As mentioned previously, there is no default plural in Wao Terero. Particularly for non-sentient entities, a bare nominal is underspecified for number. In example (31), the translation was plural in Spanish despite the fact that there is no plural marking.

(31) *a·wẽ* *pa-wẽ-ta-bo-pa*
∅·LS.plant chop-CLF.plant-PST-1-DECL
'I was cutting trees.'

For sentient third persons the *-dãdi* plural may be used, but is not always. In example (32a) the plurality of women is evident in the pluralization of the verb, but the pluralization of the noun is not necessary. As a note on the felicity of 'listening to hear a dance', within the Wao cultural context, a dance would be accompanied by chanted music. In example (32b) there is plural marking on the nominal. Both patterns appear to be fairly common.

(32) a. *okiye* *awa-dãdi* *ẽyẽ-bo-pa*
woman dance-3.PL listen-1-DECL
'I listen to what the women are dancing.'

b. *okiye-dãdi* *to-dãdi*
woman-3.PL laugh-3.PL
'The women laugh.'

The *-idi* suffix is often used for a class of similar things, or a group typified by a member. The implication in (33a) is that one is talking about animals, in general, rather than some specific group of animals. The pluralization on the verb is likely not necessary. This is a translation from Spanish, which may have influenced its use. Peeke (1979) also provides examples. In example (33b), the name of an individual receives the affix. This refers to a group of which Giketa is a typical member or leader, for instance, his family. In example (33c), the word for 'chicken' receives the suffix, indicating the class of domestic fowl.

(33) a. *kĩ-dõ kẽ-dani õi·kã-idi*

what-Q eat-3.PL animal·LS.body-COL

'What do animals eat?'

b. *giketa-idi*

giketa-COL

'Giketa's bunch' (Peeke, 1979, p. 16)

c. *tawadiya-idi*

chicken-COL

'domestic fowl' (Peeke, 1979, p. 17)

The suffix *-koo* is also used as a plural for groups or clusters, and has uses as a lexical suffix and classifier. See §3.3.4.20, for more information on classifier uses. An example was provided in (5a), repeated here as (34). As mentioned in the introduction, the lexical suffix usage for cloth and clothing may be diachronically motivated by the grouping of thread used in weaving. Peeke (1979) provides some examples that suggest that at the time of her fieldwork with a population recently contacted, this connection was relatively transparent. She consistently glosses *-koo* as 'conjunto', which can be translated as 'aggregate', 'assemblage', or generally, 'group'. I take the use of *-koo* in (34) to approximate the sense of 'flock', 'herd', or 'school'. This would mean that *-koo* may be used with animates, which is not a general, productive property of lexical suffixes. The quantifier *bakoo*, 'many', should perhaps be analyzed as being related to the suffix.

(34) *a·wẽ-de deye-koo bõ-pa*

∅·LS.plant spider.monkey-CLF.group sleep-DECL

'The spider monkeys sleep in the tree.'

Possibly due to Spanish contact, enumerable plural uses are also found for *-idi* and *-koo*. For instance, in (35), the notion is that there are some specific animals one is burying, rather than it being one's job to bury animals, the latter of which would be more consistent with a class interpretation. A question is why particular plural affixes are chosen for particular nominals when non-count plural meanings are absent. Why is *-idi* used in (35) and not *-koo* or *-dãni*? There may be semantically conditioned form classes developing for nominal plurals. Since animals are a general, abstract group, they may receive the *-idi* plural, even when talking of three specific animals.

(35)	*tee bõde-bo-pa*	*oi·kã-idi*	*õtato-de*
	bury-1-DECL	animal·LS.body-COL	ground-LOC

‘I am burying animals in the ground.’

3.2.3 Negation

A basic understanding of negation in Wao Terero is useful for assessing examples found in this work. Negation is the primary diagnostic that I use for proffered content (Roberts, 2012), essentially what is asserted in an assertion. This is explained further below. A more thorough discussion of negation is found in Peeke (1991).

There are two primary ways of expressing negation. The first, using the suffix *-dãbaĩ*, scopes over a clause. An example can be seen in (36), where the scope is over a subclause.

(36)	*dẽbo*	*go-yõ-te*	*tĩye*	*go-dãbaĩ*	*ĩ-kã*	*tõye*	*wẽ-kã-ĩ-pa*
	nemo	go-SIM-GER	quickly	go-NEG	COP-3.H	toñe	die-3.H-COMPL-DECL

‘While Nemo was going, she didn’t go quickly, and Toñe died.’

The example is taken from a narrated oral history. Nemo (Rachael Saint), is going to where Toñe, one of the first native missionaries, is attempting to convert a resistant group of Wao, but arrives too late to save him from murder. The construction utilizes a verb that carries the negative suffix. The negative marked verb co-occurs with the copula, which may be inflected with person marking. Tense and aspect occur on the copula when the negative clause is finite. An example of a finite usage of the copula can be seen in the elicited translation in (37), where the copula is marked for past tense.

(37) *Ayãbo ẽ-te go-dãbaĩ ĩ-kã-ta-pa õ·yabo teẽbe wedĩke*

bird gather-GER go-NEG COP-3.H-PST-DECL ∅·LS.leaf$_1$ heavy because

'The bird didn't take the leaf because it was heavy.'

Peeke (1991) reports that the negative marked verb may also carry the future tense affix *-kĩ*. An example is provided in (38). It is unclear if the parentheses indicate that the example was elicited both with and without the copula, or whether Peeke was indicating where the copula would have been, had it occurred. The copula is often absent when context makes it clear who someone is talking about, especially for third persons. There is reason to question the translation of (38) as future tense. The *-kĩ* affix is also used to express desire and uncertainty.

(38) *kẽwẽ-kĩ-dãbaĩ (ĩ-kã-pa)*

live-FUT-NEG (COP-3.H-DECL)

'It isn't one that will live.' (Peeke, 1991, p. 26 (18))

Peeke (1991) claims that no other preceding or following affixes are allowed with the negative affix, which she calls a deverbalizer. This may be the case for inflectional affixes, but derivational affixes are allowed.

(39) *põ-dõ-dãbaĩ ĩ-ta-bo-pa cinco dolores bo-to badã bewe*

come-CAUS-NEG COP-PST-1-DECL five dollars 1-PRO mother for

'I didn't send five dollars to my mother'

I have not determined whether lexical suffixes may occur on a verb with negation marking.

The negative affix is also used for negative commands and questions, as seen in (40). The negative commands do not require the copula. According to Peeke, the copula may occur, and then must include the imperative affix *-ẽ*, 'IMP'. The *ĩ* in (40b) is glossed as a demonstrative. It is ambiguous as to how it should be interpreted, due to object ellipsis being common, but with the absence of the imperative affix, the demonstrative interpretation is preferred. In (40c), the copula is also absent.

(40) a. *a-dãbaĩ ĩ-bi*

see-NEG COP-2

'Do you see?'

b. *a-dãbaĩ ĩ*

see-NEG DEM

'Don't look at it!'

c. *tao ke-dãbaĩ daa*

touch do-NEG thorn

'Don't touch the thorn!'

The second form of negation uses a particle *wĩĩ*. I have few examples of this particle in my corpus. Peeke presents it as essentially parallel to many of the uses of the affix *-dãbaĩ*. An example where the verb is being negated, like (41), seems to support the idea that, at least with the population Peeke worked with, *wĩĩ* could be used in place of *-dãbaĩ*. I believe that *wĩĩ* is a negative focus particle. Perhaps, given a larger context, it would be clear why a more narrow focus was placed on the verb in (41).

(41) *wẽyẽ-kã ĩ-te tõbẽ-kã (õõĩkã) wĩĩ kẽ-kã-pa*

child-3.H COP-GER PRO-3.H (meat) NEG eat-3.H-DECL

'Being a child, he didn't eat (meat).' (Peeke, 1991, p. 32 (43))

The notion that *wĩĩ* is a negative focus particle is supported by its use in contrastive constructions and its ability to narrowly negate items below the level of the predicate. Peeke claims that (42) is an example of negating only the adverbial modifier *kĩkẽ*, though the example does not provide much evidence of this, in and of itself, due to the lack of context.

(42) *wĩĩ kĩkẽ kekõ-kã*

NEG fast walk-3.H

'They don't walk fast.' (Peeke, 1991, p. 36 (64))

3.3 Description of the Lexical Suffixes

In this section a description of the meanings and distributions of most known lexical affixes is provided. I describe almost all lexical suffixes listed in Peeke (1968). I also describe some listed only in Fiddler (2011). There are also suffixes that have not been listed in previous works. Examples of usage with all parts of speech that host lexical affixes are provided. The differences in behavior across parts of speech are described. I also provide information on my in progress investigation of anaphoric properties of lexical suffixes, which I do not believe are anaphoric signals.

As stated in §1.1 and Chapter 2, there are two broad trends in lexical suffix usage. There is the compound-like usage, primarily on nominals, and there is the classifier-like usage on other parts of speech. Classifier-like usage is not obligatory for lexical suffix status. In Wao Terero, there are items that are only found on nouns, which may have no use as classifiers. I focus on *classificatory* lexical suffixes, the class of affixes that show both productive compound-like and classifier-like usage. There are some marginal items that I will mention below, and certain "candidate" classificatory lexical suffixes that may prove to have the basic properties of interest given further investigation, but I do not discuss all non-classificatory suffixes mentioned in Peeke (1968) and Fiddler (2011). All previously described affixes are presented in Appendix D.

There may be pure classifiers that do not occur in the nominal domain. I have encountered one potential instance, which is the affix *-gadẽ* (CLF.throat), discussed in §3.3.4.5.

There are systems that overlap with lexical suffix usage that are usefully considered distinct, such as person marking. There are also ambiguous lexical suffixes that have atypical uses. The *-kã*, 'LS.body' (§3.3.4.16), and *-koo*, 'LS.group' (§3.3.4.20), affixes are either identical to, or the same as inflection-like affixes. Most issues in determining clearly delineated categories within the affix space can be achieved through known diagnostics.

3.3.1 Verbal Usage as a Diagnostic

Verbs are used to test classificatory lexical suffix status. The verbal domain is morphotactically and semantically restrictive in comparison to adjectives and demonstratives. Only a highly productive core of classificatory affixes can be used with verbs. There is a distinct verbal position for lexical suffixes, while adjectives and demonstratives put lexical suffixes in competition with person marking and other affixes.

Non-classificatory lexical suffixes do not occur productively in verbs. Additionally, except for some ditransitives, verbs allow only one lexical suffix, which is not true of nominals and (possibly) numerals. This provides valuable information about potential clusters. Peeke (1968) noted that some potential suffixes look like multiple suffixes, yet the cluster may or may not behave like a single suffix. She offered no diagnostic for determining the difference. The verbal restrictiveness can be applied in such situations. When an affix occurs in a verb, it has passed a significant hurdle, and one can be confident in its status as a single, classificatory lexical suffix. If something works with a verb, it will work anywhere.

3.3.2 About the Data and Methods

This section provides a brief orientation to the types of data that informed the description below. It is not a methods section, since methods were too diverse for easy summation.

My initial fieldwork was with two consultants, one in her mid-thirties, Mary Boyotai, and the other, Ramiro Boyotai, in his early twenties. The work focused on elicited translations. Some of this work was performed with the late Pieter Muysken. The consultants were also engaged by Tod Swanson to transcribe and translate video recordings of narratives and interviews he had collected. Much of the narrative-derived examples in this section are from videos collected by Tod Swanson.

There are multiple transcriptions and translations of Swanson's videos. Here I use versions from Rubén Boyotai, who was the largest contributor of data for this thesis.

Tod Swanson was primarily concerned with cultural and historical studies. For that reason, his recordings were of older speakers. The speech of younger speakers is significantly different. It may not be as confusing as Shakespeare to modern English speakers, but Wao in their 40s and younger (perhaps older) have difficulty with the vocabulary, and refer to it as the speech of the *pikanani*, or 'old ones'.

Many examples are from elicited translations of sentences. These sentences are simpler than are commonly found in narratives, and were sometimes designed to produce minimal pairs. Translations have their issues. As an example, in Wao Terero, there is an instrumental affix -*ka*, and a conjunction *tõdo*. One speaker would sometimes accidentally use *tõdo* for the instrumental when translating from Spanish, likely because the free Spanish *con*, 'with', serves both functions in that language. She would later correct the error when reviewing her work, but similar artifacts may have made their way through the review process. One must have mitigation strategies in translation work, and be realistic about the empirical claims they can make with the results. Context sometimes helps, so that the speaker can focus on a picture, or a scenario, rather than getting two languages to “square up”. To ensure that the translations were understandable translations, after collecting the data, I later went through verification steps, where after some days, I would return to the translation and present it to the same speaker, or some other speaker, without presenting the original stimulus. Rejected sentences were removed. If the explanation of the sentence differed significantly from the original stimulus, I would ask whether the original made sense. This does not remove distortions due to language comparison, but it ensures that Wao Terero speakers agree that the phrases are grammatical and meaningful. Rubén Boyotai was asked to verify his own translations, as well as translations from Abraham Boyotai, who performed some self-verification, as well. Later, Flora Boyotai reviewed all the translations a second (or in some cases a third time) to ensure they were reasonable. Where such verification steps were not taken, I make a note of it.

Picture tasks of many kinds were used in order to avoid issues with translations, and to provide a fixed point of reference. In particular, with Rubén Boyotai, I borrowed an instrument from John R. White’s research on crop wild relatives (White and Monteros Altamirano, 2024), which featured photos of a number of domesticated plants that are used for various purposes in the Amazonian region. I interviewed Rubén Boyotai on the characteristics of the plants. Plant topics were a rich area for exploring productive noun formation using lexical suffixes.

There was also judgment-based elicitation, which involved six additional speakers for some protocols. I present examples of generated data intermixed with translation or narrative data when they are relevant. When I do, I provide a warning. Generated data is better confined to discussions of specific protocols, since it is not a Wao production.

Unfortunately, only a small portion of data that was collected during my fieldwork was sufficiently analyzed in time for inclusion in the thesis. I avoided data provided only under restrictive terms, such as Dickinson et al. (2013). I did not wish to contaminate open data I collected with claims that it relied on proprietary information. All Wao Terero consultants were aware of Tod Swanson’s videos on YouTube. They understood that their contributions would be public.

All speakers, except one, were from a single extended family, based in Geyepadi They all spent a large portion of their time in urban environments in Napo and Pastaza. The oldest speaker is currently 46, the youngest at the time of participation was 18. The majority had an Ecuadorian bachillerato, equivalent to US high school. Some had higher education in the form of vocational studies. Two speakers, both in their twenties, were finishing their bachillerato while working with me. All were literate and familiar with common orthographic conventions.

The vast majority of the data below comes from work with Rubén Boyotai.

3.3.3 A Note on Same and Different

Individual lexical suffixes are often associated with more than one meaning. An affix like *-po* may be associated with both 'hand' and 'canoe', among other meanings. As discussed in Chapter 2, this is not unexpected for lexical suffixes, especially when they have classifier functions.

In §4.3.5, I discuss formal justifications for the claim that affixes such as *-po* should be treated as a single affix with multiple meanings. I elaborate on the topic within the context of my formal framework in §6.5.2.1.

Based on the realizational stance in this work, I do not believe that meaning alone determines the morphological categories that correspond to affixes. Yet, neither does phonological form alone determine such categories. In §4.3.1 and below in §3.3.4.16, I discuss the case of *-kã*, where the third person sentient and a lexical suffix are demonstrably homophones, which in the context of the formal framework means that they correspond to distinct morphological categories.[4] In the case of *-kã*, there are distinct morphotactic positions that correspond to distinct meanings.

There may be more subtle homophones, where distributions are a question of part of speech, rather than clearly distinct positions. In particular, there may be homophonous lexical suffixes, where one is classificatory and the other is not. This would confine one homophone to the nominal uses, while the other is productive across parts of speech. One would not wish to predict that the classificatory suffix could realize the meanings of the non-classificatory suffix. Unfortunately, I have not pursued the question at depth, though there are diagnostics for such cases. These take advantages of properties of the demonstrative *bãdĩ*.

The demonstrative is useful because the stem does not restrict lexical suffix interpretations. If there is a lexical suffix with multiple meanings, all of those meanings will be available when used with the demon-

[4]For those familiar with Paradigm Function Morphology (PFM) (Stump, 2001), distinct morphological categories for homophones is similar to realizational rules with homophonous results being assigned to different rule blocks.

strative. Therefore, when one suspects that there is a homophonous affix, which only occurs with nominal stems, the demonstrative may be used to construct tests. One can test whether the nominal affix's meanings are available when the affixed demonstrative is used alone. One may also examine the effect of using the affixed demonstrative and the affixed noun in the same noun phrase to determine if the result is infelicitous, or confusing.

An example where demonstrative tests may demonstrate homophony is in §3.3.4.7 for *-dẽ*. Diachronically, the ambiguity is the result of a vowel merger. Plausibly, the phonological merger resulted in a morphological merger of two lexical suffixes. Another option is that their distributions and roles were distinct enough that they did not merge. One of *-dẽ*'s meanings is associated with food, and the other with abdomens. I have only seen the food meaning with a few nouns. The abdomen meaning is found with verbs and other parts of speech. I discuss what such homophony would mean for a formal analysis in §6.6.

Since nominals may take more than one lexical suffix, another test is whether homophonous affixes can co-occur. The *-dẽ* example makes it clear why that option may not always be available. There may be some plausible compound-like constructions that combine the food and abdominal meanings of *-dẽ*, but it is not obvious what they would be, and it would be difficult to both come up with such a word and get a speaker to accept it with the desired interpretation.

Other parts of speech are not suited to such tests. Nominal and verbal stems do not allow the full range of meanings associated with an affix due to idiomatic characteristics, semantic incompatibility, and class constraints of verbs. As an example of the latter, which is discussed in §3.2.1.1, is that some verbs may allow only body-part meanings, even if the verbal meaning would seemingly logically allow other argument types. This does not mean that noun and verb constraints signal divisions among affix identities. Noun and verb stems do not reflect the full productivity of lexical affixes. This is apparent in the fact that nouns are often frozen forms, and do not adhere to general patterns. Likewise, not all verbs take lexical suffixes. The restrictiveness of verbs serves as an excellent test of whether a potential lexical suffix is an actual lexical suffix. Yet, restricted, semi-productive, and idiomatic cases obscure the potential generality of an affix.

Adjectives do not have many lexical restrictions on classifying lexical suffix use, but the concreteness of their meanings does result in issues of semantic incompatibility with lexical suffixes at times. See §3.3.5.3.1 for a discussion of the adjective *yõbeĩ*, 'rotten', which is specific to the decomposition of animal matter, and so is incompatible with plant meanings. No such issue exists for the demonstrative, *bãdĩ*, which has an abstract meaning.

The *bãdĩ* test may not be the only necessary or sufficient test for demonstrating homophone status of two or more affixes in the lexical suffix domain. For instance, *-koo* may have homophonous uses where each is compatible with demonstrative suffixation. Additional tests are needed in this case. For more on the *-koo* issue, see §3.3.4.20.

Due to the fact that I have not performed demonstrative-like analyses on the items below, I gloss lexical suffixes as the same based on sound identity. I cannot provide firm evidence for a division in those cases, yet, so it isn't indicated in the gloss. This is not an issue for the core claims of my study, which focuses on items such as *-wẽ*, *ka*, and *-po*, which are highly productive affixes. In these cases I have used adjectives and demonstratives with the meanings I document.

3.3.4 Affix Descriptions

In this section I provide descriptions of individual suffixes in the system. As stated above, the list is not comprehensive. There are some items listed in Peeke's sources that are not covered, though there are also additional items. Peeke's lists are reproduced in Appendix D.

The organization of suffix descriptions is only semi-alphabetic. I often group affixes that are related. For instance, the first description below is first because of the affix *-ba*, but is grouped with two affixes beginning with 'y' due to a shared 'leaf' meaning. This is to avoid large numbers of tedious cross-references. The benefit of alphabetical listings is diminished with searchable electronic documents. See also Appendix E for an index of all affixes mentioned in the thesis.

3.3.4.1 *-yabo*, *-ba*, and *-yo*: Leaf Suffixes

There are three 'leaf' lexical suffixes. All three are productive in compound and classifier uses. I do not have verbal uses for any of them. Testing them with a cut verb, such as *wao toki*, would be informative.

There is *-yabo*, 'LS.leaf$_1$'; *-ba*, 'LS.frond'; and *-yo* 'LS.leaf$_2$'. The first of these is the most general, and can often subsume the use of the latter two as a classifier. The *-yabo* affix was listed in Fiddler (2011) as *-bo*. In (43), the result of an acceptability test is provided. I don't know how to gloss the *-ga* portion. It is a necessary element when using classifiers with ‘two’, but there may be exceptions. Sometimes, instead of *-ga*, there is *-ka*, and I do not know if there is a meaning difference.

The context for (43) involved two individuals. The first has a bag, and the second can’t see what is inside it. The second individual asks what is in the bag, and receives a single word response. The speaker consultant is asked if the response is acceptable. If the consultant accepts the response, they are then asked what is in the bag. I only performed the protocol with one speaker, though translations of the classifier constructions were verified with other speakers. See §3.3.6.3 for more information about the goals of the protocol.

When asked what was in the bag, the consultant’s immediate answer was “fruit.” To follow up, I asked whether the contents of the bag may have been other things, such as eggs, or balls, which were both stated to be compatible with *bẽboga*. These were accepted. When I asked if the contents could be leaves, the form was rejected. Unfortunately, I never tested *bẽ-yabo-ga*, in order to get a minimal pair, but I have frequently used *-yabo* in demonstrative and adjective-based elicitations.

(43) a. # *bẽ-bo-ga*
two-CLF-?
‘two leaves’

b. *bẽ-bo-ga*
two-CLF-?
‘two fruit’

There can be some uncertainty concerning the exact semantic boundary between the three leaf affixes, especially based only on picture matching tasks. This is likely because the boundary has some gradience, but also because pictures may not contain key features of the associated meanings. Generally, the explanation for the difference between the affixes given by speakers is based on how the different leaf-types are used within traditional Wao culture. This is interesting given the intuition of Denny (1976) that classifier systems are tied to aspects of material culture. I do not endorse all aspects of Denny’s view, but there is a tendency by Peeke and other authors (e.g. Biedny et al. (2021)) to focus on features such as shape or substance, rather than function in the classifier literature, which may not capture important details.

The affix *-ba*, ‘LS.frond’, is appropriate for leaves used for roof thatching, though a compound shape is also important. It primarily references palm fronds. In Fiddler (2011) there is a gloss ‘thing with parallel parts’, but I have no evidence that this abstract shape meaning is correct.

The affix *-yo*, ‘LS.leaf$_2$’, is appropriate for large leaves used to wrap food for cooking. For those unfamiliar with Amazonian cuisine, it is common to wrap fish or other meat and seasonings in a large leaf and cook it over a fire, called, maito in Ecuadorian Spanish – which comes from Kichwa *maytu*, ‘wrapped thing’ (Gualapuro et al., 2018). The large size of *-yo* leaves is also an important factor. A typical *-yo* leaf would be a banana leaf.

(44) a. *õ-ba*
∅-LS.frond
‘thatching material’

b. *o-yo*
∅-LS.leaf$_2$
‘leaf for maito’

c. *õ-yabo*
∅-LS.leaf$_1$
‘leaf (general)’

3.3.4.2 *-be*: ‘Territory’

The *-be* affix is a candidate classificatory lexical suffix. Before the /æ/-/e/ vowel merger it was /bæ/. It was noted by Peeke with two listed nominal occurrences. As can be seen in (45), these have both been verified in my investigations. The gloss for the affix in Peeke’s lists is ‘territory’. In addition to the concept of a territory, country, province, etc. the word *õ-be* can also refer to the countryside, forest, or workable land. The collocation *õ-be ke-kĩ*, ‘∅-LS.territory do-INF’, means ‘to work’. The canonical work is specific to agricultural activities. It may be understood as general labor in any setting by some speakers.

(45) a. *e-dõ õ-be geyepadi ĩ*

which-Q ∅-LS.territory geyepadi COP

'In which province is Geyepadi?'

b. *bi-to wa-be-ka õgõ-bi*

2-PRO other-LS.territory-? be.located-2

'Are you in another country?'

In (45b), I see no obvious interpretation for the *-ka* suffix.

These are the only two words I have unambiguously seen the affix in. I have no clear examples of its use as a classifier. Despite this, there are reasons to suppose that the affix is used more broadly. It may be related to affixes used for general directions and other geographic formations. In (46) I do not use the gloss LS.territory because it may be an incorrect identification. An 'x' is used. The *ĩ* in *ĩ-dõ* is glossed as a demonstrative. The word may be translated as 'side'. The demonstrative designation may be incorrect. There are a number of *ĩ*-stemmed (pro)nominal/demonstrative forms, which is why I made the choice.

(46) a. *a-ẽ dipẽ-be ĩ-dõ*

see-IMP left-x DEM-LS.road

'Look to the left!'

b. *a-ẽ toõ-be ĩ-dõ*

see-IMP right-x DEM-LS.road

'Look to the right!'

c. *ei-be*

above-x

'above/high'

d. *godã-be*

island-x

'island'

e. *wãteĩ-be*

island-x

'island of an oxbow lake'

Given ambiguities in nasalization patterns, the affixes in (46), if they are affixes, may be misidentified. If they are all the same affix *ei-be*, [eibe], indicates that either there is inconsistent nasalization across instances, or that the affixes are most consistent with *-be*, rather than other options such as *-bẽ* or *-pẽ*.

There is an issue with this interpretation. Adjectives only take one lexical suffix. Words like *eibe* behave like adjectives, and take classificatory lexical suffixes. This probably indicates that if *-be* is suffix-like in the words above, it is not a classificatory lexical suffix. See §3.3.5.3.1 for further discussion of *-be*.

3.3.4.3 *-bẽ*: 'cord'

The affix *-bẽ* (LS.cord) is a productive lexical suffix with both nominal and classifier uses. The *-bẽ* meanings are associated with ropes, vines, long bones, and stick-like objects. As a classifier, *-bẽ* has an additional abstract meaning of 'long'. The simplest nominal use corresponds to a general word for cord. It may also be used for 'vine'.

(47) *õ·bẽ*

∅·LS.cord

'cord'

(48) *ye·gĩ-ka-bẽ*

∅·LS.string-LS.stone-LS.cord

'rope'

There are other lexical suffixes that have related meanings. The suffix *gĩ*, ‘LS.string’, is associated with strings or cords. It arguably co-occurs in words with *-bẽ* in (48). I have no idea what *-ka* is contributing to the word, though see (51b), where *-ka* and *-gĩ* appear together in reverse order. It is also the case that *-gõ*, ‘LS.thorn’, is associated with long thin things.

I have no unambiguous examples of *-bẽ* in a verb.

3.3.4.4 *-bo*; *-bõ*: ‘egg’; ‘eye’

A case of affixes that are easy to confuse in terms of both sound and meaning are *-bo* and *-bõ*. Both appear to be productive with nominals and classifier constructions. The affix *-bo* surfaces as both [mo] and [bo]. The former sounds similar to the single surface form of *-bõ*, which is [mõ]. This can make it unclear which is which in some contexts, especially when both affixes have similar meanings. Peeke (1968) gave *-bo* the gloss ‘egg’, because it appears in *eke·bo* (∅·LS.egg), ‘egg’. The affix *-bõ* was glossed ‘eye/face’. The word for eyes and faces among modern speakers I worked with is not the one Peeke provided, but I have verified that *-bõ* can be used in classifier contexts for eyes and faces, as can be seen in (49).

(49) *awĩka dãta-bõ-bo-pa bo-to*
eye hurt-CLF.seed-1-DECL 1-PRO
‘My eye hurts.’

The original glosses from Peeke placed the two affixes in completely distinct domains of meaning. Yet, in Peeke’s later listing, provided in Fiddler (2011), *-bo* is glossed as ‘round thing’, and *-bõ* is labeled as ‘small round thing’. Lexical suffixes are sometimes associated with a set of concrete meanings in addition to conventional abstractions over those meanings, which may reduce to things like general shapes. It is unclear to what extent *-bo* is distinct from other affixes in being primarily about shape, as the gloss by Peeke would suggest. In addition to the egg noun, the affix also occurs in *pei·bo*, ‘∅·LS.egg’, ‘ball’, which appears to support a roundness meaning. Despite this, my guess is that *-bo* would be equally appropriate for a deflated ball, or a fried (even fantastically square) egg. I have yet to test this for *-bo*, but it would be consistent with *-ka*, which also has a ‘round’ meaning.

(50) *wĩyẽ-dãdi* *õkõ-do* *owẽpo-ketãte* *ke-te* *go-bõ* *a-yõ-dãdi* *die* *oto-bo*

children-3.PL house-out play-PURP do-GER go-toward see-SIM-3.PL ? basket-AUG

bẽ bõkẽ-te *põ-kã*

carry-GER come-3.H

'The children, who were playing outside, saw the woman as she arrived with the big basket.'

The affix *-bõ* is appropriate for smaller round things. The smallness contrasts with a homophonous use of *-bo*, which is also the augmentative suffix (50).

There are relatively concrete uses of *-bõ* that I do not believe are based on abstract shape, alone. One use of *-bõ* is to refer to seeds. It is productive in this usage.

(51) a. *pãto·bõ*

∅·LS.seed

'Seed for use in necklace making.'

b. *pãto·ka-gĩ*

∅·LS.stone-LS.string

'necklace'

The pairs (51a) and (51b) can be used to show that it is *-bõ* and not *-bo* at the end of the word because, when pronounced, there is no nasalization effect on *-ka*, which one would expect if the final 'o' of the stem was the cause of the nasalization of the bilabial of *-bõ*. Though it was not clear what *-ka* contributed in (48), in (51b) it may be that the interpretation of *-ka* could be 'seed', since *-ka* overlaps with *-bõ* for seed meanings. Necklaces use beads made of seeds in traditional Wao jewelry.

A verbal classifier use of *-bõ*, with a seed interpretation, can be seen in (52).

(52) *pe-bõ-kã-pa*

perforate-CLF.seed-3.H-DECL

'He/she is perforating a seed.'

(53) a. *bẽde·bõ*

fruit·LS.seed

'fruit (which one may suck on)'

b. *bẽde·ka*

fruit·LS.stone

'fruit (typically an orange)'

c. *bẽ-bo-ga*

two-CLF.egg-?

two (referring to fruit, eggs, balls)

Both *-bo* and *-bõ* may refer to fruit, which puts them both in overlap with *-ka*, as can be seen in (53). Both (53a) and (53b) represent generic names for fruit. The speaker that described the (53a) meaning emphasized that it was fruit for sucking on. There is some small fruit that is almost entirely pit, where one may place in their mouth in order to chew and suck the outer layer. The word in (53b) is not limited to a particular class of fruit, necessarily, but it does have a default meaning of 'mandarin orange'.

It is possible that *-bo* is compatible with fruit *only* due to the typical shape of fruit. Further investigation is needed. The numeral classifier example (53c) was chosen for presentation because the final *-ga* had no nasalization, and makes the status of the affix as *-bo*, and not *-bõ*, clearer. The example comes from the same elicitation as (43).

I have no unambiguous instances of *-bo* in verbs. I had a number of failed elicitation attempts, when trying to use it with *kẽkĩ*, 'eat', likely because of its non-body-part status. There are some verbs that only take body-part suffixes.

3.3.4.5 *-bõka*; *-gadẽ* : 'ear'; 'throat'

The *-bõka* suffix is a productive body-part suffix. It is used in the noun *õdõ-bõka*, '∅LS.ear', 'ear', and in verbal constructions, *dãta-bõka-bo-pa*, 'hurt-CLF.ear-1-DECL', 'My ear hurts'. It can be used with a number of other verbs as well.

(54) a. *audiophono õ toge-bõka-ẽ*
ear.phone ? remove-CLF.ear-IMP
'Take out your earphones.'

b. *bi-to audiophono õ dawe-bõka-ẽ*
bi-to ear.phone ? put-CLF.ear-IMP
'Put on your earphones.'

c. *bi-to audiophono õ we-bõka-ẽ*
bi-to ear.phone ? put-CLF.ear-IMP
'Put on your earphones.'

The use of *-bõka* is evidence that multi-syllable lexical suffixes can be used with verbs. This indicates that in cases where multi-syllable lexical suffixes do not appear to be allowed in verbs, as in discussed in §3.3.4.24, it is not due to phonological form.

It is unknown whether there is some historical linkage between *bõka* and *-bõ* and *-ka*. Ears on people are little round things on a head, and such compound meanings are not unheard of in Wao Terero. The fact that the suffix is used in verbs is an indication that it is no longer conventionally decomposable.

A very interesting coincidence is the similarity between *-bõka* and the noun for 'throat' *ẽbõka*, also *ẽbõkata*. One might suspect that the lexical suffix for throat may be homophonous with that for ear. Yet, for non-nominal uses, the lexical suffix for throat is *-gadẽ*. I have not found a nominal use of the affix, which may make it a pure classifier.

An interesting aspect of this is that *ẽbõka*, the shorter form of 'throat', could be mistaken for the verbal classifier construction with the verb *ẽ*, 'have', seen in (55a). Note that *ẽ* is not a general form of 'have'. There is some stem variation in that realm, which I don't completely understand, yet. The verbal construction for having a throat in (55b) uses *-gadẽ*.

(55) a. *kidi adoke õdõ·bõka ẽ-bõka-pa*

cat one ∅·LS.ear have-CLF.ear-DECL

'The cat has one ear.'

b. *iwa yẽdẽ ẽbõkata ĩ-kã ẽ-gadẽ-kã-pa*

howler.monkey big throat PRO-3.H have-CLF.throat-3.H-DECL

'The howler monkey has a big throat.'

3.3.4.6 *-de*; *-bode* : 'mouth'; 'butt'

With the collapse of /æ/ and /e/ the *-de* affix may cover the two glossed meanings for *-dæ* in Fiddler (2011), 'hemisphere' and 'tail of bird', as well as *-de*, 'edge/opening'. The *-de* affix is also homophonous with the locative.

It is definitely the case that *-de* is used for 'mouth' or 'orifice', *õdõ·de*, '∅·LS.mouth'. Further evidence that *-de* may be associated with a general 'opening' meaning is that *õtatode*, 'hole', ends in /de/. I am not complete convinced that it is a suffix in the latter case. There are many words that end in /de/ that do not appear related to the meaning 'opening', nor other proposed *-de* meanings. I would like to see a more established pattern.

The body part specific meaning is used verbally, as can be seen in (56). It is also possibly non-coincidental that the stem for 'talk' and 'language' is *tede-*.

(56) *ĩ-kã õdõ·de e ẽ-de-kã-ta-pa*

PRO-3.H ∅·LS.mouth ? have-CLF.mouth

'He still had it in his mouth.'

There are a number of bird names that contain /de/, but there are also other animal names with the same ending. There is an affix for posteriors, seen in *õdõ-bode*, '∅LS.butt' – 'back part', 'butt', 'rump', etc. – which may indicate that the 'tail of bird' meaning may be based on a misunderstanding. The rear end of a bird would be only a particular instance of a more general usage. It is possible that *-bode* can actually be analyzed as *-bo-de*, where the *-de* contributes something about 'rear'. See §3.3.4.23 and §3.3.4.24, for additional discussion of multi-syllable nominal lexical suffixes. I have no verbal examples of *-bode*.

I haven't seen the *-de* affix used for 'hemisphere'. Peeke's listing gave 'bowl' as an example of what a hemisphere meaning might denote. One could ask whether the concave or convex dimension motivated the usage. If the former, the hemisphere would form a depression, which would be like a hole or opening. There is another affix, *-kade*, 'LS.vessel', which would very likely be used for things like bowls, but I have not seen *-de*, alone, used for bowls or similar items.

3.3.4.7 *-dẽ* : 'food', 'abdomen'

The meanings 'food' and 'abdomen' are shared for the *-dẽ* form due to the vowel merger. I haven't seen the food meaning as a classifier, which may indicate homophony, rather than identity with *-dẽ* for abdomen. I have encountered three nominal food related examples. Two are very common words: *kẽ·dẽ*, '∅·LS.food', 'manioc tuber', and *peẽ·dẽ*, '∅·LS.food', 'plantain'. I have also encountered *we·dẽ*, '∅·LS.food', 'platanillo', which is likely some kind of banana-like food, though I am unfamiliar with it.

The body part meaning is productive, with verbal uses.

(57) a. *dãta-dẽ-te* *we-kã-ta-pa*
hurt-CLF.food-GER cry-3.H-PST-DECL

'She was crying with stomach pain.'

b. *õdõ·dẽ*
∅·LS.food

'stomach'

See §3.3.4.11 for information about other suffixes that may be associated with stomachs.

3.3.4.8 *-do* : 'forehead'

The affix is found in the word *õdo·do*, '∅·LS.forehead', 'forehead'. I know of no meaning other than 'forehead' associated with it. I have yet to see it used as a classifier.

I have only a written form, but since it is not written *onono*, I assume, based on the consistency of the nasalization rules, that if the affix is *-do*, as reported in Fiddler (2011), the stem must not end in a nasal. So, rather than the normal *õdõ·*, used for many body-part terms, there is *õdo·*. I believe that given the regularity of the system, otherwise, further verification is warranted.

3.3.4.9 *-dõ* : 'road', 'way', 'river', 'side'

The affix was historically related to a word for 'river'. Peeke (1968) has *oõdõ*, '∅-LS.liquid', 'river'. I have not had the word accepted by a speaker. The common word for 'river' is now *e-pẽ*, '∅-LS.liquid', which also means water, or liquid.

As pointed out to me by Tod Swanson and Pieter Muysken, the affix is a likely a portion of many place names in Ecuador. There are a number of rivers and places in the Amazonian region that have endings of *-no* or, perhaps with Kichwa influences *-nu*. They also frequently have endings of *-ro*. At least some of these are likely related to *-dõ*. Some examples are *Ahuano*, *Dayuno*, *Nushino*, and *Kiwaro*. Some place names may be coincidental.

Tod Swanson (p.c.) believes it may be possible to use the place name pattern in research on historical Wao settlements. I mention this aspect of the affix here it is not common knowledge, and may be of interest to historical research.

The affix also appears on nouns with non-river meanings, such as *taa·dõ*, '∅·LS.road', 'road' or 'path'. An older speaker, likely in their 70s, accepted *di·ka-dõ*, '∅·LS.stone-LS.road', as an offered word for a gravel path. A younger speaker, aged 36, found the usage unacceptable.

The affix is also associated with the meaning 'side'. It appears to function in a classifier-like distribution with this meaning.

(58) a. *mẽ-dõ-ga*
two-CLF.road-?
'two sides'

b. *a-ẽ* *dipẽbe* *ĩ-dõ*

see-IMP left DEM-CLF.road

'Look left.'

3.3.4.10 *-ga* : 'tooth'

The word *ba·ga*, '∅·LS.tooth', can be used for teeth and a bird's beak. A verbal example appears below.

(59) *ba·ga* *bakoo* *ẽ-ga-kã-pa*

∅·LS.tooth many have-CLF.tooth-3.H-DECL

'He has a lot of teeth.'

Peeke also associates the affix's meaning with tools. I have not been able to verify this.

3.3.4.11 *-gade*; *-kade* : 'stomach', 'jaw'; 'vessel'

There is uncertainty concerning the correct description of the *-gade* suffix. It is listed in Fiddler (2011) as both 'stomach' and 'jaw', without examples. I doubt the 'stomach' meaning.

Considering a *-gade*, 'stomach', meaning, the only examples that I have are adjectival examples with 'big', *yẽdẽ-gade*, 'big-CLF.jaw', which is used to refer to a person who is fat, or very large. There is an association with the stomach in this case, but it isn't clear whether this is coincidental, due to the fact that fat is carried in the abdominal region. The *-gade* suffix may be an augmentative, rather than a lexical suffix.

There are also two other terms that are used for 'stomach', which sound similar. The first, *kẽkade*, is for the interior of the general abdominal region. The second, *kẽkabode*, is for the exterior. There is also *kẽka*, which refers to the large intestines. The form *kẽkade* would seem like it should be related to *-kade*, 'LS.vessel', which is generally associated with vessels, hollows, and perhaps due to their structure, musical instruments. The similarity may be coincidental, or purely diachronic.

The general word for 'inside' is the same as the word I was told means 'inside of a stomach', as seen in (60).

(60) *adoke daa gi-po-ta-bo-pa kẽkade õ·ta-de*

one thorn enter-CLF.canoe-PST-1-DECL inside ∅·LS.shell-LOC

'A thorn got inside my fingernail.'

The *-kade* suffix occurs in a number of non-stomach related nominals, as in (61).

(61) *e-wẽ-kade*

∅·LS.plant-LS.vessel

'the hole in a tree where a macaw sleeps'

I have not found the *-kade* affix with verbs, but it does occur with numerals.

(62) *bẽ-kade-ga põ-dõ-ẽ*

two-CLF.vessel-? come-CAUS-IMP

'Give me two glasses.'

The vessel meaning seems related to the meaning of *-de*, while the stomach meaning seems related to the meaning of *-dẽ*. It would be worth looking at whether the *-kade* ending is always oral in documented examples.

The use of *-gade* for 'jaw' is productive in classifier uses. An example of verbal usage can be seen in (63).

(63) *õdõ·gade ao tõbe-gade-kã*

∅·LS.jaw ? break-CLF.jaw-3.H

'He broke his jaw.'

3.3.4.12 *-gẽ*; *-ye* : 'penis', 'vessel', 'spouse'; *vulva*, *vagina*

The *-gẽ* affix was listed in both Peeke (1968) and Fiddler (2011). It is listed as 'genitals' in the earlier work. The later work listed it under 'spouse', though it contained a 'genitals' gloss for an example. I believe that only the male genital meaning is correct for the lexical affix. The word provided for penis by speakers I worked with was *õ·gẽ*, '∅-LS.penis'. There is another word that was translated as vagina, which I believe is best analyzed as *õ·ye*. The nasality of the final vowel is ambiguous. I do not know how anatomically specific either term is. It may be that they both refer to the general genital region, or that they are specific to particular anatomy.

In terms of the productivity of the two genital lexical suffixes, Peeke (1968) provides a verb for circumcision, but given its age, I do not believe it stands as good evidence for the behavior of modern speakers without verification. It is likely that the affixes are used outside the two nouns provided. Given the general coverage of body-part lexical suffixes, it would be odd to have the semantic gap.

There is a homophonous morphological element *gẽ* that is associated with spouses. I do not believe it is a suffix. It can be found with or without gender, as seen in (64). I am not sure that the inclusion of *-kã* specifies a masculine spouse in all uses, though I provided the gloss of 'husband'.

(64) a. *dãtõ-gẽ*
?-spouse
'spouse'

b. *dãtõ-gẽ-dã*
?-spouse-3.F
'wife'

c. *dãtõ-gẽ-kã*
?-spouse-3.H
'husband'

If *dãtõ* were the stem in these cases, it would appear that *-gẽ* is an affix, which could be identified with the genital affix, as was claimed by Peeke. I do not believe this to be the case. I believe *dãtõgẽ* is a lexicalized compound.

At least historically, *dãtõ* was a relative pronoun (Peeke, 1968, p. 130). The meaning of *dãtõ* has either shifted since the time of Peeke's research, or the function that I have observed was undocumented. In my investigation, I have found that *dãtõ* is used with kinship terms when referencing a person of a kin-type, where no relationship to a particular individual is specified, for instance, 'the son' versus 'my son'. The pattern is shown in (65).

(65) a. *bo-tõ wẽ-kã*
1-PRO child-3.H
'my son'

b. *dãtõ wẽ-kã*
? child-3.H
'son'

c. *dãtõ tõdiya-ka wẽ-dã*
? sibling-3.H child-3.F
'niece'

In (64), *dãtõ* does not serve the same role as seen in (65). It is frozen in those cases, as demonstrated in (66), where a possessive pronoun may co-occur.

(66) *bi-to dãtõ-gẽ-dã*
2-PRO ?-spouse-3.F
'your wife'

Given the comparison with other kinship terms, *dãtõgẽ* is most likely a lexicalized compound, and it is doubtful that 'spouse' and 'genitals' share a synchronic morphological identity.

3.3.4.13 *-gẽdewa* : 'tongue'

The suffix *-gẽdewa* is a classificatory lexical suffix. There is a nominal *õdõgẽnewa*, 'tongue'. I have also elicited a verb, *dãta-gẽdewa-bo-pa*, 'hurt-CLF.tongue-1-DECL', 'My tongue hurts.' It is a long suffix, demonstrating that the commonality of one and two syllable affixes does not indicate a phonological limit.

3.3.4.14 *-gĩ* : 'string', 'hair'

The *-gĩ* suffix may not be a classificatory lexical suffix. It arguably appears in *yegĩkamẽ*, as seen in (48). It is also arguably on the ending of 'necklace', *pãtokagĩ*. It is prominent in the word for 'hair', *o·ka-gĩ*, '∅·LS.stone-LS.string'.

Like some other lexical suffixes, there are signs that the affix is confined to noun formation. In the verbal domain, other affixes are used for classificatory purposes. The verb *toki*, which is often given as *wao toki*, is a verb that can mean 'cut'. Even though the hair, and not a head is being cut, the *-ka* affix is used as the classifier in (67a). I don't know whether cutting fur or hair that is not on the head would require a different affix. Example (67b) is provided for comparison, and to demonstrate that the *-ka* is serving a classifier role and is not part of the verb stem.

(67) a. *kĩdãte o·ka-gi to-ka-bi*
why ∅·LS.stone-LS.string cut-CLF.stone-2
'Why did you cut your hair?'

b. *wao to-po-ta-bo-pa bẽpo ĩ-kã-te*
cut-CLF.canoe-PST-1-DECL father ACC-3.H-GER
'I cut the father's hand.'

3.3.4.15 *-ka* : 'stone', 'fruit', 'head'

The *-ka* affix is one of the most productive classificatory lexical suffixes. It is associated with a number of meanings, particularly related to 'stone', 'fruit', and 'head'. Some examples are listed in Table 3.10. I believe it may be associated with more meanings than this.

The affix has semantic and morphological overlap with *-bõ*. In (68), there is a word for 'eye' or 'face', *awĩka*, which seems like it should be related to the *-ka* pattern due to the ending, and the relation to heads. In that case the proper classifier is *-bõ*. The word *okabõ*, 'head', on the other hand, ends in *-bõ*, but the appropriate classifier is *-ka*. Peeke (1968) documented a word *õdõbõ*, for 'eye', which was not considered valid by my consultants. Notably, there is no 'head' noun that ends in *-ka*. The *-ka* affix usually directly

Stone	
dika	stone
awẽka	axe (historically of stone)
yõgaka	sharpener (historically of stone)
Fruit	
awaka	papaya
bẽyeka	fruit, mandarin orange
eketika	wild grapes (pourouma cecropiifolia)
kegĩka	avocado
kaka	achiote (bixa orellana)
Head	
okabõ	head
okabogata	hat, crown
okagĩ	hair
okata	cranium

Table 3.10: A listing of words with *-ka* and a gloss.

follows the root for head uses.

(68) a. *awĩ·ka* *dãta-bõ-bo-pa* *bo-to*

eye·LS.stone hurt-CLF.seed-1-DECL 1-PRO

'My eye hurts.'

b. *o·ka·bõ* *dãta-ka-te* *ĩ-bo-pa*

∅·LS.stone·LS.seed hurt-CLF.seed-GER COP-1-DECL

'My head is hurting.'

Both *-ka* and *-bõ* can be used for round things, in general, when used as classifiers. Though *-bõ* is more appropriate for smaller items. There is some overlap in the fruit domain. For instance, *bẽye-bõ* was reported to mean 'fruit'.

There are a lot of unexplained *-ka*s that one finds in Wao Terero words. Some of these may potentially be lexical suffix uses. That would imply that the meaning domain is quite large.

3.3.4.16 *-kã* : 'flesh'

The *-kã* affix is associated with flesh, meat, or the body as a whole. It is also homophonous with the sentient third person. Peeke (1968) considered the suffixes to be a single suffix. Perhaps that was true at the time, but synchronically there are two affixes, which follow two distinct, established distributional patterns, with correspondingly distinct roles and meaning domains. Despite this, due to the probable shared history of the affixes, there exist constructions where some notion of person, in a concrete sense, might be argued to be associated with a lexical suffix usage. I discuss two instances involving question words below, and discuss why they reflect idiomatic or frozen usage. At issue is my claim that lexical suffixes are associated with inanimate meanings in all cases.

Speakers have stated that *-kã* indicates that one is talking about a person. I believe this is only conventionally true for person marking. The lexical suffix meaning is conventionally inanimate. In cases such as (69), which I generated and a speaker accepted, no cannibalism is implied by the classifier use of *-kã*, but the subject should be sentient in the sense described in §3.2.1.1. That means the subject must be person-like, similar to what is implied by the use of English *he, she* or singular *they* in comparison to *it*. In the example, the lexical suffix and the person marking inhabit distinct positions, which is always the case for verbs. Though, for a simpler form, like *kẽ-kã-pa*, where there is only one suffix, the identity of the affix is ambiguous. It could mean 'He/She is eating', or it could mean 'It eats meat.' Despite this, I have seen no evidence that it may specifically mean 'It eats a person.'

(69) *kẽ-kã-kã-pa*

eat-CLF.body-3.H-DECL

'He eats meat.'

There are some apparent exceptions to my claim that person-hood is only conventionally associated with *-kã* as a person marker. A person meaning is arguably evident in examples such as (70). I do not believe that to be the case. The *who* meaning is idiomatic. In the example, the general question word, *eedo*, translated as 'which (cuál)', is compared against *eekãdo*, translated as 'who'. *Eekãdo* could plausibly have a semantic analysis of 'which person', despite the fact that *-kã*, the lexical suffix, is not conventionally associated with person-hood, elsewhere.

(70) a. *ee-do*

which-Q

'Which?'

b. *ee-kã-do*

which-CLF.body-Q

'Who?'

To be clear, the *-kã* in (70b) is not the third person sentient. An answer to *eekãdo* questions can be a first or second person, which indicates a more general set of alternatives. Likewise, one would expect other person marking to occur with *eedo* in that case. I have seen no evidence of this.

The appearance of *-kã* in (70b) is consistent with lexical suffix distributions. Lexical suffixes productively occur in *eedo* usage. See §3.3.5.5 for examples. For that reason, I associate the *-kã* in the example with the lexical suffix. The gloss is 'CLF' according to the general pattern of lexical suffix usage with *eedo*. Despite this, *eekãdo*, in its 'who' sense, does not signal any of the conventional meanings of *-kã* found in classifier uses more generally. I do not know whether (70b) could, in the right context, be interpreted as 'Which meat?' or 'Which body?'. Since the meaning of *eekãdo* is synchronically unpredictable, with *-kã* corresponding to neither of the conventional meaning domains for either affix, it is idiomatic. Yet, there is almost a pattern, when another question word is considered, indicating that there may have been a robust association with person-hood in the past.

Another example of a potential 'human being' association with *-kã* can be seen in (71). Questions with the *kĩ* root are generally translated as 'what (qué)'. The popular way of asking the equivalent of *How are you?*, is (71b), which is literally *What are you doing?* There is variation in whether speakers consider this to be the better way of asking after a person's health than (71a), where the *-kã* is absent. This could be seen as a preference to associate the *-kã* form with questions concerning a person, but there is a complication.

(71) a. *kĩ-do ke-bi*

what-Q do-2

'What are you doing?/How are you?'

b. *kĩ-kã-do* *ke-bi*

what-CLF.body-Q do-2

'What are you doing?/How are you?'

In (72) it can be seen that speakers also use *kĩkãdo* as the general form of 'what', rather than *kĩdo*. In asking about the time, there is no sense in which humanness is an issue. It is also the case that modern speakers do not accept lexical suffixes with *kĩdo* in general. Unlike *eedo*, there is no productive system. This does not mean that no speakers see a person–non-person contrast in *kĩdo* versus *kĩkãdo*, but it does indicate that *kĩkãdo* is lexicalized.

(72) *dẽke* *ki-kã-dõ* *hora* *ĩ*

sun what-CLF.body-Q hour COP

'What time is it?'

The *kĩkãdo* form, together with *eekãdo*, almost form a pattern, but not quite, given their lexicalized natures. Therefore, these examples are not convincing if one is to propose a conventional 'human-being' meaning for the lexical suffix *-kã*. The type of evidence I would find convincing is that adjectives and demonstratives with a *-kã* classifier are compatible with 'human being' meanings. That would mean that a construction like *bãdĩ-kã*, 'DEM-CLF.body', would be usable in roughly the same situations as *bãdĩ wao*, 'this person'. I have not seen this, nor any other unambiguous evidence that the lexical suffix *-kã* may be associated with animate meanings.

Relevant to the animate-inanimate distinction is *-koo* (§3.3.4.20), which as a plural affix is compatible with animate hosts.

3.3.4.17 ***-kado*** **: 'nose'**

Currently, I only have *õdõ·kado*, 'nose', attested. As to whether the suffix patterns more like *okagĩ*, 'hair', where only *-ka* is used in classifier constructions, or is more like *õdõbõka*, 'ear', where all of *-bõka* is used as a classifier, is unknown.

3.3.4.18 *-kapo* : 'knee'

The *-kapo* affix shows evidence of classifier uses. The word for knee is *õdõ·kapo*, '∅·LS.knee'.

(73) *õkõ-kapo-te*
wait-CLF.knee-GER
'wait' (E. G. Pike and Saint, 1988, pp. 163)

An older verbal use is seen in (73), which a modern speaker found acceptable, and easy to interpret. The lexical suffix indicates the posture of the waiting action.

3.3.4.19 *-kõ* : 'house'

I do not know if *-kõ* is rightly considered a suffix. I have only encountered it as part of *õkõ*, 'house'. I discussed the uncertainty I have concerning the phonological form of *õkõ* in §3.1.3.6.

3.3.4.20 *-koo* : 'cloth', 'group'

The *-koo* affix is a classifying lexical suffix. It occurs frequently in the word that means both 'clothing', 'cloth', and more specifically, 'woman's dress', *we·koo*, '∅·LS.group'. It occurs productively with demonstratives and adjectives. I have not seen it with a verb, yet.

(74) *yõwo bãdĩ-koo wẽtabõ-koo ĩ-pa*
now DEM-CLF.group black-CLF.group COP-DECL
'Now this (clothing) is dirty.'

The plural usage was described in §3.2.2. I have repeated example (25b) in (75), below.

(75) *a·wẽ-de deye-koo bõ-pa*
∅·LS.plant spider.monkey-CLF.group sleep-DECL
'The spider monkeys sleep in the tree.'

If *-koo* is correctly analyzed as a lexical suffix with its plural uses, there are unambiguous examples of its use with animate hosts. I have not investigated whether the plural meaning is compatible with adjectives or demonstratives. I have not seen it used in verbs. The notion that the plural-like meaning and the cloth meaning may have some connection was discussed in §3.2.2. There are some interesting generated adjectival examples in §3.3.5.3.2. At this point, I have too little data on the affix(es) to say whether the plural and cloth uses should be considered accidental homophones, distinct but historically related affixes, or two meanings for the same affix.

3.3.4.21 *-pa* : 'board', 'dart'

The *-pa* affix is a productive classificatory lexical suffix. The 'dart' meaning is evident in *õ·pa*, 'dart', which is generally associated with the small thin darts used in blow guns. Both the 'dart' and 'board' meanings are productive. The board meaning appears compatible with a general meaning of 'flat'.

(76) a. *ã·wẽ-pa* *wido ke-ke* *ke-kã-pa*
∅·LS.plant-LS.board throw-FUT do-3.H-DECL
'He's going to toss out the board.'

b. *bãdĩ-pa* *wiwa* *ĩ-pa*
DEM-CLF.board bad COP-DECL
'This (board) is no good.'

I do not have verbal examples, yet.

3.3.4.22 *-pade* : 'river'

The *-pade* suffix is easily recognizable in many Wao community names, such as *Geyepade*, the community where the family I worked with was from. I have seen no evidence that it is productively used for classification. The general word for 'river' is *epẽ*. There are river related nouns with *-pade*, such as *õ·pade*, '∅·LS.river', 'stream'.

3.3.4.23 *-pẽ*; *-pẽdẽ*; *-kapẽ* : 'liquid', 'arm'; 'leg'; 'elbow'

The *-pẽ* suffix is a classificatory lexical suffix. As already mentioned it appears in 'water', or 'river', *e·pẽ*, '∅·LS.liquid'. It also appears in the word for 'blood', *we·pẽ*, '∅·LS.liquid', as well as other liquid words.

The use of the affix for 'arm' (LS.arm) is also productive. The *-pẽ* sequence occurs in other lexical suffixes related to limbs and joints, which according to their verbal uses, appear to be morphologically indivisible. For instance, *-pẽdẽ* occurs in the word *õdõ-pẽdẽ*, '∅-LSLS', 'leg', which is used verbally, as see in §3.3.4.27. Such indivisibility may not always be the case. I have little data on *-kapẽ*, which Fiddler (2011) glossed as 'elbow'.

Alone, *-pẽ* meanings range over all manner of forelimbs, such as wings, arms, and forelegs. As a body-part affix, *-pẽ* occurs with verbs, as can be seen in (77a).

(77) a. *ao* *tõ-pẽ-te* *ĩ-bi* *õdõ·pẽ*

? break-CLF.arm-GER COP-2 ∅·LS.arm

'Are you cutting your arm?'

3.3.4.24 *-po*; *-depo*; *-gõ* : 'hand', 'canoe', 'instance'; 'year'; 'thorn'

The *-po* affix is a classificatory lexical suffix, with a number of associated meanings.

I am aware of only one noun that uses the *-po* affix explicitly for watercraft, *wi·po*, 'canoe'. The word appears to be related to *wi·po-ta*, '∅·LS.canoe-LS.shell', used for the base of palm fronds, which tend to be buoyant and form a hollow. *Wi·po-ta* may also be used in a joking manner for small kayaks. The use of the affix with adjectives and demonstratives is productive with the watercraft meaning. I have not encountered it in verbs.

The 'hand' meaning covers the vicinity of the hand. The generic 'hand'-word is, *õdõ·po*. It may also refer to animal paws. It is reported to also be appropriate for clusters, such as clusters of grapes, which may provide a general shape-based meaning by analogy between such cluster and hand-shapes. The affix may be used with verbs with the 'hand' meaning, as seen in (67b), repeated here as (78).

(78) *wao to-po-ta-bo-pa* *bẽpo* *ĩ-kã-te*

cut-CLF.canoe-PST-1-DECL father ACC-3.H-GER

'I cut the father's hand.'

There is also a time related meaning. An instance where it was used with a numeral can be seen in (79).

(79) *bẽ-po-ga* *be-bo*

two-CLF.canoe-? drink-1

'I drink two times.'

Such usage relates the affix to *-depo*, which is for years and the future, such as *wa·depo*, 'year'. The *-depo* affix was documented by Peeke (1968) as a lexical suffix. I have not seen other examples of *-depo*. It may be that the *-po* ending of *-depo* is coincidental. Though, another synchronic analysis of the 'year' word may be *wade·po*, which would be my first choice if it were not for Peeke's claim. The extremely limited distribution of *-depo* is why I do not consider it to be a lexical suffix.

There is an interesting instance of variable morph order when *-po* combines with *-gõ*, which is a classificatory lexical suffix for long pointed things (Fiddler, 2011). It was also glossed as a 'corn' suffix in Peeke (1968). The ending occurs with *kagĩ·gõ*, 'corn'. It also combines with the word for 'thorn', *daa*, for words such as 'tack', *daa-gõ*. I frequently use it with adjectives in generated data, indicating it is a classificatory affix. It plays a modifier role in uses with *-po* in the two forms of 'finger' seen in (80). See §3.3.6.3 for more information on the (80) examples.

(80) a. *õdõ·po-gõ* *kẽ-po-ta-bo-pa* *yebẽ-ka*

∅·LS.canoe-LS.thorn cut-CLF.canoe-PST-1-DECL machete-INS

'I cut my finger with a machete.'

b. *õdõ·gõ-po* *kẽ-po-ta-bo-pa* *yebẽ-ka*

∅·LS.thorn-LS.canoe cut-CLF.canoe-PST-1-DECL machete-INS

'I cut my finger with a machete.'

c. # *õdõ·gõ-po* *kẽ-gõ-ta-bo-pa* *yebẽ-ka*

∅·LS.thorn-LS.canoe cut-CLF.thorn-PST-1-DECL machete-INS

'I cut my finger with a machete.'

d. # *õdõ·gõ-po* *kẽ-gõ-po-ta-bo-pa* *yebẽ-ka*

∅·LS.thorn-LS.canoe cut-CLF.thorn-CLF.canoe-PST-1-DECL machete-INS

'I cut my finger with a machete.'

In (80), one can see that the two lexical suffixes are acceptable in either order on the 'finger' nouns. The examples were generated, but has proven extremely robust. The variable morph order, by itself, is interesting. I have yet to document similar cases.

Another interesting aspect of the example is that *-po* is the only valid verbal classifier. I provide only one of the 'finger' options, but I have tested that it isn't a question of order in the nominal or verb. This example demonstrates that some potential lexical suffixes are the combination of more than one suffix, and that they do not behave as a unit in a classificatory context. The example demonstrates the utility of verbal usage in determining the divisibility of potential multi-affix constructions.

3.3.4.25 *-ta* : 'shell', 'paper'

The *-ta* affix is a classificatory lexical suffix. I have had a great deal of success using it in elicitation with adjectives. It is associated with things like shells and paper. I do not have any verbal examples.

There are several body parts with *-ta* endings. The word for 'skull', *o·ka-ta*, is likely associated with the -*ka* classifier, given other examples, such as *o·ka-gĩ*. The word for 'lung' is *wĩdõ·ta*. The word for 'fingernail', or 'claw', is *õ·ta*. I have not determined if these meanings have classifier uses.

In the plant domain, there is an association with the hard shell of items. In particular, it is used for the fruit of the calabash tree, *owe·ta*, which due to their use as drinking vessels and other tools, may allow some analogical extension.

The paper association seems to have come from the use of *bẽkayõ·ta*, 'wasp nest', which also means 'paper', and 'workbook'. The word for 'student', and 'to study' is a collocation of *bẽkayo·ta* and the verb *aki*, 'to see'.

(81) *bo-to* *bẽkayõ·ta* *a-bo* *ĩ-bo-pa*

1-PRO ∅·LS.shell see-1 COP-1-DECL

'I am a student.'

3.3.4.26 *-tawe* : 'chest'

I have only encountered *õdõ·tawe*, 'chest'.

3.3.4.27 *-ti* : 'thigh'

I have a number of examples of *õdõ-ti*, 'leg'. It isn't translated in a way that makes it clear that it may only refer to the upper part of the leg.

It may or may not be usable as a classifier. In a verbal example that involves having a legs, the *-pẽdẽ* affix is used, instead of *-ti*.

(82) *tawadiya* *ẽ-pẽdẽ-pa* *bẽa* *õdõ·ti*

chicken have-CLF.leg-DECL two ∅·LS.thigh

'A chicken has two legs.'

3.3.4.28 *-to* : 'basket'

The nominal *o·to* exists in my corpus, and is translated as 'basket'. I have no adjectival or verbal examples, yet.

3.3.4.29 *-wa* : 'foot'

The *-wa* suffix is a classificatory lexical suffix. It occurs in the word for 'foot', *õdõ·wa*. It also appears in words that indicate position, such as *ye·wa*, 'below', and *ya·wa-pẽ*, '∅·LS.foot-LS.liquid', 'down river'. There are also metaphorical extensions to things related to feet. For instance, *taa·dõ-wa*, '∅·LS.road-LS.foot', is 'stairs'. In (83a), the denotation is a 'footprint'.

(83) a. *ee-kã-do* *ẽ-wa-kã*

which-CLF.body-Q have-CLF.foot-3.H

'Who's footprints (are these)?'

b. *bo-to* *tõdiya-da* *dãta-wa-da-pa*

1-PRO sibling-DU hurt-CLF.foot-DU-DECL

'Both my sibling's feet hurt.'

An example of verbal usage is (83b).

3.3.4.30 *-wẽ* : 'plant', 'pole'

The *-wẽ* affix is a classificatory lexical suffix. It is associated with trees, plants, poles, columns, and branches. The noun *a·wẽ*, may be associated with all of those meanings. Example (84) provides a verbal use.

(84) *õ·be* *a·wẽ* *pa-wẽ-ta-bo-pa*

∅·LS.territory ∅·LS.plant cut-CLF.plant-PST-gls1:l-DECL

'In the garden, I cut a pole.'

The *-wẽ* affix often results in synecdoche. A word like *kẽ·wẽ*, which one may translate as 'manioc plant', may also refer to a branch, or stalk of the plant, due to the polysemous nature of the affix.

3.3.4.31 Summary

Table 3.11 lists the affixes without reference to associated meanings. The idea is to focus on general morphotactic properties. There are three columns, one for whether there are clear nominal uses, another for adjectival uses, and a final one for verbal uses. I assume that any affix that occurs with a verb will also occur with an adjective, demonstrative, or numeral, but adjectival uses are noted because I may not have a documented verbal usage to provide evidence of classifier status. I also have a column for whether the affix is associated with body-parts, and another for whether they are associated with plant anatomy, since those qualities seem to affect the distribution of the affixes. These last two columns provide a gloss.

There are a number of notes referenced in Table 3.11.

	Nouns	Adjectives	Verbs	Body	Plant
ba	x	x			palm frond
be	x				
bẽ	x	x			vine
bo	x	x			[3.3.4.31]
bõ	x	x	x	eye	seed
bode	x			butt	
bõka	x		x	ear	
de	x		x	mouth	
depo	x				
dẽ	x		x	stomach	[3.3.4.31]
do	x			forehead	
dõ	x				
ga	x		x	tooth	
gade	x		x	jaw	
gadẽ			x	throat	
gẽ	x			penis	
gẽdewa	x			tongue	
gĩ	x			hair	
gõ	x	x			[3.3.4.31]
ka	x	x	x	head	fruit
kã	x		x	body	
kado	x			nose	
kapo	x		x	knee	
kõ	x				
koo	x	x			
pa	x	x			
pade	x				
pẽ	x		x	limb	
pẽdẽ	x		x	leg	
po	x	x	x	hand	cluster
ta	x	x		nail	husk
tawe	x			chest	
ti	x			thigh	
wa	x	x	x	foot	
we	x	x	x		plant
ye	x			vulva	
yabo	x	x			leaf
yo	x	x			cooking leaf

Table 3.11: The table contains a listing of the lexical suffixes discussed in this section. The first column is the suffix. The second contains an ‘x’ if the suffix occurs in nouns. The third indicates if it is documented with adjectives. The fourth indicates if it is documented with verbs. The fifth contains a gloss if it has a body-part meaning. The sixth contains a gloss if it has a plant meaning. There may also be a numerical reference to a note in the final column.

- The *-bo* suffix did have some evidence of usage for fruit, but it is not clear that the usage was due to a concrete meaning correspondence, or some shape-based compatibility.
- The *-dẽ* suffix is associated with plantains and manioc tubers, but I believe that the association has more to do with function, texture, and other factors, especially because plantains and manioc are berries and roots, respectively. These are radically different parts of a plant.
- The *-gõ* affix may be said to be associated with fingers, and thorns, but it does not clearly serve as a classifier for either in my data.

It should be clear that *-bõ*, *-ka*, *-po*, and *-wẽ* have certain advantages when investigating properties of the class. Namely, they are unlikely to be rejected on morphotactic grounds alone. Though I worked with a number of other affixes in my research, these provided a reliable core. These core affixes provided a helpful base of comparison. If the core affixes all work with an adjective, then an affix that does not work with the adjective is different in some way from the core affixes. It is not that the adjective is incompatible with lexical suffixes. This allows for pinpointing subclasses of affixes. Given that the core affixes all have fairly broad meaning domains, they also help in determining the extent to which a stem limits the interpretation of an affix. For instance, it is clear that a single affix may be associated with a variety of meanings. These may not intersect at some underspecified core. For instance, *-po* may be used for hands and canoes. All of the meanings for *-po* may be available when used with a demonstrative. In contrast, with the verb *dãtakĩ*, 'to ache', which takes lexical suffixes productively, only the body-part interpretation is available.

There are not very many lexical suffixes. There are likely unknown affixes. Yet, given the attention given to the language by Peeke, and to a lesser extent by myself, the set is not likely much larger. Among the relatively small list in Table 3.11, there are notably a still smaller number of affixes that can be said to be classificatory, about 23 given current documentation. Of those, it isn't clear how many have uses that extend beyond body-part reference.

Body-part meanings make up a notable core of the lexical suffix system, but it is clear that there are a significant number of affixes that are not documented as having body-part meanings. It is also clear that even those that do have body-part meanings may have additional meanings.

In the next section I discuss the ways that lexical suffixes interact with their hosts, which is highly dependent on part of speech.

3.3.5 Descriptions of host types by parts of speech

A rare aspect of the Wao Terero lexical suffix system is the number of parts of speech that serve as hosts. Though not all affixes distribute equally, there is a core of affixes that show productive use with roots that are specific to diverse parts of speech. Meanings seen with one part of speech carry over to the others. The available meanings may be constrained by the host, but there is a recognizable enumeration of meanings that a particular affix may be associated with across host contexts.

3.3.5.1 Nouns

Nouns are an important part of speech in the lexical suffix system. They have unique qualities in comparison to other parts of speech. As has already been emphasized, the meanings are compound-like, rather than classifier-like. The range of meanings available in a nominal construction may be constrained. Out of several meanings that may be associated with an affix, only one may be evident. Though, this is not always the case.

There are also bound root hosts in the nominal system, which require a specific suffix, such as *a·wẽ*, 'tree', where *a·*, as a nominal, only occurs with the *-wẽ* suffix. This extreme selectivity is not characteristic of other parts of speech. There are also bound roots that take many suffixes, such as *õdõ·*, used for many body parts. I know of no *õdõ·*-like base among other parts of speech. Despite the lexicalization these bound roots imply, lexical suffixes are productive in nominal constructions. Unlike other parts of speech, there is no evidence of competition. The norm is one or two lexical suffixes, but there is convincing evidence that this is not a hard morphological limit.

In this section I address a number of questionable claims made in the literature. Derbyshire and D. L. Payne (1990) claimed that classifiers are shortened versions of nouns, which are verbally incorporated. I show that this is not the case.

I also discuss the issue of animacy in the nominal system. There is a divide in the morphotactics of inanimate and animate nouns. Except for the ambiguous *-koo*, animates do not take lexical suffixes. Unlike inanimates, animate nouns take person marking. In Chapter 6, I consider the animacy divide to be stable enough to propose two broad form classes for the nominals.

3.3.5.1.1 Productivity and constraints on interpretation

Nouns constrain the meanings associated with lexical suffixes more than other hosts. I believe this is for two reasons. First, nominal meanings are often concrete, limiting available interpretations. In productive use, a plant-related stem will imply a plant-like interpretation of an affix. Second, nominal meanings logically, and likely diachronically, provide the source of available interpretations for a particular suffix. I say logically because, in a classifier-like usage, an affix like *-po* is associated with a plurality of meanings, such as *canoe*, *instance*, 'hand', etc. If one only knows that plurality, it is difficult to predict the single meaning associated with the noun *wi·po*, 'canoe'. Yet, in determining what may be within the collection of meanings associated with a classifier-like usage, an enumeration of specific nominal examples will be informative. Essentially, it is the difference between saying, "Some cats are gray, therefore, Garfield the cat is gray.", versus, "Garfield the cat is gray, therefore some cats are gray." The specific may inform the general, but there is no automatic mechanism to go from the general to a specific nominal meaning.

I assume that words like *wi·po* and *di·ka* are not synchronically formed by morphological processes. One can recognize a signal in their forms, but their roots are non-free and have no independent meaning. The lexical "suffix" on these words is somewhere between a phonestheme, like the English *gl*, which is associated with light due to many words like *glitter* and *glare*, and an actual suffix, due to the status of the form in other contexts, such as *yẽdẽ-po*, 'big-CLF.canoe'. It is possible that the words were not perceived as carriers of lexical suffix related signals in the past. In a process of semantic generalization of the classifier affixes, the meanings of the unanalyzable pieces may have been incorporated into the aggregate of classifier meanings. That is to say that *-po*, at one point, was associated with most of its current meanings except for 'canoe'. Then there was a reanalysis of *wipo*, where *-po* was perceived as an existing signal, similar to the perception of agentive *-er* in English *burglar* (Becker, 1993). Due to the existing flexible plurality of *-po* meanings, 'canoe' could be added to the set with no disruption of existing suffix interpretations. The quasi-agreement of Wao Terero may have encouraged this. A phrase like *bãdĩ-po õdõ-po*, 'this hand', may have encouraged an analogy to *bãdĩ-po wi·po*, 'this canoe'. As a non-historical linguist, this is not my area of expertise, but the scenario is plausible. The accrual of meanings by lexical suffixes and classifiers in diachronic analyses

	-po		*-ka*		*-wẽ*
õdõpo	hand	okabõ	head	awẽ	tree
?	grapes	bẽdeka	fruit	awẽ	pole
?	cluster	x	round	awẽ	branch
wipo	canoe	dika	stone	awẽ	plant

Table 3.12: Comparing three lexical suffixes, which have multiple meanings associated with them as classifiers, *-po* and *-ka* have evidence of documented nouns being specific only to one meaning type. The *-wẽ* affix, has a documented nominal, *awẽ* that is compatible with all meaning types. A question mark indicates that a nominal has not been documented with the affix for the meaning. An 'x' indicates that the meaning is expected to only be compatible with a classifier. The listed meanings are not, necessarily, exhaustive.

is well established (Mithun, 1984; Mithun, 1986; Seifart, 2007). There must be mechanisms for diachronic lexical suffix meaning expansion, and though the relationship between canoes and hands may have had some cultural significance, none are currently evident. Yet, the formal similarity is evident.

The notion that lexical suffix meanings logically depend on the meanings of specific nominals is not intended to imply that every lexical suffix meaning can be paired to a noun that ends in the suffix. Words that are associated with only one lexical suffix meaning, like *wi·po* and *di·ka*, represent only one nominal pattern. In Table 3.12, *a·wẽ* is compatible with all known *-wẽ* meanings.

It is not necessarily the case that every meaning associated with a lexical suffix can be pared to a nominal that hosts the suffix. In the *-po* column of Table 3.12, there are a number of meanings that have no known noun pair. Elsewhere, the noun associated with the lexical suffix meaning may not contain the lexical suffix. For instance, *-bõ*, 'CLF.seed', rather than *-ka*, is the correct lexical suffix for 'eye', as seen in (49). Though Peeke (1968) documented *õdõ·bõ*, 'eye/face', modern speakers use *awĩ·ka*, and do not recognize *õdõ·bõ*.

The *-ka* column of Table 3.12 demonstrates that lexical suffixes may not be the final suffix of words associated with their meanings. In general, the word final position is salient in Wao Terero morphology.

The *-ka* column also demonstrates that some abstract meanings of classifier-like uses have no nominal pair. The affix *-ka*, possibly due to its association with a number of prototypically round things, has a 'round' meaning when used as a classifier, which is not associated with any nominal.

At the time of writing, I have no extended listing of single lexical affix words such as *wipo*, *dika*, and

wi · po	canoe	LS.canoe
di · ka	stone	LS.stone
a · wẽ	plant	LS.plant
pei · bo	ball	LS.egg
eke · bo	egg	LS.egg
taa · dõ	road	LS.road
e · pẽ	water	LS.liquid
we · pẽ	blood	LS.liquid
bẽkayõ · ta	notebook	LS.shell
owe · ta	cup	LS.shell
ba · ga	tooth	LS.tooth
kagĩ · gõ	corn	LS.thorn
oĩ · kã	animal	LS.body
we · koo	dress	LS.group
peẽ · de	plantain	LS.food

Table 3.13: A sample of nominal roots that occur with only one lexical suffix.

awẽ. A sample listing of bound roots that take only one affix are listed in Table 3.13. Some of these items may be found to take additional affixes given a larger data set. As stated above, bound items that take only a single affix are not the result of a morphological process on a stem. I refer to the endings as affixes because they are affixes elsewhere.

There are a number of patterns that can be seen in Table 3.13. For cases such as *dika*, the lexical suffix is associated with the meaning of the whole. Synchronically, the *di·* carries no meaning that is not a meaning of the suffix. As can be seen with examples such as *peibo* and *ekebo*, an affix may occur with more than one root, even if the root may not occur with more than one affix.

A very clear stem constraint, where the root signals a type of meaning association, even if the root cannot be said to have any independent meaning, are body-part nominals formed with *õdõ·*.[5] It may seem appealing to gloss *õdõ·* as 'body'. Consider that the composition with *-po*, *õdõ·po*, would then be 'body hand'. Neither does *-po* fail to mean hand (of a body) when used without *õdõ·*. Clearly, as a bound root, *õdõ·* does not mean anything by itself. Instead, of carrying an independent meaning *õdõ·* constrains the meaning of the lexical suffix to only body-part interpretations.

No other bound nominal root is comparable in the number of lexical suffixes *õdõ·* occurs with. I provide

[5]For *õdõ ·*, I maintain the use of '·', despite a more stem-suffix-like relationship, because *õdõ ·* is still bound. Arguably, it deserves a different symbol than *di · ka* due to differing qualities.

õdõ · bode	butt	LS.butt
õdõ · bõka	ear	LS.ear
õdõ · bẽka	throat	LS.throat
õdõ · de	mouth	LS.mouth
õdõ · gade	jaw	LS.jaw
õdõ · gõ-po	finger	LS.thorn.canoe
õdõ · po-gõ	finger	LS.canoe.thorn
õdõ · ka-gĩ	tail	LS.stone.string
õdõ · kado	nose	LS.nose
õdõ · kapo	knee	LS.knee
õdõ · po	hand	LS.canoe
õdõ · pẽ	arm	LS.arm
õdõ · pẽdẽ	leg	LS.leg
õdõ · ti	thigh	LS.thigh
õdo · do?	forehead	LS.forehead
õdõ · wa	foot	LS.foot
õdõ · yabe	spine	LS.spine

Table 3.14: A sample of nouns formed with the root *õdõ ·*.

a sample listing in Table 3.14. The list is not comprehensive.

The *õdõ·* root takes a mixture of lexical suffixes that compose, as can be seen with *õdõ·po-gõ* and *õdõ·gõ-po*, as well as single affixes. It is sometimes unclear whether the affixes are combinations or not. I listed 'tail' as a combination. That may be wrong. The meaning is not transparent, but given *o·ka-gĩ*, '∅·LS.stone-LS.string', 'hair', which is a composition based on the verbal classifier diagnostic, it seemed reasonable. As discussed above, the lexical affixes seen in Table 3.14 may or may not correspond to classifiers. I included *õdo·do* with a question mark because I am uncertain about the phonological form. See §3.3.4.8 for more on *-do*.

There are other bound stems that take more than one affix. A sample is listed in Table 3.15. The *bẽye·* examples result in overlapping meanings. I do not mean to imply that the *-ka* and *-bõ* instances are perfect synonyms, but both may be translated to 'fruit'. A specific meaning of *bẽyeka* is 'mandarin orange'. The word *bẽyebõ* is associated with fruit that one may suck on.

The *kẽ·* examples map to related meanings for different aspects of the manioc plant. Remember that the word *kẽwẽ* may also be translated as 'manioc branch' or 'manioc stem'. The *te·* meanings may be historically related. Chonta palm fruit is used to make a particular kind of chicha. Chicha is a generic term used in Spanish

bẽye · ka	fruit	LS.stone
bẽye · bõ	fruit	LS.seed
kẽ · wẽ	manioc plant	LS.plant
kẽ · dẽ	manioc tuber	LS.food
te · wẽ	chonta palm	LS.plant
te · pẽ	chicha	LS.liquid

Table 3.15: A sample of nouns formed with a bound root that take more than one affix.

kẽ · wẽ	manioc plant	LS.plant
kẽ · wẽ-yabo	manioc leaf	LS.plant.leaf$_1$
†kẽ · wẽ-yeka	manioc root	LS.plant?
†kẽ · wẽ-po	manioc field	LS.plant.canoe
kẽ · wẽ-kode	manioc garden	LS.plant?
bĩgi · ka	arazá fruit	LS.stone
bĩgi · ka-wẽ	arazá plant	LS.stone.plant
bĩgi · ka-yabo	arazá leaf	LS.stone.leaf$_1$
†bẽye · ka-tapo	fruit peal	LS.stone?
†bẽye · bõ-pẽ	fruit liquid	LS.seed.liquid

Table 3.16: A sample of productive nominal lexical suffix usage.

to refer to native fermented beverages. Among current speakers, the more specific name for chonta based chicha is *dage kapẽ*, and *tepẽ* is for manioc based chicha. Perhaps it was always so, but the name similarity is suggestive of a historical relation.

In addition to the primary lexical suffix combination, secondary combinations are possible, which modify the stem as a whole in a reasonably transparent manner. I list some examples in Table 3.16. Some of the examples listed reveal possibly new lexical suffixes. I have not studied these in depth, so a meaning gloss is provided, but the analytic gloss may be left as '?'.

Not all items in Table 3.16 were checked with multiple speakers. I have put a '†' in front of some that were provided during work with only one speaker. Especially where the meaning looks fairly transparent, such as *bẽye·bõ-pẽ*, I feel reasonably certain that such an interpretation is available. There is always a question of whether a particular speaker will consider it to be a *real word*, whether it is interpretable or not. This is an important issue in exploring productive lexical suffix uses given that different speakers have different linguistic attitudes. Some speakers enjoy word play. Others feel that there is a correct Wao Terero, and that words should not be invented. Some will reasonably point out if there is a more specific word that should be used in place of a transparent combination, like *bread slicer* instead of *bread dividing apparatus*.

An interesting entry in Table 3.16 is *kẽ·wẽ-po*. The *-po* affix is listed with an additional meaning in Fiddler (2011) of ‘earth’.

There is a very strong trend of headedness in the Tables above. This needs to be taken with a grain of salt. The word *bĩgi·ka-wẽ* is ‘arazá plant’ (eugenia stipitata), but *bĩgi·ka*, through synecdoche, may be used to refer to the fruit plant, as well as the fruit – especially if the fruit plant is the mandarin orange plant. This means that even though a headed meaning may be available in some cases, the meaning of the whole may be conventionalized in a non-transparent fashion. This should not surprise many linguists, but there is a tendency to overemphasize the semantic transparency of the notion of *head* and supposed word-internal scope (Harley, 2012).

The majority of the nominal examples in my data contained bound roots. This is not a necessary characteristic of nominal lexical suffix constructions. I believe that it has more to do with lexical suffix usage being sparse, in general. The forms that necessarily use lexical suffixes are bound roots. In Table 3.17, I have a listing using the word *daa*, ‘thorn’. The examples were constructed, and placed in a large randomized listing, where many additional items, with different stems intervened. I asked for translations for each item. Some were rejected. It wasn’t always obvious what would be rejected. The fairly reasonable looking *daa awẽ* was rejected but *daa-wẽ* was accepted. Perhaps, there was a phonological explanation in that case. I included some true noun-noun compounds in the listing as a point of comparison with the lexical suffix constructions. I have no general conclusions to draw from these comparisons, except that there is evidence that meanings may vary depending on whether a true compound or lexical suffix construction is offered.

All of Table 3.17 was provided by a single speaker. There is one item that was reviewed by more than one speaker *daa-gõ*, ‘tack’.

In all data above, the number of lexical suffixes per word does not exceed two. This is not a hard limit, though the clearest examples of three or more affixes are lexicalized. I have occasionally probed whether longer forms are acceptable with transparent meanings. One speaker accepted the generated form *bẽye·ka-wẽ-yabo*, ‘∅·LS.stone-LS.plant-LS.leaf$_1$’. The form was presented in writing, without explanation, except that I wanted a translation. The immediate answer was that it meant ‘abió leaf’ (pouteria caimito). I asked if it could mean ‘fruit tree leaf’, and received a positive answer. A second speaker stated that it didn’t mean ‘abió

daa-wẽ	thorny tree	LS.plant
# daa awẽ	thorn tree	
daa-ka	fruit with thorns	LS.stone
daa bẽyeka	fruit of a thorny plant	
daa bẽyebõ	fruit with thorns	
daa-bõ	seed of thorny plant	
daa-bẽ	spiny vine	LS.cord
daa-yabo	leaf of a thorny plant	LS.leaf_1
daa-gõ	tack	LS.thorn

Table 3.17: The word *daa*, 'thorn', with lexical suffixes.

leaf', which they stated was *bẽye·ka-yo*, '∅·LS.stone-LS.leaf_2'. When asked if it could mean 'fruit tree leaf', they provided *bẽyegowẽ* as an alternative, which I do not know how to analyze. Given these mixed results, it is clear that *bẽye·ka-wẽ-yabo* is not a conventionally established word, which I already knew due to its provenience. It is also clear that there is variation in its acceptability. The variation does not completely invalidate the intuitions of the speaker who accepted the form, which I believe argues for a more organized study.

As previously mentioned, there are lexicalized forms with more than one affix. Examples include *awẽkata*, 'shoe', which looks suspiciously like a lexical suffix formation of 'LS.plant.stone.shell'. *Awẽ* is associated with wood and wood-like materials, which may be related to shoe soles. There is also the more clearly analyzable *a·wẽ-pa*, '∅·LS.plant-LS.board', 'board', which is almost certainly related to *a·wẽ-pa-do*, 'table', though I do not know what to think of the final *-do*.

I have previously mentioned that *yegĩkabẽ*, 'rope', may be *ye·gĩ-ka-bẽ*, '∅·LS.string-LS.stone-LS.cord'. One translation by Rubén Boyotai gives it as 'rope made of knots'. That the form is morphologically analyzable, even if the meaning is idiomatic, is supported by similar form-meaning correspondences, elsewhere. There is *õ·gĩ-bẽ*, '∅·LS.string-LS.cord', 'cord made from the chambira palm'. There is also *pãto·ka-gĩ*, '∅·LS.stone-LS.string', 'necklace', which pairs with *pãto·bõ*, 'bead made of a seed'. The word for 'hair', *o·ka-gĩ*, also provides some evidence that *yegĩkabẽ* is analyzable.

Peeke (1968) believed that lengthy lexical suffix sequences are possible. She presented the example *o·ka-bo-ga-ta*, '∅·LS.stone-LS.egg-LS.tooth-LS.shell', 'crown'. She described *-ga* as possibly meaning place or surface, but only in the one example. The word *o·ka-bo-ga-ta* is familiar to modern speakers, and is also used for 'cap', like a baseball cap.

There is good evidence that nouns may contain more than two affixes. It remains unclear if this can be done productively. Notably, the situation is unlike what is seen in the morphotactics of other parts of speech, where lexical suffix competition is unambiguous. Given the available examples, I do not morphotactically constrain the number of nominal affixes in the formal analysis.

3.3.5.1.2 Classifiers are not shortened forms of nouns

Derbyshire and D. L. Payne (1990) made the claim that classifier uses of lexical suffixes are shortened forms of nouns. The implication is that verbal uses of classifiers are a form of incorporation, the combination of two free forms. According to Derbyshire and D. L. Payne (1990), verbal classifiers in Wao Terero are a dominant source of classificatory usage, which extends to numeral classification as a secondary manifestation. No mention was made of other parts of speech. Implicit in the claim were the works of Mithun (1984) and Mithun (1986), which described instances of verbal classifier genesis. Those works describe a diachronic process where verb-noun compounds gradually develop into classifier systems through a series of stages. Exactly how the lexical suffix system developed in Wao Terero is unknown. The fact that verbal uses are not currently the most productive, does not eliminate the possibility that verbs may have been involved in the earliest stages of lexical suffix development. That is an open question. What can be eliminated as a possibility is that current lexical suffix usage is a form of incorporating free nouns in a morphologically shortened form.

It may seem that my claim that lexical suffix meanings are logically dependent on noun meanings would support Derbyshire and D. L. Payne (1990). This is not the case. My idea is that classificatory lexical suffixes are an established class, where individual affixes exhibit an enumeration of meanings. These meanings are informed by nouns, which plausibly contain the same affix. I do not think that *dika* incorporates to provide a 'stone' meaning to verbs or adjectives.

Clearly, nouns do not incorporate as a whole to form classifier-like constructions. There are no examples. This does not eliminate the possibility of systematic morphological shortening. The argument against this is that one would expect the process to be predictable. Consider the possibility that the productive shortening process is root deletion. This works well for a subset of the *õdõ·* nouns. By deleting the root, forms like *bõka* 'LS.ear' may be isolated, and incorporated into verbs like *dãta-bõka*, 'hurt ear'. This clearly does not work for the variant finger words *õdõ·gõ-po*, and *õdõ·po-gõ*. First, consider that the 'hurt' form is *dãta-po*, and cannot contain *-gõ* in addition or instead. Even if one were to add a condition that "modifying" affixes should be deleted, the condition could not be stated as a morphological rule that depends on morphological form patterns. A powerful morphological theory may propose a notion of deletion based on semantics. Then one has to explain why *-ka* occurs verbally in (85), rather than *-gi*. The *-ka* is either semantically irrelevant, or serves as a modifier expressing something like 'head hair'.

(85) *kĩdãte o·ka-gi to-ka-bi*
why ∅·LS.stone-LS.string cut-CLF.stone-2
'Why did you cut your hair?'

Even the concept of removing some portion of the noun based on semantic information presupposes the affix-like nature of the forms. In fact, even recognizing a root versus non-root portion of the word strongly implies that the non-root contents are affixes. The alternative is a bound root compounding analysis. For those who harbor such notions, the question is why the affixes consistently distribute in a manner distinct from *õdõ·*. The *õdõ·* root is bound, yet it never occurs in an affix-like position. Neither does *bõka* occur in a primary stem position. A productive bound root should behave in non-affix-like ways.

Even with a relatively regular pattern as is seen with *õdõ·*, there are considerable difficulties for an incorporation analysis. Considering other nominal patterns, the problems increase. For instance, *bẽye·ka* is the most generic fruit word, and could be said to shorten to *ka* to incorporate. Yet, there is also *bẽye·bõ*. It would be odd for the word to reduce to the only portion that is variable.

The argument could be made that I'm picking the wrong nouns. There may be particular nouns that incorporate, and that even the affix-like things on other nouns are just shortened versions of those. I see no evidence of this. Table 3.12 demonstrates some gaps in noun to classifier meaning correspondences. This may be due to lack of documentation in two cases, but there are some meanings that are not related to any noun, such as the roundness associated with *-ka*.

I see no other reasonable arguments for a noun-shortening, incorporation analysis.

3.3.5.1.3 Animacy and lexical suffixes

I have already described in the discussion of *-kã* (§3.3.4.16), *-koo* (§3.3.4.20), and plurals (§3.2.2) the fact that except for the *-koo* affix, there is no lexical suffix usage with animate nouns. For the *-koo* affix, I only accept it tentatively, since at the time of writing I lack evidence for differentiating a 'cloth' *-koo* from a plural-like *-koo*.

As a recap, the *-kã* lexical affix has some history of being associated with person-hood, and there is some evidence that it was once the same affix as the third person sentient. In the case of *-kã* serving as a modifier of animates, there is no evidence of systematicity, even if there are some lexicalized forms that have person-hood associations.

The *-koo* affix has a use as a plural-like affix. In that capacity, it may be used on animate nouns to express that they are a plurality, or group.

At the moment, I feel that the *-koo* examples make it impossible to firmly state that a lexical suffix will never occur with an animate noun. On the other hand, *-koo* is a highly notable exception to an otherwise regular pattern, and the very fact that the plural use deviates from the general lexical suffix pattern argues for splitting the plural and the 'cloth' affix.

In contrast to systems such as Burmese numeral classifiers (Burling, 1965), which clearly can be used with reference to animals and people, there is almost no evidence in my corpus or works such as Peeke (1968), Peeke (1991), and Fiddler (2011), that lexical suffixes are compatible with animate nouns.

There are cases where inanimate noun hosts are used to refer to people, but only as parts of collocations with verbs. An example appears in (86a), where *bẽkayõ·ta* translates to 'workbook', and the *-ta* has an established association with paper. A student is literally a 'workbook looker'.

(86) a. *bo-to bẽkayõ·ta a-bo ĩ-bo-pa*

1-PRO ∅·LS.shell see-1 COP-1-DECL

'I am a student.'

b. **bo-to* *bẽkayõ·ta* *ĩ-bo-pa*

1-PRO ∅·LS.shell COP-1-DECL

'I am a student.'

c. **bo-to* *bẽkayõ·ta-bo* *a-bo* *ĩ-bo-pa*

1-PRO ∅·LS.shell-1 see-1 COP-1-DECL

'I am a student.'

d. *bẽbẽ-bo* *ĩ-bo-pa*

grandfather-1 COP-1-DECL

'I am a grandfather.'

The nominal element in the construction above cannot be used alone (86b), or take person marking (86c). Animate nouns for people may take person marking (86d).

The nominal patterns evident in my data suggest a dichotomy. In addition to what is seen with lexical suffixes, there is a sensitivity to animacy and sentience in the inflectional system (§3.2.1). This sensitivity extends to the inappropriateness of *-dãdi* (3.PL §3.2.2) as a general plural. It can only be a plural for living things. In Chapter 6, I propose that there are two noun form classes, the animates and inanimates. They are treated as similar to inflection classes. I do not consider whether they could be said to constitute a gender system.

3.3.5.2 Verbs

Verbs are unique in the system in allowing person marking and lexical suffixes to occur in the same word. This provides further evidence that lexical suffixes are a different class than person marking. I have provided examples of verbs that demonstrate this, so I do not highlight it in following discussion. The lexical suffix corresponds to a verbal object in most cases. I have not studied the grammatical roles associated with the objects in any depth. In general, the objects are patients, which get cut or broken, or stimuli, which cause pain. The object that the lexical suffix shows concord with is always inanimate. I have not seen a lexical suffix show concord with a subject in my research, though see §3.3.5.2.4 for a possible exception from

Peeke's data.

Verbal uses of lexical suffixes are generally classifier-like. There are exceptions, and evidence of some compound-like usage. The meanings of the classifiers are often restricted in a manner similar to the semantic restrictions seen with nouns. The restrictions are influenced by meaning compatibility, but there are also verbs that may only take body-part suffixes, which cannot be motivated by the verbal meaning. Nearly all of my verbal examples involve body-part suffixes. It was only recently that I began to see productive non-body part uses with verbs. For this reason, the relevant verbs were not prominent in elicitation.

As discussed in §3.3.4, verbal usage is restrictive. Only one lexical suffix may be used per word, which delineates compound-like aggregates that are seen with nouns from isolatable suffixes. Though, there is, at least, one ditransitive that allows two suffixes. I have used the verbal restrictiveness to help determine a set of affixes that constitute the "core" of the classificatory lexical suffixes. If they can be used with verbs, they can be used anywhere.

To be clear, not all verbs take lexical suffixes. I have never succeeded in using a lexical suffix with a verb of motion, even though it is realistic to move to or from a tree, for instance.

3.3.5.2.1 Compound, idiomatic uses of lexical suffixes with verbs

An example idiomatic, compound verbal use of a lexical suffix can be seen in (87a). A speaker, without the aid of Peeke's translation of a narrative found in E. G. Pike and Saint (1988), provided both the idiomatic translation of 'listen', and a more transparent meaning, of literal ear-pulling. A possible compound-like combination may be the use of *-de* in (87b), since speaking is done with the mouth. I have no pair for (87b) using a *te·* stem alone or with another suffix, so it may be a coincidence.

(87) a. *godõ-bõka-te*

pull-LS.ear-GER

'listen' (E. G. Pike and Saint, 1988, p. 161)

b. *te·de-dãdi-ta-pa*

∅·LS.mouth-3.PL-PST-DECL

'They spoke.'

3.3.5.2.2 Verbal body-part form class

Evidence that there are verbs that take only body-part suffixes due to grammatical, rather than meaning-based constraints comes from verbs such as *kẽkĩ*, 'to eat'. This is because one can eat things that are non-body-parts, yet the verb is restricted to body-parts. Notably, *kẽkĩ*, as a verb, may also mean 'to cut', which works only as a reflexive, and carries the same semantic restriction. The meanings are differentiated by the occurrence of an explicit cutting instrument, such as a machete, within the clause. The reflexivity of the 'cut' meaning in (88a), would explain why it is specific to body-parts. I cannot cut my own leaf. The eating example (88b) is not reflexive. The action of eating a finger is not odd, since monkey is not an uncommon meal in traditional Wao communities. For this reason, it seems like some other kind of food should be compatible with the verb. Yet, a perfectly plausible scenario, of a turtle eating a manioc plant was rejected. The example below is only one of my efforts involving eggs, leaves, and more, all of which consultants felt could be eaten, but rejected as lexical suffix additions to the 'eat' verb. It is not that *-wẽ* cannot be used with verbs, as can be seen in (88d). For this reason, there is evidence for a grammatical class of verbs that only take body-part affixes.

(88) a. *õdõ·gõ-po* *kẽ-po-ta-bo-pa* *yebẽ-ka*

∅·LS.thorn-LS.canoe cut-CLF.canoe-PST-1-DECL machete-INS

'I cut my finger with a machete.'

b. *õdõ·gõ-po* *kẽ-po-ta-bo-pa*

∅·LS.thorn-LS.canoe eat-CLF.canoe-PST-1-DECL

'I ate a finger.'

c. # *ẽkete* *kẽ·wẽ* *kẽ-wẽ-pa*

tortoise ∅·LS.plant eat-CLF.plant-DECL

'The tortoise is eating the manioc.'

d. *õ·be* *a·wẽ* *pa-wẽ-ta-bo-pa*

∅·LS.territory ∅·LS.plant cut-CLF.plant-PST-1-DECL

'I cut a branch in the garden.'

3.3.5.2.3 Some Ditransitives Allow Two Lexical Suffixes

I encountered one possible counter example to the rule that verbs may take only one lexical suffix. This has since been verified by data in Fawcett (2023). In example (89), the word *gi-po-de-da* has a *-de*, 'mouth', suffix.

(89) *tobẽ-da* *õdõ·de* *õdõ·po-gõ* *pede* *gi-po-de-da*

PRO-3.F ∅·LS.mouth ∅·LS.canoe-LS.thorn ? put-CLF.canoe-?-3.F

'She puts her fingers in her mouth while eating.'

The use of two suffixes with a ditransitive does not invalidate the value of verbs as a diagnostic. The additional suffix corresponds to an additional object, rather than a compound-like cluster, as are seen with nouns. The construction is consistent with the finding that *-po-gõ* should be considered two suffixes, since only *-po* is used in reference to 'finger' in example (89).

3.3.5.2.4 Lexical suffixes are not deverbal affixes

A questionable claim of D. L. Payne (1987) and Derbyshire and D. L. Payne (1990) is that verbal lexical suffixes serve as deverbal affixes. Peeke (1968) listed a number of verbal forms with noun-like glosses. Some of these were conventional nouns. For instance *kẽkĩ*, 'food', is clearly related to the verb 'kẽkĩ', 'to eat', yet it is a highly conventional noun. The *-kĩ* suffix appears to serve as a general deverbalizer. It is often used in translations of a Spanish infinitive. The word *kẽkĩ*, for food, is so common that it is best to consider it frozen. An example with *kẽkĩ* given by Peeke (1968) was *kẽkĩ-gõ*, where *-gõ*, ‘LS.thorn’, is associated with ‘corn’ in this case, and the translation provided is ‘corn to be eaten’, but could also be ‘food corn’. The base may be analyzed as a nominal in that example. No deverbalization needs to be posited.

Other examples in Peeke’s list showed evidence of alternate deverbal morphology. For instance, the affix *-ĩ* may result in a participle-like meaning, as seen in the form provided as a translation for (90a). It is likely the same suffix as used in (90b), listed by Peeke.

(90) a. *yõga-ĩ*
sharpen-PTCP
‘sharpened’

b. *go-dẽ-ĩ-kõ-de*
go-?-PTCP-LS.house-LOC
‘in the dwelling where he had gone’

It is also the case that subordinate clauses, such as relative clauses, may have no inflectional endings, or only person marking, as in (91a). It may also be that a bare verbal form may be translated as a nominal in some cases, as in (91b). I have not seen such forms with lexical suffixes in my data. But if they occurred, it would not be evidence of lexical suffix deverbalization. Verbs without lexical suffixes have deverbal uses independently. Consider also (86a).

An interesting aspect of some of Peeke’s examples would be that the lexical suffix may be showing concord with a subject, which is interesting, but not necessarily evidence that the lexical suffix is necessary for the deverbalization. For instance, Peeke gives (91c), but it isn’t clear whether it is intended to be causative or inchoative.

(91) a. *tokode po ẽ-ta-bo-pa*
money come gather-PST-1-DECL
‘I got the money that arrived.’

b. *pe-kã-ta-pa* *yãgĩ* *dãta*
shout-3.H-PST-DECL great pain

'He shouted in great pain.'

c. *pe-wẽ*
grow-CLF.plant

'growing manioc'

Given alternative explanations and open questions, stronger evidence than the terse listing found in Peeke (1968) is required to support a deverbalization claim. It is hard to completely assess her examples without some additional context. I have seen no evidence of lexical suffixes as productive deverbalizers among modern speakers.

3.3.5.2.5 Verbal classifiers are optional

Like all parts of speech that take lexical suffixes as classifiers, classifier use is optional. The same entailments are expressed by both examples in (92).

(92) a. *bo-to* *õdõ·po* *dãta-po-bo-pa*
1-PRO ∅·LS.canoe hurt-CLF.canoe-1-DECL

'My finger hurts.'

b. *bo-to* *õdõ·po* *dãta-bo-pa*
1-PRO ∅·LS.canoe hurt-1-DECL

'My finger hurts.'

3.3.5.2.6 Final note on verbal analysis

In Chapter 6, I consider only body-part verbs in the analysis due to the fact that I have less data on verbs that are not restricted to body-parts.

3.3.5.3 Adjectives and demonstratives

Adjectives and demonstratives are extremely productive bases for classifier uses of lexical suffixes. I include them under a single heading because I have noted few differences in morphotactic behavior. As noted above, there does not appear to be any affix that occurs with verbs that cannot occur with adjectives and demonstratives. Though, I have not documented all the affixes seen on verbs with an adjective or demonstrative, I have never had an affix that occurred with a verb rejected when used with an adjective or demonstrative. There are affix meanings, and some affixes (which may not be true lexical suffixes) that do not occur with verbs, and appear to be in competition with adjectival and demonstrative classifiers.

I include in the set of adjectives positional words such as *eibe*, 'above'. Whether these are syntactically distinct categories is not known. In my data they appear to overlap in distribution with adjectives. Positional words take lexical suffixes similarly to adjectives such as *yẽdẽ*, 'big'.

The affixes that may appear on demonstratives and adjectives are not necessarily an identical set. Person marking appears on both, but I have never encountered first or second person with a demonstrative, though, this depends on how *ĩ* is analyzed. The stem *ĩ* was described as the distal demonstrative by Peeke (1968). In §3.2.1.2 I discuss collocations used for signaling the object of a verb, and gloss *ĩ* in those uses as ACC. The accusative uses allow first and second person marking. Perhaps some analysis could plausibly unify the category of distal demonstrative uses, some pronominal uses, and the accusative role. I do not and treat them as distinct homophones since they distribute in a distinct enough manner to justify division, and are not a focal point of the analysis. I do not discuss *ĩ* further in this section, focusing only on the proximal demonstrative *bãdĩ*. This is due to the fact that my efforts to perform elicitation with *ĩ* as a demonstrative too frequently ran into issues with homophone ambiguity.

Demonstratives have some additional affixes for indicating location and time, which I will not cover here. An example adjective affix that may or may not occur with demonstratives is non-'LS.jaw' *-gade*, which I proposed was an augmentative in §3.3.4.11. There are likely other affixes that are either specific to either part of speech, or are semantically incompatible with one of the parts of speech. Here I focus on the intersection of morphotactic properties within the context of the lexical suffix system, which is consistent enough that I feel it is justified to propose that demonstratives and adjectives share a parent morphological form class.

The fact that demonstratives and adjectives appear to take suffixes that are not part of the lexical suffix system is the reason that verbal usage is the best diagnostic of classificatory lexical suffix status.

Unlike verbs, lexical suffixes are in competition with person marking on adjectives and demonstratives. This is not because adjectives and demonstratives are limited to only one affix of any kind. Lexical suffixes may occur with the instrumental suffix *-ka*. This can be seen in example (93).

(93)	*bãdĩ-ka-ka*	*bẽbe-idi*	*yiga-dãdi-ta-pa*
	DEM-CLF.stone-INS	grandparent-COL	tree.cut-3.PL-PST-DECL

‘With this (ax) our ancestors cut down trees.’

In the verbal system person marking shows concord with subjects, and classifier usage shows concord with inanimate verbal objects. Demonstratives and adjectives are not transitive. They also do not have meanings that encode a variety of grammatical roles that may canonically map to inanimate or animate. For instance, an agent or experiencer is canonically animate in the various cutting, hurting and breaking verbs. The verbs that demonstrably allow lexical suffixes tend to have canonically inanimate patients and stimuli, which cause pain, or get cut and smashed. Though I have not encountered any clearly intersective adjectives in Wao Terero, there remains a uniformity in what it means to be ‘big’, or ‘dark’, whether the argument is a person, bird, or stone. For this reason, demonstratives and adjectives have fewer distinctions to signal, explaining a lack of articulation for the affix positions.

To be clear, I do not believe that verbal morphotactics are *caused* by transitivity or grammatical roles. English has both transitivity and grammatical roles, and English verbs do not have Wao Terero-like morphology. Due to history, Wao Terero verbs provide particular signals for qualities associated with their lexical semantics and syntactic category, which may or may not have correlates in other languages. My point is, those qualities are absent in adjectives and demonstratives, so it makes sense that they are not signaled. Meaning being present does not explain why verbs signal certain meanings and not others. Contrariwise, a meaning being absent *is* explanatory of a *lack* of signaled meaning.

Given the animacy divide, one would never expect person marking and a classifier to refer to a single argument, except, perhaps, for the ambiguous *-koo* affix. Yet, morphotactic competition, rather than meaning, is the more likely explanation for the pattern. This is because, like verbs, demonstratives and adjectives allow only one lexical suffix. An example can be seen in (94), where the bad demonstrative and adjective uses are paired with a speaker suggested replacement. The same rejection of *-po-gõ* as a single affix seen with verbal examples is seen with adjectives and demonstratives. Unfortunately, I did not collect an example of *-po* use with a hand meaning with *giita*. Given the nominal, the *-po-gõ* combination obviously makes sense. Rejection is based on morphotactic constraints.

(94) a. **bãdĩ-po-gõ kẽ-po-ta-bo-pa yebẽ-ka*

DEM-CLF.canoe-CLF.thorn cut-CLF.canoe-PST-1-DECL machete-INS

'I cut this (finger).'

b. *bãdĩ-po kẽ-po-ta-bo-pa yebẽ-ka*

DEM-CLF.canoe cut-CLF.canoe-PST-1-DECL machete-INS

'I cut this (hand).'

c. *bãdĩ õdõ·po-gõ kẽ-po-ta-bo-pa yebẽ-ka*

DEM ∅·CLF.canoe-CLF.thorn cut-CLF.canoe-PST-1-DECL machete-INS

'I cut this finger.'

d. **giitã-po-gõ kẽ-po-ta-bo-pa yebẽ-ka*

small-CLF.canoe-CLF.thorn cut-CLF.canoe-PST-1-DECL machete-INS

'I cut the small one (finger).'

e. *giitã õdõ·po-gõ kẽ-po-ta-bo-pa yebẽ-ka*

small ∅·CLF.canoe-CLF.thorn cut-CLF.canoe-PST-1-DECL machete-INS

'I cut the small finger.'

3.3.5.3.1 On the productivity of adjectival uses

That adjective uses are productive has been determined using a number of probes. I present the data from one here that was useful due to the large number of minimal pairs it produced. In early work, I generated an 11 adjective by 12 lexical affix listing of 132 items. For each item, a speaker was asked to provide a translation, if they could. The goal was to determine a qualitative base-line expectation for adjectival classifier productivity in the absence of an overt nominal. The lack of nominal meant that the interpretation of the affix was not explicitly given to the speaker. They needed to think about what a form *could* refer to.

Lexical suffix usage with adjectives is not frequent in corpora (Fiddler, 2011). I believe this is less likely to be true in natural conversations, but that remains to be seen. I had reason from the first to suppose that productivity in adjectives is higher than a *naive* corpus study might predict. By, naive, I mean only looking at frequency of a construction type as an indication of its grammatical salience and interpretability. Lexical suffixes are used across many parts of speech. They occur with many bases. This may reinforce productivity across the system. I believe this hypothesis is consistent with productivity measurement theories such as Hay (2001), where frequency of derivations relative to a base, rather than token or type frequency alone is indicative of productivity.

As stated above, I wished to make a qualitative assessment of morphotactic productivity, and the semantic transparency of forms. For this reason, a generated listing was ideal, since it would not be based on only those forms that are so frequent that they had already been found in small Wao Terero corpora. The generated sentences had the form *adjective*-LS *ĩpa*, where *ĩpa* can be translated as 'It is.', in English, and tended to be translated as *es* in Spanish. The listing was randomized and presented to the speaker in a spreadsheet. They were aware that I had generated the list. We worked through it together, and talked about each item. I noted any explanation that the speaker wished to add. I estimate we spent an average of 1 minute on each item. I have not reproduced the task with another speaker, but subsequent work has reinforced key findings.

Only 19 of the listed items were said to have no meaning. These were largely due to a number of understandable factors. I did not yet know that *-pẽ* surfaces as [mẽ] when following a nasal vowel, which was responsible for 9 rejections. Some uses of an intended *-be*, which I did represent as properly nasalizing, were interpreted a *-pẽ* and *-bẽ* in some instances. For this reason, I have a couple of examples of those affix meanings.

Another source of many rejections was that I did not realize that *yõbeĩ*, 'rotten', is not appropriate for leafy plant matter and wood. The meaning issue may have been responsible for another 5 rejections. I think it was also confusing that I added a final n on the written form *ñomein*, which may have thrown the speaker off. The issues with *yõbeĩ* demonstrate that stem associated meanings are relevant to adjectival lexical suffix interpretation and compatibility, though not to the extent seen with verbs and nouns, since many adjectival meanings are general, such as ‘big’ and ‘small’.

In §3.3.4.3, I brought up whether *-be* (LS.territory) is a classificatory lexical suffix with the spacial, territory meaning. In my protocol, there were only four adjectives that did not end in a nasal vowel, where the suffix surfaced as *-be*. In two of those it was rejected, for *eibe*, ‘above’, and *gobe*, ‘far’ (**eibebe*, **gobebe*). It may have simply seemed like a repetition in those cases. The other instances had problematic translations that I find too confusing to present, but the affix was not associated with territory or space in those cases.

The *-po* affix was rejected a number of times. One instance was with *wepei*, ‘red’. The word *we·pe*, ‘∅·LS.liquid’, ‘blood’, has an adjectival status when given the *-i* affix. I do not understand the affix, and have little knowledge of how denominalization works in Wao Terero. An interesting aspect of *wepẽi* is the internal *-pẽ* lexical suffix. This indicates that given some additional derivational morphology, such as *-i*, classificatory lexical suffixes may be added to a derived base that already contains a lexical suffix. This may be relevant to some patterns discussed in §3.3.5.2.4. Another instance when *-po* was rejected was with *kidabẽ*, ‘folded’. I have no idea why *-po* was rejected with either base. *Wepei* may be a dispreferred term for ‘red’. In later work, I used *obatowe*, which ends in *we*. Peeke’s list in Fiddler (2011) listed a lexical suffix *-we*, with the meaning ‘red’. I have tried and failed to verify this. The difficulty speakers sometimes appear to have with using *wepẽi*, in general, may explain rejections.

Another rejected item was also with *kidabẽ*, the written form was *kidameme*, which may have been rejected due to a repetition, but I do not know. The example (93) indicates that repetitions of form do not necessarily cause any issues.

Given the size of the set, the rejections were very few, and most had easy explanations. There are some strange translations in the mix, and some puzzles that I leave aside here, but the data are fairly consistent, otherwise. A sample of every adjective is provided in Table 3.18.

In the translations in Table 3.18, the meaning associated with the classifier is in parenthesis. This was suggested by the speaker who told me that it was something being referred to. Sometimes, he would add ‘things’ to the classifier translation, indicating that the listing was non-exhaustive.

An interesting comment on *eibe-ka* and *eibe-bõ* was that the speaker stated that it was as if there were

ãbĩbõ-ba ĩpa	dry-CLF.frond	It's dry (leaf for making house).
ãbĩbõ-pẽ ĩpa	dry-CLF.liquid	It's dry (lake bed, floodplain).
dooyẽ-bo ĩpa	long-CLF.egg	It's long (egg, things).
dooyẽ-po ĩpa	long-CLF.canoe	It's long (canoe).
eibe-ka ĩpa	high-CLF.stone	It's far. It's from the big tree (fruit).
eibe-bõ ĩpa	high-CLF.seed	It's far. It's from the big tree (fruit).
giitã-yabo ĩpa	small-CLF.leaf$_1$	It's small (leaf).
giitã-yo ĩpa	small-CLF.leaf$_2$	It's small (leaf).
giitã-ta ĩpa	small-CLF.shell	It's small (shoe, sandal).
gobe-koo ĩpa	far-CLF.group	It's far (things, animals).
gobe-wẽ ĩpa	far-CLF.plant	It's far (pole, plant).
kidabẽ-pa ĩpa	folded-CLF.board	It's folded (board).
dãbẽta-pa ĩpa	white-CLF.board	It's white (spear).
yẽdẽ-pa ĩpa	big-CLF.board	It's wide (board).
yõbeĩ-koo ĩpa	rotten-CLF.group	It's rotten (meat, fish).
teẽbõ-po ĩpa	hard-CLF.canoe	It's hard/dry (tracks/footprints).
wẽtãbõ-pẽ ĩpa	black-CLF.liquid	It's dirty (water).
wepei-ka ĩpa	red-CLF.stone	It's red (fruit, things).

Table 3.18: A sampling of a task where a speaker provided translations of 132 adjectival items.

two trees, and the fruit was in the taller one. I indicated this intention by using *the* in the translations. As is common among speakers of Spanish in the Amazonian region where I worked, articles were often absent in their Spanish, likely influenced by the lack of articles in Wao Terero and Kichwa. Tod Swanson (p.c.) believed that in some cases this was due to not being native speakers. Pieter Muysken (p.c.) felt that it was a developing Amazonian Spanish dialect.

The *gobe-koo* example involved some commentary. The speaker said that it would be an answer to questions such as 'Where is the clothing?', or 'Where do the wild boars live?' This solidifies the notion that the *-koo* affix acts as animate-compatible plural.

The data from the translation exercise was extremely valuable and demonstrated that speakers find adjective-classifier combinations easy to interpret. This finding is reinforced in protocols using other methods with other speakers. There was one speaker who sometimes rejected decontextualized examples such as those in Table 3.18. She has stated that she would like an explicit nominal. I believe that brute force translation protocols are not ideal in revealing aspects of the context that are important in natural speech, and that her rejections reflect this. A common approach in my elicitation work is to provide pictures or other contextual cues. I talk more about such methods in §3.3.6, and discuss other instances where speakers felt that

awẽ	tree, plant, pole
awẽkata	shoe
bẽkayõta	notebook
daa	thorn
dika	stone
ebo	airplane
ekebo	egg
okabogata	hat
oyo	large leaf used for cooking
peibo	ball
tepẽ	chicha (fermented drink)
wekoo	clothing, dress
wipo	canoe
õyabo	leaf

Table 3.19: The nouns used in the adjective-noun translation task.

more explicit context was necessary. I stand by a the brute force approach of working through large listings of generated content, as well as engaging in meta-linguistic discussions with consultants. What is important is being aware of limitations of such techniques, and using them within a multi-method framework, where the goal is to have findings converge, and not to rely exclusively on any particular method.

I have not executed similar protocols for demonstratives, but have seen no evidence that they have more restricted productivity.

3.3.5.3.2 Adjectives with noun arguments

Due to my focus on anaphoric properties, I often used adjectives in protocols without explicit noun arguments. I have investigated basic behaviors of adjectives with classifiers relative to noun arguments. I had a particular interest in determining the extent to which the system behaved like a noun class system with grammatical agreement. In another translation protocol, I chose 14 nouns, and 12 endings with the single adjective *yẽdẽ*, 'big', to generate 168 sentences of the form *yẽdẽ*-LS *noun ĩpa*. The listing of nouns is in Table 3.19. The listing of affixes intended is in Table 3.20. I say "intended" since it is clear that the speaker interpreted some of the affixes differently.

As in the prior task described, I worked through the randomized listing with a speaker and asked for translations. In this case, the majority (107) were considered meaningless. I do not believe that the rejections

koo	LS.group
ka	LS.stone
be	LS.territory
pẽ	LS.liquid
yabo	LS.leaf$_1$
wẽ	LS.plant
pa	LS.board
bo	LS.egg
yo	LS.leaf$_2$
ta	LS.shell
po	LS.canoe
ba	LS.frond

Table 3.20: The suffixes used in the adjective-noun translation task.

serve much purpose in describing the findings. The vast majority were obviously strange. For instance, *yẽdẽ-wẽ tepẽ ĩpa* uses the plant affix with the word for a liquid beverage. In fact, no *tepe* sentence was accepted, which indicates that the adjective noun combination, independent of classifier usage, was an issue for the speaker. Another source of many rejections is that I did not understand how *-pẽ* surfaces when following a nasal vowel.

Below, I provide only a subset of the accepted phrases. I emphasize sentences that matched with later findings. The major takeaway is that the system does not look like grammatical agreement. There is a clear semantic matching pattern.

(95) a. *yẽdẽ-ba* *o·yo* *ĩ-pa*
big-CLF.frond ∅·LS.leaf$_2$ COP-DECL
'It's a big leaf. (It refers to a leaf to make a house.)'

b. *yẽdẽ-ba* *õ·yabo* *ĩpa*
big-CLF.frond ∅·LS.leaf$_1$ COP-DECL
'It's a big leaf.'

c. *yẽdẽ-yabo* *o·yo* *ĩ-pa*

big-CLF.leaf$_1$ ∅·LS.leaf$_2$ COP-DECL

'It's a big leaf.'

d. *yẽdẽ-yabo* *õ·yabo* *ĩ-pa*

big-CLF.leaf$_1$ ∅·LS.leaf$_1$ COP-DECL

'It's a big leaf.'

e. *yẽdẽ-yo* *o·yo* *ĩ-pa*

big-CLF.leaf$_2$ ∅·LS.leaf$_2$ COP-DECL

'It's a big leaf.'

f. *yẽdẽ-yo* *õ·yabo* *ĩ-pa*

big-CLF.leaf$_2$ ∅·LS.leaf$_1$ COP-DECL

'It's a big leaf.'

The examples in (95) demonstrate compatibility between the various leaf-related suffixes and different leaf words. There are times when a speaker will reject some matches, and state that they prefer that *-ba* goes with *õba*. Yet, the *-yabo* affix is fairly general, and can usually be used for any leafy thing. I believe the differences between the affixes would come into relief if one were attempting to contrast different leaf types. The important thing is that the leaf affix to leaf word pattern does not look like grammatical agreement, but semantic concord.

(96) a. *yẽdẽ-bẽ* *eke·bo* *ĩ-pa*

big-CLF.cord ∅·LS.egg COP-DECL

'It's a long egg.'

b. *yẽdẽ-bẽ* *di·ka* *ĩ-pa*

big-CLF.cord ∅·LS.stone COP-DECL

'It's a long stone.'

In (96) there are two examples that I analyze as involving the *-bẽ* (LS.cord) affix due to *length* being invoked in the translation. The classifier affects the dimension of largeness in this case.

(97) a. *yẽdẽ-bo daa ĩ-pa*
big-AUG thorn COP-DECL
'It's a big thorn.'

b. *yẽdẽ-bo di·ka ĩ-pa*
big-AUG ∅·LS.stone COP-DECL
'It's a big stone.'

c. *yẽdẽ-bo we·koo ĩ-pa*
big-AUG ∅·LS.group COP-DECL
'It's fatty's clothes.'

The examples in (97) were analyzed as involving the augmentative, which does not occur in verbs as a lexical suffix. It has no clear effect on the interpretation of the sentence. Perhaps, it would be useful in a comparative context, to indicate that something is bigger. In the last of the three augmentative examples, I provide an instance where the adjective is interpreted as referring to a fat person, who owns the clothing.

There are a number of instances where the adjective was interpreted as a possessor. In some cases this was due to a homophonous name.

(98) *yẽdẽka ebo ĩ-pa*
ñenenka airplane COP-DECL
'It's Ñenenka's airplane.'

In (99), it is shown that the *-koo* affix on the adjective is interpreted as a plural. For the clothing word in particular, it is interpreted as singular.

(99) a. *yẽdẽ-koo* *ebo* *ĩ-pa*
big-CLF.group airplane COP-DECL
'They're big planes.'

b. *yẽdẽ-koo* *di·ka* *ĩ-pa*
big-CLF.group ∅·LS.stone COP-DECL
'They're big stones.'

c. *yẽdẽ-koo* *we·koo* *ĩ-pa*
big-CLF.group ∅·LS.group COP-DECL
'It's big clothing.'

There were instances where the relationship between the affix and the argument was not one of concord. Example (100a) looks like the possession examples from above. Example (100b) has the opposite relationship implied. I would like to look for more examples like (100b). I'm not convinced by the translation. I think that it may be that since a shell is an intimate property of an egg, it may be better translated as 'It's a big shelled egg.' It could be that the Spanish interferes with a more direct translation.

(100) a. *yẽdẽ-wẽ* *õ·yabo* *ĩ-pa*
big-CLF.plant ∅·LS.leaf_1 COP-DECL
'It's the leaf of a big tree.'

b. *yẽdẽ-ta* *eke·bo* *ĩ-pa*
big-CLF.shell ∅·LS.egg COP-DECL
'It's the shell of a big egg.' (Orig. 'Es cascara de huevo grande.')

Notably, in the possession instances, I believe one could argue that a compound-like meaning, rather than an orthogonal classifier meaning exists, though a more contextually rich protocol would need to be used to determine this.

The examples above are nice because they are simple minimal pairs, but due to the methods used, one may ask if affix-noun mismatches occur in natural speech. I have too few examples in corpora to say how frequent affix mismatches may be. Yet, I have used sentences of these kinds frequently in my research on anaphoric properties. The speakers reject cases where they do not think affixes are compatible. Patterns of acceptance and rejection are reproducible. There are more examples in §3.3.6.

3.3.5.4 Numerals

Numerals are productive lexical suffix bases. Like adjectives, demonstratives, and essentially all parts of speech except verbs, the suffixes are in competition with person marking. Numerals may have a noun-like ability to take some multi-affix lexical suffix patterns. There may be a question of whether Wao Terero numerals are true numerals, given that there are only two, but the ability to use them to count indicates that they are signals in a one-initial base-two system. There are also some number terms that are deverbal constructions, which use the affixes for hands and feet. Aikhenvald (2000), despite the absence of any such claim in Peeke (1968), her only primary source, claimed that lexical suffixes in numerals are infixes. This claim is revisited.

(101) a. *ado-bo-ke kẽwẽ-bo-pa*
one-1-LIM live-1-DECL
'I live alone.'

b. *ado-ka-ke di·ka dẽ-bo-pa*
one-CLF.stone-LIM ∅·LS.stone have-1-DECL
'I have one stone.'

c. *bẽ-koo-ga dẽ-kã-pa*
two-CLF.group-? have-3.H-DECL
'He/she has two (dresses).'

In (101), the first example shows a use with person marking. With the *ado-ke* construction, these are often translated as 'alone', 'solo'. The second examples is a simple classifier usage with *ado-ke*. The third is a similar example with *bẽ-ga*. In the last example, I was told to place the word for 'dresses' in parentheses,

1	adoke
2	bẽa
3	bẽa go adoke
4	bẽa go bẽa
5	bẽa go bẽa go adoke
5	ẽbepoke
10	bẽa go bẽa go bẽa go bẽa go bẽa
10	tipẽpoga
20	tipẽpoga go tipẽwa

Table 3.21: Different ways of saying numbers in Wao Terero.

since it was not explicitly mentioned.

The final *-ga* in *bẽ-* constructions does not always appear to be necessary. The neutral word for 'two' is *bẽa*. Two women can be referred to using *bẽ-dã* (two-3.F). I have seen *bẽ-ka* (two-CLF.stone) used for two stones, or fruit. There are also cases where *-ga* might be expected but *-ka* occurs, such as *bẽ-kade-ka* (two-CLF.vessel-?). The two uses have some significant unknowns. It is possible that *-ga* is something like an inanimate dual, or some other indicator of paired items.

The one formation, *ado-ke*, never appears to occur without *-ke*, which also occurs widely as the limitive.

There is at least one instance of a multi-suffix construction that may be used with numerals. There is a word *tigi·ta-bõ*, '∅·LS.shell-LS.seed', 'coin'. The *tigi·* root may be onomatopoeia. The affixes suggest a composition, since *-bõ* is associated with small round things, and *-ta* with hard shell-like things. The affix combination may be used with *ado-ta-bõ-ke*, and *bẽ-ta-bõ-ke*, for 'only one cent/coin' and 'only two cents/coins', respectively. I have not yet been able to use *-tabõ* with anything other than nominals and numerals. This may indicate a slightly more nominal behavior for numeral lexical suffix morphotactics.

Table 3.21 demonstrates that one can use the base-two system to count. I do not know the provenience of *ẽbepoke*, *tipẽpoga*, or *tipẽwa*. I believe they are deverbal constructions. The *-po* and *-wa* are definitely transparent to Wao Terero speakers, who have explained that five means 'just one hand', ten 'two hands together', and twenty 'two hands and two feet'. I do not know if the words were an SIL invention for educational purposes or not. Perhaps *-po* (LS.canoe) came to be associated both with hands and instances due to the hand's role in counting. The use of *-ke* on *ẽbepoke* is consistent with the general limitive use. The use of *-ga* on *tipẽpoga* reinforces the idea that some kind of pair meaning is associated with the affix.

Aikhenvald (2000) made the claim that the numeral uses of the lexical suffixes were infixes. The implication would be that there are two kinds of lexical affix processes, or that an item and arrangement system would need to do something distinct based on the lexical category of the base. I do not want to get into the weeds in terms of definitions of infixes. I don't believe it is necessary in this case. There are reasons to doubt that *bẽ-* requires the *-ga* affix. As noted above, sometimes it does not occur, or a different affix occurs. There are also instances where *-ga* occurs that do not have *bẽ-* as a stem. It is true that *ado-* appears to require *-ke*, but the limitive *-ke* is a productive affix. I see no reason to treat it as anything other than a word final suffix with *ado-*.

3.3.5.5 Quantifiers and question particles

I am combining the sections on quantifiers and question particle because I have only a few examples, and have not studied either part of speech in detail. In §3.3.4.16, I already discussed the use of *-kã* in *eekãdo*, 'who'. I won't repeat that here. Though Peeke (1968) discussed productive use of classifiers with the 'what'-root, *kĩ·*, I have seen no examples except for the previously mentioned used of *-kã*. Only the *ee·*, 'which', stem appears to productively take the suffixes. A simple use is seen in (102), with a pair that demonstrates that it is not obligatory.

(102) a. *ee·ka-do* *di·ka* *wa* *ĩ*
which·CLF.stone-Q ∅·LS.stone good COP
'Which stone is good?'

b. *ee·do* *di·ka* *wa* *ĩ*
which·Q ∅·LS.stone good COP
'Which stone is good?'

I have found one instance of a possible two affix use with *ee·*. Two separate objects appear to be referenced.

(103) *bẽye·ka* *ee·wẽ-ka-do* *we*

∅·LS.stone which·CLF.plant-CLF.stone-Q fall

'From which tree did the fruit fall?'

At the time of writing, I feel that it is uncertain whether quantifiers take lexical suffixes. There are instances such as *ba·koo*, 'many', which may be analyzed as containing *-koo*, 'LS.group'. Other examples that I have found are too questionable to present. I may lack examples simply due to a lack of focus on quantifiers.

3.3.6 Lexical suffixes and anaphoric properties

At the moment there are many open questions concerning the nature of the relationship between classifier constructions and context. There does not appear to be any strong evidence that classifiers signal the properties of entity anaphora, such as definite descriptions, demonstratives, or pronouns. Lexical suffix usage, in general, cannot be said to be anaphoric, due to the use of the suffixes on ordinary nouns. This clearly establishes that the locus of an anaphoric signal is not the lexical suffix, itself. The extent to which lexical suffixes might be claimed to signal anaphoric properties depends on the construction they occur in.

The meanings of classifier constructions are generally complex. For instance, a verb will have tense, valency, mood, grammatical roles, and more, any of which may potentially interact with a classifier meaning. Such complications are considerable, and I do not expect them to be homogeneous across parts of speech. For instance, verbs differ from demonstratives. Nor do I expect a particular part of speech to exhibit only one pattern. Adjectival expressions with nominal arguments may differ from adjectival expressions without nominal arguments. For this reason, each kind of construction type must be looked at in turn. I focused primarily on adjectives without nominal arguments. They were considered optimal for a number of reasons. First, for adjectives, a nominal argument would allow for local semantic concord, I wished to study phenomena where the local context was less informative. Second, for nouns, the compound meanings do not allow for any special, orthogonal role for the lexical suffix. Third, demonstratives are anaphoric by definition, which presents a number of issues, one of which is the inability to test whether classifiers are compatible with non-anaphoric uses. Fourth, numerals are too simple, with only two options. Finally, verbs were too complex. There would have been too many variables to control for.

Within the adjectival system, there is good evidence of classifier discourse dependencies. If an adjectival construction is about some discourse element, and contains a classifier, the classifier meaning should match with whatever the adjective is about. I say *about* to be neutral, since the adjectival construction might be relevant to some topic, without referring to an established entity. In English, someone can say, *I saw a little tree on someone's roof yesterday.* Someone might follow up and state, *Yeah, it is a marketing ploy. Last week I saw a big one on that new building downtown.* In Wao Terero, one may be able to use *yẽdẽ-wẽ*, 'big-CLF.plant', for uses similar to *big one*. In such cases, one would expect the classifier to match some salient, thematic property. This kind of indefinite usage is characteristic of a family of phenomena variously called *one-anaphora*, *null-head anaphora*, and *noun ellipses* (Nerbonne, Lida, and Ladusaw, 1989; Nerbonne and Mullen, 2000; Gardiner, 2003; Gũnther, 2011; Khullar, Bhattacharya, and Shrivastava, 2020). I am still investigating whether this is the right analysis for Wao Terero adjectives. I believe I will have success. For that reason, I say *about*, rather than *referential to*. Adjectival expressions, may refer to entities, which is not inconsistent with *one-anaphora*-like patterns, which are neutral to definite status in English. The discourse sensitivity that a classifier demonstrates in such cases is clear. Referring to a rock with a plant suffix doesn't make sense, and one can easily demonstrate infelicity in such cases.

Before getting in to the Wao Terero data, some clarifications need to be made. The word *anaphor* is thrown about loosely in the classifier literature, and does not seem to intersect with sound observations from formal semantics and pragmatics. I do not plan to invoke any detailed theory. In fact, I am explicitly avoiding a deep engagement with formal semantic theory in this section, so as not to get too far off course from purely descriptive goals. Despite this, there are some properties of anaphora that one must acknowledge in any coherent study. The next subsections explain what it would mean for a classifier to signal anaphoric status, rather than simply occur in anaphoric expressions. Note, for nouns we already know that lexical suffixes do not signal anaphoric status. *Dika*, 'stone', does not have a non-anaphoric pair. Also, it is clear that lexical suffixes are not anaphoric signals in demonstratives. The demonstrative, itself, is the signal. Yet, one might imagine that with some other part of speech, such as an adjective, the anaphoric status might be signaled by whether there is a classifier suffix or not.

3.3.6.1 Terms and concepts for talking about anaphora

In this subsection, I introduce some terms to make it easier to talk about anaphora. Some terms are necessary to provide a coherent description. The first of these is proffered content, which is semantic content presented for acceptance or rejection by a speaker, and may be canceled under negation (Roberts, 2012). For an anaphor, the proffered content is an anaphoric linkage. Anaphoric linkage isn't a technical term, but a descriptive one that doesn't assume formal theory. Consider a scenario where I am talking to the police about a hit-and-run on my street, and I say, *Yup, I saw the car.*. A nosy neighbor might yell out *That's not true!*. Assuming their objection is the expression *the car*, the negation does not entail that I didn't see a car. The neighbor could follow up with, *Noah saw a blue one.* What is called the *descriptive content*, signaled by *car*, is not canceled. Neither does the negation cancel the *familiarity* (Heim, 1983) of the discourse referent. It is still understood that there is some car that the interlocutors are referencing. The negation cancels only the anaphoric linkage. Whatever I saw, according to my neighbor, it didn't match with what the previous discourse established as the familiar car.

The way that proffered content behaves under negation allows for a number of revealing probes. I describe two.

First, it allows for delineations of meaning, which may not be obvious if one is focused only on a formal unit. When used with English pronouns, we can see that they are multi-meaning elements, similar to definite descriptions, despite the lack of multiple formal elements. If I say, *I saw him.*, and it is negated by my busybody neighbor, neither the maleness of the referent, nor the idea that I saw someone masculine is canceled. The delineation provided by negation tests reveals that anaphors, and other types of content, may have multiple, orthogonal component meanings. Things like *car* and the masculinity of a pronoun are both descriptive content.

Second, behavior under negation is an important diagnostic for anaphoric properties. As described above, when a linkage is contradicted, more than the descriptive content fails to cancel. There are also a number of other characteristics of the anaphor that remain intact, such as the notion that there is some familiar referent. Something can be false without invalidating familiarity. When familiarity is violated, the result is something other than *false*. Imagine someone calling general campus tech support. After a standard greeting, they are asked *Why are you calling the IT helpdesk today?* The caller responds *I clicked on the button, but it wouldn't* *open.* The statement isn't necessarily false. Despite this, the tech support employee would likely feel that there is information missing. Saying *That's not true!*, doesn't get at what is wrong. The issues is that there is no familiar referent for *the button*, or *it*. It isn't falsity. It is infelicity. By looking at what is or is not canceled under negation, one can determine properties of an expression, which may not be obvious otherwise. When familiarity holds, one may be able to assess a question of truth or falsity given a negated statement. When familiarity fails, the sentence simply cannot be interpreted.

I do not go any further into a theory of anaphora than this. The above is enough to discuss my data. I have introduced the terms proffered content, the concept of an anaphoric link, descriptive content, and familiarity. Below, I also talk about *anaphoric signals*, which aren't meanings, but an overt grammatical element that signals an anaphoric linkage, for instance, a word like *the*.

In the next subsection, I discuss what does or does not constitute evidence of an anaphoric signal.

3.3.6.2 Evidence and non-evidence of an anaphoric signal

The occurrence of a classifier in an anaphoric expression does not mean that the classifier is an anaphoric signal. Otherwise, all English nouns would be anaphoric signals, since all English nouns can occur in definite descriptions. There are works that discuss anaphoric properties of classifiers in the literature (Mithun, 1986; Downing, 1986; Aikhenvald, 2000), but I have never seen evidence of a classifier serving as an explicit anaphoric signal.

Classifiers are always referred to as verbal, adjectival, numeral, etc. Part of their nature is that they depend on some larger construction type. There is, therefore, a fundamental issue that must be overcome when looking at classifiers as anaphoric signals. One must always take into account properties of their hosts. Classifiers may be obligatory or non-obligatory. For instance, Burmese (Burling, 1965) requires classifiers to be used with numerals. I do not know the facts for Burmese, but if there were a language where an expression with only a numeral and its classifier could serve as an anaphor, it would not be trivial to make a case that the classifier was the anaphoric signal. This is because the classifier always occurs with a numeral, and a numeral with a classifier. In cases such as Wao Terero verbal classifiers, where classifier use is optional, one would need to show that some anaphoric property is absent if the classifier is absent, and only exists when the classifier is present.

Argumentation in Mithun (1986) and Downing (1986) that classifiers have anaphoric properties relies largely on demonstrating that classifiers repeat throughout a text. A key piece of evidence is that a more explicit nominal introduces a referent before the classifier is used later. I find this inadequate, since it isn't clear to what extent the classifier is necessary to signal an anaphoric status for a construction. It may be that it is only used for descriptive content.

There is evidence in the psycholinguistic literature that the tendency to use more specific, or detailed, descriptive contents early, and less specific descriptive contents later, has motivations that are independent of signaling anaphoric status. Almor (1999) looked at pairs of sentences in English, both containing definite descriptions that were expected to refer to the same antecedent. Their expectation, which was supported by the study, was that processing is facilitated when the second definite description has a less specific meaning than the first, and that there is a cost in assimilating equally specific or more specific information for an established antecedent. For instance, when the first reference is more specific, such as *the robin*, and following references are less specific, such as *the bird*, they found that the processing load was lower than when the first reference is less specific, or the references are equally specific. The key point is that the use of more general terms following more specific terms were found to be beneficial to processing when an anaphoric signal *the* remained constant. One cannot say that a more general descriptive content signals anaphoric status *because* it follows a term that has a more specific descriptive content.

I do not know of any experimental works on non-entity anaphor patterns, but the Gricean Maxim of Quantity (Grice, 1975) would predict that in clear felicitous discourse that the trends noted in Almor (1999) should follow for non-anaphoric references to thematic properties in discourse, as well. That is to say, that after stating someone states 'I saw a little tree.' a follow-up by an interlocutor is less likely to be perceived as flouting the Maxim if they say 'Too bad you didn't see a big one.', than 'Too bad you didn't see a big tree.' This is all to say, classifiers, as more general, less specific terms, in general, would be expected to be used following more specific information, independent of entity reference.

I do not wish to claim that classifiers in other languages do not signal anaphoric status. The point is that the necessary evidence is usually missing.

Peeke (1991) claimed that classifier usage was anaphoric in Wao Terero. Her single example was illustrative of issues one may find in claims of anaphoricity in the general literature. The example is seen in

(104). The clear issue is that a demonstrative is used. If a demonstrative is used, the anaphoric signal is carried by the demonstrative, itself. Perhaps, the classifier could be said to provide an additional signal, but it is difficult to see how this could be measured when the demonstrative alone is always anaphoric.

(104) *bãdĩ-koo* *dooyeo-koo* *bẽ* *õyõ-pa*

DEM-CLF.group long-CLF.group exist lie-DECL

'This long one (dress) lays (here).'

In the next subsection, I discuss some initial evidence that informed my thinking on the role of familiarity in Wao Terero classifier uses.

3.3.6.3 Strong familiarity

As discussed above, anaphors require some familiar referent. Familiarity may be supplied in a variety of ways. One conception of familiarity is *strong familiarity* (Roberts, 2003), which requires an explicit previous mention. Initial data on the verbal system lead me to believe that it may be the case that if not *strong familiarity*, some very explicit signal of what was being referenced was needed for felicitous classifier use. In the end, I found the initial evidence unconvincing, but I feel that it has descriptive value, and points to methodological issues that need to be accounted for. For these reasons, I present it here.

I presented a scenario to four speakers where two friends are talking on the phone. I often emphasize phone conversation when I want to ensure that the speaker understands that information must be passed back and forth linguistically. In the scenario, one of the individuals is injured, and the other asks why. Various response options were presented The speaker was asked whether they were acceptable. In (105), a fraction-like notation is used to indicate the number of rejections out of the four speakers. Each speaker was presented the protocol in complete separation from others. Only the first speaker was presented the stimuli in an order other than presented here, where the least specific answer is provided first. One speaker was provided (105c) before (105a) and (105b). She rejected (105b), despite the additional clue to the intended meaning.

(105) a. $\frac{\#}{4}$ *kẽ-po-ta-bo-pa*

cut-CLF.canoe-PST-1-DECL

'I cut my hand.'

b. $\frac{\#}{2}$ *kẽ-po-ta-bo-pa* *yebẽ-ka*

cut-CLF.canoe-PST-1-DECL machete-INS

'I cut my hand with a machete.'

c. *õdõ·po* *kẽ-po-ta-bo-pa* *yebẽ-ka*

∅·LS.canoe cut-CLF.canoe-PST-1-DECL machete-INS

'I cut my hand with a machete.'

The speakers found (105a) odd, since it implied that some kind of hand eating was taking place, which didn't seem to answer the question. Notably, the first two speakers I worked with disliked (105b). The first felt that it was too vague. She knew that a hand related thing was being talked about, but it seemed unclear to her. The second speaker to reject it didn't feel like he could explain what bothered him, and wasn't completely committed to his rejection. When these speakers had an explicit body part mentioned, as in (105c), they found the answer completely understandable. I was expecting to continue to see the same pattern a year later with two brothers of the first speaker. Yet, they had no problem with (105b). To them it was completely clear.

The data above was tested with different body parts, as well. I have used some of that data in previous examples.

What I suspect, though I cannot prove, based on examples such as (105), is that the availability or lack of information provided by discourse is important. Despite the importance of context, some speakers are more ready to fill in gaps than others.

Given that lexical suffixes are fairly vague in their meanings at times, I provided a protocol for one of the brothers, Rubén Boyotai, to see how he would respond to a scenario with little information. I described this in §3.3.4.1 in the description of (43).

The context for (43) involved two individuals. The first has a bag, and the second can't see what is in it. The second individual asks what is in the bag, and receives a single word response. The speaker consultant is asked if the response is acceptable. If the consultant accepts the response, they are then asked what they think is in the bag. I only performed the protocol with one speaker, though translations of the classifier constructions were verified with other speakers. When asked what was in the bag, the consultant's immediate answer was 'fruit.' To follow up, I asked whether the contents of the bag may have been other things, such as eggs, or balls, which were both stated to be compatible with *bẽboga*.

Example (43) is repeated here as (106).

(106) a. # *bẽ-bo-ga*
two-CLF-?
'two leaves'

b. *bẽ-bo-ga*
two-CLF-?
'two fruit'

The quickness with which the speaker jumped to an answer of 'fruit', is notable. Given that they accepted a wide variety of other items after the initial answer indicates that one might have assumed they would not have a ready answer, except, *It depends.* Clearly, the speaker filled in some background from world knowledge or linguistic frequency. It is similar to saying *He bores easily* in English, and likely finding that people assume one is talking about a person's quick disinterest, rather than their facility with a drill. Yet, I feel that it is different from how an English speaker would respond to an answer like *It is.*, for *What's in the bag?*. The lack of familiarity would simply be confusing. Lexical suffixes may have a wide number of meanings, but they are still rich enough in information that speakers appear to be able to accommodate their vagueness.

Rubén Boyotai also provided much of the data in §3.3.5.3. He was consistently ready to provide possible interpretations for lexical suffix constructions. I expect, based on experiences with other speakers, that such readiness is not universal, and that vagueness may not always be accommodated so easily. The speaker's niece, Flora Boyotai, often rejects adjectival constructions without an overt nominal, for instance. I suspect some prescriptive tendencies, made by analogy with Spanish, but I also believe that such examples are potentially confusing.

In my argument-less adjective data, I often use pictures instead of explicit mentions of referents. Flora generally behaves similarly to her uncle and other speakers in these tasks.

The acceptance and rejection patterns in examples like (105) are revealing of why one must avoid naive interpretations. I initially felt that my data indicated that strong familiarity was necessary. Therefore, the issue was likely an anaphoric issue. Yet, consider that in order to get the 'cut', rather than 'eat' interpretation of the sentences, some instrument is necessary. The machete needs to be mentioned, or presented in some way, in order to disambiguate the sentence. Rubén Boyotai, who has proven ready to provide interpretations of context-poor lexical suffix constructions, has consistently had the same intuitions as other speakers on this judgment. An intuition that information is missing is not necessarily an indication of anaphoric properties. For that reason, even for those speakers that felt that some explicit mention was important, a much more detailed study of the construction would be needed to show that *-po* is an anaphoric signal, and not simply vague descriptive content.

An issue with body parts is that they are also contextually entailed. When talking about a person, or many animals, things like heads, feet, etc. are known to exist in the context. This adds to the notion that *-po* was simply vague, since in a normal context, there would be no question that a person's body parts are within the context, and thus familiar.

I have no examples that provide evidence of strong familiarity. In the next subsection, I provide data that suggests that classifiers are used in adjectival expressions that refer. It also provides some documentation of how classifiers relate to contextually supplied meanings.

3.3.6.4 Adjectives as anaphors

As an example of classifier discourse sensitivity, a common elicitation paradigm that I have used is the true, false or error (sometimes stated as nonsense) paradigm. Error is ambiguous between felicity and grammaticality in and of itself, but this is always true when a sentence is rejected without further information. In the examples below, I consider the error/nonsense category to be infelicity due to supporting evidence. The paradigm is used to ascertain the locus of infelicity when sentences containing classifier constructions are rejected. False corresponds to statements where reference succeeds, but the statement in question is contrary to facts that were previously established. An error or nonsense response indicates a failure of compatibility with the context, which is arguably a violation of familiarity.

(107) **context:** Given the image, are the following statements true, false or nonsense?

a. T *giitã-ka* *eibe* *ĩ-pa*
small-CLF.stone above COP-DECL
'The small one is above.'

b. F *yẽdẽ-ka* *eibe* *ĩ-pa*
big-CLF.stone above COP-DECL
'The big one is above.'

c. # *giitã-wẽ* *eibe* *ĩ-pa*
small-CLF.plant above COP-DECL
'The small one is above.'

d. # *yẽdẽ-wẽ* *eibe* *ĩ-pa*
big-CLF.plant above COP-DECL
'The big one is above.'

In (107) an image was presented to speakers, where a small stone is above a larger stone. This particular context and sentence paradigm was presented to 3 speakers. It was part of a larger protocol using the same sentences in all cases but varying the pictures in terms of which size of object is above the other and in terms of the pictures, themselves. Extremely similar protocols have been presented to 8 out of 10 of the speakers who participated in elicitation at some point in the course of fieldwork from 2019 to 2024 with highly consistent results. In (107), a number of sentences using different classifiers were presented to the consultants which stated either that the larger object was above or the smaller object was above. In cases where the classifier was deemed felicitous, answers followed a true or false pattern, indicated by T or F in the example. When the classifiers were not compatible, the speakers judged the sentence to be an error. Speakers were asked why they responded as they did. Explanations verified that the rejection was due to the sentence not matching the context. A different sentence with a compatible classifier or no classifier was sometimes offered as a correction. Explanations for false responses asserted that the referenced object was not located above but below, correcting the location of the pictured object but also indicating that reference had been established.

In (107) the classifier *-ka* was found to be appropriate in the context but the classifier *-wẽ* inappropriate.

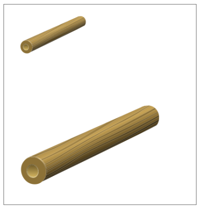
(108) **context:** Given the image, are the following statements true, false or nonsense?

a. $\frac{T}{\#}$ *giitã-bõ* *eibe* *ĩ-pa*
small-CLF.seed above COP-DECL
'The small one is above.'

b. $\frac{F}{\#}$ *yẽdẽ-bõ* *eibe* *ĩ-pa*
big-CLF.seed above COP-DECL
'The big one is above.'

Speakers vary somewhat in what they feel is an appropriate classifier. In (108) I provided the same context, but used the affix *-bõ*, which is suitable for small round objects. I presented this to the same three speakers. Two treated the sentences with the new affix identically to the *-ka* examples. One rejected the sentences stating that they were more appropriate to small round things.

(109) **context:** Given the image, are the following statements true, false or nonsense?



a. T *giitã-wẽ* *eibe* *ĩ-pa*

small-CLF.plant above COP-DECL

'The small one is above.'

b. F *yẽdẽ-wẽ* *eibe* *ĩ-pa*

big-CLF.plant above COP-DECL

'The big one is above.'

In example (109) it can be seen that the use of *-wẽ* was rejected only because of the context, not because the sentences were badly formed. Changing the image allowed for felicitous reference.

(110) **context:** Given the image, are the following statements true, false or nonsense?

a. T *giitã-bõ* *eibe* *ĩ-dãbaĩ* *ĩ-pa*

small-CLF.seed above COP-NEG COP-DECL

'The small one isn't above.'

b. F *yẽdẽ-bõ* *eibe* *ĩ-dãbaĩ* *ĩ-pa*

big-CLF.seed above COP-NEG COP-DECL

'The big one isn't above.'

I also used negated sentences. Examples in (110) demonstrate that the descriptive content of the classifiers are not negated, otherwise both sentences would be false. In order to attempt to target the adjective, and thus the classifier, as near as possible, I used a variation of (110) with *wii*, which I believe is a negative focus particle. This can be seen in (111). I only tested this with one speaker. As mentioned before, they tend not to use Spanish articles. For that reason, their translations were of the form 'No es pequeño arriba. Es grande.' and 'No es grande arriba. Es pequeño.' Notably, they did not add anything corresponding to the classifier in the translations.

(111) **context:** Given the image, are the following statements true, false or nonsense?

a. T *wii* *giitã-bõ* *eibe* *ĩ-dãbaĩ* *yẽdẽ-bõ* *ĩ-pa*

wii	*giitã-bõ*	*eibe*	*ĩ-dãbaĩ*	*yẽdẽ-bõ*	*ĩ-pa*
NEG	small-CLF.seed	above	COP-NEG	big-CLF.seed	COP-DECL

'It isn't the small one above. It's the big.'

b. F *wii* *yẽdẽ-bõ* *eibe* *ĩ-dãbaĩ* *ĩ-pa*

wii	*yẽdẽ-bõ*	*eibe*	*ĩ-dãbaĩ*	*ĩ-pa*
NEG	big-CLF.seed	above	COP-NEG	small-CLF.seed

'It isn't the big one above. It's the small'

The above are solid anaphoric uses of adjectives. As well as the classifier meaning failing to cancel, the adjectival meaning fails to cancel. It would be odd for the adjective alone to signal anaphoric status. I have seen nothing to suggest that there is an anaphoric and non-anaphoric way to say 'big'. The question is whether the classifier signals the status.

In (112), I provide some examples of using no classifier. The reference to the big and small rocks is not affected by negation. The classifier was not signaling an anaphoric status.

(112) **context:** Given the image, are the following statements true, false or nonsense?

a. T *giitã eibe ĩ-pa*

small above COP-DECL

'The small one is above.'

b. F *yẽdẽ eibe ĩ-pa*

big above COP-DECL

'The big one is above.'

c. F *giitã eibe ĩ-dãbaĩ ĩ-pa*

small above COP-NEG COP-DECL

'The small one is above.'

d. T *yẽdẽ eibe ĩ-dãbaĩ ĩ-pa*

big above COP-NEG COP-DECL

'The big one is above.'

3.3.6.5 Concluding remarks on anaphoric status

There is evidence of discourse dependencies with classifier usage. Consider that when the wrong classifier is used, as *-wẽ* was in (107), the classifier meaning could be interpreted as an English-like indefinite, e.g. *a plant*. This would allow for sentences with classifiers to introduce a plant, which would enter the discourse. That would mean that rejected *-wẽ* examples in (107) could be interpreted as simply false. I would be like stating, *A big plant is above.* In English, this could be stated to be false, rather than nonsense. With the classifier, there is some notion of mismatch. It is an error. Despite this, there was clearly no indication that the classifier, itself, necessarily signaled anaphoric status in terms of entity reference.

3.3.7 Summary of lexical suffix properties

Not all lexical suffixes are necessarily classificatory. There are some that appear to only occur with nominals. It is likely that the set of lexical suffixes is reasonably heterogeneous. Yet, there are some affixes, such as *-ka*, *-po*, and others that occur in many contexts, and are associated with a relatively stable enumeration of meanings. Lexical suffix hosts may filter the meanings due to semantic compatibility, or due to strictly lexical factors, as is the case with nominals and some verbs.

Morphotactic qualities depend on the host. Nouns allow multiple lexical suffixes. Numerals allow some combinations. Verbs have a specific lexical suffix position, distinct from person marking. Adjectives, demonstratives and numerals have a shared lexical suffix-person marking position.

Nouns are the most restrictive when it comes to which lexical suffixes a host may take. This is especially true when the roots are bound, and there may arguably be no *affixation*, so much as a form-meaning relation to productive affix usage. There are classes of verbs that take only body-part suffixes. It is unknown whether some verbs may accept any classificatory lexical suffix. Verbs also encode an animacy divide, where lexical suffixes correspond to non-animate objects. Adjectives and demonstratives appear to be relatively unrestricted hosts, which take any classificatory lexical suffix, though in some cases, an adjectival meaning may be incompatible with some lexical suffix meanings.

3.4 Conclusion

This chapter has provided essential information about Wao Terero. The section on phonology serves as a guide to essential measures used in diagnostics. Nasalization patterns help delineate bound and free forms. It is clear that the phonology is under-described. There are some puzzles. Some of these may affect morphological analysis, such as which lexical suffix occurs where. Despite this, the description demonstrates that there is a core of certainty. Errors in phonological analysis are not likely to be catastrophic to an understanding of the system, even if they affect some details.

The person and number systems do not look as rigid as Spanish. There is some wiggle room in agreement patterns, and choices that a speaker may make, such as whether to explicitly mark femininity. There is nothing in this that is incompatible with a grammatical treatment of the person affixes, when used with verbs, and other parts of speech. One can look at it as a form of overabundance. A feminine subject may trigger more than one verbal agreement pattern. Yet, it may also indicate that the person marking is not entirely grammatical. A combination of grammatical agreement and a more loose notion of semantic concord is not out of the question.

The discussion of the lexical suffix system was recently summarized above. My claims, made in the introduction, have been supported by examples. It is now possible to move on to theoretical issues in the following chapter.

Chapter 4

Issues for a Formal Analysis

This chapter is concerned with relevant formal literature. I discuss Distributed Morphology (DM) (Halle and Marantz, 1993), and Paradigm Function Morphology (PFM) (Stump, 2001), since these are the most influential realizational frameworks. The focus is on issues that theories have in handling lexical suffixes. One issue is that lexical suffixes behave like normal grammatical suffixes, but are associated with concrete meanings. I begin by describing the divide between lexical and functional content, and breaking down arguments that lexical information should be associated with one kind of morphological form, such as a stem, and functional information should be associated with another, such as an affix. After concluding that such a divide has very little value, I discuss previous and potential treatments of lexical suffixes in popular realizational theories. Finally, I outline the approach that I take in this thesis, though an analysis will wait until Chapter 6.

4.1 The Functional-Lexical Divide

Providing an analysis of lexical suffixes in Wao Terero strikes at a number of fundamental issues in morphological theory. Wiltschko (2009) went so far as to describe the existence of lexical suffixes in Halkomelem Salish as paradoxical given common theoretical assumptions. This is exciting language, but it speaks more to the quality of particular assumptions than to qualities of the data. The primary assumption in question is the correspondence between lexical and functional meaning with particular morphological units, what I'll refer to often as *the divide*.

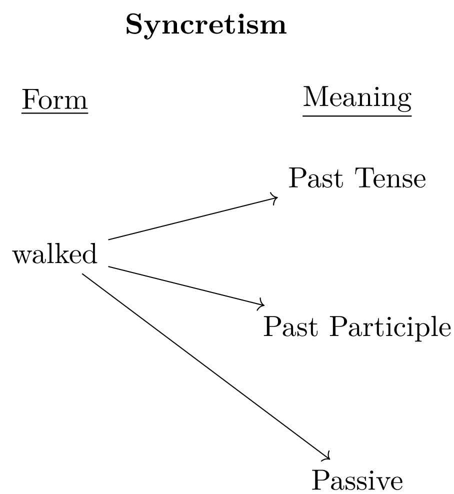


Figure 4.1: There is no function from forms to meanings in morphological systems as demonstrated by syncretism. Here the *-ed* form of WALK may be associated with either the past tense, past participle or the passive.

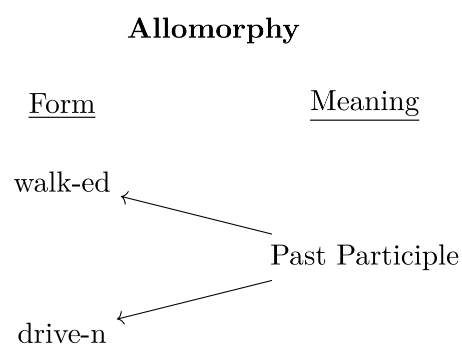


Figure 4.2: There is no function from meanings to form in morphological systems as demonstrated by allomorphy. Here *-ed* and *-n* are allomorphs of the past participle in English.

Meaning classes	**Form classes**
concrete (lexical)	free form
abstract (functional)	affix

Table 4.1: Some examples of classes of meaning and classes of forms. The rows align with the assumptions of the divide.

A goal of this work is to look beyond the divide from the point of view of separationism. There are a number of conceptions of separationism that are explored in this work. At its core, separationism is the notion that given a set of meanings and a set of morphological forms, the relationship between the two sets is non-functional, in either direction. An inflected form may be associated with multiple meanings, as in syncretism as schematized in Figure 4.1. A meaning may have multiple forms, as in complementary allomorphy (Figure 4.2) or freely varying allomorphy (generally called overabundance.)[1] There is no function from meanings to forms or vice versa.[2] Separationism therefore describes a readily observable fact.

What is somewhat less accepted is that classes of forms do not correspond to classes of meaning. A class of meaning is concrete lexical meaning, or abstract functional meaning. A class of forms is affixes or free forms. These are listed for clarity in Table 4.1. When it comes to such classes, the notion of separationism is not adopted by all authors, justifying the comment by Wiltschko (2009). A closed class affix, according to the divide, should not have a concrete meaning.

In this chapter I will demonstrate that the concept of separationism also applies to classes of meanings and forms. It is not the case that given a class of meaning, such as a concrete or abstract meaning, that there is a necessary relationship to a particular morphological form class, such as a free form or affix. Likewise, the opposite is also false. Given a particular class of form, such as an affix or root, one cannot say that an associated meaning must be abstract or concrete.

In order to make this point concerning separationism, I examine the notion of lexical-functional divide from the perspective of modern morphological theories. I focus on post-syntactic realizational theories, in particular DM (Halle and Marantz, 1993; Bobaljik, 2017), PFM (Stump, 2001; Stump, 2016). I explain what it means to be post-syntactic and realizational with increasing detail as the concepts become relevant. For

[1] Overabundance (Thornton, 2012) describes cases where there is more than one valid inflected form for a lexeme, such as *leaped* and *leapt* for the past tense of LEAP.

[2] Inflection class allomorphy is sometimes treated using functions from meaning to form, but this generally requires reference to inflection class categories. So the function is actually from inflection class-meaning pairs to forms.

now, it is enough to know that these theories posit that the concrete morphological forms that correspond to abstract morphological information require some reference to syntactic objects of a fully formed phrase. All such theories adopt some notion of separationism. This section focuses more on the specifics of DM, because the lexical-functional divide is at issue in the theory. In PFM it is a conventional assumption.

It isn't always clear to how to characterize the stance of these theories. In PFM, the lack of analyses relevant to the divide makes it necessary to extrapolate what a reasonable analysis parallel to DM would be. A benefit of PFM is that the definition of the morphological component of the theory is rigorous, with a canonical work by a single author (Stump, 2001). This is in contrast to DM. DM is a subprogram of Minimalism (Chomsky, 1995), rather than a single theory. I discuss the difference between theories and programmes in §6.2. There is a great deal of diversity within the sub-programme. For this reason, I will generally attribute the core ideas of DM to early works by Alec Marantz (Marantz, 1996; Marantz, 1997; Halle and Marantz, 1993), or discussions of his works by other authors. Despite this, there are important and highly relevant works that significantly deviate from Marantz. In particular, the *late insertion* literature (Harley, 2014; Acquaviva, 2009) has convergent qualities with my proposal, and a very different conception of √roots. Also in contrast to PFM, there is no formal definition of DM. This may surprise some readers since Minimalism and DM are often referred to as *formal syntax*. For some clarification of the details needed for a truly formal specification of a Minimalism-like framework, see Collins and Stabler (2016) and cited works in the Minimalist Grammar framework. The lack of formal specificity requires extrapolation, but I provide a fair characterization.

In DM and PFM there are separate treatments of concrete lexical content, and content that is functional or abstract, expressed in the domains of inflection and derivation. In fact, this is largely the case throughout the realizational literature. Lexical content is generally associated with free forms, roots (Halle and Marantz, 1993), or the stems of inflected forms (Stump, 2001). These qualities are considered representative of open classes of forms. Inflectional and derivational functional content is associated with affixes and other morphological processes (Aronoff, 1994; Beard, 1995; Halle and Marantz, 1993), which constitute closed classes of processes and/or morphological items. There are also distinct theoretical constructs. A theoretical construct is a concept or object that cannot be directly observed, but is hypothesized to explain empirical patterns. For lexical content there is the DM notion of √root, and the widely adopted concept of lexeme used in PFM. For

	lexical	functional
meanings	concrete meaning	abstract meaning
constructs	$\sqrt{}$root, lexeme	morphosyntactic features
exponent types	root/stem/free form	affix/ablaut/etc.
class characteristics	open	closed

Table 4.2: The putative lexical-functional divide.

functional content, the notion of morphosyntactic feature is by far the most popular type of covert object. Notably, these abstract objects are intended to either closely correlate to, or even determine the exponent types. For instance, PFM (Stump, 2001; Stump, 2016) provides a system where the lexeme determines the stem(s) for inflection while morphosyntactic features determine processes, such as affixation.

Table 4.2 provides a schematization of the lexical-functional correspondences. Although DM has evolved and diverged over the years, a highly explicit manifestation of this divide was expressed in seminal works within the theory. Marantz (1997) introduced the concepts of I-morphemes versus L-morphemes, where I-morphemes fit the functional side of the table and L-morphemes the lexical side. It is unclear whether the correspondences of the initial theory have any remaining substance given how analyses have developed with respect to non-conformant data, but I am unaware of any DM practitioner who has explicitly claimed that the I versus L-morpheme contrast is now irrelevant to the theory. Aronoff (1994) and Beard (1995) were also very explicit about the difference between the meanings and class properties of lexemes on the one hand, and other types of morphological content on the other, which is also consistent with Table 4.2. In PFM (Stump, 2001) I am not aware of an explicit stance but in practice PFM is largely consistent with theories where the divide is explicit. It is relatively easy to demonstrate that the correspondences represented by the divide are flawed.

4.1.1 Counter Examples to the Divide

Many of the obvious counter examples to the divide arguably lack systematicity. I mention a few of these before providing a more extended discussion of the DM notion of "light" phenomena, and cases where there is general agreement that words have functional content. I then discuss how descriptions of lexical suffixes have dealt with the divide before discussing the issues with these treatments, and their implications for morphological theory.

4.1.1.1 Non-systematic Counterexamples

It is known that lexical-functional correspondences leak. Consider suppletion, as in English *go/went*, where a distinction otherwise indicated by an unremarkable inflectional pattern is provided by two unrelated forms. Theories may have differing analyses of this example, but a fairly straight forward description is that phonologically unrelated word-forms signal a functional distinction that is normally within the domain of affixation or alternation within the English language. Yet, in English at least, this is a rare phenomenon. There is no systematicity. If one's theory sees the lexical-functional divide as explanatory, it is worthwhile to look to alternative explanations for inflectional suppletion, rather than scrap the program. In the discussion of DM and "light" phenomena, we'll see a popular approach to suppletion that does treat it as part of a systematic explanation for morphological phenomena.

As another example of a leak, derivation, even when formally productive, may be associated with unpredictable, thus lexical meanings. For example, consider deverbal *-dor* in Spanish *hablar* 'to speak' goes to *hablador* 'someone who talks too much', while *comer* 'to eat' goes to *comedor* 'dinning room'. Despite this, these leaks can be thought of as marginal exceptions. The suffix *-dor* has some productivity as a deverbalizer, but beyond that functional commonality in its various uses, the meaning of particular items that contain the affix could be argued to be idioms. Again, there is an arguable lack of systematicity.

There are also non-affix elements of compounds that are not derived from an open class, such as cranmorphs. These are seen in compounds such as *cranberry*, where *cran* has no independent meaning. Lack of systematicity can again be used to argue that they are special cases.

4.1.1.2 Functional Words and "Light" Phenomena

(113) *wĩyẽ-dãdi go-dãdi*
child-PL go-PL
'Children go.'

A more systematic challenge to the notion that roots or lexemes are necessarily associated with lexical content comes from functional verbs such as the copula, or closed syntactic classes such as English prepositions, which may be devoid of lexical meaning content in some cases. Below I discuss different ways a

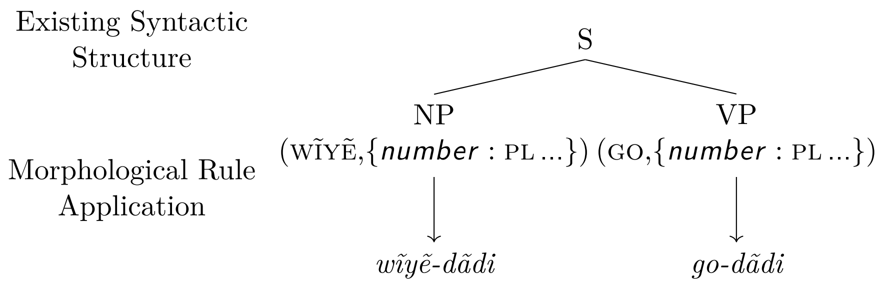


Figure 4.3: Post-syntactic realizational theories apply rules to nodes in syntactic structures to calculate form-meaning correspondences. The arrows at the bottom of the syntactic nodes NP and VP represent a realizational morphological rule applied to the node's contents.

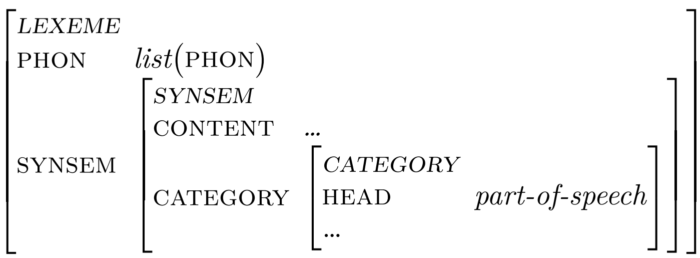
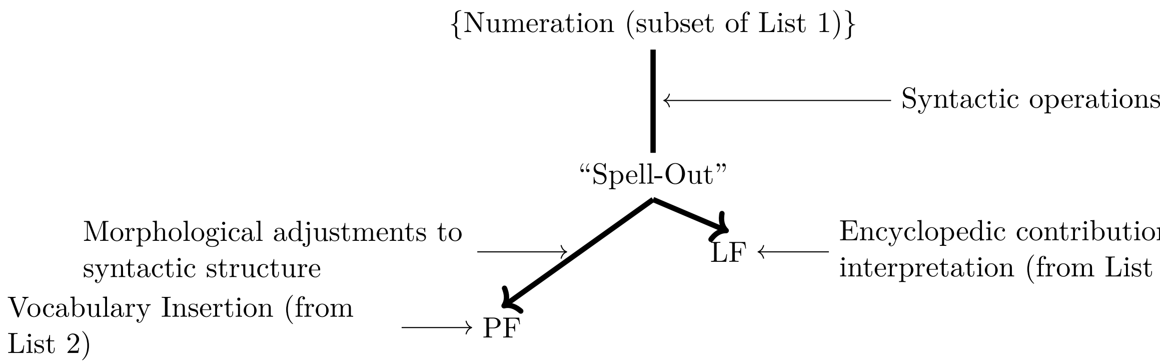


Figure 4.4: A simplified schema for a Head Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994) lexical entry. The lexical entry for a lexeme contains all of the information needed to combine with other elements in syntax.

morphological theory may deal with such phenomena. As stated above, I focus on popular *post-syntactic* realizational theories, where morphology is thought of as rules that are applied to nodes in syntactic structures. The basic idea behind these theories is schematized in Figure 4.3, where a sentence that is glossed in example (113) is represented as a syntactic structure. The content of the syntactic nodes in post-syntactic realizationalist theories are assumed to be some kind of abstract information that can feed morphological rules. In the figure I follow PFM-like assumptions where the information retrievable from a node is a pair of a lexeme and morphosyntactic features.

4.1.1.3 Two families of post-syntactic morphology

Post-syntactic realizational morphological theories fall into two families based on their assumptions concerning syntax. This has implications for the data that a theory may be responsible for. The first family of



List 1 Feature bundles/terminal nodes/abstract morphemes: Syntactic primitives

List 2 Vocabulary Items: Instructions for pronouncing terminal nodes in context

List 3 Encyclopedia: Instructions for interpreting terminal nodes in context

Figure 4.5: The model of a DM (Halle and Marantz, 1993) derivation where rules and data are distributed throughout a syntactic calculation. Figure based on Harley (2014, p. 228).

theories assumes a lexicalist syntax, where lexical entries compose in order to form larger phrasal structures. Lexical entries are sign-like units in the sense of de Saussure (2011), pairings of form and meaning that may also include additional information, such as a syntactic category or morphosyntactic features. Given a collection of lexical entries, there are rules that specify how they combine. HPSG (Pollard and Sag, 1994) and categorial grammars (CGs) are examples of such syntactic theories. A simplified lexical entry schema from HPSG appears in Figure 4.4. The phon(ology) would have a value that represents the pronunciation of a lexeme or word. The synsem groups semantic and syntactic information under content and category, respectively. Theories such as PFM assume such a syntax.

The assumption that there are lexical entries is one of the fundamental differences a theory like PFM shares with DM. DM does not assume such entries, except that their notion of √root may be sign-like in some instances of the theoretical program. Instead, DM assumes a form of Minimalist syntax (Chomsky, 1995; Hornstein, Nunes, and Grohmann, 2005) that applies different sets of information, rules or constraints at different points in a syntactic derivation, gradually transforming relatively unstructured grammatical data, called a *numeration*, into highly structured parallel phonological and semantic representations. The fact that different sets of rules and conditions apply at different points in the derivation "distributes" lexical

DM	**HPSG**
List 1	CATEGORY
List 2	PHON(OLOGY)
List 3	CONTENT

Table 4.3: With reference to Figure 4.4 and Figure 4.5, one can see how information within a lexical entry in HPSG is distributed throughout a DM derivation.

information rather than consolidating it in a single lexical entry. Table 4.3 helps make the comparison explicit. I assume that the reader has had some experience with the notion of lexical entry and the Y-architecture so I do not discuss all of the details of the figures provided.

As a side note, I provide a hybrid framework in this work, where syntactic lexical entries exist, but many properties of word-forms are listed independently in a relational network of information.

The major theoretical difference between the lexical and distributed approaches, relevant to post-syntactic morphology, is that in lexicalist syntax there are lexical entries that act as axioms or theorems of the syntactic theory. In the DM case there are not. This will be argued to have implications for syntactic compositionality in §5.2. In DM, the vocabulary insertion (List 2) portion of the morphological component is responsible for all non-lexical mappings of morphosyntactic meaning to form. For lexical meaning, there is a split in DM. There is a debate in the DM literature concerning how the form-meaning mappings of √roots are determined. There is the early insertion view (Embick and Halle, 2005; Embick, 2000), which states that √roots are sign-like units with forms and meanings entering the syntax prior to the realizational calculation. There is also the late insertion view (Harley, 2014; Haugen and Siddiqi, 2013), where forms for √roots are determined at the same point in the derivation as functional content. The lexical syntax case is more consistent with early insertion. There are two systems for matching forms and meanings, directly in the lexical entry, and additionally, in a specialized morphological component.

The theory put forward in this work assumes, like the late insertion group, that there is a unified realizational mechanism, whether meanings are lexical, or functional. Yet, none of this is post-syntactic, which places my approach within a distinct theoretical paradigm.

4.1.1.4 Approaches to functional free forms and lexemes given a lexicalist syntax

Given the post-syntactic lexicalist syntax assumption, closed class, free items with functional meanings, such as English prepositions are allowed to lie outside the domain of morphological description. Within theories of morphology that assume a lexicalist syntax, like PFM, one can assume that there are syntactic lexical entries that specify form-meaning correspondences that morphology does not calculate. For instance, in Figure 4.4 there is a feature phon that may be defined for a word like ON without any post-syntactic calculation. This may be desirable in English, due to the fact that prepositions do not vary in form in the same way that inflected verbs and nouns vary in form. Verbs like CONJUGATE have forms like *conjugate*, *conjugates* and *conjugating*, but the form of a preposition like ON is always *on*. Therefore, the morphological apparatus could be argued to not come into play. There is simply no co-varying form-meaning correspondence to describe in the morphology. Except, of course, that relative to the static form of a preposition, the meaning and category may vary in a pattern that shares attributes with other members of its class. The variation of meaning and category relative to a static form still constitutes a type of co-variation of form and meaning. It is simply one that some morphologists may consider irrelevant to their conception of morphology.

Treating the form-meaning correspondences of items such as ON, FROM, etc. as something like axioms of the grammatical theory might make sense according to certain theoretical positions, but the treatment would not help preserve the lexical-functional divide seen in Table 4.2 (nor is it claimed to). If the grammatical theory maintains that prepositions, like verbs or nouns, are lexemes, then there are lexemes that are members of a closed class with functional meanings. To preserve the concrete, open class correspondence with lexemes, a grammatical theory might divide out prepositions and similar phenomena from the class of lexemes, providing some other abstract theoretical primitive used in the syntax. This is not unprecedented, as the discussion of DM strategies, below, demonstrates. Given such a move, this notion of lexeme would weaken its correspondence to form, since there could be other abstract elements associated with free forms. If prepositions are treated as ordinary lexemes, a form correspondence is maintained at the cost of a strict association between lexemes, open classes, and concrete meanings.

There are also items that lexicalist post-syntactic theories place within the morphological description that many would classify as functional. The verb BE may be an example. Due to its suppletive nature, one could argue that it should be treated as a group of lexical entries with form-meaning correspondences that

	first	second	third (non-sentient)
ĩ 'to be'	ĩbo	ĩbi	ĩ
go 'to go'	gobo	gobi	go

Table 4.4: The Wao Terero copula inflects like other verbs.

are not calculated by morphology. If we entertain such a treatment, we can look to other languages where the argument would not hold. The Wao Terero copula ĩ provides an example. It is not suppletive, and largely inflects in the same way as any other lexical verb for person and tense. It is used in constructions that are comparable with English BE, where it contributes no concrete meaning to a sentence.

Table 4.4 provides a small sub-paradigm for the copula and another verb to demonstrate the shared co-variation of form and meaning between the two items. It would be difficult to motivate treating the Wao Terero copula in a manner distinct from other verb-like items in the morphological system. Therefore, if a theory stipulates that a lexeme must be the input for the calculation of inflection, as is the case in PFM, items with functional meaning, such as the Wao Terero copula, would be a clear instance where such a theory would need to admit a contradiction to the lexical-functional divide.

A theory like PFM has little other option than to assume a lexicon that contains lexemes with functional meanings that may or may not be associated with closed classes. For PFM, specifically, this isn't necessarily an issue. As far as I am aware, the divide is more of an expression of convention than an explicit tenet of the theory, which focuses primarily on inflection. This is not the case in DM, which has the explicit I and L-morpheme concept. DM also has to deal with functional words, yet their strategy is somewhat different than what I described above for PFM. In PFM lexemes lose a correspondence to meaning, retaining a correspondence to form. In DM the relationship between √roots and concrete meanings is considered of central importance.

4.1.1.5 A non-lexicalist treatment of functional words

As noted, DM does not assume a lexicalist syntax. Neither does DM have a concept of lexeme. The closest correlate is the concept of √root. This is not the general notion of root, which corresponds to the surface observable, unanalyzable core of a word, but a theory internal notion, which is why it is written √root. Concrete meanings are associated with √roots, which have no syntactic category or subcategorization information associated with them. At least, this is how the classic theory works. As will be seen below, there are often complex √root structures, which must have their conventions specified in some way. There are authors that utilize various projections that embed √roots, such as a .√*P* (√root phrase), to provide syntactic structure that constrains √root combination (Harley, 2014). Others simply provide recursive √root schemata (Wiltschko, 2009), without explaining constraints on combination. Presumably the latter analyses rely on the notion that PF and LF should filter out overgenerated constructions. In any case, despite some variation in how complex √root structures are constrained, syntactic structure is claimed to explain the distribution of √roots. By definition, a √root is non-functional. The question then is how words which are largely functional are treated in such a system.

In the DM literature various types of functional, "light" phenomena have special status as free forms. The notion of light verb belongs to an established literature (Butt, 2014), which some may believe is a motivation for the concept of "light" in DM. Yet, it is important to note that DM generalizes the concept. Not everything that is considered "light" by DM theorists is necessarily considered to be so outside the theory, thus the scare quotes, which I will consistently use to differentiate the theory internal notion of "light" used in DM from a more general use. As functional content, English prepositions are "light" phenomena in DM. Rather than focus on prepositions, here, it is more interesting to consider "light" elements that receive morphological analyses in other theories. These would include verbs such as *be*, *go*, etc., which have treatments in PFM and other theories. In a PFM-like theory, they may be treated as normal lexemes with inflected forms, even if their content is functional.

In DM, "light" verbs and other "light" phenomena are considered to be either functional or to have meanings attributable to Universal Grammar (UG) (Marantz, 1996; Harley, 2014), where UG, from this perspective, is a rich innate grammar that underlies all human language. "Light" phenomena, which would include functional prepositions in addition to traditional concepts such as light verbs, are not considered to be √roots in the theory, and instead are realizations of features on a functional head in the syntactic structure (Marantz, 1996; Harley, 2014). This requires some explanation.

DM is a theory that treats morphology as the arrangement of morphemes within the phrasal syntax. The general pattern for lexical nouns and verbs, as well as other items with lexical content in DM is to position a √root in a syntactic tree structure below some governing category, which provides the syntactic category

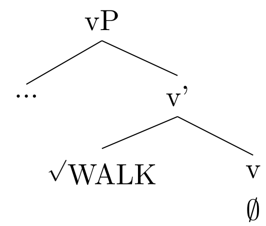


Figure 4.6: A tree representation of a lexical verb in DM.

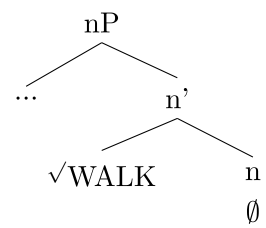


Figure 4.7: A tree representation of a lexical noun derived from verbal *walk* in DM.

of the √root. √roots are considered to have no category relevant to syntax. This can be seen in Figure 4.6, where the vP structure provides the syntactic category for a √root element, using the DM standard √ notation. The head of the vP must have morphosyntactic features, which I do not provide in the diagram. In the *walk* case, the features do not correspond to an affix so ∅ is given. Since there is a zero derivation relationship between verbal *walk* and nominal *walk*, the structure in Figure 4.7 would correspond to the nominal use. In the nominal example, the √root is the same but the nP structure provides the nominal category.

In some cases, the features on v will correspond to a derivational affix, as in the denominal example in Figure 4.8. In contrast to the above, a verb with no lexical content, or with only UG content, will have a structure as in Figure 4.9, where the form of the word is associated with features on the v-node, rather than a √root. Note that full DM analyses may involve additional transformations and structure. These trees are intended to give readers the gist of the analyses with minimal distractions.

The consequence of the DM analysis of "light" phenomena is that such items only look, on the surface, like lexical words. Underlyingly, they occupy the sames space as a functional affix would in terms of syntactic structure.

The DM analysis of "light" phenomena explicitly contradicts a lexical-functional, form-meaning correspondence. It preserves the notion that √roots are not associated with functional meanings, but asserts

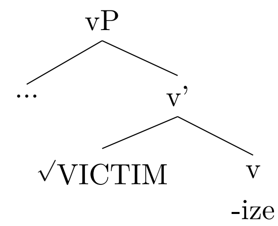


Figure 4.8: A tree representation of a denominal verb in DM.

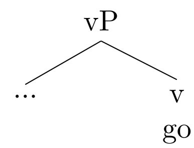


Figure 4.9: A tree representation of a "light" verb in DM.

that non-complex free forms need not correspond to √roots. The question remains whether in DM "light" phenomena may be associated with open classes. I believe the answer, in some cases, is *yes*, but the question of class-hood for morphological items is complicated. Avoiding details of that question here, I'll provide a brief argument that some "light" items in DM are members of open classes. I use English GO as an example.

In many manifestations of DM, English GO is necessarily a "light" verb. From a theory internal perspective there are two criteria that make this status sufficient, and one that makes it necessary. First, it is sufficient based on an *arguable* lack of concrete meaning for GO. Second, GO may be argued to correspond to a universal meaning. The third criteria is that English GO is suppletive. Suppletion necessarily makes GO a "light" verb in the theory because lexical content cannot be suppletive in DM (Marantz, 1996). The reason for this is fairly esoteric, and is not worth detailing here, but see Harley (2014) for a clearly stated explanation and an alternative theory of √roots that allow lexical content to also be suppletive. According to conventional DM, only functional items may have phonologically unrelated allomorphs, such as the *-en* in *oxen* versus the *-s* on *cats* for plural. Therefore, "light" words, which are underlyingly identical to functional affixes, may also have phonologically unrelated allomorphs. According to DM, that is what one is seeing with *go* versus *went*.

Having established the status of GO in DM, there is the question of whether it is an open-class item. An independent means of determining class-hood is by looking at whether there are items that share distributions.

For instance, we might ask if GO is replaceable with other items in syntax such that it establishes category membership. This is trivial to accomplish, and places GO into a subclass of English verbs. It is also generally understood that once a class is established, if one can freely add to it, that is the definition of being open. English verbs, including verbs of motion, are well known to be an open class. By this measure GO is a member of an open class. The implication is that morphosyntactic features can correspond to open class items in DM. This is a fair characterization because items associated with √roots are considered open class items due to the same diagnostic. If the diagnostic works for items associated with √roots then the same logic should apply to items associated with morphosyntactic features.

It is true that the feature values themselves may be finite, with a finite number of licit combinations, but the fact remains that an open class may be composed of items that are either associated with √roots or morphosyntactic features in DM. One cannot use a diagnostic for open class-hood, and determine from that result whether or not a member of a class is associated with a √root, or a morphosyntactic feature structure. From the DM perspective, further diagnostics are needed to determine abstract morphosyntactic representation, such as whether the item is suppletive, or has a meaning associated with UG. Given late insertion versions of DM, where lexical items may be suppletive (Harley, 2014), the suppletion diagnostic is not definitive, either.

4.1.2 Taking stock of the divide

The notion of the divide is now diminished. It is not clear what criteria should be used to define functional or lexical items across theories. In the beginning there was a form, meaning, construct, and class type that aligned on both sides of the divide. Now there is not. So the question is, what is primary about something being lexical or functional? Is it a meaning, a form type, a construct, or a class type? In PFM, lexemes remain the abstraction that binds together *lexical morphemes* in the sense that a lexeme determines both a stem (or free form) and an associated meaning. Yet, that meaning may be an abstract meaning, rather than a concrete meaning. In DM the √root remains the clear construct for lexical meaning content. To provide coherence to the notion of a divide, I will consider the abstract constructs, lexeme and √root, as primary in determining the properties of lexical versus functional content. This is schematized in the Tables 4.5 and 4.6 for DM and PFM-like theories, respectively.

	lexical	functional
meanings	concrete meaning	abstract meaning
constructs	√root	morphosyntactic features
exponent types	root/free-form	any
class characteristics	open	any

Table 4.5: The putative lexical-functional divide after taking into account functional words with a DM treatment.

	lexical	functional
meanings	any	abstract meaning
constructs	lexeme	morphosyntactic features
exponent types	stem/free-form	affixes/alternation etc.
class characteristics	any	closed

Table 4.6: The putative lexical-functional divide after taking into account functional words given reasonable lexical syntax assumptions.

There are a number of contrasts worth noting. In both tables, morphosyntactic features remain associated with abstract meanings. Only in the lexicalist syntax-based table do they remain associated with bound, non-root or stem forms, though it is worth noting that a deeper look at the literature of PFM-like theories would reveal that morphosyntactic features may correspond to collocations of free forms, as in periphrasis (Sadler and Spencer, 2001; Bonami, 2015), or single word sized alternations, as in suppletion (Stump, 2016). There is no explicit restriction on the form types that may be realized in PFM, but these intricacies are beyond the current discussion. It is important to PFM-like theories to maintain a closed class characteristic for analyses involving morphosyntactic features due to their conception of inflectional paradigms. In DM, features, though finite, themselves, may be used, without √roots, to realize members of open classes. DM also lacks a necessary correspondence between morphosyntactic feature and form type.

On the lexical side of the divide, the DM correspondences are the same as in Table 4.2. Given a free form or a bound root form, one cannot predict based on the form alone whether it is associated with a √root or morphosyntactic features, but if the meaning is concrete, it must be a √root. For a PFM-ish theory, a free form, or the stem of a free form corresponds to a lexeme, whatever the meaning type.

In both cases the lexical-functional correspondences have been diminished. It is difficult to argue that functional words do not erode the divide. Lexical suffixes will help push that further.

4.2 Treatments of lexical suffixes relative to the divide

The lexical-functional form-meaning correspondence has its problems. Given the above, it is significantly diminished. Yet, the idea that concrete meanings are associated with only open classes would be a useful observation, if true. Lexical suffixes disrupt this correspondence. They are closed class items with concrete meanings. This has resulted in uncomfortable categorizations of lexical suffixes in the literature. Sapir (1911) chose to emphasize the formal characteristics of lexical suffixes over the meaning characteristics in his categorization. He considered lexical suffixes to be grammatical content, which implies that they are more like inflection. Wiltschko (2009) went in the opposite direction, which is to treat lexical suffix meanings as the important characteristic. In her DM analysis, she treats lexical suffixes as $\surd$roots. This would mean that lexical suffix constructions are a form of compounding or incorporation, despite the fact that the "incorporated" elements are not derivable from open classes.

In the next subsection I discuss the implications of the Wiltschko (2009) analysis on the divide. In anticipation of an examination of the perspective of Sapir (1911) within the context of PFM, I provide a discussion of morphosyntactic features. Afterward, I provide an analysis in PFM of Wao Terero verbal inflection to demonstrate how such a theory works on the data it was designed for. I then present evidence that feature-based analyses of Wao Terero-like lexical suffixes do not work, due to both representational issues of features, and the notion of competition. The latter issue affects DM late insertion analyses as well.

In general, I believe that Sapir (1911) had a more reasonable perspective than Wiltschko (2009), in the sense that lexical suffixes should not be treated as formally distinct from other affixing systems. The lack of formal distinction results in there being no strict divide between the lexical and functional, which requires a unified approach to both types of content. This raises the question of what the underlying constructs should look like in a unified system. Since lexical suffixes behave in a manner similar to other affixes, except for their meanings being concrete, their licensing conditions should look like other affixes. This is problematic in current theories. When looking at the ability of a theory such as PFM to deal with closed class affixes with concrete meanings, we see a roadblock, which is the concept of morphological features as driving the realizational calculation. This leads to the question of whether morphosyntactic features merit their special role. One question is whether they are sufficient for that role. I show that the are not sufficient for lexical suffixes. It is also reasonable to ask if they are sufficient in other cases, such as inflection. This is not a question that can easily be answered definitively, since what constitutes a sufficient analysis may depend on varying goals. I provide argumentation that, at the very least, such analyses have significant empirical and descriptive issues. Another question is whether morphosyntactic features are necessary. If they are not, one is free to propose an alternative. Chapter 6 demonstrates that there are alternatives with positive qualities. The core contribution of this work is providing a theory that is able to model realizational form-meaning correspondences, whether those meanings are concrete or abstract. It does so without relying on certain enshrined constructs. My belief is that, at the very least, this encourages a questioning of why those constructs are used so uncritically. I also believe that the alternative constructs I propose are an unambiguous improvement. They allow for a formally well-defined system of lexical realization, which does not otherwise exist, even if there is independent literature that argues for its necessity (Harley, 2014; Acquaviva, 2009; Haugen and Siddiqi, 2013).

4.2.1 Treatment of lexical suffixes as $\sqrt{}$roots

Wiltschko (2009) provides a DM-style analysis of lexical suffixes in Halkomelem Salish. In terms of differences between Halkomelem Salish and Wao Terero, Halkomelem Salish has lexical suffixes that were presented as being more productive with verbs than I have so-far encountered in Wao Terero, where not all verbs appear to be able to host lexical suffixes. Another difference is that Wao Terero lexical suffixes appear in constructions involving a wider range of categories than was presented for Halkomelen Salish. Halkomelen has fewer part of speech categories overall. Despite these differences, the semantics and syntax of the constructions are comparable. Example (114a) provides an example of a noun-noun compound-like meaning on a nominal stem and (114b) provides a classifier example with a verb, both from Halkomelen Salish.

(114) a. *tale*-**áwtxw**
money-building
'bank'

b. *th'éxw*-**wíl**-*t-es* *te* *ló:thel*

wash-dish-TRANS-3.SG DET dish

'He washed the dish.' (Wiltschko, 2009, pp.208, 211)

Wiltschko (2009) noted that lexical suffix constructions involving verbs had syntactic and semantic characteristics very similar to what is called classifier incorporation (Sara T. Rosen, 1998; Mithun, 1986), a form of incorporation where the nominal element does not semantically or syntactically play the role of an argument to its host, but does narrow the range of possible lexical arguments, where here *lexical* refers to the notion that there is an overt nominal surface argument. The primary evidence for this is called *doubling*, and is evident in the Halkomelen Salish example (114b), where there is a classifier element glossed 'dish' and a lexical noun labeled 'dish' in the same clause. Such examples also occur in Wao Terero as seen in (115). The import of this is that Halkomelem and Wao Terero phenomena are comparable. Little in my analysis of Wao Terero relies on verbal constructions uniformly behaving like Wiltschko describes for Halkomelem, otherwise.

(115) *õdõ·po* *kẽ-po-ta-bo-pa* *yebẽ-ka*

∅·LS.hand cut-CLF.hand-PST-1-ASS machete-INST

'I cut myself on the hand with a machete.'

Chapter 2 discusses more subtle aspects of Salishian languages, and diagnostics for classifier-like status.

Wiltschko (2009) sought to provide an analysis of lexical suffixes that both places the affixes on the correct side of the lexical-functional divide, and explains doubling. From a DM perspective, there really is no other option except to consider that content with concrete meanings are √roots. Therefore, lexical suffixes must be √roots. The obvious choice is therefore to treat the host of a lexical suffix construction as a compound of √roots. The issue is that √roots generally exist under syntactic structures that provide categories and have certain subcategorization requirements and consequences for syntax. Syntactic structures *must* have syntactic consequence, otherwise there is no evidence of their syntactic reality. Syntactic consequence is exactly what must be avoided to allow for doubling. The lexical suffix cannot be an argument of its host.

The consequence of stem and lexical suffix combination should be consistent with lexical integrity or anaphoric island hypotheses (Di Sciullo and Williams, 1987; Postal, 1969), which state that word-internal structure does not interact with the syntax. To achieve this, Wiltschko (2009) proposes a structure where √roots may be composed of additional √roots, as schematized in Figure 4.10. Such syntactic treatments without syntactic consequence have precedent as "complex root domains" and are not uncommon in DM (Acquaviva, 2008). As already mentioned, there are authors that utilize various projections that embed √roots, such as a $\sqrt{P}$ (√root phrase), to provide a quasi-syntactic syntactic structure that constrains √root combination (Harley, 2014), but may not otherwise affect syntax. This is not used by Wiltschko, but more recent ideas concerning the projections of roots would ameliorate the run away recursion her treatment otherwise suggests.

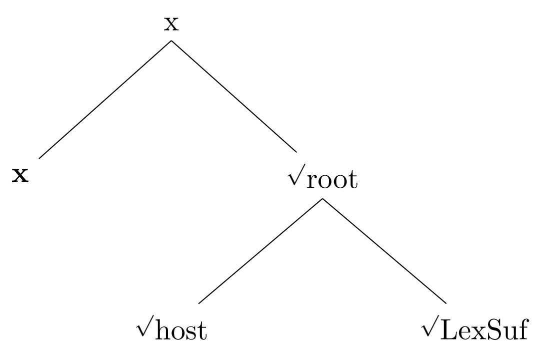


Figure 4.10: Wiltchko's root incorporation structure.

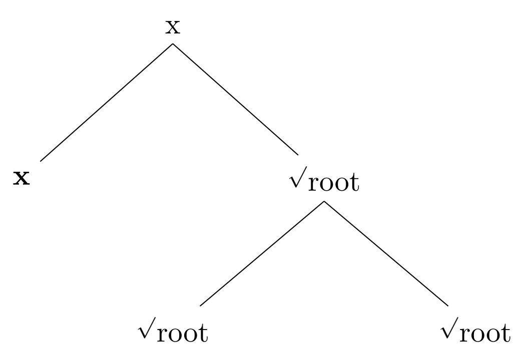


Figure 4.11: Wiltchko's root incorporation structure emphasizing the recursive nature of the schema.

Note that the words "root", "host" and "LexSuf" for lexical suffix are not of formal import. What we are seeing is a recursive √root structure. The words "root", "host" and "LexSuf" are merely intended to label particular portions of the schema relative to what is seen in Halkomelen Salish, but each node label may be

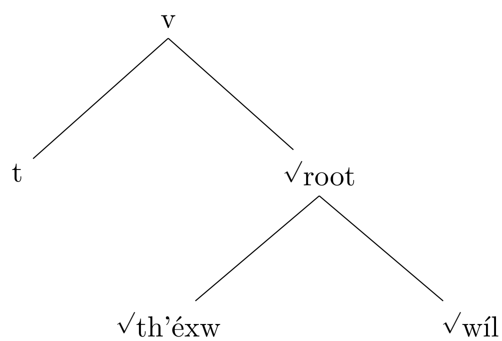


Figure 4.12: This is a structure corresponding to an uninflected version of (114b).

	lexical	**functional**
meanings	concrete meaning	abstract meaning
constructs	√root	morphosyntactic features
exponent types	any	any
class characteristics	any	any

Table 4.7: The putative lexical-functional divide in DM after taking into account lexical suffixes.

replaced with √root as in Figure 4.11.

An example where the schema is applied to the verb in (114b) is provided in Figure 4.12, where the verbal stem fills in the space for "host" and the lexical suffix for "LexSuf". This example appears absent of details needed in a full analysis, but is complete enough for our purposes.

In terms of the consequence of this analysis to the discussion of the lexical-functional divide, the proposal of Wiltschko (2009) further weakens the predictive power of the divide. Here √roots are associated with a closed class of affixes. Lexical suffixes by definition are not derived from open class items. Given this, the only distinguishing feature of lexical versus functional content that doesn't cross the divide are that concrete meanings are associated with the abstract concept of √root. This is depicted in Table 4.7. What we can now say is that lexical meanings are concrete and functional meanings are abstract. The lexical-functional divide is entirely a semantic divide at this stage for DM.

There is a question as to why DM uses both roots and features in their architecture.

One might question the utility of the √root concept at this point. In post-syntactic realizational morphology features are the default representational mechanism. Sometimes this approaches a type of representational monism, where syntactic categories are features, meanings are features, etc. Why is the line drawn at concrete meanings? One issue is that analyses of concrete meanings do not reduce to a finite set of features. Neither do concrete meanings exist in clear opposition to one another. A core prediction made when a feature is used is that two or more categories exist in absolute opposition. Given that feature representations are a poor fit, some stand-in is necessary. I discuss the issue further, below. The takeaway is that it is largely a technical issue of meaning representation. There is a consensus that concrete meanings simply do not fit the feature mold.

To conclude, the disconnect between exponent types and types of meaning or class characteristics is consistent with the notion of separationism that I advocate. The tenets of DM naturally lead to this conclusion. Though I do not borrow the concept of √root or syntax driven morphology, I am in complete accord with this result.

Despite this, there are a number of issues with the Wiltschko (2009) analysis. Even though I have noted in my discussion of the divide that the author's analysis requires that √roots may be associated with closed class items, that point was not made by the author. In Wiltschko (2009), lexical suffixes are aligned with the concept of incorporation. The √roots are still open class items. This is not consistent with empirical evidence in either Halkomelem Salish or Wao Terero. The entire motivation for the recognition of such affixes as deserving a distinct treatment was due to the fact that they are not derived from open class items (Sapir, 1911). As mentioned above, there have been trends since Wiltschko (2009) to elaborate √roots into complex phrasal objects. This could limit the distribution of √roots in important ways. Yet, my impression is that, in DM, it is seen as a good thing to ignore the closed class nature of items like lexical suffixes in order to provide unified incorporation analyses of a wide variety of phenomena (Harley, 2012; Johns, 2017). This is an instance of comparative concerns outranking descriptive accuracy. Even so, it is plainly suboptimal on comparative grounds, since there are a number of languages with lexical suffixes, which share the closed class characteristic. A good comparative theory provides contrasts to capture language subtypes in addition to mechanisms that describe similarities. In this case, there is no mechanism for expressing a contrast. The analysis is an overgeneralization.

The question is, if we accept Sapir (1911) and respect the closed class nature of lexical suffixes, what issues will that raise for modern theories? I discussion this point in the following subsections.

4.2.2 Morphosyntactic features

Outside DM, there has not been a lot of work in post-syntactic realizational theories on lexical suffix-like phenomena. The general consensus in these theories is that closed class affixes should be treated using morphosyntactic features. These are the bridge, so to speak, between the morphological form, a syntactic pattern and a meaning. In a later section, I describe how morphosyntactic feature analyses work, using a PFM-like approach. I follow that with a discussion of the issues encountered when attempting to provide a feature based analysis of lexical suffixes. As a lead up to this I provide a relatively in-depth discussion of features and how they are commonly used. I argue against their utility and soundness in the areas where they are commonly applied.

4.2.3 What do morphosyntactic features represent?

In this section, I focus on what features represent in popular formal theories. I do not extend it to the stochastic, computational, or psycholinguistic realm. I focus on DM and PFM-adjacent systems, with some commentary relevant to HPSG-like systems. I also discuss features as used from a typological, or comparative perspective. This is a broad class of study. It is not my intention to review it in a comprehensive manner. I discuss typology and non-UG comparative studies because feature proposals in PFM and other theories are often based on comparative categories. Therefore, some commentary on the source of evidence is needed.

The discussion of features is a lead up to a sample PFM-like analysis of Wao Terero verbs. It is difficult to understand choices that are made in the use of features in PFM without first understanding common practice. What features represent is often left implicit in analyses since, superficially, the features just work. In the case of using morphosyntactic features to explain lexical suffix patterns, I aim to demonstrate a negative outcome. This requires a deeper understanding of features since what is often taken for granted, or left implicit must be exposed to examination. Additionally, even for data normally considered to be well analyzed using features, a critical reader will have many questions concerning why certain choices are made when it would appear that a myriad of options are available. For these reasons, before diving into a more concrete analysis, it is worth looking at the basic analytical tool under consideration.

A feature, as used in morphological theory, is a labeled set of related values that are mutually exclusive. I the values are in opposition. A feature structure is a collection of feature values, which designate concepts such as paradigm cells, and information at syntactic nodes. A paradigm structure begins as a Cartesian product of feature values associated with a class of lexemes. In PFM, there are additional constraints imposed on the structure so that certain coordinates do not exist. Please be aware that I overload the word paradigm when I use it for my own framework, in a manner that is not inconsistent with usage in Word and Paradigm (WP) theories. I define taxonomies with classes. Those classes are also called paradigms.

I assume the reader is familiar enough with morphosyntactic features that they have seen notation such as [+plural, −masculine]. These are popular in DM, where more features are used, but with each feature having only two values. These technically form paradigms (orthogonal taxonomies), but DM scholars are often skeptical of the explanatory utility of such taxonomies (Harley, 2008; Bobaljik, 2002). I don't talk much about the conventions of purely binary features below. Another notation that is likely to be familiar is {NUMBER: plural, GENDER: feminine}. I will be using the latter notation in discussing PFM. I follow the convention of only including the feature name, such as NUMBER, when it is necessary, and otherwise display only the values, for instance {plural, feminine}.

4.2.3.1 The meaning of features

Despite having a fairly consistent use in multiple formalisms, the conceptualization of morphosyntactic features across theories is not consistent. This inconsistency takes many forms. One important aspect of inconsistency is the differences one sees in their explanatory role. There is a basic dichotomy in theories of features. On the one hand there is the notion that features are shared between languages, and so one should look to typology or UG for evidence of a feature, and an understanding of its properties. On the other is the notion that features simply express contrasts, and naming them is an exercise in providing a mnemonic for what a researcher believes they (indirectly) relate to in terms of meaning (speaker versus addressee), or some other aspect of grammar (AUX for auxiliary verb). I believe it is safe to say that the former is, by far, the dominant practice.

The notion that this divide exists in linguistic scientific practice is discussed in works such as Haspelmath (2019). There the author discusses *comparative concepts*, which correspond to cross-linguistic feature sets, or UG features. He contrasts this with *descriptive linguistic categories*, which are a means for expressing contrasts within a particular grammar. The author advocates for the need for descriptive linguistic categories, rather than assuming comparative concepts as the driver of language specific analysis.

Theories that see a relationship between morphosyntactic features and UG are diverse. This is an additional reason why DM is best described as a programme, since the interpretation of basic representational units have vastly different interpretations within the literature. Sometimes morphosyntactic features are seen as basic primitives of UG, as in a DM subprogramme called Cartography (Cinque and Rizzi, 2004). According to Cartography, a closed set of cognitive features directly explain the similarities and differences across all languages, as well as what languages may exist. Essentially, the programme states that features are substantive. A substantive feature corresponds to something concrete, like an anatomical tongue position feature in phonology. According to this view, features are inherent, and encoded physically in the human mind. Others in the DM literature argue that the features themselves may be language specific, but are constrained by UG (Cowper and Hall, 2022). This implies that the features are an epiphenomenon with respect to an explanation of *why* there are commonalities among languages. That is to say, the particular features of a language are a secondary phenomenon caused by a primary phenomenon, the constraints of UG. To the extent that features are necessarily explanatory from this point of view, it is only in their role of distinguishing categories and distributions on a per language basis. They may also be considered explanatory in their role of representing delineations found in comparative studies.

UG research is not the only area where comparative concepts inform feature validity. In theories that are agnostic or skeptical of UG, authors may take into account concerns stemming from linguistic typology in determining what should be considered a valid feature. These researchers may or may not be typologists themselves, but the justification for feature choice or relationships among features may align with comparative patterns (Brown and Hippisley, 2012; Corbett, 2005). Typological patterns are often conflated with UG, but in the UG case, the patterns are presupposed to be cognitive, while more broadly in typology, explanations for patterns shared across languages may be more diverse (Andrea D Sims et al., 2022; Greenberg, 1963). From such a stance, features are not explanatory of similarities and differences in languages, necessarily, but features are shared among languages. Within a theory of language types, they represent cross linguistic patterns. For instance, if there is a clusivity distinction in one language it is the same clusivity as found in other languages (up to certain considerations of types and subtypes of clusivity). The features that represent clusivity then have an identity across such languages. A discovery of some property of clusivity in language A should therefore influence the feature analysis of clusivity in language B. From this perspective there will be concerns about what the features mean in a cross-linguistics sense. This may involve proposals of natural classes, markedness, implicative relations and other properties within the feature theory. Though the feature itself is not considered to explain why a phenomenon exists in a language, the existence of a feature with particular properties in one language is thought to explain the existence of features with the same properties in another language. That is to say that features explain *how* languages form a system of types, though not necessarily *why*. This thinking is explicit in PFM analyses (Stump, 2001, pp. 235–241), though not always mentioned as a justification for feature choice. The stance is more overt in related theories such as Network Morphology (NM) (Brown and Hippisley, 2012) and, especially, Canonical Morphology (CaM) (Corbett, 2005), a typological framework. That features are informed by typology, and syntax is a system of features, in effect, makes syntax an object of typological theory. The place of a language within a taxonomy of languages is expressed by the features that are proposed to exist in the syntactic structure. Without understanding this typological view, it is difficult to understand why certain proposals concerning features are made in post-syntactic realizational theories, since the perspective may result in less parsimonious analyses than one would want.

In contrast to those who appeal to comparative categories, Pullum and Tiede (2010) see no necessary correspondence between features used in one language with those in another. Their practice corresponds to what Haspelmath (2019) referred to as descriptive linguistic categories. In their view, if there is a grammatical contrast, there is a grammatical feature that represents that contrast and that is the extent of their explanatory role. The features are formal objects for distinguishing categories within a specific grammar, and cross-linguistic patterns are not explained by languages sharing similarly named morphosyntactic features. I am basically in accord with this stance, despite not finding morphosyntactic features to be the best way to express such contrasts. It should be noted, that Pullum and Tiede (2010) recognize a tension, even then, between practice and aim. A large portion of their work examines with gender features values such as feminine and masculine are not the same feature values across languages. One has to ask why such argumentation is necessary. The answer is that, in practice, a comparative template is always implicit in feature analyses. Pullum and Tiede (2010) essentially argue that the comparative template does not always fit, but they recognize that it is standard practice.

The takeaway is that even though there is a rough consensus concerning how feature are used as formal objects in different theories, they are by no means the same type of theoretical object in one theory or another.

4.2.3.2 How features relate to syntax and semantics

All post-syntactic theories agree that the features relevant to realization are primarily, if not entirely, syntactic. This is consistent across foundational works, such as Stump (2001), Zwicky (1985a), Aronoff (1994), Beard (1995), Carstairs-McCarthy (1987), and Halle and Marantz (1993). Morphology in these theories is a non-generative system that is largely determined by syntactic information. There are features that are purely morphological, which may determine certain allomorphy patterns, but not all theories use these, and how they are used is variable. I discuss them in more detail later in this work. None of the cited theories allow for other kinds of non-syntactic features during realization.

The issue with features being purely syntactic is that it creates strange distortions in the role of syntax. In modern theories, syntax is rarely a system limited to modeling and explaining word order and co-occurrence restrictions as one might expect. Constituency tests, replaceability, and other syntax-based discovery mechanisms rarely, if ever, figure in the discussion of the validity of a feature within a post-syntactic analysis. None of the works cited in this section do so.

A common trend in post-syntactic theories is to embed semantic information in syntactic structures. For instance, in DM, morphosyntactic features and √roots are essentially abstract meaning objects that are interpreted at LF. If there is a deterministic relationship between a morphological form and a meaning in DM, the meaning cannot be introduced at LF because there is no direct communication of information between LF and PF branches after spell-out. The meaning must be in the syntactic structure prior to spell-out to be considered in the realizational calculation. PFM, likewise, claims that semantics is mediated by the syntax, such that a pair of morphosyntactic features and a lexeme, each of which are considered to be objects of the syntax, determines meaning for an inflected form. This is unambiguous in the description of the Hypothesis of Paradigm-Based Inflectional Semantics (Stump, 2001, p. 248), where the meaning of morphological forms are provided by paradigm cells, which *must* correspond to features at syntactic nodes. What this means is that meanings may be considered as relevant (if not more) than purely syntactic patterns in proposing morphosyntactic features. As a particularly clear DM example of this, see Acquaviva (2009), where semantically motivated features, such as a concrete notion of person-hood, are introduced into the grammar to avoid overabundant analyses of irregular English plurals, e.g. *oxes* versus *oxen*, where the former is (supposedly) sensitive to a positive person feature. This approach is criticized in Panagiotidis (2024), though not due to any issue with the notion of embedding meaning in the syntax. The claim is only that √roots should be used in place of such features more often.

Consider also that in the description of the Wao Terero person system in §3.2.1 there are plausibly semantically motivated patterns. I choose to provide a grammatical analysis of Wao Terero person for reasons described in Chapter 6. I explicitly state there that there is more than one valid hypothesis at this stage of Wao Terero description. I am able to make a choice as to whether I treat the person system as purely syntactic, semantic, or some combination. This is not a choice that can be made in the most popular post-syntactic theories. If the meanings are not available in syntax, then there is no story as to how they can affect the realizational calculation in PFM or DM.

There is, of course, diversity in realizational theory. Other theories may allow means of expressing realization in a manner distinct from DM and PFM. Unfortunately, the review of feature use in this work is less than comprehensive, which is to be expected given the long history and centrality of feature-based approaches in linguistic theory. My goal is to focus on the most famous theories, and expose their negative consequences.

4.2.3.3 Orthogonality and necessary opposition

That multiple, orthogonal meanings are associated with word-sized units is a basic observation. Orthogonal taxonomies of words are elegantly expressed using features. Despite this, orthogonal meanings do not require features for expression, and their use introduces properties to taxonomies that may not be desirable. The use of features makes a strong prediction concerning grammar, which is that morphosyntactic categories are *necessarily* in opposition. At the least, there must be a pair of feature values. A single valued category is not a feature, since the necessary opposition is not expressed. Non-feature categories are things like √roots or lexemes, where no opposition is implied.

In early morphological theories, the notion of feature was not as universally codified as in current theories (Bloomfield, [1933] 1984; Hockett, 1947; Z. S. Harris, 1942). Yet, even when the formal mechanisms were absent, feature-like ideas were sometimes present. An important correlate to the predictions made by feature use are found in Z. S. Harris (1942). Harris asserted that morphemes exist in oppositional relationships, called parallels, which I am calling necessary opposition. As an example, in English the plural morpheme was expected to be in necessary opposition to a singular morpheme, this drove the need for zero morphemes on bare nominals. This notion of necessary opposition is evident in modern feature analyses. In post-syntactic realizationalism, English nouns are, likewise, considered to have a singular and plural cell due to the inherent opposition within the NUMBER feature. This is reasonable given English agreement patterns. Yet, the consequences of necessary opposition may not always be positive. Hockett (1947, p. 322) warned that Harris's concept of necessary opposition potentially creates spurious grammatical categories. This is evident in modern feature use.

The notion of necessary opposition suggests that words and word-forms exist within taxonomies defined by meaning oppositions. The first effort at supplying formal definitions of feature-based linguistic taxonomies, that I am aware of, was in a comment by Paul Kay in Colby et al. (1966). The idea of a feature based paradigm, defined by the Cartesian product of feature values was proposed. As an example Cartesian product of features, consider a situation where there is one set of values for gender, for instance, $\{\text{masculine}, \text{feminine}\}$, and another for number, $\{\text{singular}, \text{plural}\}$. These are then multiplied together to form a four cell grid $\langle\text{masculine}, \text{singular}\rangle$, $\langle\text{masculine}, \text{plural}\rangle$, $\langle\text{feminine}, \text{singular}\rangle$ and $\langle\text{feminine}, \text{plural}\rangle$. The utility of features is that they may resolve to such grids, called paradigms, also known as orthogonal taxonomies. As the realizational literature demonstrates, such structures allow for a great deal of abstract reasoning over inflectional patterns.

From a Harris-like stance, one would expect perfect grid-like paradigms to be common. This is because zero morphemes provide formal units that correspond to cell meanings, even when there is no formal exponent. According to the stance, one would expect a distinct theoretical object corresponding to all of the persons and numbers found in English for each verb. If there is a third person singular *walks*, there must be a form for first, and second, as well, each with orthogonal plural and singular values. This would predict a grammatical second person plural in English. Contrary to this, Paul Kay stated that it appeared that orthogonal taxonomies are actually very rare. This is not to say that there may not be orthogonal information, but there is also sparseness and non-orthogonal dependencies, which lead to more tree-like taxonomic structures in most cases.

Despite embracing orthogonal taxonomies of the type described above, PFM avoids some spurious oppositions through mechanisms such as *co-occurrence restrictions*, which allow for items to be deleted from the taxonomy based on certain patterns. Despite this, the nature of features and the Cartesian approach to taxonomy definition still results in issues that invoke Hockett's warning. The clearest example is what (Stump, 2001, p. 215) calls *unstipulated syncretism*. Unstipulated syncretism covers exactly those morphosyntactic categories with no overt morphological expression anywhere in the grammar. For instance, unlike the past participle–past tense syncretism of forms like *talked*, which can be compared with non-syncretic *drive/driven* to demonstrate the validity of a syncretic analysis, there is a uniform lack of formal contrast. This is not necessarily problematic. There is legitimate syncretism that cannot be said to be marked by a morphological contrast anywhere in the grammar. Evidence for such cases comes from other grammatical cues. For instance, the Spanish *usted* forms of verbs are identical to third person forms. In that case, pronominal patterns provide evidence. Unstipulated syncretism is not confined to such cases. As discussed above, necessary oppositions may predict entirely spurious categories. An example of a spurious distinction is stating that there are distinct grammatical English second person plural and singular categories. There is no external evidence of the grammatical distinction, only a prediction made by necessity due to the nature of features, which require grammatical opposition.

4.2.3.4 Common practices in feature discovery

It was clear from Kay's discussion that there was no single feature discovery methodology at the time. This remains true today. Kay discusses the notion that feature discovery may be performed with or without an etic grid, which is a kind of template of features found outside the language under study that is matched to contrasts within the language. This indicates that the dichotomy between comparative and language specific descriptive features was already well understood at that time. Feature analyses reinforce etic grid analyses, and place a focus on comparative categories, which has negative effects on linguistic description. Aside from this, post-syntactic analyses often utilize opaque feature discovery procedures, which put in doubt the empirical validity of the features.

This section is written from personal experience of feature discovery practice. I have worked on a va-

I see her.	Botõ abopa idã idã.
She sees me.	Ĩdã adãpa botõ ĩbo.
He sees me.	Ĩkã akãpa botõ ĩbo.
I see you.	Botõ abopa bitõ ĩbi.
You see me.	Bitõ abipa botõ ĩbo.
You all see me.	Bĩditõ abĩdipa botõ ĩbo.
They see me.	Dãditõ adãdipa botõ ĩbo.
We see you.	Bõditõ abõdipa bitõ ĩbi.

Table 4.8: An example of basic translation data centered on the word 'see'.

riety of linguistic documentation projects, and have received training as a fieldworker under a number of scholars. I cannot speak to all practices, but I believe that personal experience speaks more to actual practice than idealized textbook procedures. A more comprehensive discussion of feature discovery would include these, but I expect that my discussion of practices matches with the experience of other scholars engaged in linguistic description, even if there may be disagreements concerning my conclusions.

The standard field practice for meaning and category correspondences for words focuses on minimal pairs. This is what is taught to any introductory linguistics student. Pairs of differing word-forms are compared, which are believed to share some meaning, or that are known to share some form. On my first day working on Wao Terero, I began my own fieldwork by asking for translations of Spanish phrases into Wao Terero. I wanted to discover information about pronouns, person marking, etc. I kept the verb in the phrase constant, and asked questions like "How do you say, 'She sees him'." or 'I see her'. Through this process, I was able to discover a variety of forms corresponding to 'see', where it was very likely that they differed only in information related to person. My data looked much like Table 4.8. The data is not glossed, because it is an example of primary data. The idea is to look at it and figure out the patterns.

Looking at Table 4.8, it is easy to see what remains constant, and what changes from one line to the next. There is no obvious change in the verb depending on the object. There is a dimension of person, another of number, and one of gender. As a first approach, if I am looking to characterize the data in terms of features, each of these dimensions would correspond to a feature. The values would correspond to what is observed, three persons, two genders, and two numbers. The analysis would be expanded on finding further pairings.

It should be noted that an analysis of complete orthogonality among dimensions is in question, immediately. For instance, there is no gender for the first and second person. This implies a dependency of gender on third person, which suggests a tree-like structure. Later, I encountered clusivity patterns. This brings into question whether the persons are actually in opposition, since clusivity can be considered a combination of first and second persons. Inclusive also intersects with plurality. With two persons, there is logically more than one person.

If I were a feature theorist, I may take two courses of action when I make such observations. If the dependencies result in difficulties for analysis, I may decide that some form of restriction needs to be placed on the taxonomy, in order to encode the dependencies. These are used in PFM, and are called co-occurrence restrictions. If the dependencies do not interfere with the analysis, I may choose not to place unparsimonious restrictions on the taxonomy, this may result in unstipulated syncretism, but would simplify the definition of the taxonomy. I may also chose not to use co-occurrence restrictions if I actually believe that masculine and feminine genders are covertly active for first and second persons. Essentially, my choice may be heuristic, based on a preference for a particular type of model or parsimony, or it may be based on a scientific hypothesis or theoretical stance.

The methodology described above is simplistic. It is subject to certain biases and distortions. Whether I wanted to or not, I was imposing an etic grid. In particular, I was constrained in my investigation by my own language. There were questions that I would not have thought to ask for this reason. I did not think the verb form would change if I was talking to my mother, for instance. Yet, there is a second person form for speaking to one's mother. I was also likely adding things to the data due to my bias. For instance, Spanish gender agreement may be echoed in the Wao Terero translation in a way that does not represent normal usage. As described previously, the feminine is not used in a Spanish-like fashion. For these reasons, I augment naive elicitations with other methodological techniques, such as corpus studies. Even then, I am a creature of my training. I am looking for agreement morphology. I expect certain patterns.

The training is good. It facilitates my understanding of patterns, when correct. Comparative linguistic science has made very useful observations about cross-linguistic patterns that can and should be leveraged when appropriate. Yet, the training is also a trap, since it may blind me to particular details, and encourage me to jump to certain conclusions. Frequently, when presenting data to other scholars, where lexical suffixes recur in a sentence, there is an automatic response that it is grammatical agreement. That pattern has been drilled in to us as linguists to an extent that it is difficult to ask why else a formal pattern may occur multiple times in a sentence.

Immediately jumping from data to a feature analysis entrenches such thinking. By pidginholing an analysis early, it may blind one to alternatives. Features, by their nature, contain a comparative bias. Where do the feature names and concepts come from? At the initial stages of analysis in primary field research they never come from the particular language under study, they are always outside categories that one fits to a language.

I approach my work with skepticism toward comparative categories. The etic influence in my work is often implicit, and something I attempt to question. Yet, it is common to embrace a comparative approach. The way that morphosyntactic field research is often taught involves explicit etic templates. In the initial training of a field researcher, a text book such as T. Payne (1997) is used, where there is a chapter on person and number, a chapter on valency, one on tense, mood and aspect, etc. In each chapter there is discussion of cross-linguistic variation found within the area of focus, and information on various diagnostics to place a language within a typological taxonomy. One can follow the text as a template for writing a grammar. When following this approach, minimal pairs are also used for good evidence. That core concept is not lost. An issue, or feature, depending on one's perspective, is the emphasis on fitting the language within particular slots. An extreme example is the *The Lingua Descriptive Studies Questionnaire*[3]. This tool contains a series of questions to ask about a language. It is designed with the idea that one can create grammars according to a rigid template, which can then be easily compared with other languages using the tool. An example of a grammar that uses the Questionnaire for Imbabura Kichwa is Cole (1985). All of the grammars using this technique have a uniform structure, which hypothetically makes them excellent for comparison.

The way that features emerge from these studies is that by matching the language of study to a template, there are a number of features that have already been proposed in the literature for the various comparative categories the template encodes. It makes it easy to fit the language into an established typological or UG framework. One issue is that by not having diagnostics, or ready made features, for something unique to the language, such traits may remain hidden. Another is that by steering the linguist toward questions concerning phenomena that may not exist in a language, a language may be assigned to a typological category based on best fit, rather than a truly meaningful identification. Such issues can be and are ameliorated through careful

[3]https://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/linguaQ.php

scholarship and awareness of the issues involved. There must always be an interplay between comparison and a study of the internal categories and relationships of a language.

Despite this, I am often concerned that comparative studies may be suffering from systematic confirmation bias, given that grammars are written to match comparative theory, and then comparative studies justify their claims by reference to grammars that were written to conform to the theory. Consider projects such as the Online Database of Interlinear Text.[4] The project allows for decontextualized mining of interlinear glosses found on the web. Many of these conform to Leipzig Glossing Rules (2015), a comparative theory of interlinear glosses, which linguists are trained to conform to. The prominence of such a standard is scientifically odd. What if there had been a hypothesized periodic table, which all chemists necessarily needed to conform to in presenting results, instead of a real periodic table that had been proven true? The standard exists because the proponents believe that they are describing something comparable. That all interlinear glosses then conform to their standards validates that belief. I do not believe there are any linguists that believe that one should not deviate from the standard if they feel it does not fit a language, but the training is to go to the standard first, and argumentation for deviation may sometimes be arduous. If there is agreement that sometimes aligns with sex, perhaps it is easier to simply call the values masculine and feminine, rather than fight an uphill battle against reviewers, and pointing out the many other relevant dimensions of the category. I believe that something similar occurs with explicit feature analyses, since utilizing feature names that do not refer to standard comparative categories is contrary to training, and necessitates a great deal of additional work when justifying the practice to other scholars.

Another common approach to proposing features is to read a grammar, and then decide on the features that allow for one's secondary analysis. A very informative example comes from a comparison of Potawatomi analyses from three famous theoretical works Stump (2001), Anderson (1992), and Halle and Marantz (1993). I will not describe these analyses. They are in well-known works, and Potawatomi is far outside the core concerns of this thesis. I simply wish to highlight some aspects of the choice in features, which the reader can verify. No primary research was performed in any of these studies. They are all a secondary interpretation of Hockett (1948). The analyses focus on verbal inflection. All share commonalities, such as person and number features, though they are represented differently across analyses. Beyond this, there are a variety of

[4]https://depts.washington.edu/uwcl/odin/

differences and innovations.

Halle and Marantz (1993) propose in their feature system that Potawatomi has a nominative-accusative case system. This is a bold move considering that there is no case marking on nouns in the language, nor a particularly strict word order. It is contrary to what many scholars have described for languages of the family (Dryer, 1997; Wolfart, 1973), and was not a claim made in the Hockett analysis. It seems very likely that the choice was made to fit the language into a comparative mold. The authors also propose a feature *obv*, which has some unique traits. Despite the fact that the authors proposed a binary $+/-$ feature system, it has the unique property of carrying three values, one of which is a null value. It provides the reader with the sense that constraints within the framework are only meaningful until they are inconvenient. It also raises unanswered questions concerning the feature discovery procedures used by the authors. Though the obviative is important in the Algonquian literature, no analysis, before or since, has accepted that it is anything other than a binary feature, or unary category.

Anderson (1992) is relatively conservative in the features proposed. One innovation is that unlike the other two authors, there are features $+/-$Verb and $+/-$Noun, used for the inflection of verbs. The notion is that at some point in the inflectional calculation of a verb, the realization is driven by some form of nominal status. This was criticized by Halle and Marantz (1993), demonstrating that despite their own practice, they believed, unlike Pullum and Tiede (2010), that features should have some meaningful interpretation, rather than being mere signals of contrast. It is true that there are patterns shared by nouns and verbs in Potawatomi, but it is odd to make a claim that much of verbal morphology is based on something like a nominal feature. It was not consistent with Hockett's analysis.

Stump (2001) does the best at paying tribute to Hockett's original analysis by borrowing the terminology such as *major reference* from the original, which he proposed as a feature that works similarly to Hockett's concept. There are a number of innovations used in his analysis. Feature structures in the analysis have embedded, tree-like structures. Remembering the discussion of Paul Kay's initial formal work, this kind of structure implies that the use of an orthogonal taxonomy was not the right choice for the language, and that there is some shoehorning taking place.

I leave it to the reader to review the famous analyses above in depth. The only point that I wish to make is that this is how features are often discovered. The features used by the three authors were so radically different, that I strongly suspect that none of the authors were rigorous in their feature discovery procedures. If the chosen features represent valid theoretical constructs, they should converge when properly measured, even when very different methods are used (Campbell and Fiske, 1959). Radical divergence signals that the constructs themselves may be invalid, the methodology deeply flawed, or both.

There are good ways and there are bad ways to perform feature discovery. Despite the claims of Cartography, one thing that has never been demonstrated is that there are substantive morphosyntactic features. There is never a concrete manifestation. Therefore, even when features conveniently group, they do not form natural classes, in the sense of nasal phones. Features in formal morphosyntax are always building blocks of theoretical constructs and hypotheses, of what may be true, but they cannot be measured directly. Their use, as seen in the formal literature, expresses an attitude, or position, never an established fact. Note that given the discussion above, it is also the case that despite similar practices in feature use, a scholar may believe the features to be cognitive, typological, or simply convenient naming for entirely language internal categories. The fact that one may not be able to differentiate which of the three sources a feature is born from, except by a stated attitude, reinforces the empirical disconnect of feature analyses.

I see this as a problem that requires solving. For this reason, I take a form-oriented position. Form can be concretely measured. It is also inherently emic, specific to the language under study. It is better to proceed from the measurable to the abstract.

Before I can demonstrate how this works, I want to show not only that feature theories make questionable assumptions, but that there is data that they can't handle. As a precursor I perform an analysis of data features do handle. In the next section I provide a PFM analysis of Wao Terero verbal inflection. I focus on a verbal paradigm that looks a lot like inflection in other languages. The analysis will show how theories that use morphosyntactic features are supposed to work.

4.2.4 Wao Terero verbal inflection in PFM

Table 4.9 displays a sub-paradigm of the Wao Terero verb GO, 'to go'. I largely follow the paradigm as drawn by Peeke (1968), with some tweaks that are explained throughout this section. The table excludes forms for non-present tense, non-indicative/interrogative mood, and other characteristics focusing on person and number morphology. Given that a relatively detailed description of the data was supplied in §3.2, I will

	Singular	Plural	Dual
1	gobo	gobõdi	gobõda
1 Inclusive		gobõ	
2	gobi	gobĩdi	gobĩda
2 Maternal	gobĩ	**gobĩdi**	**gobĩda**
3 Non-sentient	go	go	go
3 Sentient	gokã	godãdi	goda
3 Feminine	godã	**godãdi**	**goda**

Table 4.9: A sub-paradigm of GO, 'go'.

be brief in going back over the data. The basic description follows:

1. A lack of person marking is the most common case for the third person. It is found across the non-sentient items and in the third person dual.

2. The *-dã* suffix occurs in the feminine singular and in the sentient third person plural.

3. The suffix *-kã* is the most restricted third person affix in terms of distribution.

4. The non-nasal *-bo* and *-bi* are specific to singular non-maternal first and second persons, with nasal *-bõ* and *-bĩ* occurring more generally.

5. If not for the third person dual, it would look like *-dã* distributes much like *-bõ* and *-bĩ* as a more general third person form for sentient persons.

6. The affixes *-di* and *-da* correspond to the plural and dual respectively, based on their alignment with those features and their independence from person marking.

There are a couple of forms in Table 4.9 that are in bold. This is to point out the effect of *full specification* in post-syntactic realizational theories. Essentially, these forms would not be proposed as being associated with the rows for maternal or feminine without this concept. They are instances of unstipulated syncretism. Full specification is of central importance to understanding post-syntactic realization. It dictates that there must be one feature value specified for each feature relevant to the paradigm as a whole in every cell. The features I will use are NUMBER, PERSON, INCLUSIVE, MATERNAL, SENTIENT and GENDER. Their possible values are listed in Table 4.10.

NUMBER	singular, plural, dual
PERSON	1, 2, 3
INCLUSIVE	inclusive, exclusive
SENTIENT	sentient, non-sentient
GENDER	feminine, neutral
MATERNAL	maternal, non-maternal

Table 4.10: The features to be used in the Wao Terero analysis.

Position 1	Position 2
bo	di
bi	da
bõ	
bĩ	
kã	
dã	

Table 4.11: The affixes in Table 4.9 can be divided up into two position classes.

Given full specification, this means that there should be a cell for each combination of the values above, for $3 \times 3 \times 2 \times 2 \times 2 = 72$ cells. This is not the case in Table 4.9. I explain why and go into greater detail on the concept of full specification below.

The effect of comparative categories and meaning should be evident in the choices of features that I have made. For instance, I use 1, 2, 3 for person, following a general comparative assumption for grammatical person distinctions. This causes the sentience and gender distinctions to be sub-distinctions of persons rather than something on a par with what we call persons, despite the fact that it would potentially simplify the analysis. Even though GENDER functions somewhat differently in Wao Terero, I chose to align it with the broader notion of gender through the feature name. I may have chosen the feature name FEMININE and values specified and unspecified, for instance. Distinctions such as GENDER, SENTIENCE and MATERNAL are given their own feature and values. The choice to divide them up so is based largely on the meaning domain they are concerned with, not a syntactic analysis. I do not feel that any of this is unusual. Few of my choices are necessary choices, but I believe I am being faithful to common practice.

Given the paradigm above, it is clear that *-di* combines with some exponent of all persons and should therefore be considered independent of *-bĩ*, *-bõ* and *-dã*. It is also clear that the dual *-da* is similarly independent, especially given that it occurs without a preceding person suffix in the third person. Both *-di* and *-da* follow *-bõ* and *-bĩ* while never co-occurring. Given this, it seems reasonable to propose that they form a class. The other affixes *-bo*, *-bi*, *-bõ*, *-bĩ*, *-dã* and *-kã* also do not co-occur. Given a more complete paradigm, it is demonstrable that they largely pattern together when other affixes are involved. For instance, they always precede the assertive affix *-pa* and follow affixes such as *-kĩ* used for desires and future events. What I am describing has been commonly modeled using position classes, where *-di* and *-da* would be in one position class and the other affixes in another. This is diagrammed in Table 4.11.

PFM does not have position classes but instead a concept of rule block, where items that can be grouped into classes are in so called *competition*. Competition is a concept that explains a lack of co-occurrence relative to some class of feature values. Therefore, with respect to Table 4.11, given that items in position 1 and 2 may co-occur with items in the other position, but not with items listed within the same position, and because the person and number features they are sensitive to arguably form classes of features, there should be two rule blocks that correspond to the two positions in the table.

Note, PFM rule blocks are not a perfect analog for position classes. Position classes are more intimately descriptive of morphotactic patterns, the syntagmatic dimension of morphology. A rule block is more concerned with competition, which is to say non-co-occurrence relative to some class of features. The distinction will be clear once rule blocks are presented. Before doing so, it would be helpful to talk about full specification in slightly more depth so as to understand the context of the analysis better.

4.2.4.1 Full specification

In Table 4.9 there are various forms in bold. I have provided a brief explanation for why these exist, which is full specification. Here I describe what full specification is intended to represent, and what its empirical claims are. Explaining this is somewhat hard to do with Wao Terero since the reader will be unfamiliar with the grammar. For that reason it will be good to look at an example that is well known. The example will be returned to for the purposes of theory comparison in later chapters. It is clearly a toy example, but I only intend it as an aid in discussion. The example is salient in the literature due to some exchanges between Blevins (1995) and Stump (2001).

In Table 4.12, we see the English present tense sub-paradigm for WALK. According to the standard PFM view, there are two sets of feature values relevant to the paradigm, person feature values $\{1, 2, 3\}$ and number

	SINGULAR	PLURAL
1	walk	walk
2	walk	walk
3	walks	walk

Table 4.12: A fully specified sub-paradigm of the English verb WALK.

values {singular, plural}. The Cartesian product of these values:

$$\{\langle \text{singular}, 1\rangle, \langle \text{singular}, 2\rangle, \langle \text{singular}, 3\rangle, \langle \text{plural}, 1\rangle, \langle \text{plural}, 2\rangle, \langle \text{plural}, 3\rangle\}$$

defines two dimensions of the paradigm structure for a verb. Each element of the paradigm structure that corresponds to one of these coordinates is called a cell. Each cell corresponds to feature values that exist on syntactic nodes. The node may have no more, nor less features than the cell provides, otherwise the cell is not fully specified (Stump, 2001, p. 12). It is important to understand that the paradigm structure is not a statement concerning the morphological form-meaning correspondence directly. *The cells are a specific proposal concerning syntax.* I emphasize this statement because the definition of the cells are easily mistaken for a morphological claim. The feature values are syntactic primitives. Nodes corresponding to each cell that carry the feature values must exist in the syntactic theory. This is a meaningful prediction made by post-syntactic realizational analyses. Essentially, a reasonable syntactic analysis, performed independently of morphological concerns, should require no greater nor fewer syntactic distinctions than indicated by the cells. Otherwise the proposed paradigm lacks a verifiable foundation.

The question is how well the syntactic claim is born out. There are reasons for doubt. It is clear from Table 4.12 that full specification overgenerates syntactic distinctions. The second person plural in English does not exist as a syntactic agreement distinction. Additionally, only for the copula is there an agreement distinction for first person singular. Using the distribution and agreement patterns of the copula for verbs in general would be an odd choice in English.

A popular counter proposal (Pollard and Sag, 1987; Andrews, 1990; Blevins, 1995) to the fully specified paradigm is provided in Table 4.13, which is that English verbal agreement exhibits a paradigmatic opposition between the third person singular and an elsewhere case. Importantly this is the view of early HPSG (Pollard and Sag, 1987; Pollard and Sag, 1994), which is a syntactic theory commonly assumed by PFM practitioners.

3 Singular	Elsewhere
walks	walk

Table 4.13: A counter proposal to full specification.

This indicates that the fully specified paradigm is not a necessary syntactic analysis of English, though perhaps there are syntacticians who would argue that it is a superior analysis.

Stump (2001, p. 287) states that the proposal in Table 4.13 is contrary to realizationalism. The realizational calculation of PFM expresses a similar generalization while still maintaining the fully specified paradigm. I do not feel that the result of this type of realization is always so meaningful. I will return to this point after discussing how realizational rules function in post-syntactic theories. It also assumes that the only way to conceive of realization is by reference to features and post syntactic analysis. I discuss whether my approach is therefore realizational in Chapter 7.

It may be clear that I take a negative attitude toward the syntactic interpretation of full specification. Yet, I absolutely feel that it attempts to express something useful. Consider the claim of Harris, that there is a zero morpheme for English singular nouns. This communicated two things, a kind of proto-feature analysis, where related meanings form sets of values in opposition, which I question, but also that a form with no overt affix for singular is the correct form for singular usage. It is more restrictive, and accurate, than saying that a bare noun may be associated with any meaning and is simply "elsewhere". That restrictiveness needs to be communicated somehow. Harris chose to tag forms with zero morphemes to do so. This is not ideal, since a zero morpheme has the same status as a non-zero, but the discovery procedure is not the same. So long as one can posit a meaning, they may add a zero, which is then purported to have the same ontological status as a form-bearing object discovered through minimal pair analysis. What the zero morpheme usefully achieves is to acknowledge the multiplicity of meanings that may be associated with identical forms in a constrained manner.

Full specification has the same aim, though without the ontological complication of the zero morpheme. All else being equal, Table 4.12 is more constraining and descriptive than Table 4.13. The first says, for the bare form, these meanings are available, no more no less. The second says, use this form whenever the meaning isn't third person singular. Of course, all things are not equal, and it is expected that syntactic and semantic constraints exist in other areas of the system if one assumes Table 4.13.

The issue with full specification is not that it attempts to list the valid meanings, it is the way it does so. The taxonomy of meanings is fundamentally flawed. Paul Kay was right. Such perfect orthogonal taxonomies rarely exist in nature. This is an opinion shared by post-syntactic realizationalists. CaM (Corbett, 2005) is an explicitly typological theory that focuses on systematic ways in which taxonomies diverge from the orthogonal ideal, which is thought to rarely, if ever exist. For the analysis of a particular language, even if one has comparative concerns, beginning from such an idealistic assumption is nearly begging for analyses that poorly fit the data. This is the case in Table 4.12. The multiplication forces items to match up, which perhaps do not need to be restricted from doing so, but may simply not occur. Since features are the only way of expressing meanings, perhaps they are being used in cases where the values are not in grammatical opposition, but are incompatible for other reasons. It would be good to understand those and model them appropriately in terms of semantic entailments, perhaps. Later in this work, it will be shown that syntacticians actually require distinctions that are not found in paradigms, such as Table 4.12. It is almost certainly the case that Table 4.12 is failing to anticipate licensing conditions that are less than obscure, but, again, may not be very feature-like.

It is not enough to simply abandon the descriptive value and restrictiveness that full specification attempts to communicate. I take a far different approach, which is more fine grained and better integrated with syntax and semantics.

4.2.4.2 Property co-occurrence restrictions

The discussion of full specification above explains the significance of the bold forms in Table 4.9. The values for MATERNAL and GENDER are part of the Cartesian product that defines the paradigm structure. Therefore there must be cells for these values combined with all NUMBER values, whether it is obvious that the distinction is relevant in the plural or not.

Despite this, Table 4.9 does not have 72 cells. In PFM there are some limits on full specification. So called *property co-occurrence restrictions* may be used to carve out a subset of the Cartesian product, placing a limitation on possible cells and by extension possible syntactic nodes. For instance Stump (2001, p. 88) proposes such a restriction on a binary feature for inclusive, such that the positive value for inclusive cannot co-occur with various persons and numbers. I follow this for Wao Terero. I do not know the clusivity patterns perfectly, so I assume that they are relevant only for the plural. Likewise, it seems logical that values maternal and feminine as well as the interlocutors, are incompatible with non-sentience. It also seems logical that for all cases of positive maternal values there is also a feminine value. This still leaves us shy of the proper number of cells. These were omitted in the table so as to not fill the page with cells that are irrelevant to the analysis. Despite this, things like {1, dual, sentient, non-maternal, feminine} and {2, dual, non-sentient, non-maternal, non-feminine} must exist as cells under full specification.

Returning to an earlier point concerning the use of comparative information and appeals to semantics, note that the property co-occurrence restriction concerning sentience and persons appeals to world knowledge, rather than grammar. Comparative convention dictates that the contrasts among persons are 3. Yet, if I were to place four values under person, one for first, second, third sentient and a fourth for the non-sentient, I wouldn't need a property co-occurrence restriction and therefore could more concisely express the grammatical contrast. Relying on comparative and semantic justifications for feature choice may complicate an analysis in some cases, and obscures the notion of grammar as independent from meaning.

4.2.4.3 Rule blocks for Wao Terero verbs

The feature values that define the cells of a fully specified paradigm, together with a lexeme, provide the input to post-syntactic realizational calculations. A lexeme is associated with a stem. The stem and features values for a cell are fed to a sequence of rule blocks. A rule block for the *-di* and *-da* suffixes is provided in Figure 4.13. I start with this rule block because it is simple. In the rule block there are three rules, called *rules of exponence*. On the left is some subset of possible feature values for particular cells of the paradigm. The input feature values of a cell are compared against each rule of exponence in order. The first rule that specifies a valid subset of the input features is applied to the input stem and no others. If the rule block in Figure 4.13 is fed the feature values {1, exclusive, dual, sentient, non-maternal, feminine}, it will fail to match the first rule in the block, since {plural, exclusive} is not a subset. It will then be compared against the second rule, where there is a match. To the right of the arrow is the effect of the rule, which is to append the suffix *-da* to the stem, represented by X.

There is always a default rule in a rule block, sometimes called an elsewhere case. Sometimes this may have an effect, like placing a suffix on the stem. Despite this, it is often an identity rule that has no effect on

{plural, exclusive}	→	Xdi
{dual}	→	Xda
∅	→	X

Figure 4.13: A rule block for number marking in Wao Terero.

{3, singular, sentient, neutral}	→	Xkã
{2, singular, non-maternal}	→	Xbi
{3, plural, sentient}	→	Xdã
{3, singular, feminine}	→	Xdã
{1, singular}	→	Xbo
{1}	→	Xbõ
{2}	→	Xbĩ
∅	→	X

Figure 4.14: A rule block for person marking in Wao Terero, where there is a "duplicate" rule for *-dã*.

the stem.

The ordering of the rules of exponence within the block is according to the cardinality of the subset of features the rule is sensitive to, called Pāṇinian rule ordering in PFM. This can be seen in a more elaborated manner in Figure 4.14. The notion is that more specific rules block more general or default rules. This implies little or nothing about generality or frequency of distributions. It is often the goal of an analysis to express that more general rules corresponds to more general distributions, but it is trivial to provide examples where this is not the case. For instance in 4.14, the *-dã* associated with third person plural, which occurs in more cells than *-bo*, is still a more specific, less general rule. This could be considered a criticism of the analysis but the point is that such a system does not automatically lead to items with broader distributions being more general.

The rule block in Figure 4.14 feeds the rule block in Figure 4.13. The blocks are generally ordered such that earlier blocks contain affixes that are nearer to the "root" of the stem. In this way they resemble position classes. This is fairly superficial. The effect of a rule of exponence is a function on a stem. Concatenation, i.e. affixation, uses an implicit concatenation function. A rule can specify either prefixation (αX) or suffixation (Xα) without regard to which rule block it is in. Additionally, the function on the stem is not limited to concatenation. Any function is allowed. This means that morphotactics is not, necessarily, expressed by rule block order. The individual rules of exponence encode this information.

$$
\begin{array}{lcl}
\{3, \text{singular}, \text{sentient}, \text{neutral}\} & \rightarrow & \text{Xkã} \\
\{2, \text{singular}, \text{non-maternal}\} & \rightarrow & \text{Xbi} \\
\{3, \text{non-dual}, \text{sentient}\} & \rightarrow & \text{Xdã} \\
\{1, \text{singular}\} & \rightarrow & \text{Xbo} \\
\{1\} & \rightarrow & \text{Xbõ} \\
\{2\} & \rightarrow & \text{Xbĩ} \\
\emptyset & \rightarrow & \text{X}
\end{array}
$$

Figure 4.15: A rule block for person marking in Wao Terero, where there is a separate DUAL feature.

A notable trait of the rule block in Figure 4.14 is that there are two rules for *-dã* suffixation. This is because of the third person dual having no person marking. If this were not the case, it would be possible to express that *-dã* has a general distribution for sentient third persons. Though, I arguably have a better description by not being able to resolve *-dã* to a single rule. I think it is actually a very nice result, when reading the rules, to see that *-dã* has this disjoint pattern. It is true that it is both associated with the sentient plural, and, separately, with the feminine singular. The lack of person marking for the dual divides these two concerns in a way that forces the analyst to acknowledge some truth. Despite this, it is not a valid option in PFM. Having "duplicates" in a rule block, where an affix (or lack of affix) is associated with non-intersecting subsets of feature values, violates a fundamental concept of PFM, *The Pāṇinian Determinism Hypothesis*. This states that competition among members of the same rule block is resolved by an ordering according to the cardinality of the sets of feature values the rules are sensitive to. Essentially, the idea is that the block will resolve to contrastive forms, each ordered according to the specificity of their subsets. Providing multiple rules for the same exponent is obviously contrary to the Pāṇinian Determinism Hypothesis.

Post-syntactic analyses can be front loaded to address the types of issues seen in the Wao Terero verbal analysis. For instance, there are theories of number where dual is not an individuated feature value but results from two values occurring in combination (Nevins, 2011). The dual is often spoken of as "marked." Given the comparative argumentation in play for determining feature values, I may have decided to capitalize on such literature and state there is an independent feature DUAL, with values dual and non-dual. NUMBER would then have no dual value. A property co-occurrence restriction could make the value dual incompatible with singular. Then the proposed rule block could be as in Figure 4.15.

This works, but it is dissatisfying. Consider that the plural, dual, and singular, whatever the inconveniences thrown up by including them under a single feature, remain mutually exclusive. There can be little argument that dual and singular should not co-occur. Likewise, it is clear that the dual and plural exponents are in competition. This creates a requirement for the co-occurrence restriction. Yet, the use of the co-occurrence restriction is doing exactly what the feature is supposed to do. The feature is supposed to encode mutually exclusive related values.

I am caught between two bad options in the analysis. Either I violate the Pāṇinian Determinism Hypothesis in one case, or I rely on co-occurrence restrictions, where features should be doing the work. There are feature rewriting mechanisms in the theory, called rules of referral, but I would be abusing them here. I am somewhat dissatisfied because I feel that the proposed features are good choices, despite my skepticism of comparative categories, and feature theory in general. Yet, a reasonable, legal analysis does not fall out of a reasonable empirical description. I blame the Pāṇinian Determinism Hypothesis. I think it is correct to say that *-dã* is simply disjoint in its licensing contexts. I think that is a truth that is erased by Pāṇinian competition.

I am not convinced that reducing the features down to some subset in order to allow competition to succeed is telling me anything about the data. The feature structure $\{3, \text{non-dual}, \text{sentient}\}$ is not as informative. Note, that it relies on the negative non-dual value, whereas the previous subsets were only about positive values. I do not know what it really means to be "sensitive" to an absence in this case.

A point that I discussed earlier, but did not provide an example for, is that making every relevant licensing condition into a feature is weird. In order to express that *-bĩ* is a form associated with motherhood, I needed to cast it as a feature. Yet, by doing so, I also need to have an odd non-maternal value, otherwise there is no opposition. Features values never come free, there must always be a balance paid, an opposing value. This insistence that all grammatical concepts and categories are in such balance is a very strong claim that is not based on empirical evidence. It creates values and categories out of nothing, which then have real effects in the grammar. I am not aware of any form that is explicitly sensitive to absence of maternity in reality. I see no actual opposition, as is arguably the case for number. It is as reasonable to have the non-maternal feature value as it would be to have a non-dog feature value, which would be active in the grammar whenever the conversation is about anything but a dog. Strangely, in the rule block, in order to get the analysis to work right, I need to make a form sensitive to non-maternal, *-bi*, even though I don't think that it is a valid value, while the form I feel *should* be sensitive to maternal, *-bĩ*, does not reference the feature at all. I would prefer a different representation that allows me to express the actual meaning relationships, but I have no choice, because everything must be a feature.

In my own analysis, the purely grammatical aspect of the patterns treated here, which drives purely grammatical agreement, simply references the form itself. I don't need subsets in the rule blocks. It is already obvious which forms are in competition. I can take that as a primitive observation. The notion of intermediary features are just a distraction. The meanings, when appropriate, are actual semantic meanings, such as a property $\mathit{mother} : e \rightarrow t$, not features in the syntax. I have no reason to posit that all non-maternal things are explicitly $\neg\mathit{mother}$.

To close this section, the PFM analysis demonstrated that features were sufficient in providing a means of calculating the licensing of morphs using reasonable feature choices for the data. Despite this, I question the adequacy of the feature treatment on empirical grounds, given my questions concerning full specification, and the validity of characterizing all meanings relevant to licensing in terms of grammatical oppositions. I also question the descriptive utility, since the analysis does not capture disjoint licensing conditions for morphs due to the framework design. To the extent that PFM is representative of feature-based realizational theories in general, which I believe to be the case, I have demonstrated that even if it is ambiguous whether such theories are sufficient for all inflectional analyses, there are prominent questions concerning their appropriateness, and that there is merit to investigating alternative approaches.

In other words, in abandoning features in order to provide a realizational treatment of lexical meanings, I am not by any means throwing the baby out with the bathwater.

4.2.5 Features and competition cannot capture lexical suffix patterns

There is very little to say now concerning lexical suffixes and features. I am not going to dwell much on the conceptual issues of whether features are a good match for lexical meaning. There is a consensus that they are not in the realizational literature, which is why there are lexemes and $\sqrt{}$roots. As of yet, authors that have sought to cast human meaning in terms of feature values have not provided a sufficiently developed feature set to have encompassed all human meaning (Katz and Fodor, 1963; Katz, 1981; Jackendoff, 2002). For that reason, we cannot say all of the features of *kitten*, and how their values intersect with the feature values of *George Bush Sr.* We are not yet able to observe the Cartesian product of the full specification of human concepts. For that reason, I will simply be demonstrating the limits of our current theoretical science.

In this section I will focus on adjectives, where the very same rule blocks as used above will serve as the basis for a series of attempted analyses of lexical suffixes. I will assume a co-occurrence restriction with plural and dual values and any lexical suffix feature, as a simplification. The choice to reuse blocks is due to the fact that with adjectives, person marking is in competition with classificatory lexical suffixes.

Logically, the lexical suffixes are all associated with an ANIMACY feature value of inanimate. The inanimate value is incompatible with a positive sentience value. The animate value is incompatible with lexical suffix values.

The features for the lexical suffix can come in one of two types of forms, combined or individuated. One may place all meanings that may be associated with the lexical suffix under a single feature, or a binary feature may be supplied for each meaning.

Let us consider only the lexical suffixes *-po* and *-ka*. An example of the first option, where there is a single feature with all licensing meanings as values is presented in Table 4.14. The feature is called LS. In the table, the plural and dual are ignored, and the animate value for sentient persons is left as understood. The first thing to notice is that the third non-sentient inanimate, without an FS value cannot occur, since there must be an FS value according to full specification, unless we wish to make a very specific property co-occurrence restriction. In fact, the bare form should be an overabundant variant form for all of the values of FS. A solution may be to provide an "any" value to FS. Yet, these are not the biggest problems that the lexical meanings present. For canoe, hand, and cluster, as well as fruit, round, rock, and head, there is syncretism, due to polysemy, along a single dimension. Since these are syntactic features, this really is syncretism, according to the framework. Unfortunately, there is no way to represent the syncretism in a rule block. The intersection of either group, *-po* or *-ka*, is 3 Non-sentient. This will not work.

What may allow for a little more traction is defining some additional features, such as FS_2 and FS_3, up to some suitable number of features, where more general features are also supplied. This would be done under the assumption that all lexical suffix meanings resolve to some intersection. This can be seen in Table 4.15. The issue then is that for *-ka*, that the more general feature values may not match the full range of specific meanings that license the lexical suffix. For instance, if there is a shared round value, then non-round rocks are not matched. Additionally, other affixes may match against round, such as *-bõ*,

	Singular
1	yẽdẽbo
2	yẽdẽbi
2 Maternal	yẽdẽbĩ
3 Sentient	yẽdẽkã
3 Feminine	yẽdẽdã
3 Non-sentient animate	yẽdẽ
3 Non-sentient inanimate	yẽdẽ
3 Non-sentient canoe	yẽdẽpo
3 Non-sentient hand	yẽdẽpo
3 Non-sentient cluster	yẽdẽpo
3 Non-sentient fruit	yẽdẽka
3 Non-sentient round	yẽdẽka
3 Non-sentient rock	yẽdẽka
3 Non-sentient head	yẽdẽka

Table 4.14: A sub-paradigm of YẼDẼ, 'big', which includes lexical suffix values under a single feature.

3 Non-sentient fruit round	yẽdẽka
3 Non-sentient any round	yẽdẽka
3 Non-sentient rock round	yẽdẽka
3 Non-sentient head round	yẽdẽka

Table 4.15: A sub-paradigm of YẼDẼ, 'big', which includes lexical suffix values under a multiple FS features.

which specifically expresses roundness, though for smaller round things. This could be differentiated with the addition of still another feature, with a value like small. Yet, there may still be conflicts, since some items may be underspecified for these values, such as a round piece of meat, which would match *-kã*. One would be stuck with *-po*, from the beginning, since it is difficult to think of an intersection for hands and canoes.

Another option would be to split the different meanings for a lexical suffix out into additional rule blocks. With the proper arrangement of rules, one might be able to ensure that for a value like rock that only one of the rule blocks does not resolve to an elsewhere, zero case. Yet, this would not capture the notion of competition, and would be contrary to the goals of PFM.

So long as any polysemy is evident, the Pāṇinian Determinism Hypothesis will make legal rules impossible. The same issue will occur if there is a binary feature per value, such as ROCK, with values rock, non-rock.

	Singular
1	yẽdẽbo
2	yẽdẽbi
2 Maternal	yẽdẽbĩ
3 Sentient	yẽdẽkã
3 Feminine	yẽdẽdã
3 Non-sentient animate	yẽdẽ
3 Non-sentient inanimate any	yẽdẽ
3 Non-sentient canoe	yẽdẽpo
3 Non-sentient fruit	yẽdẽka

Table 4.16: A sub-paradigm of YẼDẼ, 'big', which includes lexical suffixes with grammatical values that do not encode polysemy.

Relaxing the Pāṇinian Determinism Hypothesis is not desirable, since it affects all of the system, not just one ad hoc case, but also because, without it, the system fails to capture the polysemy in a manner consistent with a realizational treatment of syncretism. Even if a relaxation of the hypothesis is restricted only to lexical meanings, the approach results in a listing of homophonous morphs, that differ based only on their licensing conditions, which is contrary to the general program.

Another approach, which side steps the issue, is to say that there are grammatical lexical suffix values, and the meanings are taken care of off in some interface. Then there can be only one value for each of the lexical suffixes, as in Table 4.16. This is essentially how I handle the interlinear glosses. It is equivalent to providing a label per form. It is a kind of analysis of defeat for the system. Whatever meanings actually license the form, the analysis says they are beyond the scope of the theoretical framework. It does allow for coherent rule blocks to be defined.

The approach is not fundamentally different than providing opaque √roots. For this reason, DM hits similar walls. An option in an early insertion model of DM (Embick and Halle, 2005), where √roots are similar to lexical signs, is to provide a distinct √root for every meaning of a particular lexical suffix. Relatedness can only be expressed by form identity in such a case. Due to the fact that √roots are not realized in early insertion, the realizational system does not come into play in expressing polysemy. It is therefore outside the morphological system. In a late insertion model (Harley, 2014), √roots are realized. For that reason, one would expect that there is some potential for polysemy to be handled. Due to the fact that a similar conception of rule competition exists in early insertion theories, the result is largely the same as it would be when treating the items with features.

4.2.6 Concluding remarks

In this section I began with a further weakening of the lexical-functional divide. I discussed how treatments of lexical suffixes as √roots equate the suffixes with open class items, despite the fact that they appear to be normal suffixes. I then discussed features, what they are, and various conceptions evident in the relevant morphological literature. I provided a criticisms of issues introduced by features. I then demonstrated an analysis of Wao Terero verbal inflection, where I concluded that there are sufficient flaws in a feature-based realizational approach to consider an alternative. Following this, I demonstrated that for lexical realization, there is no obvious road forward given feature theory. Since feature theory, and √roots, are not sufficient for a realizational treatment of lexical suffix-like content, and have sufficient weaknesses even in areas such as inflection, there is good motivation to provide a different set of abstractions.

A full description and analysis using the abstractions will need to wait until Chapter 6. In the next section, I will provide a brief preview, that is relatively non-technical. It will also clarify goals of the final analysis.

4.3 Concepts of the Formalism

The full presentation of my alternative to post-syntactic realizational theory is provided in Chapter 6. In this section a conceptual overview is provided, as well as a reminder of core goals.

I start with a description of what it means to be form-oriented, which is my alternative to a feature-driven approach. Following this, I provide an architectural preview and a description of how the paradigmatic and syntagmatic dimensions of word-form validity are expressed. I then describe the generalities of how hierarchies that constrain the paradigmatic and syntagmatic dimensions are defined, and compare them to inheritance hierarchies. I also review some aspects of the polysemy of lexical suffixes, describing what I intend to capture in my analysis, and justifying the approach. I provide a final statement on what I expect a WP system to provide, before closing.

4.3.1 Form-oriented categories

The most important contrast between my approach and other realizational theories is that it is form-oriented. I expressed skepticism of feature-based categories in the previous sections. My goal is to provide something that is more concrete and measurable. I also wish to provide a means of working with language specific descriptive categories, where cross-linguistic comparison becomes possible given abstractions over those categories, but where the foundation is language specific. A principle in doing so is that there are clear methodologies associated with category discovery. There should be no substantial disagreement concerning their grounding among practitioners. For this reason, I focus on surface forms of words.

In §7.3, I provide arguments that words-forms are an empirically sound concept, despite claims that the related concept of *word* is immeasurable (Haspelmath, 2017). For now, I simply assume that word-forms are a basic unit of the grammar. By word-form, I do not mean lexemes, √roots, or anything of that kind. A word-form is a freely occurring element of the syntax. I accept that there are corner cases, items that sometimes are bound and sometimes free. There are patterns such as 'I believe in pre, rather than post-syntactic realizational theories.' In assessing these, I am strict. If *pre* and *post* occur freely, that use is free, and they are word-forms in those cases. If *pre-* and *post-* occur as prefixes, they are parts of word-forms in those cases. I am not presenting a theory of morphemes. Items such as *pre* do not need to correspond to a single lexical morpheme. If there is ambiguity as to whether some morphological item is free or bound, and speaker intuitions suggest variation in that status, then the item corresponds to both free and bound manifestations. I expect that when there is variation in free or bound status that there is systematicity. It is not random. One can express such patterns with rules, and the behavior will conform to measurement. It therefore poses no threat to my program. By focusing on word-forms, my enterprise rests on the soundness of the concepts of free and bound content. I feel comfortable with that commitment.

I discussed the outline of a feature discovery procedure using basic Wao Terero data earlier in this chapter. I presented some data in Table 4.8. I repeat it in Table 4.17, without the English translation. This is not to say that the meaning is not important. I rely on translation, and therefore meaning, when identifying morphs. There are a lot of things that coincidentally sound the same in a language. The lack of translation is only to make a point, here.

Given the nice linguistics 101 line up of data in Table 4.17, some clear patterns emerge. There are repeat-

Botõ abopa idã idã.
Ĩdã adãpa botõ ĩbo.
Ĩkã akãpa botõ ĩbo.
Botõ abopa bitõ ĩbi.
Bitõ abipa botõ ĩbo.
Bĩditõ abĩdipa botõ ĩbo.
Ĩdãdi adãdipa botõ ĩbo.
Bõditõ abõdipa bitõ ĩbi.

Table 4.17: An example of untranslated sentences using the verb 'see'.

ĩbo	[ĩ, bo]
ĩbi	[ĩ, bi]
ĩdã	[ĩ, dã]
abopa	[a, bo, pa]
adãpa	[a, dã, pa]
botõ	[bo, tõ]
bitõ	[bi, tõ]

Table 4.18: A listing of word-forms and their form based categories.

ing forms, such as *bo*, *tõ*, *a*, etc. There are also relationships between recurring forms. As two examples, the form *bo* appears to be in complementary distribution with *bi*. Additionally, the first element of each phrase shares formal characteristics of the second element, while the third and fourth elements share a different formal element. From the examination of only the forms, meaningful patterns emerge. These form patterns are the basic signals of categories used to begin a form-oriented analysis.

The goal of an analysis within my system is not the categorization of individual morphs. The primitives of categorization may often align with individual morphs, but they need not. What I categorize in an analysis are word-forms. Just as when features are used to categorize cells, a number of value primitives are used. In this case they are about form, not meaning or syntactic category. I name the primitives after morph forms. An example categorization is in Table 4.18. I use different notation in Table 4.18 than Chapter 6 because Chapter 6 provides a formalism that introduces details that would be distracting, here.

Just to be clear, the categories are items such as [ĩ, bi], and [a, bo, pa]. They correspond to the entire expression with brackets. Technically, a category primitive, like an 'ĩ' or 'bi' in [ĩ, bi], without brackets, corresponds to a signal when used in a category, but it is not a category in isolation. There are singleton categories, such as [ĩ], with brackets, which in this case would correspond to the bare form of the Wao Terero copula. The principle is analogous to feature structures. In PFM, a feature structure defines a cell/category. A feature structure is a collection of feature values. The structure may technically contain only one value, but a feature value outside a structure does not define a category.

It is not wrong to view the categories in Table 4.18 as parse objects. A parse object is an abstract representation of structured information. The parse allows one to reason about the information in an abstract manner. The objects within the category are named after formal items, but they are not strings. The names are useful mnemonics, but could be a series of numbers, or some other encoding.

Within the framework the categories are examined in different contexts for patterns that are relevant to those contexts. In terms of morphotactics, if a word-form has the category [a, bo], it will be examined to determine if there is another form that can be derived from it. In terms of meaning, if the category of [a, bo] is examined, one or more meanings may be provably related, such as ‘I see’, or ‘I find’. In terms of syntactic category, the same category may provably correspond to a verbal category that requires first person agreement.

The elements within the categories are not features. If there is some element that occurs only once in the language, such as a hypothetical, unanalyzable word-form *wiki*, it may be associated with the singleton category [wiki]. The wiki value does not require that there is some negative wiki value. The category need not exist in any contrastive grouping by default. Unlike a feature it is not a set of values in opposition.

I begin an analysis by assigning every similar morph (signal) a category sub-component of the same kind, which allows for a concrete starting point. The result is that some forms that should be differentiated may initially share a category. The form *kẽ-kã* is an example. It may be glossed as either ‘eat-LS.body’, ‘eat meat’, or ‘eat-3.H’, ‘He eats.’. If there were no evidence that the *-kã* affix signals a necessary category contrasts, the word-form would maintain a single category [kẽ, kã] during the process of subsequent category refinement. The realizational calculation is intended to handle the mapping of categories to distinct meanings and syntactic categories. For this reason, I favor leveraging the realizational system and reducing the number of proposed categories.

In actual fact, *kẽ-kã* serves as an example of why form alone is insufficient in determining category primitives. Distributions also need to be considered. It is true that *kẽ-kã* is ambiguous between a person and classifier usage, but this is only because both classifiers and overt person marking may be absent on verbs. Additional structure makes it clear when *-kã* is person marking and when it is a lexical suffix. For instance, if a person morph is added, *kẽ-kã-bo*, then *-kã must* be interpreted as a classifier. If a lexical suffix morph is added, *kẽ-po-kã*, then *-kã must* be interpreted as person morphology. Likewise, some items systematically allow only person marking or lexical suffixes. There is no ambiguity in such cases. This demonstrates that there are two distinct distributions for *-kã*, and that those distinct distributions correspond to distinct interpretations. This has morphotactic consequences. For this reason, proposing distinct category primitives for *-kã* is the right choice.

In contrast, the distinct meanings that may be associated with *-po*, such as 'hand' and 'canoe', do not justify multiple category primitives. These meanings are not associated with systematically distinct distributions. On adjectives, demonstratives, and numerals both interpretations are available, even when there are additional affixes. In verbal and nominal uses, there is no ambiguity as to whether they belong to one category or another. Whatever the meaning associated with it, *-po* acts as a lexical suffix. It is true that nominal and verbal stems may only allow a subset of *-po* interpretations, but such restrictions are either idiomatic, in cases such as *wi·po*, 'canoe', or correspond to a property of the stem, as in cases such as *dãta*, 'hurt', where only body-part interpretations are allowed. Proposing that different category primitives should be assigned to some formal pattern based on meaning alone represents a misunderstanding of their purpose within the framework. The primitives delineate formal, distributional distinctions. Meaning may play a role in providing evidence for differences when it aligns with a distribution, but it is not the determining factor in and of itself. There is a realizational system that communicates that a formal pattern may corresponds to a plurality of distinct meanings. That is not the role of the category primitive, itself.

My practice in category primitive assignment is analogous to the general approach of post-syntactic theories. Distinct category primitives associated with homophonous morphs are justified in my system in cases where a PFM analysis would place rules of exponence with homophonous results in distinct rule blocks. The decision to place items in different rule blocks is not driven by meaning alone. If potential homophones do not demonstrate distinct distributions, the item is licensed within a single rule block, even if it is associated with a variety of meanings. Within that block, the Pāṇinian Determinism Hypothesis rigidly enforces that there is only one rule for licensing the form, which corresponds to my practice of assigning only one category primitive to a formal pattern.

All that being said, in my framework, which lacks a Pāṇinian Determinism Hypothesis, the analogous practice of assigning a single category primitive to a form associated with many meanings is a choice. Technically, one could choose to have many category primitives for the same form, even when there are no distributional distinctions. The result of such a quasi-Item and Arrangement (IA) approach would be an inelegant, unparsimonious analysis, and would be observationally equivalent to a realizational approach. I talk about observational equivalence, below. All else being equal, fewer categories and rules are better.

I do not want to make it seem as though the categories and their sub-components can be determined without considering meaning. That doesn't actually work. No system has been able to reliably induce morphs through pure induction on forms. There is simply too much ambiguity (Goldsmith, 2001). Efforts specifically aimed at Wao Terero in order to computationally induce paradigms, admittedly from an extremely small and sparse dataset, were complicated by there being small, frequent stems, such as *a, ã, ĩ, ẽ*, which may also be initial parts of larger, unanalyzable stems (Copot et al., 2022). During the discovery procedure, I am not looking for morphemes, which lock a form and meaning together, but I am looking for signals, in a realizational licensing relationship with other categories of the grammar. Therefore, I use fairly standard procedures to discover valid morphological segmentations. The important point is that, in this theory, due to the separationist principle, two distinct categories are not defined for the same homophonous signal, unless it is by appeal to form-patterning. Semantics is not an explicit category determiner at the morphological level.

A consequence of this is that apart from certain morphotactic ambiguities, morphological categories tend to be form-contrastive. Recalling the rule blocks of PFM, the idea was to take some large set of features that are proposed to be relevant to words in a language, a fully specified feature set, then through an analysis that squeezes these down to competitive subsets, one has a rule block where competing forms are reduced to single, non-redundant rules. The system I propose takes this set of contrastive forms as a starting point. Though, not on a morph-by-morph basis as in PFM. Only free word forms are categorized. The morphotactic system forms a pool of (mostly) contrasting forms. This set is supplied by a calculation, which is intended to prove word-form validity according to purely morphological criteria.

Such forms are then related to meanings and syntactic categories according to observable patterns in the language. A form is not considered to have its meaning domain reducible to a subset, as is the case in defining a rule of exponence in PFM. What this means is that there is only one category [yẽdẽ, po]. If that form is relatable to a meaning including 'canoe', it is true. If that form is relatable to a meaning including 'hand', it is true. The point is to accurately map the systematic relationships between co-variation of form and meaning. I therefore allow disjoint meanings to relate to single word-forms.

This makes it trivial to interface to lexical semantic models of polysemy. As stated in the introduction, I do not provide a sophisticated theory of lexical semantics. The goal is to provide the means for different lexical semantic theories to be used. The interface only requires some means of matching primitives of the lexical semantic model to the parse objects of the morphology. Yet, the design of the interface is not entirely neutral. It is intended to accommodate a specific theoretical position. I do not believe that polysemy can always be reduced to prototypes, or intersections of meanings. For instance, authors have convincingly argued for network-like relationships between classifier meanings (Aikhenvald, 2000). An example are the radial categories adopted in cognitive linguistics (Lakoff, 2008). Berlin (1968) utilized a discovery procedure for meanings associated with Tzeltal numeral classifiers that resulted in exemplar-like representations. It is also widely accepted that a common diachronic trend in classifier grammaticalization is that the morphs pick up diverse meaning associations over time (Aikhenvald, 2000; Grinevald and Seifart, 2004; Mithun, 1986; Seifart, 2007). The notion of a single affix being appropriate for hands and canoes is not odd in that context. For such a process to take place, polysemy must be allowed to be disjunctive. I do not explore the lexical semantics of lexical suffix meanings in any depth, but I do accommodate analyses where a diversity of lexical meanings may be associated with a single form. This is innovative and unique within the context of realizational theory.

4.3.2 Architecture preview

The above discussion naturally leads to the question of what the framework architecture looks like. A diagram is provided in Figure 4.16. There will be more details in §6.3.

The basic idea is that there are two kinds of paradigms, really two kinds of taxonomies containing paradigms. First there are form paradigms, which are collections of forms discussed above. I am reusing the term paradigm for something that is quite different than the orthogonal taxonomy assumed by post-syntactic realization. Form paradigm entries are produced by generative morphological rules, which do not reference

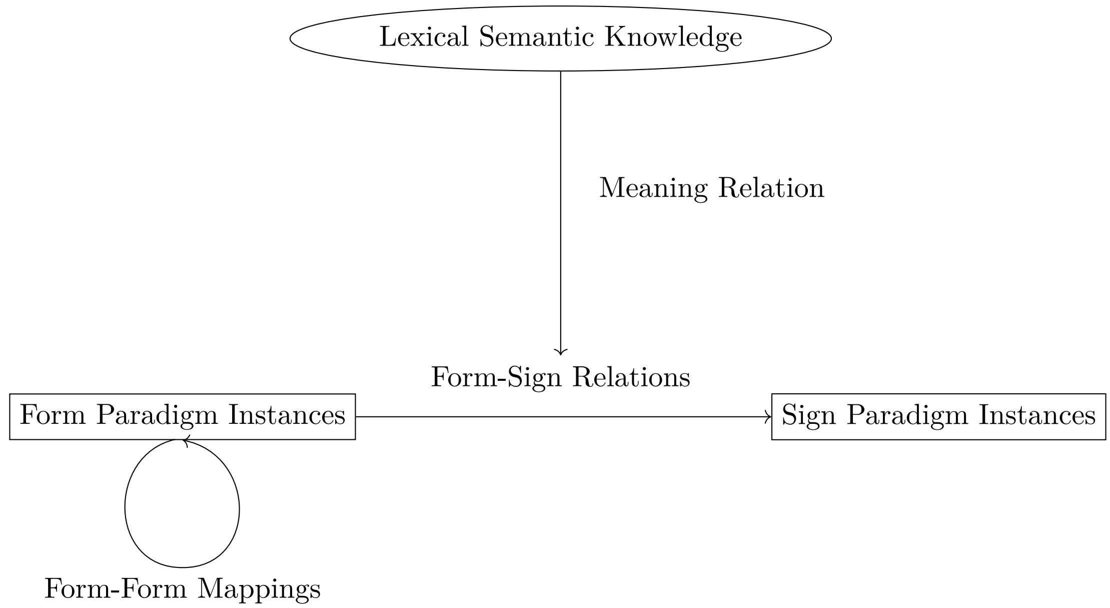


Figure 4.16: An overview of the architecture of the formal theory.

meaning or syntactic category in a direct manner. Though, there are some indirect ways that meaning and syntactic category could be said to influence the morphological level. For instance, verbal word-forms have certain patterns not seen with adjectives. Inanimate nouns have distinct patterns from animate nouns. These classes only become associated with syntax and semantics at the interface. In the morphology, they are treated neutrally, like any other morphotactic pattern, which may or may not have such a correspondence. The items found in the form paradigm are whole word forms and only whole word forms, categorized as described above.

There is a second type of paradigm, sign paradigms, which are composed of syntactic lexical entries. The arrow between the two paradigms indicates that the forms of the form paradigm are mapped to lexical entries. This relationship is augmented with lexical semantic information. This means that a mapping from a form like *yẽdẽpo* may have some set of syntactic categories, Adj, NP, etc., depending on whether it is used alone, or with a nominal argument, for instance. Within each of these syntactic categories, due to the meaning relationship, there may be various combinations of meanings that are associated with *yẽdẽ* and *-po*, such as ‘big, tall, fat’, combined with ‘hand, canoe, grapes’. The information from both domains provides the equivalent of full specification. In certain instances, as may be the case with the demonstrative, which does not have enough concrete meaning to have a lot of interaction with a classificatory lexical suffix, the meanings may be quite orthogonal. Often, the stem and affixes will interact, and there will be some non-trivial meaning relation between them. Expressing these relationships requires at least a simplified lexical semantics, which is all I provide for portion of the architecture labeled Lexical Semantic Knowledge. Notably, meaning and category may be represented as orthogonal in the system, but other types of relationships may also be expressed.

4.3.3 An Alternative to Inheritance Taxonomies

An important aspect of formally modeling the lexical suffix system is allowing for the productivity and broad distribution of the suffixes while maintaining sufficient restrictive mechanisms so as not to overgeneralize. Authors have generally used inheritance for similar ends (Flickinger, Pollard, and Wasow, 1985; Flickinger, 1987; Riehemann, 1998; Brown and Hippisley, 2012; Corbett and N. Fraser, 1993). The insight is that based on reasonably straight forward structural diagnostics, one can group and categorize stems, lexemes or roots (depending on one's theoretical primitive) into categories where the same rules apply. Such a category is sometimes called an inflection class but to generalize beyond inflection, I use the term form class. One of the major motivations for such classes has been to explain systematic allomorphy patterns, especially in inflection. The AR-ER-IR distinction in Spanish, which I discussed in the Chapter 1, is a classic example. English weak, or regular verbs, and verbs that have other forms of the past tense like MEAN or DRIVE are other examples. I provide some toy examples relevant to these patterns in Appendix B.

4.3.3.1 Paradigmatic Categories

As mentioned, form classes are often used to model patterns in inflectional allomorphy. Even though there is little non-phonologically conditioned allomorphy in the Wao Terero morphological system, such categories provide an explanation of the distribution of affixes. Form classes are expressive of the paradigmatic dimension of the morphological system (de Saussure, 2011). The paradigmatic dimension describes something like members of a position class. An example from Spanish can be seen in Figure 4.17, where the syntagmatic axis specifies something like linear order and number of positions. The paradigmatic axis represents different options available at any particular position along the syntagmatic axis. The classes indicate patterns

AR	habl	aba	n
		ara	n
		aro	n
ER	com	ía	n
		iera	n
		iero	n

Paradigmatic Axis (↑)

Syntagmatic Axis (→)

Figure 4.17: The paradigmatic versus syntagmatic dimension with the Spanish AR verb HABLAR and ER verb COMER. All are third person in the imperfect indicative, imperfect subjuntive and preterite as examples.

of allomorphic variation at different positions.

The paradigmatic dimension remains relevant to Wao Terero as well, despite not being needed for allomorphy. This is because paradigmatic options still differ depending on the stem. This can be seen in Figure 4.18, where numbers stand in for form class names. The paradigmatic options for each class are differently constrained. Class 1 only allows the *-po* ending. Class 2 allows nearly any lexical suffix. Class 3 allows at least one “compound” suffix, *-tabõ*, for coins and similar objects. These classes actually correspond to the names wipo_k, inanim_k and ado_k, in the analysis I provide. The limitive *-ke* affix is included to demonstrate that the items also share a more general class. In this limited example the three classes are in one-to-one correspondence with items that have differing syntactic categories. There is nothing wrong with one-to-one correspondences. The claim that syntactic categories do not determine morphological form classes does not mean grammatical levels may not strongly influence each other. Yet, here, the correspondence is diminished if a larger sample is considered. For instance, Class 1 is specific to a single item. Other nouns may be member of classes that inherit from the Class 1, or may even be co-members, but not all nouns have a relationship to Class 1. Class 2 would also include adjective uses and inanimate noun uses. Syntactic categories help with understanding some boundaries and relationships between classes but do not determine them.

In the system, the form classes define the schema of the paradigm taxonomy. A paradigm instances is identified by a stem and a form class.

1	wi	po	ke
2	bãdĩ	po	ke
		wẽ	ke
		...	ke
3	ado	tabõ	ke
		po	ke
		...	ke

Paradigmatic Axis (↑)

Syntagmatic Axis (→)

Figure 4.18: Even without allomorphy, differing patterns along the paradigmatic axis require a notion of category. The stems *wi-*, for canoe; *bãdĩ*, a demonstrative; and *ado-*, used in 'one' and elsewhere, are used to demonstrate this. Each differs in the lexical suffixes that may follow the stem. The ending *-ke* is called the limitive and signals a meaning of 'just' or 'only'.

AR	habl	a	ba	n
		a	ra	n
		a	ro	n
ER	com	í	a	n
		ie	ra	n
		ie	ro	n

Figure 4.19: A more fine grained division of affix parts for Spanish verbal inflection to emphasize the cross-column predictability of the system.

4.3.3.2 Complications with Splitting and the Paradigmatic-Syntagmatic Dimensions

One might make a slightly different "cut" for the Spanish forms in Table 4.17, so that it looks something more like Table 4.19. Other cuts are possible. The divisions between word-form parts, *morphs*, are sometimes a matter of heuristics. The analyst may make different choices for different reasons. One may be more concerned about units that clearly correspond to meanings. Another may be more or less tolerant of phonological redundancy between morphs. Such measurement-oriented choices lead to a lot of variety in valid analyses.

The fact that cuts in observable structure are partially due to analyst choice, and may result in a number of valid proposals, was acknowledged by early diagnostic designers. In his guide to grammatical analysis, Z. S. Harris (1951) stated that different practitioners would arrive at different results using his methods because they would emphasize different diagnostic criteria. Similar points were made by Hockett (1947).

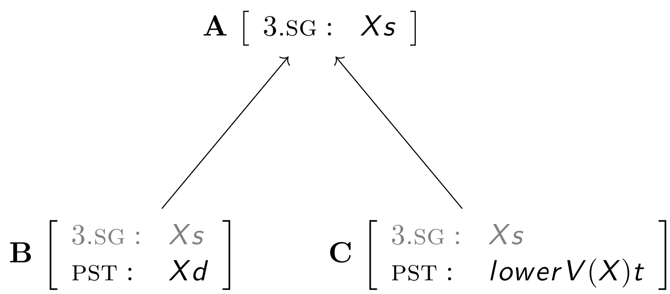


Figure 4.20: In inheritance systems, complex objects rely on hierarchical relationships to determine how they should be fully constructed. Here only a rule for PST is explicitly defined in the leaf nodes. The rule for 3.SG is inherited.

I do not find this to be greatly concerning, so long as the differences are all in reference to concrete manifestations of form. I think that what is most important is that measurement is performed relative to concrete objects, with consistent, reproducible methods.

Another point I wish to emphasize is that the items in the various columns are not, necessarily independent. Obviously, the stem goes some way in determining the affixation pattern, which is why the table is divided in two. There is also a great deal of left to right predictability after the stem. The form class does not just capture the fact that a particular column has different allomorphy patterns but that the patterns of allomorphy extends across columns, so to speak. This means that the paradigmatic and syntagmatic dimensions are not orthogonal but interdependent. It is useful to treat syntagmatic and paradigmatic categories as somewhat distinct but they can and should make reference to one another.

4.3.3.3 Inheritance Hierarchies

Form classes tend to fall into hierarchical taxonomies. Certain general rules are shared by each form class. An example of using inheritance to define a portion of two English form classes is schematized in Figure 4.20, where a third person singular rule is shared between two classes that differ in their past tense rules. The function $lowerV$ lowers a vowel in the stem X. As usual, affixation is represented by implicit concatenation. The labels A, B and C are ad hoc form class names. The B pattern is consistent with English weak verbs. The C pattern is consistent with verbs like MEAN.

Obviously, Figure 4.20 is not intended to be a complete example of English verbal inflectional patterns,

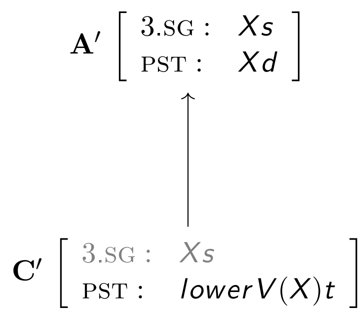


Figure 4.21: In some inheritance systems, rules may be overwritten.

but it exemplifies one choice that may be made in such a schema, which is that nodes higher in the hierarchy are more abstract, or incomplete, as seen with A. In such a case one may think of the hierarchy as providing rules to construct leaves. Another possibility is allowing inheriting nodes to overwrite defaults, as seen in Figure 4.21. Although not all modern morphological theories explicitly model inflection or form classes as objects, default overrides, where default rules are overwritten by more specific rules, are the dominant model for compressing many rules into few rules in modern morphological theories. It is the mechanism used in PFM rule competition. In the English example using form classes it might seem reasonable, since there is a specific past tense ending that has a wider distribution. It could be argued that it is more explanatory to model other past tense forms as exceptions to a default.

The idea that default overrides are more natural, or explanatory, just because one may model many linguistic systems in terms of them, is not well founded. There are morphological systems where no defaults exist (Dąbrowska, 2001). This means that for some data it is strictly false and misleading, even if the data may be shoehorned in to such a system.

That default overrides can often be leveraged, does not reveal anything special about human language. Default overrides are a common strategy for increasing rule parsimony in a wide variety of domains. There isn't anything particularly linguistic about it. It is a general approach taken in object oriented programming, which is used to model the structure of computer programs for essentially any purpose. Such programming languages are usually chosen because of familiarity or taste, rather than some fundamental match between the domain of the data and the default taxonomic tools provided by the programming language. All linguistics form class taxonomies could doubtless be modeled in SQL, a non-objected oriented database programming language. This could, perhaps, be done elegantly, but that doesn't make SQL particularly natural.

There are clear limitations of inheritance. One of the most prominent are situations where inheriting from more than one parent may cause conflicts. These are serious enough that a number of authors have sought alternatives (Koenig and Jurafsky, 1994; Crysmann and Bonami, 2016; McConville, 2006; Guzmán Naranjo, 2019), which are not, strictly speaking, inheritance. The notion of conflict enters the picture because inheritance, as implemented in morphological theory, is not only a means of expressing hierarchical relationships but also constructing objects. The objects are what are labeled A, B and C, in the previous figures. These objects have well-formedness rules. Only some objects are legal objects. A very common rule is that given some value, such as PST, in the previous Figures, there should be only one value for the property per object. A *conflict* is a way of describing a situation that would result in that well-formedness condition being broken.

A simple example of an analysis that would involve a conflict due to multiple inheritance involves form class conditioned overabundance. Overabundance (Thornton, 2012) is a phenomena where more than one morphological form of a lexeme is compatible with the same inflectional category. An example of form class conditioned overabundance is that *leapt* and *leaped* are past tense forms of LEAP. Each form of the past tense follows a distinct form class pattern found in English, both the B and C introduced above. Notably, LEAP is not just an isolated case. Words like DREAM, BURN, SPILL, and others follow the same pattern for many speakers of English. At the same time, there are similar sounding words like SEEM and SLEEP that exhibit only one form of the past tense. Even if the class is small, its synchronic explanation cannot be reduced to meaning or phonology.

Since B and C have been defined already, for weak verbs and MEAN-like verbs, the logical thing to do, given that words like LEAP, DREAM and others exist that follow both patterns would be to propose an additional class for these items that inherits from both B and C. Given the assumptions of an inheritance system, this is the only solution to the issue consistent with the explanatory purpose of the model.

Despite being the only solution within such a system, modeling the overabundance relationship as inheritance results in a conflict as seen in Figure 4.22. The question marks before PST in D indicate that the object is not well formed.

This is a serious flaw. I am unaware of a language without some form of overabundance, even if it is not

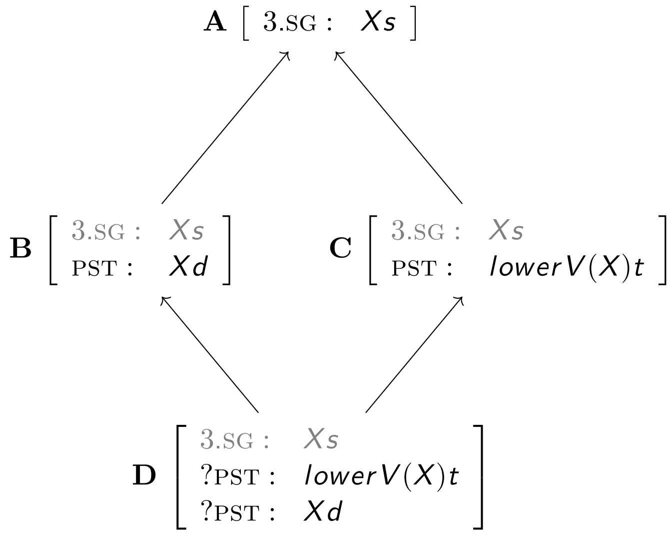


Figure 4.22: Here D is inheriting from both B and C, which each have distinct definitions of a PST rule. Therefore there is a conflict in D.

always due to "conflicting" form classes.

4.3.3.4 Non-determinism in inheritance systems

Inheritance systems are more complex than necessary. As the arrows between objects in Figure 4.22 suggest, the hierarchy one would want to express is very simple. A is the intersection of B and C. B and C are disjoint. D is the union of B and C. Where the difficulty comes in is that this hierarchy is unnecessarily conflated with the construction of the objects, which have their own well-formedness conditions that are orthogonal to the ideas one wants to express with the hierarchy.

Simply relaxing the well-formedness conditions to accommodate reasonable hierarchies has consequences. For instance, one could allow a double valued property for D but then they would lose a restriction on possible object types that may be essential elsewhere. The system disallows such objects so that it is easier to reason about the formal system. When it is sometimes possible to deviate from the one value per property constraint, any previous reasoning that depended on the constraint becomes more complex. Especially in a system where one may overwrite defaults, it is unclear how overwritting and multiple values per property should interact. The best way to deal with what is sometimes described as *non-determinism* in inheritance hierarchies is an open question.

Person×Number

	Singular	Plural
1	Xo	Xamos
2	Xas	Xan
3	Xa	Xan

B→A

$$\mathbf{A}\begin{bmatrix} 3.\textsc{sg}: & Xs \end{bmatrix}$$

$$\uparrow$$

$$\mathbf{B}\begin{bmatrix} 3.\textsc{sg}: & Xs \\ \textsc{pst}: & Xd \end{bmatrix}$$

Figure 4.23: Two different types of taxonomies are common in morphological theory. On the left is the orthogonal taxonomy, which can be expressed in terms of the product of sets, where members of each set are expected to have few interdependencies. The string rules in the cells correspond to Spanish present tense for an AR verb. On the right is a hierarchical taxonomy, which can be expressed in terms of logical implication or subsets, where items are expected to have important interdependencies that are modeled by the hierarchy.

There are systems that have been used to express non-determinism in object construction, but they use alternatives to inheritance, such as online type construction (Koenig and Jurafsky, 1994; Crysmann and Bonami, 2016). I am not aware of any attempts to use online type construction to solve inheritance conflicts in form classes, in particular, but it is not difficult to explain how it might be done. I'll provide an example here, but I ignore details of how my exposition might be adapted to a particular theoretical framework.

An important point that I would like to make at the onset is that online type construction expresses something more like the product of sets than the subset-like relationship expressed by the edges of an inheritance graph. A comparison can be seen between the types of orthogonal taxonomies expressed by products and the hierarchical taxonomies expressed by implication in Figure 4.23. This is just a reminder, since the idea was described earlier in the chapter. The form class relation has dependencies, and a tree-like structure. One can see that the two taxonomy types specialize in expressing quite different relationships between information.

The basic idea behind online type construction can be seen in Figure 4.24. The boxed items that are labeled Past and Shared are like two sets of objects of a certain kind. The online type construction produces a filtered product of the two sets. A normal product of sets of natural numbers $\{1, 2, 3\}$ and $\{4, 5\}$ may result in a set of ordered pairs, $\{\langle 1, 4\rangle, \langle 1, 5\rangle, \langle 2, 4\rangle, \langle 2, 5\rangle, \langle 3, 4\rangle, \langle 3, 5\rangle\}$. For online type construction, the properties of the combined objects are unioned, resulting in a new object with all of the properties from

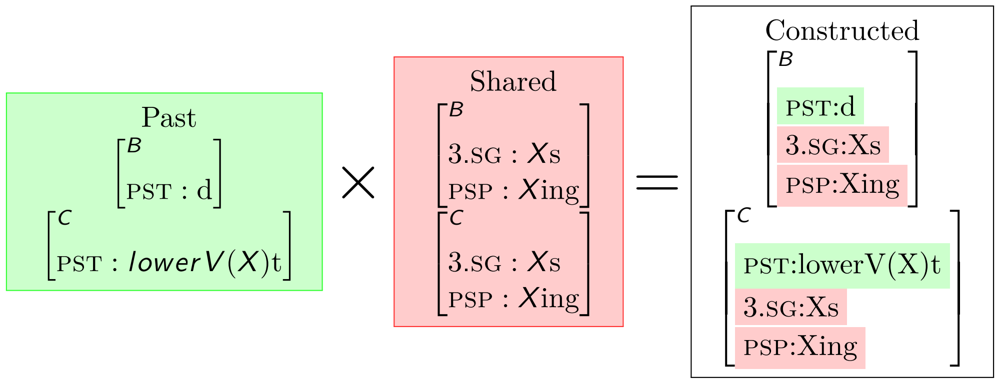


Figure 4.24: Online type construction is performs a product-like operation on sets of objects. Here, Past and Shared are two sets. One constructs complex objects through a union of the properties of compatible objects. I am indicating compatibility with the annotations B or C, which must match.

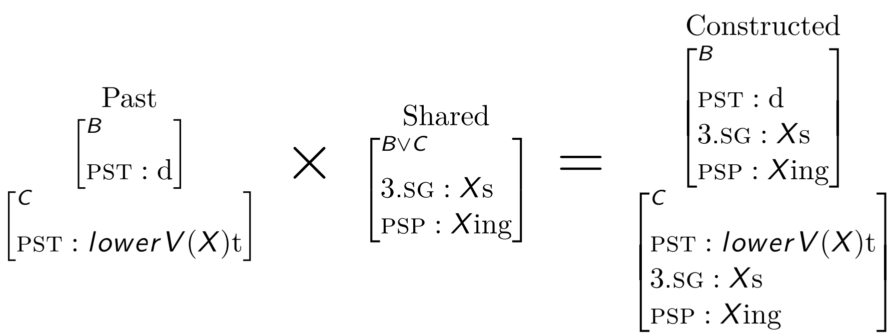


Figure 4.25: Notational redundancies can be reduced using $\vee$ on the annotations.

both of the combined objects. This can be seen in the color coding of Figure 4.24.

By a filtered product, I mean that some condition exists on the output set such that only a subset of the product exists in the final result. For instance, one may have a condition such that only evens are paired with evens or odds with odds, resulting in $\{\langle 1, 5\rangle, \langle 2, 4\rangle, \langle 3, 5\rangle\}$ for the sets used earlier. Online type construction uses a notion of object compatibility, such that the result of the operation only outputs the combination of objects that are compatible. I annotated the objects in the set to indicate their compatibility. In Figure 4.24 the annotations B and C indicate which objects in each set are compatible. It is kind of like having an automatic property co-occurrence mechanism, as was described for PFM.

The Shared set looks a bit redundant with two objects that are essentially identical. Figure 4.25 demonstrates how one can use the $\vee$ notation to indicate compatibility with more than one object type. This is

$$
\underset{\text{Past}}{\left\{ \begin{bmatrix} {}^{B\vee D} \\ \textsc{pst} : \text{d} \end{bmatrix}, \begin{bmatrix} {}^{C\vee D} \\ \textsc{pst} : \mathit{lowerV}(X)\text{t} \end{bmatrix} \right\}} \times \underset{\text{Shared}}{\begin{bmatrix} {}^{B\vee C\vee D} \\ 3.\textsc{sg} : X\text{s} \\ \textsc{psp} : X\text{ing} \end{bmatrix}} = \underset{\text{Constructed}}{\left\{ \begin{bmatrix} {}^{B\vee D} \\ \textsc{pst} : \text{d} \\ 3.\textsc{sg} : X\text{s} \\ \textsc{psp} : X\text{ing} \end{bmatrix}, \begin{bmatrix} {}^{C\vee D} \\ \textsc{pst} : \mathit{lowerV}(X)\text{t} \\ 3.\textsc{sg} : X\text{s} \\ \textsc{psp} : X\text{ing} \end{bmatrix} \right\}}
$$

Figure 4.26: By specifying that the both of the past objects are compatible with D, two D objects can be constructed with one of either of the past rules.

used in Crysmann and Bonami (2016) for a system that focuses on morphotactics, a syntagmatic dimension, rather than a paradigmatic, form class behavior.

Figure 4.26 provides an example of a non-deterministic addition. The D form class in Figure 4.22, results from adding D to the compatibility disjunction of the objects in the Past set. The result is that there is the equivalent of four objects in the Past set. The Shared set must also include a D compatible object. In the Constructed set, there is the equivalent of two D annotated objects. Each object has a different value for PST, which solves the issue of needing to have two values for a single property in one object. Either constructed D object provides a valid pattern for stems in a D compatible class.

One might imagine an alternative to this. Somewhere in the system one will need to list the stems that these rules apply to. Instead of producing two D objects, the stem could be listed as being compatible with B or C. The issue with this is that it would not express that there is a class of stems like D as a fact of English grammar, where not just LEAP behaves like this but also a number of other verbs, such as DREAM. For some reason speakers are comfortable with the variation of these items, whereas similar sounding items such as SLEEP or SEEM follow only the C or B option, respectively. The purpose of drawing out the hierarchies is supposed to communicate such facts.

One reason to suspect that online type construction is not the most elegant approach is that one must specify the additional constraints on the product-like operation, through annotations. This implies that the taxonomy does not have a grid-like, orthogonal organization. There is an interdependence between objects in each set, an interdependence that needs to be simulated through the annotations, but would simply fall out of an explicitly hierarchical structure. Therefore, the online type construction approach sacrifices the explanatory simplicity of the hierarchy in Figure 4.22 to maintain the well-formedness of the constructed objects as a trade-off.

It seems like it would be more simple to separate the concerns. One needs a way to express that certain categories have behaviors that are subsumed by others. One also needs a way to describe the rules themselves and how they relate to other grammatical categories, for instance that some rule is associated with PST. Conflating these issues causes a great deal of complexity in inheritance systems and is completely unnecessary.

To separate these concerns, I refine a technique utilized by McConville (2006), which has an analog in Guzmán Naranjo (2019). The system uses axiomatized relationships between categories that are referenced by lexical rules. This is schematized in Figure 4.27. The fundamental difference between this and inheritance is that the axioms that define the hierarchy are independent of rules that manipulate strings. In Figure 4.27 there is a listing of three rules. In each case there is a condition, which relies on a variable α. This variable represents the form class of some stem that is input to the rule. A value for α might be A, B, C or D. Rules that reference categories higher in the hierarchy are more general as they also apply to categories that are lower. The rules do not specify what is higher or lower in the hierarchy, themselves. The rules simply have a condition on their domain of applicability. Due to the fact that $\leq$ is a reflexive relation, it is only clear that if α is A then $\alpha \leq \mathsf{A}$ rules are applied, and when B then $\alpha \leq \mathsf{B}$ etc. There is no further information relevant to the hierarchy that can be determined based on the conditions of the rules. The hierarchy, itself, is defined separately by a list of axioms. It is by reference to this hierarchy that it becomes clear that if α is D then any of the three listed rules may apply. It is simply stated as a fact, in a parsimonious and straight forward manner. This approach which emphasizes listing facts about a morphological system, with each kind of fact in its own place, avoids a lot of unnecessary complications.

It is important to remember that inheritance, in and of itself, does not make any strong scientific predictions. It is simply a way of organizing information. I provide a more versatile approach to information organization, which allows me to express patterns in the data without conflating unnecessary issues of object construction.

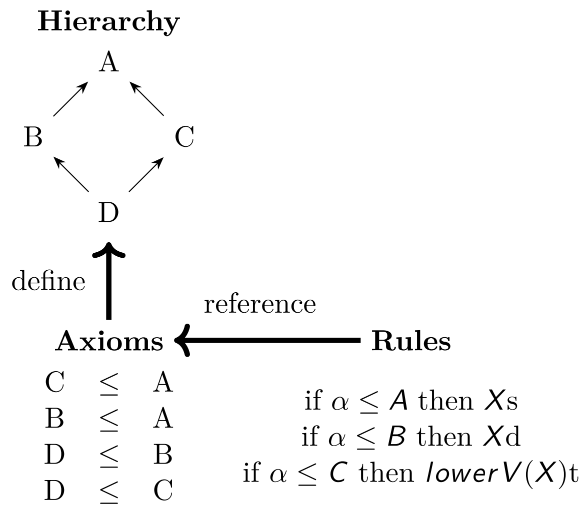


Figure 4.27: As an alternative to inheritance, relationships between categories (A,B,C) are defined by axioms. Rules reference these relationships to determine the applicability of a rule.

4.3.4 A hierarchical morphotactics

I use hierarchies that are defined in the same manner to describe morphotactics. I provide two mechanisms of providing proofs that word-forms exist in the system. One of them is analogical. One can make inferences concerning forms that exist in one paradigm instance by comparison with forms that exist in another paradigm instance.

The other mechanism for proving a word-form is incremental. If some unaffixed word-form has been proven to exist, then a rule, appropriate to the form class of the stem, may be applied, resulting in a new word-form. Rules are only between word-forms. There are never incomplete word-forms. Anything proven must be demonstrably free in the syntax.

It isn't exactly true that the incremental form building is always about building forms. Hypothetically, there could be subtractions in form. There are also special lateral rules, which are discussed in Chapter 6.

The notion that hierarchies are involved in word-formation has existed since, at least, Lashley (1951), though it is implicit in works on languages with derivational processes that are earlier (Hockett, 1948). It is sometimes conflated with syntactic or semantic scope. Hierarchies in morphotactics simply allow for groupings of routine actions. Remember that morphotactics is descriptive of a series of routine actions taken in time. They are certainly routine given the reuse of morphs and stems. It also appears that morphs can group into certain domains, such as the domain where derivational affixes commonly apply, versus inflection. There are also complex, component systems, where multi-morph components compose in Algonquian languages (Hockett, 1948), and taken as a unit, compose into larger structures. In Algonquian languages, there are units of a verb called the initial, medial, and final. A verb will generally have at least an initial and final, in addition to inflectional affixes. The initial is root-like. The medial is a unit of one or more lexical suffix-like affixes. The final tends to categorize the verb within the form classes of the system. This is only intended as a very loose description. Each of the initial, medial, final may be morphologically complex. This is routine-like, where certain routines span over smaller routines.

Hierarchical models have been used in areas such as psychology and robotics to describe groupings of certain reusable actions into hierarchies (Braitenberg, Heck, and Sultan, 1997; Braitenberg and Schüz, 2013; Grupen, 2023). Pointing this out isn't intended to provide some argument for hierarchies being more psychologically plausible. It is simply to point out that describing morphotactics with hierarchies isn't any different than describing making a pot of coffee using such models, as is done in Cooper and Shallice (2006). There is no notion of phrasal syntax or semantics that is automatically expressed in parallel when describing the routine movements of making coffee, neither is it necessarily the case in morphotactics.

It is also possible to provide an expression of such hierarchical systems as regular languages (Harel, 1987). This may be advantageous in ultimately providing formal constraints on valid morphotactics, though I do not leverage it here.

PFM and Information-Based Morphology (IbM) favor more template-like systems. This is superficial in PFM, since the rule blocks only express domains of competition. In IbM, the idea is expressed in a more rigid manner. It appears that if one is only considering inflection, that such a model is largely adequate. Neither system attempts to extend it to derivation.

The representation of morphotactic hierarchies uses a similar mechanism as described above for form classes. That is to say, it uses some external hierarchy, that is referenced by rules.

Beyond this, I believe that most linguists have seen hierarchical approaches to morphotactics. It is a common approach. Mine is unique in many characteristics, but these will be much more clear given a full example, so I will wait until Chapter 6 to describe them.

-wẽ	plant, branch, stem, stick, log, pole
-ka	fruit, rock, head, round-thing
-po	hand-part, cluster (like grapes), canoe

Table 4.19: Meanings associated with three lexical suffixes.

4.3.5 Multiple meanings per form

Throughout this work, I have emphasized that multiple meanings may be associated with lexical suffixes. A summary of meanings associated with three lexical suffixes is provided in Table 4.19. In classifier constructions there are few limits on potential interpretations within the range of established convention. Context provides bounds on felicitous interpretations. A base such as *bãdĩ*, a demonstrative, may be used with any of the affixes in Table 4.19 and have any of the listed interpretations.

Having a range of possible interpretations is not specific to the use of lexical suffixes as classifiers. For instance, non-classifier nominal occurrences may be polysemous. A nominal word-form like *a·wẽ*, in (116) may be associated with all of the meanings listed for the suffix *-wẽ* in Table 4.19.

(116) *a·wẽ*

∅·LS.plant

'plant/tree/pole/etc.'

Despite this, in nominals, the range of meanings available may depend on the stem to a greater extent than seen in classifier uses. An example of this is that the form *wi·po* in (117a) is only associated with the meaning 'canoe' or 'boat', not 'cluster'. Likewise, (117b), also ending in *-po*, is only interpreted as having to do with hands. It does not refer to plant anatomy or canoes.

(117) a. *wi·po*

∅·LS.canoe

'canoe'

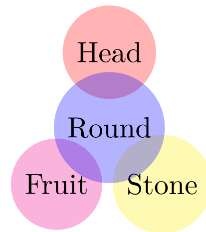


Figure 4.28: The various dimensions of meaning associated with *-ka* depicted as a Venn diagram. Each circle represents a dimension of meaning, where non-intersecting concepts such as heads, fruit and stones intersect with roundness but are not subsumed by roundness.

b. *õdõ·po*

∅·LS.canoe

'hand'

c. *dãta-po-bo-pa*

hurt-CLF.canoe-1-DECL

'My hand hurts.'

Verbal construction types may also limit suffix interpretation to a greater extent than seen in adjectival or demonstrative classifier use. The verb DÃTA, 'hurt', can take *-po* or *-ka*, but only the body-part interpretations are available, as in (117c), which cannot mean 'My canoe hurts.'

The form class allow exactly the level of constraint required for these cases. General patterns where the entire range of conventional meanings may be associated with an affix may be expressed. Yet, it is always possible to clamp down on such productivity, where it does not exist.

4.3.6 Patterns in Lexical Meaning

Sometimes there are plausible relationships between a lexical suffix's meanings. These often take on more than one dimension of relatedness. For instance, *-wẽ* meanings contain a pattern of synecdoche, a whole to part relationship, between plants and their stems and branches. At the same time, stems, sticks and poles have a shared shape property, adding an additional dimension to the pattern of meaning relations.

The *-ka* meanings also have a shape-base component to their pattern. This seems motivated by the fact that other meanings associated with *-ka*, such as heads, stones and fruit, are often round. Despite this, it is important to understand that fruit, rocks, and heads do not need to be round for felicitous *-ka* usage. Each referent type seems to open new avenues of analogical usage. The relationship between the meaning domains for *-ka* is schematized as a Venn diagram in Figure 4.28.

The import of these multi-dimensional meaning relationships is that the meanings of *-we* or *-ka* cannot be reduced to some common intersection of meanings or a prototype. The meanings are related but there truly are multiple meanings. These meaning relationships are very interesting in their own right but my goal is simply to point out that they exist. In assigning meaning to lexical suffixes, it might be tempting to try to find some core shared abstract attribute that could be assigned to the affix. This would not be descriptively adequate. Additionally, when it was attempted in the context of a PFM analysis, it was shown to simply not work.

This lack of intersecting meanings is important to capture. Lexical suffix polysemy cannot be reduced to single abstract or prototypical meanings. Beyond providing a means to represent disjoint polysemous meanings, this work does not spend a lot of time on modeling relationships between meanings. I do not draw linkages between lexical meanings within a formal system, as might be seen in Cognitive Linguistics (CL) (Lakoff, 2008) or Generative Lexicon (GL) theories (Pustejovsky, 1991; Asher, 2011; Chatzikyriakidis and Luo, 2017). Relating meanings to meanings is outside the scope of this work. Instead the focus is on relating meanings to morphological forms. I only specify a morphological interface to lexical semantics. The interface allows for subsequent articulation using a theory of lexical semantics but that is not be necessary to achieve the goals of this thesis.

4.3.7 On instances of questionable polysemy

Polysemy among related meanings does not always offer an explanation of the meanings associated with a lexical suffix. As Table 4.19 indicates, where the *-po* lexical suffix is associated with meanings as unrelated as hands and canoes, there is no clear synchronic relationship between all *-po* meanings. Despite this, due to the closed class nature of the system, where there are a small, finite number of affixes, this is not an instance of mere homophony.

Affixation Rules			
…			
wẽ rule	X	⇒	Xwẽ
po rule$_1$	X	⇒	Xpo
po rule$_2$	X	⇒	Xpo
ka rule	X	⇒	Xka
…			

Context of Application

…

When pole apply **wẽ rule**

When tree apply **wẽ rule**

When canoe apply **po rule**$_1$

When hand apply **po rule**$_2$

When grapes apply **po rule**$_2$

When fruit apply **ka rule**

…

Figure 4.29: Given a fixed inventory of affix sound sequences in a language, represented by the Affixation Rules, providing separate affixation rules to apply in different meaning context such as "po rule$_1$" and "po rule$_2$" is not only redundant, but results in observational equivalence.

Homophony is coincidence. The fact that the past tense of English *rise* sounds the same as the word for the flower *rose* is only made more likely than chance by the fact that both exist within the phonological system of English, which provides constraints on possible word-forms. Wao Terero has a closed inventory of affixes. There is no way to freely derive new affixes. If any two affixes sound alike, the coincidence is not only decreased by phonology but by their coexistence in a fixed inventory.

This means that coincidence is decreased and less homophony-like. Yet, I will go further and claim that there is an identity relationship between the *-po* associated with one set of disjoint meanings and the *-po* associated with the other. Consider an abstraction of the rules that are responsible for affixation from rules that associate meaning to an affix. Let us consider the hypothesis that there are two *-po* affixes that are homophonous. When we try this out we find that specifying two rules to perform this affixation would be redundant. Figure 4.29 schematizes the situation. If one were to represent the affix inventory of a language in terms of affixation rules, failing to identify two rules with the same outcome as the same rule results in the redundancy. This is not just because two rules do the same string manipulation but because having two rules has no consequence. This lack of consequence makes a proposal of multiple affixes more than merely redundant. The lack of consequence means that the redundant option is observationally equivalent to the non-redundant option. Both predict the same thing, that when a morph that sounds like /po/ occurs, there are particular meanings that may be associated with it. To illustrate this, the figure includes a *context of application*. The context of rule application is a listing of what rules should be applied given a meaning. It is not at all clear what a theory would gain by having multiple rules to apply *-po* in a ‘canoe’ versus ‘hand’ context. The result of applying either rule would be exactly the same.

As stated above, there are instances where different categories are assigned to the same form. In the case of *-kã*, this was due to the fact that there were two distinct distributions. That is not the case for *-po*. There is no *-po* that distributes like a person marking, tense, or some other morphological system.

4.3.8 Word and paradigm properties

As has been emphasized above, the basic units of the form paradigm are word-forms. Specifically, they are categorized word-forms. These are grouped into hierarchies, called form classes, which describe a paradigmatic dimension. Some stem within a particular form class will have exactly the same affixation patterns as other members of the class. The morphotactics, is defined in terms of rules, which are directional, often from an unaffixed item to an affixed item. These relate two word-forms with the same stem, and are also constrained using a hierarchy. The realizational calculation is distributed according to a relation from a morphological form paradigm to a syntactic sign paradigm. The mapping utilizes patterns between word-forms and syntactic categories to provide lexical entries, which may be used in the phrasal syntax. Lexical meanings are listed in a separate component that is referenced by the relation between paradigms. This is able to supply disjoint lexical meanings for lexical suffixes and other items. Together the lexical semantic and sign paradigm provide the full range of meanings and categories that may be associated with a particular word-form.

All of this provides a way of being able to describe the co-variation of the forms of words and meanings. Given the mechanisms described above, the system is already nearly a WP framework. It is already word-form based, though not necessarily word-based, in the sense of a concept of lexeme. The form classes and morphotactic hierarchies provide important elements of paradigm structure. Yet, a necessity for a true WP theory is the ability to leverage paradigmatic patterns to make inferences concerning one paradigmatic instance given another of the same kind. This allows bidirectional rules, yẽdẽpo $\leftrightarrow$ yẽdẽ and yẽdẽpo $\leftrightarrow$ yẽdẽka, where $\leftrightarrow$ is if and only if. It also allows one to state, for another adjective stem *giita*, ‘small’, that giitapo $\leftrightarrow$ yẽdẽpo, or even giita $\leftrightarrow$ yẽdẽpo. This is a necessary goal for a true WP theory. I supply the means to do so at the end of Chapter 6.

Chapter 5

Formal Preliminaries

In this chapter I provide an overview of formal systems that this work relies on. Inherent in the notion of preliminary, I discuss work that is prior to my own. The goal is to summarize and explain ideas. My innovation, relative to what is presented in this chapter, comes from how these systems are interpreted within the context of morphological theory, which may not be fully apparent until Chapter 6. The ideas I borrow from type theory and Linear Categorial Grammar (LCG) are, themselves, innovative. Type theory and categorial grammar (CG) are influential, though they may be unfamiliar to those who have little experience with formal theoretical models. I chose to build on the a foundation of type theory and LCG for specific reasons. There are other available options. For this reason, there is also discussion of why these systems are good systems, in addition to simply explaining them.

There are three sections. The first covers basic dependent type theory for non-technical readers. As stated in the introduction, the morphological framework formalism is written in Calculus of Inductive Constructions (CiC) (Coquand and Huet, 1988; Luo, 1990; Coquand and Paulin-Mohring, 1990). For those who wish to go deeper, I have found that the clearest reference to CiC is found in the reference documentation for the Coq proof assistant (The Coq Development Team, 2019).[1] I will go over some type constructors and basic concepts. Patterns that I use frequently in Chapter 6 will be explained. There will also be some additional discussion of type theory in the section on semantics.

The second section covers LCG (Pollard, 2015; Worth, 2014; Worth, 2016; Plummer and Pollard, 2012;

[1]Coq will soon be renamed Rocq. The continuously updated documentation can be found at: https://coq.inria.fr/doc/master/refman/.

Pollard, 2015; Martin, 2013; Martin and Pollard, 2014; Martin, 2015; Yasavul, 2017; Needle, 2024; Needle, 2022). Readers may not have a great deal of experience with CGs. This work does not require a great deal of syntactic knowledge in general, but it is good to know the advantages of such theories, as well as important differences one may find between CG analyses and phrase structure analyses in terms of agreement and other co-occurrence restrictions. Despite this, the focus will be on the structure of lexical entries, which I use to define sign paradigm entries of sign paradigms.

Lastly, there will be a section on the semantics of the system. This will not really be about semantics, but on how Needle (2024) embeds semantic types under a single type.[2] For that reason, the section is really more about additional concepts in type theory. For less technical readers, it may be enough to know that it works. For those interested in type theory and formal systems, it demonstrates the power of type theory in providing useful abstractions. I consider Needle's work in this area to be key to my ability to provide a realizational interface to lexical meanings. One might say that the integration of this approach to semantics into a morphological theory is the major contribution of this thesis, a point I return to in Chapter 6.

5.1 Type Theory

It is assumed that the reader is familiar with Higher Order Logic (HOL), which is used extensively in popular theories of formal semantics, as well as other areas of linguistics. Even if one is not accustomed to calling it HOL, it is the logic that is assumed, or roughly approximated, in most formal semantics papers. For that reason, a comparison of HOL and CiC will be helpful. CiC, like HOL, is a sequent-style natural-deduction system. It differs from HOL in a number of ways, those that need to be emphasized to understand this work will be described.

First, just to ground the reader, it is good to consider some basic terms and types that require none of the additional dependent type machinery of CiC. Whether one is a linguist or not, they many have encountered types such as $\mathbb{N}$, for the natural numbers. One can state $3 : \mathbb{N}$ to say that 3 is a natural number. It is similar,

[2] Despite the citation I have chosen, which is for Jordan Needle's thesis, the exact technique for embedding a static semantics below a single type was never published. Alternative means for abstracting over a static semantics can be found in the thesis, but it is different than what I present. The Coq code that was used in for the version of his semantics that I am using may be found at: https://github.com/needle29/needle29. My presentation of his work is simplified and incomplete so the curious should look there for something that is further elaborated.

though not the same, as stating $3 \in \mathbb{N}$, which means that 3 is in the set of natural numbers.[3] In HOL, as well as CiC, functions are expressed in the lambda calculus. I assume the reader has encountered λs in the course of their linguistic education. A function for adding two natural numbers, for instance, can be written $\lambda xy.x + y$. This can be written with type annotations in a wide variety of ways. One common way is the following: $(\lambda(x : \mathbb{N})(y : \mathbb{N}).x + y) : \mathbb{N}$. This annotates the particular arguments to the function "inline" and then provides resulting type of the function as a whole. Another common way is to provide a single type annotation, $(\lambda xy.x + y) : \mathbb{N} \to \mathbb{N} \to \mathbb{N}$. Such functions are considered *curried*. This is named for Haskell Curry. It means that a multi-argument function is a series of single argument functions. So, given $\lambda xy.x + y$, and a natural number argument 12, through β-reduction (function application), one produces the function that adds 12 to some natural number, $(\lambda y.12 + y) : \mathbb{N} \to \mathbb{N}$. An important aspect of such systems is that terms, constants, and variables may have types, but types do not have types. For this reason, if one wished to define a type of all numbers, which would subsume $\mathbb{N}$, they cannot state $\mathbb{N} : Number$. Likewise, they cannot create a type of the even numbers, through some form of quantification over $\mathbb{N}$.

Note that the '$\to$' operator, which non-coincidentally looks like a common notation for logical implication, is the most common notation for function types in the literature. There is a tendency in linguistics to use a set pair notation for function types, such that the equivalent of $\mathbb{N} \to \mathbb{R}$, may be written $\langle \mathbb{N}, \mathbb{R} \rangle$. I prefer the more standard notation. Additionally, there is a tradition in quasi-formal semantics of using the type judgment operator ':' to mean presupposition. This is done in only a small corner of the linguistic literature. Elsewhere, across disciplines, it indicates that the preceding expression has the type that follows, such as $x : \mathbb{N}$, meaning that x has the type of natural numbers. Despite these differences in conventions, the logical foundation of what I am describing and what is done in the linguistic literature, generally, is the same. When an intransitive verb, such as *sleep*, is given a type in the literature, such as $\langle e, t \rangle$, it can be written $sleep : e \to t$. I use examples with the natural numbers in this section largely to avoid getting bogged down in considerations of the legitimacy of a particular analysis. Linguistic formal treatments are areas of active research. Mathematical examples that I provide are a matter of established definitions.

Unlike HOL, where there are only terms and types, every expression in CiC has a type, including types, themselves. These types are not necessarily simple. In CiC, types can contain (or be) variables, and there are

[3]The basic difference between sets and types is that sets are model theoretic, while CiC types are logical formulas.

type constructors that bind variables in types. As an example, there is a type constructor $\Pi x : A.B(x)$. The first part of this is often read as "for all x of type A" and an alternate notation, which I will use is $\forall x : A.B(x)$. Types of this form are called dependent product or dependent function types. The notion is that B is some type constructor parameterized on x. The $B(x)$ notation can be read as B where x occurs. A realistic and common example is that B is a type constructor for polymorphic vectors, vec, such that the first argument to vec is its arity and the second is the type of its members. Then $\forall x : \mathbb{N}.vec(x, \mathbb{R})$ would be the type of a function that takes a natural number and returns a vector of real numbers of length x. Note, if B is not dependent on the value of x, the construction is a simple function type. For instance $\forall x : \mathbb{N}.\mathbb{R}$ is the same as $\mathbb{N} \to \mathbb{R}$.

There are certain types called universes and every type is an inhabitant of one of them. The two most important universes are $Prop$ and Set. Set is the universe of datatypes, for instance $\mathbb{N} : Set$. Types that are inhabitants of Set have terms, and not types as their inhabitants. That is to say that if A is a type in Set, none of the inhabitants of A will be types. Using the example from above, if we wished for a type of a function that takes a type and returns a vector of length 5 of that type, it is $\forall x : Set.vec(5, x)$. Since all types also have types, this type will be in a universe other than Set, due to the quantification-like quality of the type constructor. In CiC there is an infinite hierarchy of universes, named $Type_n$ such that $Type_1$ belongs to $Type_2$, etc. This means that $\forall x : Set.vec(5, x) : Type_1$. This is called predicativity. There is a hierarchy of types such that no type is the type of all types, thus allowing one to quantify over types without having to worry about paradoxes.

$Prop$ replaces the HOL type t of formulas. Like Set, the inhabitants of $Prop$ that are types do not, themselves, have types as inhabitants. Inhabitants of types within $Prop$ are (interchangeably) *proofs* or *witnesses*. In fact, inhabitants of Set are also considered to be proofs, but some proofs in $Prop$ are abstract calculations. An example type in $Prop$ is $4 \leq 5 : Prop$. A term of type $4 \leq 5$ would be a proof of $4 \leq 5$. Just to make sure the reader understands this, since the concept may be new, $4 \leq 5$ is a *type*.

To provide an inhabitant of the type, one needs to provide a proof.[4] To provide a very simple example,

[4]There may be many inhabitants of a type if there are many ways to provide proof. For a type like $\mathbb{N}$, this is desirable, since the goal is to have an infinite number of unequal proofs, such as 7, 34, and 10. Some may feel that having many unequal proofs of $4 \leq 5$ is undesirable, and that such a proposition should be considered only to be proven or not proven. This relates to the concept of proof irrelevance, where all inhabitants of a type are equal (Gilbert et al., 2019). This is an optional feature of types in modern type theory. If one were to define the universe Prop so that proofs are irrelevant, any proof of $4 \leq 5$ would be equal to any other.

consider two types involving equality. One is $5 = 5 : \mathit{Prop}$, the other is $5 = 6 : \mathit{Prop}$. Both of these are formulas, and may inhabit Prop. The difference in whether they are true or not depends on being able to supply an inhabitant. This is a core concept of constructive mathematics. Truth is provable. False is unprovable.

In the definition of equality, there is the notion of reflexivity, which means that something is equal to itself. For that reason it trivial to reference that definition to provide a proof of $5 = 5$. That proof, is the inhabitant of $5 = 5$. Let us say that we named the proof, in a manner similar to how one names a function, such as $5eq5 := 5$ is equal to itself by reflexivity. I'm using prose here instead of running through a proof in a technical manner. S proof often involves a calculation. For instance, $3 - 2 + 4 = 1 \times 5$ would require a reduction of the arithmetic before arriving at a point where one can appeal to reflexivity. This is why I will refer to proofs as calculations in some cases. It also means that proofs may be fairly complex objects. In the $5 = 5$ case, a name for the proof was supplied, so not all of the details of a calculation, even if it is only an appeal to reflexivity, need to be provided to supply a judgment, where a judgment is the use of ':' to signal that a type is inhabited. Using the name for the proof, one can say $5eq5 : 5 = 5$.

The other type, $5 = 6$, by default, is not false simply because there is no proof. Rather it is simply unproven. Something that is unproven is not, necessarily, unprovable. To show that $5 = 6$ is false, one needs to demonstrate that there is no possible proof. In my work on morphological theory, there is rarely a reason to go through the work of proving that something cannot be proven. The goal is to provide a predictive model, and to test those predictions empirically. In this context demonstrating constructive falsity is rarely necessary.

Providing a proof by calculation for a type under Prop is a common way of supplying an inhabitant. There is also another way. One can declare that there is an inhabitant of a type by axiom. The declaration of axioms differs between CiC and HOL. In HOL one may make a formula of type t an axiom by stating, $\vdash \phi$. In CiC, one needs to declare a constant that inhabits ϕ, $\vdash p : \phi$. As an extended example, it will be relevant to this work that partial orders are defined over certain types, say foo, that have no natural ordering. In order to specify that foo is a type and that it should be a datatype, one states $\vdash \mathit{foo} : \mathit{Set}$. In order to define the partial order over some type foo in HOL one would state $\vdash \leq : \mathit{foo} \rightarrow \mathit{foo} \rightarrow t$. In CiC this would be $\vdash \leq : \mathit{foo} \rightarrow \mathit{foo} \rightarrow \mathit{Prop}$. This means that a proof of an inequality of two foos is an inhabitant of Prop.

The properties of the order can be defined as follows:

$$\begin{aligned}&\vdash \leq_{reflexive} : \forall x : foo. x \leq x\\&\vdash \leq_{antisymmetric} : \forall x, y : foo. x \leq y \rightarrow y \leq x \rightarrow x = y\\&\vdash \leq_{transitive} : \forall x, y, z : foo. x \leq y \rightarrow y \leq z \rightarrow x \leq z\end{aligned} \tag{I}$$

In the statements above, $\leq_{reflexive}$, $\leq_{antisymmetric}$ and $\leq_{transitive}$ are constants. The $\forall$ is the dependent product type constructor discussed above. This means that $\leq_{antisymmetric}$ has the type of a dependent function. This function takes two foos, x and y as its first arguments. It then takes an argument of type $x \leq y$ and an argument of type $y \leq x$, in other words, a *proof* of each of these types. If these proofs can be supplied, the function returns a proof of $x = y$, where $x = y$, itself, is of type $Prop$. By providing the constant $\leq_{antisymmetric}$ we do not have to construct the proof that this function exists. Likewise, consider the example of equality above. There it was stated that reflexivity was part of the definition of '='. Above, with $\leq_{reflexive}$ one can see how this is done.

If one were to wish to specify that $a, b : foo$, $a \leq b$ and $b \leq a$, we would likewise need to provide constants as inhabitants of these types, since there is no natural order for the type.

$$\begin{aligned}&\vdash a : foo\\&\vdash b : foo\\&\vdash \leq_{a,b} : a \leq b\\&\vdash \leq_{b,a} : b \leq a\end{aligned} \tag{II}$$

Given this, we can see that $\leq_{antisymmetric} (a, b, \leq_{a,b}, \leq_{b,a}) : a = b$.

The primary CiC tools that I use in this work are Π-types, similar to those provided in the examples, and inductive types, which are types with an inductive definition. A simple inductive type defines $\mathbb{N}$ in (III).

$$\begin{aligned}&\vdash \mathbb{N} : Set :=\\&\quad | O : \mathbb{N}\\&\quad | S : \mathbb{N} \rightarrow \mathbb{N}\end{aligned} \tag{III}$$

Each clause in (III) is a constructor, used to provide witnesses of $\mathbb{N}$. A constructor like O provides a simple witness. One can say that $\mathbb{N}$ contains O. The constructor S provides a means of constructing inhabitants of $\mathbb{N}$. The type $\mathbb{N}$ is closed by the S constructor.

I can define the order $\leq: foo \to foo \to Prop$ inductively, as well. Instead of providing a declaration that $\leq$ has a type $foo \to foo \to Prop$, and then separately stating axioms that allow for proofs of $\leq$, an inductive definition allows one to state all of the means of providing inhabitants of $\leq$ as inductive clauses. There are actually many types that are defined in an inductive type of this kind because $\leq$ takes arguments that construct types in $Prop$.

$$\begin{aligned} &\vdash leq : foo \to foo \to Prop := \\ &\mid \leq_{reflexive}: \forall x : foo.x \leq x \\ &\mid \leq_{transitive}: \forall x, y, z : foo.x \leq y \to y \leq z \to x \leq z \\ &\mid \leq_{a,b}: a \leq b \\ &\mid \leq_{b,a}: b \leq a \end{aligned} \tag{IV}$$

In (IV), there are a number of clauses. The line that states $\leq_{a,b}$: $a \leq b$, provides a witness $\leq_{a,b}$ of the type $a \leq b : Prop$. The line $\leq_{reflexive}$: $\forall x : foo.x \leq x$ defines a means of constructing inhabitants of many types that conform to a particular condition. An inductive definition provides the same benefits within CiC as in other formal systems. Boundaries are provided to the definition and inductive proofs are also facilitated. Inductive types may also contain recursive constructors, though this is not a necessary trait. Therefore, inductive types in CiC also serve as non-dependent (ordinary) sum types, which serve to provide an enumeration of inhabitants. For instance, there is no recursion in the example below.

$$\begin{aligned} &\vdash food : Set := \\ &\mid meat : food \\ &\mid fish : food \\ &\mid vegis : food \end{aligned} \tag{V}$$

Notably, the antisymmetric property of $\leq$ needs to be defined outside the inductive definition, since it is a means of proving '$= : foo \to foo \to Prop$', not '$\leq : foo \to foo \to Prop$'.

For the semantics, in addition to Π types and inductive types, I will be using dependent sum types, or

Σ-types. I wait until that section to discuss them, since they are not used elsewhere. Little depends on the use of CiC over other modern type theories that include these features, to my knowledge.

5.2 Linear Categorial Grammar

It is necessary to know the basics of the LCG formalism before moving on to Chapter 6. This is because the formal characteristics of the interface with syntax depends on the structure of LCG lexical entries. The morphological framework defines the interface as a relationship between two paradigms. This is described in detail in Chapter 6. The first of these paradigms is the form paradigm, which consists of categorized word-forms. The second is the sign paradigm, which consists of lexical entries. The rules which map between paradigms, form-sign relations, therefore directly reference the structure of lexical entries. Essentially, they provide the means to construct, or prove lexical entries within CiC.

The fact that form-sign relations depend on LCG lexical entry structure does not mean the morphological framework is conceptually incompatible with other syntactic theories. It is only that I have adopted LCG, in this instance, that the structure of LCG lexical entries, in particular, are referenced. One may chose different syntactic theories, but they would need to alter aspects of form-sign relation definitions to account for any significant difference in the target lexical entry structure. Otherwise, any lexical syntactic theory may be used, in principle. This is not to say that the quality of a theory using another syntactic interface would be equal. There are very specific qualities of LCG that make it a superior choice as an interface partner of a morphological system. I discuss these at the end of the section so that the reader may become acquainted with the formalism first.

I assume that many readers have had no experience with CGs, much less LCG, specifically. Unfortunately, I only scratch the surface in this section. For a relatively brief paper, with interesting examples, and relatively low technical overhead, see Mihaliček and Pollard (2012), which describes the benefits of LCG (there called pheno-tecto-distinguished CG) when modeling interrogative expressions. Other works have been are cited, but Mihaliček and Pollard (2012) is a very accessible first step in leaning more.

5.2.1 Some background on CGs

CGs are a form of generative syntax that predates *Syntactic Structures* (Chomsky, 1957). The earliest examples being Ajdukiewicz (1935), Bar-Hillel (1953), and Lambek (1958). They operate in a manner that is, at least, superficially similar to a transformational grammar, to the extent that they involve a derivation to demonstrate that a sentence is grammatical.

If one has seen a CG only once, they may recognize formulas like the following: $\mathsf{NP}\backslash\mathsf{S}$, which corresponds to the category of an intransitive verb, such as *sleep*. It says that given an NP on the left, then S. If *the mice* is demonstrably an NP in the system, the category can be used to verify that the string *the mice sleep* is a grammatical sentence because NP $\mathsf{NP}\backslash\mathsf{S}$ reduces to S. The slash system does not look radically dissimilar to a phrase structure rule, such as $\mathsf{S} \Rightarrow \mathsf{NP}\ \mathsf{VP}$, which states that an S is a category formed by an NP to the left of a VP. The slash notation is a little more terse. It does not invoke an additional symbol for a VP, but otherwise constrains grammar in a similar manner.

Unlike a phrase structure rule, there is no hierarchy of categories implied. One does not provide the same kind of abstraction over verb phrases as in phrase structure systems – at least, not by default or design. One begins to see what this difference means when comparing an intransitive and transitive. An English transitive, such as *love*, can be represented as $\mathsf{NP}\backslash\mathsf{S})/\mathsf{NP}$. This says that a complement NP must occur to the right. To express this with a phrase structure rule, one would provide some additional definition of VP, such as $\mathsf{VP} \Rightarrow \mathsf{TV}\ \mathsf{NP}$, where TV stands for transitive verb. This means that there is a difference in abstraction, since there may be many kinds of VP, yet only one rule $\mathsf{S} \rightarrow \mathsf{NP}\ \mathsf{VP}$ to account for how they combine with a subject. Note, there is no difference in the number of reductions that are required in either case. It will be shown that the CG style, where one does not abstract over phrasal arity and subcategorization, is beneficial when defining an interface to semantics.

Despite the differences, until one starts considering manipulations, such as movement or other operations, the systems look like similar notation for expressing similar ideas. Prior to important work by van Benthem (1988) on the syntax-semantic interface, the most prominent advantage of CGs over phrase structure, as far as I can tell, is that CGs correspond to specific logics, which provide them well understood formal foundations. This is not to say that some phrase structure grammars were not formally well-defined. Nor is it the case that phrase structure grammars could not be defined in terms of specific logics, but to my knowledge, historically, they were not.

The types of logics used in many CGs, called typelogical grammars, are substructural logics. A substructural logic is a type of logic lacking one or more structural rules. An example structural rule is weakening. Technically, it means that if some formula is provable based on a set of premises, for instance $\Gamma \vdash B$, where Γ is the set of premises and B the provable conclusion, then the weakening rule allows an additional premise A to be added to the assumptions $A, \Gamma \vdash B$, without changing the provability of B. An example of weakening, in syllogistic style, is to begin with some premises, such as "All dogs go to heaven. Fifi is a dog." and then use that to prove "Therefore, Fifi goes to heaven." Weakening allows one to add an additional premise to the set of premises that enabled the proof, such as "God loves fish." Then the proof is "All dogs go to heaven. Fifi is a dog. God loves fish. Therefore, Fifi goes to heaven." From a Gricean point of view it sounds odd, but it is useful in some formal systems. Weakening does not work so well for syntactic formulas, as it would allow the introduction of arbitrary syntactic categories into a derivation. Avoiding going into details, it would be like saying NP NP\S reduces to S, therefore it is also the case that NP NP NP\S reduces to S. In general, the reason that a typelogical grammar is defined as a substructural logic is to avoid such issues. Syntactic categories, though often very abstract, correspond to a physical signal, which have particular properties, for instance, that something can't be two places at the same time, nor can physical things duplicate or disappear without consequence. Syntax is more concrete than is assumed by the rules of some logical systems. Substructural logics help account for this.

Grammars defined using logics have an obvious appeal for theorists concerned with formal quality and elegance. For this reason, historically, despite the popularity of phrase structure grammars, sufficient value was found in typelogical grammars during the 1960s through 80s, that they continued to be used within certain domains. This eventually lead to an observation that drew more interest in CGs, which is that there is a correspondence between the form of syntactic categories in CGs and the types of semantic expressions (van Benthem, 1988). This can be seen in Table 5.1. In the table there are various syntactic categories listed in the center. On the left is some example word they may correspond to. On the right is a semantic type. Looking at the specific category names and their arity there is a clear correspondence between categories and semantic types, which are independently motivated.

When this pattern was discovered, it resulted in a renewed interest in CG due to the elegant expression

Example	Syntactic Category	Semantic Type
he	NP	e
sleep	NP\S	$e \to t$
love	(NP\S)/NP	$e \to e \to t$
think	(NP\S)/S	$t \to e \to t$

Table 5.1: A correspondence between syntactic categories in a typelogical grammar and semantic types.

of a compositional syntax-semantics interface. It allowed for an interface based on the compositional rule-to-rule principle. Whenever an operation affects syntactic combination, a corresponding operation affects semantic combination. This can be seen in (VI) where each reduction in the syntactic category results in a reduction of the semantic type. To emphasize the correspondence I only show the types, but the '/'-elimination rules, responsible for syntactic reduction, correspond to function application of semantic terms. For instance $\mathsf{NP}; \mathit{speaker} : e\ \mathsf{NP}\backslash\mathsf{S}; \mathit{sleep} : e \to t$ reduces to $\mathsf{S}; \mathit{sleep}(\mathit{speaker}) : t$.

$$\dfrac{\text{he;NP; } e \qquad \dfrac{\text{loves;(NP\textbackslash S)/NP; } e \to e \to t \qquad \text{him;NP; } e}{\text{loves him;NP\textbackslash S; } e \to t}}{\text{he loves him;S; } t} \tag{VI}$$

For those who feel that compositionality should be the default manner of relating syntactic combination and meaning, the system is obviously all one could wish for. It is constraining and, in practice, highly expressive. An additional advantage of the system is that an expression such as $(\mathsf{NP}\backslash\mathsf{S})/\mathsf{NP}; e \to e \to t$ exposes the semantics to the same extent as the syntactic expression. When interfacing such a system with morphology the semantics is transparent. There is no need to use syntactic structure as a conduit to semantics.

This can be compared to a widely used alternative to the syntax-semantics interface, which was proposed by Lewis (1970). It was later popularized by Heim and Kratzer (1998), an introductory semantics text book. In that approach, a syntactic structure is first constructed according to syntactic rules. Afterward, there is a semantic interpretation, as a secondary step. The issue with a secondary interpretation is that it essentially results in a paraphrase. This is especially true in Minimalist theories that utilize the pattern. They have concept of a syntactic level of LF, which has divided off and is independent of further form-based syntactic structure manipulations. Operations at LF covertly reorder the syntactic tree structure prior to the interpretation stage (Hornstein, Nunes, and Grohmann, 2005) with the goal of making the tree more amenable to desired interpretations. The practice has also resulted in the need to store semantic information within the syntactic structure so that it can be interpreted at the interface, so called "interface instructions" (Panagiotidis, 2024). The issue is similar to loading non-syntactic information into syntactic nodes to facilitate morphological realization. The information isn't justifiable on syntactic grounds alone, but cannot live anywhere else due to the architecture of the framework, making syntax a dumping ground for information that rightly belongs elsewhere.

Syntactic categories in CGs are not data structures that may contain data. Their non-syntactic consequences are encoded as well defined, parallel relationships to other aspects of the linguistic system.

5.2.2 Abstract versus concrete syntax

LCG explores the proposal of Curry (1961) that natural language syntax may be fruitfully analyzed according to a concrete-abstract dichotomy. The concrete syntax is called the phenogrammar, and the abstract syntax is called the tectogrammar. I will refer to these as pheno and tecto. The tecto deals with abstract relationships between categories, such as functor argument relationships, and grammatical agreement. The pheno deals with whether the auditory correlate of the argument occurs before or after the auditory correlate of the functor. That this division exists makes sense. For instance, one might expect that the very same kind of functor argument patterns found in the aural-oral modality would exist in sign languages, even if their concrete expression is fundamentally different. On some level this separation is evident in many theories, to a greater or lesser extent, so I do not view it as particularly controversial. What is innovative in LCG, and similar theories called λ-grammars or Curryesque grammars, is the manner in which the idea is expressed. Following Oehrle (1994), a λ-expression provides a pheno function on string terms, such that $\lambda s.s \bullet \mathit{sleep}$ is a function that uses a string concatenation operator '$\bullet$' to express that its argument precedes the string sleep. These systems leverage a similar correspondence as seen between the types of syntactic categories and the types of semantic functions to provide a constraint on pheno functions. The result is that the pheno is expressed as functions on strings, while the tecto is simplified.

In the LCG tecto, the slashes seen in expressions such as $\mathsf{NP}\backslash\mathsf{S}$, in the examples above, are replaced by a single operator, called linear implication, written $\multimap$. The resulting techo looks like $\mathsf{NP} \multimap \mathsf{S}$. Such a tecto can be paired with the pheno function $\lambda s.s \bullet \mathit{sleep}$. The tecto ensures that the correct syntactic category relation holds, while the pheno ensures that strings occur in the right order. Technically, the tecto is an

Example	Pheno Type	Tecto	Semantic Type
he	str	NP	e
sleep	$str \rightarrow str$	NP $\multimap$ S	$e \rightarrow t$
love	$str \rightarrow str \rightarrow str$	NP $\multimap$ NP $\multimap$ S	$e \rightarrow e \rightarrow t$
think	$str \rightarrow str \rightarrow str$	NP $\multimap$ S $\multimap$ S	$t \rightarrow e \rightarrow t$

Table 5.2: A correspondence between tecto categories and pheno and semantic types.

expression of the implicative fragment of linear logic (Girard, 1987), a substructural logic. The implicative fragment of linear logic is the portion of the logic that contains only rules for linear implication. The implication operator by itself doesn't behave differently than implication in other logics. Given the full logic, there are significant differences between linear logic and other logics, but the implication operator requires no special understanding. The utility in using the unfamiliar operator is that it signals the type of logic being used, and qualities such as lack of structural rules. If the reader would like to know more, an accessible introduction to linear logic can be found in Wadler (1991).

As mentioned, a similar correspondence between tecto categories and pheno types holds as between those of the tecto and semantic types. This can be seen in Table 5.2. The same rule-to-rule principle holds with the introduction of pheno.

5.2.3 The rules and lexical entries of LCG

Now that the reader has some familiarity with CGs, and some high level conceptual grounding in LCG, in particular, we are ready to look at LCG more concretely. I begin this subsection by discussing the structure of an LCG lexical entry, which is also called an LCG sign. At the morphological level these signs, or lexical entries, are referred to as sign paradigm entries. The lexical entry is divided into three portions for the pheno, tecto, and semantics.

In (VII), we see the six signs needed in LCG for WALK, as well as most intransitive verbs that follow default inflectional patterns. For the first three of the below, the tectos say that given some nominal category there is a finite clause, which has category Fin. The category Nom is named for nominative. Without the subscript it is a general nominate, suitable for any person or number. The category $\mathsf{Nom_{3s}}$ is specific to singular third person. The category $\mathsf{Nom_{\neg 3s}}$ is specific to non-third person singular. Note that terms such as $\mathsf{Nom_{3s}}$ are only category names. They are not decomposable, nor are they restrictions on meaning. The names Bse, Prp, and Psp are mnemonics for base, present participle, and past participle. Their lineage as abbreviations goes back to Generalized Phrase Structure Grammar (GPSG) (Gazdar et al., 1985), which shares a lineage with LCG through Head Driven Phrase Structure Grammar (HPSG). The PRO is a stand in category used in non-finite clauses. I won't provide detailed descriptions of how the last three entries are used in analyses, as it is outside the scope of the thesis.

$$\langle \lambda \mathrm{s.s} \bullet \mathrm{walk}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, \mathit{walk} \rangle \tag{VIIa}$$

$$\langle \lambda \mathrm{s.s} \bullet \mathrm{walks}, \mathsf{Nom}_{3s} \multimap \mathsf{Fin}, \mathit{walk} \rangle \tag{VIIb}$$

$$\langle \lambda \mathrm{s.s} \bullet \mathrm{walked}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{walk}) \rangle \tag{VIIc}$$

$$\langle \mathrm{walk}, \mathsf{PRO} \multimap \mathsf{Bse}, \mathit{walk} \rangle \tag{VIId}$$

$$\langle \mathrm{walking}, \mathsf{PRO} \multimap \mathsf{Prp}, \mathit{walk} \rangle \tag{VIIe}$$

$$\langle \mathrm{walked}, \mathsf{PRO} \multimap \mathsf{Psp}, \mathit{walk} \rangle \tag{VIIf}$$

Although considerable work has been done to expand the theory of LCG signs, particularly the pheno (Worth, 2014; Worth, 2016; Needle, 2022) and semantics (Plummer and Pollard, 2012; Pollard, 2015; Martin, 2013; Martin and Pollard, 2014; Martin, 2015; Yasavul, 2017; Needle, 2024), I'll assume a simplified theory here. The pheno and semantics were originally conceived of as expressions in higher-order logic. Given the work of Needle (2022) and Needle (2024), they have received new definitions in CiC. I will not be following these treatments. I do not require the power provided by Needle's innovations in order to achieve my aims within the morphological domain. In order to treat the lexical entries as data types, the tecto, pheno and semantics are embedded under types in CiC, named τ, ϕ, and sense. The pheno and tecto embeddings are not described. In fact, in Appendix A, only ad hoc types defined for the pheno and tecto, since the focus is on the morphological theory and its relationship to lexical semantics. The pheno will be taken to be conceptually the same as described in Mihaliček (2012). The definition of sense is the topic of the following section.

5.2.3.1 Notes on phenogrammar

The expression $\lambda s.s \bullet walked$ contains a non-logical constant $walked$, which is called the *string support* of the pheno. In this interpretation of LCG, string support correspond to free words and only free words. It will be the job of form-sign relation to relate these string supports to morphological forms. Notably, the types of strings used in the pheno and the types of strings used in the morphology are not the same. The mapping between them is done via a function that injects morphological strings into the string type of the pheno ($\eta : string \to str$). η encodes the unit natural transformation of the syntax level string monad. I call the pheno string type str, and the morphology level string $string$.[5] The discussion of the two overlaps so little that there should be no confusion. If the reader does not know what a monad is, the detail is unimportant. One can simply think of η as converting between types of strings.

One might wonder how strings can be different kinds of strings. Consider that numbers can be different kinds of numbers, such as natural or rational. Certain types of calculations are possible with rational numbers that are not possible with natural numbers. If one had a natural number, and wished to use it within the context of rational computations, they could convert it to a rational with a function. The concept isn't very different here.

There are a number of reasons why the different types are needed. First, my framework explores a strong version of lexical integrity (Di Sciullo and Williams, 1987) within the context of a CG. LCG may offer different challenges and advantages to the notion of lexical integrity relative to assumptions inherent in phrase structure grammars. Consider that a criticism of lexical integrity such as A. C. Harris (2000), which discusses infix clitics in Udi, strongly relies on phrase structure assumptions and a far different morphological system than I define in this work. It is not clear that the argumentation against lexical integrity follows once the foundations are altered. Second, even if one were to relax lexical integrity, the primitives of the morphological representation in this theory correspond to phones, and few would propose that these are the concern of syntactic manipulation, whether lexical integrity holds or not. Therefore one would still need to have an abstraction over morphs, or some sub-word unit larger than a phoneme.

[5] Concerning the types of compound elements, this depends on the kind of compound and the extent to which it is analyzable. There is a cline in English from compounds that one needs a degree in etymology to recognize, on one end, to pseudo-lexical compounds, that appear to be syntactic phrases, on the other. The former would be type $string$, the latter str. In between, there are lexical combinations, which may often be variable, deserving treatments as *both* kinds.

This explains the string supports. A single means of combining pheno terms is provided in this work. There is a free monoid on str with an operator $\bullet : str \rightarrow str \rightarrow str$ and a constant $e : str$ such that $\forall x, y, z : str.x \bullet (y \bullet z) = (x \bullet y) \bullet z$ and $\forall x : str.e \bullet x = x$ and $\forall x : str.x \bullet e = x$. This is simply the standard definition of a string. One does not need to be concerned with unfamiliar vocabulary. The take away is that '$\bullet$' concatenates strings, and there is an empty string e. It means that if x is a string and y is a string then $x : str$, $y : str$, and $x \bullet y : str$, etc.

5.2.3.2 LCG rules

In most works on LCG, lexical signs are not written as in (VII), as their status as data structures is not emphasized. They generally appear as in (VIII), with semicolon separators and without brackets. This is how they will appear when used in LCG proofs.

Now that I am talking about LCG rules and proofs, I am temporarily entering a different formal context. Only to the extent that LCG may be embedded in CiC are the rules and proofs related to CiC. The turnstile in the examples below is not the CiC turnstile. The notion of LCG sequent, where context and hypotheses live, is distinct from what one will find in the CiC literature, so it is important not to confuse the two. Neither do I use CiC features within LCG. In providing a proof, below, I am not providing a witness of some type. Please keep this context change in mind.

$$\vdash \lambda \text{s.s} \bullet \text{walks}; \mathsf{NP} \multimap \mathsf{Fin}; walk \tag{VIIIa}$$

$$\vdash \text{bill}; \mathsf{NP}; bill \tag{VIIIb}$$

This raises the question of what the turnstile and sequent do. The turnstile '$\vdash$' delimits the context, containing hypotheses, and the consequent. The consequent is proven when there is nothing on the left of the turnstile. Lexical entries are axioms of LCG, for this reason they are given with nothing to the left of the turnstile.

Items end up on the left of the turnstile when introduced as hypotheses using the trace axiom schema. For now, I will introduce a simplified trace axiom, which is elaborated later. They have the basic form of $x; T; y \vdash x; T; y$, where the same information is provided on the left and the right of the turnstile. The T is a tecto and the x and y are pheno and semantic expressions containing no logical constants. I will demonstrate the use of trace axioms below.

The rule schemas of LCG are very basic. First there is modus ponens.

$$\frac{\Gamma \vdash \mathrm{f}; \mathsf{B} \multimap \mathsf{E}; g \qquad \Delta \vdash \mathrm{a}; \mathsf{B}; c}{\Gamma\Delta \vdash \mathrm{fa}; \mathsf{E}; gc}\ \mathrm{MP} \tag{IX}$$

Second there is hypothetical proof.

$$\frac{\Gamma, \mathrm{x}; \mathsf{A}; z \vdash \mathrm{a}; \mathsf{B}; c}{\Gamma \vdash \lambda \mathrm{x}.\mathrm{a}; \mathsf{A} \multimap \mathsf{B}; \lambda z.c}\ \mathrm{HP} \tag{X}$$

Modus ponens is analogous to merge in Minimalism. When I discussed the rule-to-rule principle above, I was essentially assuming a rule like modus ponens. In the rule definition above, only the terms are provided. Below I add some example types for clarity. The rule results in function application in the pheno and semantics.

$$\frac{\Gamma \vdash \mathrm{f} : \mathrm{str} \rightarrow \mathrm{str}; \mathsf{B} \multimap \mathsf{E}; g : e \rightarrow t \qquad \Delta \vdash \mathrm{a} : \mathrm{str}; \mathsf{B}; c : e}{\Gamma\Delta \vdash \mathrm{fa} : \mathrm{str}; \mathsf{E}; gc : t}\ \mathrm{MP} \tag{XI}$$

This is a good point to demonstrate the trace axiom. The way it works can be seen in (XII). A lexical entry with a techo of $\mathsf{NP} \multimap \mathsf{Fin}$ is provided on the left.

$$\frac{\vdash \lambda \mathrm{s}.\mathrm{s} \bullet \mathrm{walks}; \mathsf{NP} \multimap \mathsf{Fin}; walk \qquad \mathrm{x}; \mathsf{NP}; y \vdash \mathrm{x}; \mathsf{NP}; y}{\mathrm{x}; \mathsf{NP}; y \vdash \mathrm{x} \bullet \mathrm{walks}; \mathsf{Fin}; walk\ y}\ \mathrm{MP} \tag{XII}$$

The goal is to provide a proof of a sign with a tecto of Fin and nothing to the left of the turnstile. Given that, the move I made may in adding a trace axiom may seem a little obtuse since it adds hypotheses for no particular reason, but for now, I only wish to demonstrate the effect of a trace axiom. A trace axiom can enter a proof wherever a lexical entry may. When the two expressions combine, the pheno and semantic expressions combine through function application. The hypotheses on the left of the turnstiles also combine into a shared listing. The order of hypotheses does not matter. They form an unordered collection. The same hypothesis cannot occur twice on the left of the turnstile. This does not mean that items with the same tectos cannot both be hypotheses, but the variable names for the semantics and pheno must be distinct.

Hypothetical proof corresponds to move in Minimalism. For the pheno and semantics, it corresponds to lambda abstraction. Hypothetical proof is how one moves hypotheses back into play, so to speak. Consider (XII). The result of modus ponens with the trace is that a hypothesis is stored to the left of the turn style. Utilizing hypothetical proof we can reverse this move, and bring the NP back into play.

$$\frac{\text{x}; \mathsf{NP}; y \vdash \text{x} \bullet \text{walks}; \mathsf{Fin}; \mathit{walk}\ y}{\vdash \lambda\text{x.x} \bullet \text{walks}; \mathsf{NP} \multimap \mathsf{Fin}; \lambda y.\mathit{walk}\ y}\ \text{HP} \tag{XIII}$$

Given this simple example, it may not be obvious why this is useful. The basic idea is that, at times, some temporary filler is needed to ensure the right tecto term for further combination. By using a trace one makes a promise that, eventually, something concrete will fill the blank. A more involved example is needed to show this. A good example is quantifier scope ambiguity, following an analysis by Oehrle (1994).

The lexical entries in (XIV) are simplified to avoid discussing case and agreement for this example. I am also ignoring types. The entries will be used to prove a sentence of the form *Some bird pecks every head.* This is ambiguous as to whether there is one ambitious bird pecking all heads, or whether every head has some individual pecker. I provide a full proof of the first reading, but skip redundant aspects of the second. The proofs demonstrate the utility of traces, the power of the pheno, and the compositional quality of the grammar.

$$\begin{aligned}
&\vdash \text{bird}; \mathsf{N}; \mathit{bird}\\
&\vdash \text{head}; \mathsf{N}; \mathit{head}\\
&\vdash \lambda\text{st.s} \bullet \text{pecks} \bullet \text{t}; \mathsf{NP} \multimap \mathsf{NP} \multimap \mathsf{S}; \mathit{pecks}\\
&\vdash \lambda\text{sf.f(every} \bullet \text{s)}; \mathsf{N} \multimap (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \mathit{every}\\
&\vdash \lambda\text{sf.f(some} \bullet \text{s)}; \mathsf{N} \multimap (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \mathit{some}
\end{aligned} \tag{XIV}$$

Something to notice about the entries for the quantifiers is that they take two arguments. The first, s, looks like it should be a normal string. The second, f, is a function of type $str \to str$. This is what it means to allow lambda expressions in the pheno. Not only can one pass in simple strings, but functions over strings. Considering the correspondence, note that the second tecto argument must be $\mathsf{NP} \multimap \mathsf{S}$, which is also a complex type. The form of the pheno type is a perfect match, $str \to (str \to str) \to str$. Clearly, this means that the type of the semantics requires a similar type. The semantic type is a little different, since the type associated with an N is $e \rightarrow t$. The types don't always line up perfectly, particularly due to the types of common nouns, but they line up consistently. The semantic type is $(e \rightarrow t) \rightarrow (e \rightarrow t) \rightarrow t$.

I will take the proof step by step, separating out portions, rather than presenting the tree all at once. I will used Roman numbered references to portions of the proof so as to focus on individual steps.

In (XV) and (XVI), I use modus ponens to combine nouns with quantifiers. The resulting sign has a pheno type of $(str \rightarrow str) \rightarrow str$, again corresponding to a tecto of a similar form $(\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}$, which in turn corresponds to a semantic type of $(e \rightarrow t) \rightarrow t$. Everything moves in lock step.

$$\frac{\vdash \lambda \text{sf.f(some} \bullet \text{s)}; \mathsf{N} \multimap (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \mathit{some} \qquad \vdash \text{bird}; \mathsf{N}; \mathit{bird}}{\vdash \lambda \text{f.f(some} \bullet \text{bird)}; (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \mathit{some\ bird}} \text{MP} \tag{XV}$$

$$\frac{\vdash \lambda \text{sf.f(every} \bullet \text{s)}; \mathsf{N} \multimap (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \mathit{every} \qquad \vdash \text{head}; \mathsf{N}; \mathit{head}}{\vdash \lambda \text{f.f(every} \bullet \text{head)}; (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \mathit{every\ head}} \text{MP} \tag{XVI}$$

Crucially, at this point a $\mathsf{NP} \multimap \mathsf{S}$ is required by the quantified nouns. This is the type of an intransitive verb. The verb we want to use is transitive *pecks*. Here we may take advantage of the trace axiom, as seen in (XVII). This provides a sign of the correct kind.

$$\frac{\vdash \lambda \text{st.s} \bullet \text{pecks} \bullet \text{t}; \mathsf{NP} \multimap \mathsf{NP} \multimap \mathsf{S}; \mathit{peck} \qquad \text{x}; \mathsf{NP}; y \vdash \text{x}; \mathsf{NP}; y}{\text{x}; \mathsf{NP}; y \vdash \lambda \text{t.x} \bullet \text{peck} \bullet \text{t}; \mathsf{NP} \multimap \mathsf{S}; \mathit{peck\ y}} \text{MP} \tag{XVII}$$

The result of (XVII) may be combined with the quantified object, *every head* from (XVI). In (XVIII), in order to help everything fit on a line, I reference (XVII), rather than writing it out. In the pheno, the lambda expression $\lambda f.f(every \bullet head)$ is applied to the argument $\lambda t.x \bullet peck \bullet t$. The first has type $(str \rightarrow str) \rightarrow str$, and the second type $str \rightarrow str$. Replacing f with $\lambda t.x \bullet peck \bullet t$, we get $(\lambda t.x \bullet peck \bullet t)(every \bullet head)$. This results in further reduction, where t is replaced with $every \bullet head$, like so: $x \bullet peck \bullet every \bullet head$. The reductions can be confusing at first, if not broken down.

$$\frac{\vdash \lambda \text{f.f(every} \bullet \text{head)}; (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \mathit{every\ head} \qquad \text{(XVII)}}{\text{x}; \mathsf{NP}; y \vdash \text{x} \bullet \text{pecks} \bullet \text{every} \bullet \text{head}; \mathsf{S}; \mathit{every\ head}(\mathit{peck\ y})} \text{MP} \tag{XVIII}$$

Notably, at this point there is an S by itself, which doesn't look much like a sentence. This is fine, because we are not done. There is an undischarged hypothesis to the left of the turnstile. In (XIX), the hypothetical proof rule is applied. The NP, which was put on loan as a hypothesis in the course of building up the verbal phrase, is now due to be paid.

$$\frac{\text{x}; \mathsf{NP}; y \vdash \lambda\text{t.x} \bullet \text{peck} \bullet \text{every} \bullet \text{head}; \mathsf{S}; \textit{every head}(\textit{peck } y)}{\vdash \lambda\text{x.x} \bullet \text{pecks} \bullet \text{every} \bullet \text{head}; \mathsf{NP} \multimap \mathsf{S}; \lambda y.\textit{every head}(\textit{peck } y)}\ \text{HP} \qquad \text{(XIX)}$$

To finish the proof of the case where one bird is doing all the pecking, (XIX) must be combined with (XV), which looks a lot like the combination in (XVIII).

$$\frac{\vdash \lambda\text{f.f(some} \bullet \text{bird)}; (\mathsf{NP} \multimap \mathsf{S}) \multimap \mathsf{S}; \textit{some bird} \qquad \text{(XIX)}}{\vdash \text{some} \bullet \text{bird} \bullet \text{pecks} \bullet \text{every} \bullet \text{head}; \mathsf{S}; \textit{some bird}(\lambda y.\textit{every head}(\textit{peck } y))}\ \text{MP} \qquad \text{(XX)}$$

In order to get the second reading, everything is the same up to (XVII), when a trace was introduced into subject position. This time, instead of continuing as before, a second trace is placed in object position.

$$\frac{\text{(XVII)} \qquad \text{x}; \mathsf{NP}; y \vdash \text{x}; \mathsf{NP}; y}{\text{x}; \mathsf{NP}; y\text{w}; \mathsf{NP}; z \vdash \text{x} \bullet \text{peck} \bullet \text{w}; \mathsf{S}; \textit{peck } y\ z}\ \text{MP} \qquad \text{(XXI)}$$

This allows us to bring the subject position back out using hypothetical proof. Moving forward, trading subject for object in the following proof steps results in a reversal of the scope relationship.

$$\frac{\text{x}; \mathsf{NP}; y\text{w}; \mathsf{NP}; z \vdash \text{x} \bullet \text{peck} \bullet \text{w}; \mathsf{S}; \textit{peck } y\ z}{\text{w}; \mathsf{NP}; z \vdash \lambda\text{x.x} \bullet \text{peck} \bullet \text{w}; \mathsf{NP} \multimap \mathsf{S}; \lambda y.\textit{peck } y\ z}\ \text{HP} \qquad \text{(XXII)}$$

From this point on, it is redundant to provide the rest of the proof, since it follows the same pattern. All that was needed to get the different scope was to change the location of the trace.

At this point the reader should be able to see that LCG is both powerful, with some mechanisms that are quite unfamiliar to phrase structure approaches, but also highly constrained. Everything needs to work in lockstep. There are no rules that operate only on one of the pheno, tecto, or semantics. This is not to say that idioms are impossible, but they don't fall out of syntactic combination. They require the lexicon to provide the lexical entries.

5.2.3.3 Co-occurrence

A question that is relevant to morphology, especially given the importance of agreement patterns to the concept of morphosyntactic features, is how LCG deals with co-occurrence patterns. First and foremost, there is no concept of full specification. Tectos represent underspecified categories. There is no need to supply a bunch of extra category information for morphology when co-occurrence resolution is an essential property of an adequate syntactic theory.

Above I provided some simplified lexical entries. Here I repeat them, introducing a category Neu, which is neutral for case. As stated earlier, WALK lexical entries have tectos that specify the case of their arguments. This isn't really born of some elaborate theory of case, but is more of a mnemonic. It differentiates things like *he* from things like *him*. Of course, in English, only the pronouns differ along these lines. Most noun phrases are neutral, as is the case for BILL.

$$\vdash \lambda s.s \bullet \text{walks}; \mathsf{Nom_{3s}} \multimap \mathsf{Fin}; \mathit{walk} \tag{XXIIIa}$$

$$\vdash \text{bill}; \mathsf{Neu}; \mathit{bill} \tag{XXIIIb}$$

Observing (XXIII) it may occur to the reader that Neu does not combine with $\mathsf{Nom_{3s}}$. They are different categories. In order to allow compatibility between them we need to alter the trace axiom schema. The new schema can be seen in (XXIV). Given the '$\leq_\mathsf{s}$', it introduces the idea that tecto terms are ordered.

$$\text{x}; \mathsf{B}; y \vdash \text{x}; \mathsf{B}'; z \text{ for } \mathsf{B} \leq_\mathsf{s} \mathsf{B}' \tag{XXIV}$$

What the new trace axiom schema allows are traces of the form $\text{x}; \mathsf{Neu}; z \vdash \text{x}; \mathsf{Nom_{3s}}; z$, so long as $\mathsf{Neu} \leq_\mathsf{s} \mathsf{Nom_{3s}}$. The order on tectos is analogous to the order on syntactic categories in pregroup grammar (Lambek, 1997; Lambek, 2001). McConville (2006) also uses an order on syntactic categories in Combinatory Categorial Grammar (CCG) (Steedman, 1996) as a replacement for features. At the very least, this keeps the focus on the relationships between categories, without the distraction of needing to invent individuated features.

Given the new trace axiom schema above, one can prove the syntactic validity of the following:

$$\cfrac{\cfrac{\cfrac{\vdash \lambda s.s \bullet \text{walks}; \mathsf{Nom_{3s}} \multimap \mathsf{Fin}; \mathit{walk} \qquad x; \mathsf{Neu}; z \vdash x; \mathsf{Nom_{3s}}; z}{x; \mathsf{Neu}; z \vdash x \bullet \text{walks}; \mathsf{Fin}; \mathit{walk}\ z}\ \text{MP}}{\vdash \lambda x.x \bullet \text{walks}; \mathsf{Neu} \multimap \mathsf{Fin}; \lambda z.\mathit{walk}\ z}\ \text{HP} \qquad \vdash \text{bill}; \mathsf{Neu}; b}{\vdash \text{bill} \bullet \text{walks}; \mathsf{Fin}; \mathit{walk}\ b}\ \text{MP} \qquad \text{(XXV)}$$

One may question whether such a mechanism is capable of handling all co-occurrence constraints cross-linguistically. It may need elaboration, but the claims inherent in the approach are defensible. First, syntactic co-occurrence restrictions are a purely syntactic issue. The "morpho" in morphosyntactic doesn't need to enter the picture. Second, there is no need to use the syntax as an information dump in order to handle co-occurrence. Overloading syntactic nodes with grammatically superfluous information needs to have another justification, since it clearly isn't necessary to model co-occurrence patterns. Note as well, that features are for modeling orthogonal taxonomies. The relationship between verbal arguments and noun phrases in English appears amenable to a hierarchical treatment.

In terms of the possibility of scaling the approach, I see no concrete technical issues. There are no abstract algebraic patterns that can be defined over feature values that are dependent on their nature as sets of feature values. For that reason, I see no reason to suppose that tecto terms cannot be effectively related in CiC according to the empirical needs of an analysis.

5.2.4 Benefits of LCG for morphology

At the beginning of the section, I stated that there is no conceptual reason why the morphological framework cannot interface with another syntactic theory. Despite this, LCG, or a theory very much like it, is the best choice.

First, it has good coverage of a wide variety of syntactic phenomena. It has numerous analyses of English. There is a large fragment of Serbo-Croatian (Mihaliček, 2012). There has also been significant work in K'iche' Mayan (Yasavul, 2013; Yasavul, 2017). This is a good amount of variety and a reasonably broad proving ground for a non-mainstream theory. In addition, LCG is a descendant of HPSG (Pollard and Sag, 1994), a well established framework. It inherits many of the analyses from that system. Despite this, it is not bound by the capabilities of HPSG, but is considered an improvement over a previous design. A an example of an area where it excels in comparison to its ancestor, unlike phrase structure grammars, it provides natural and elegant analyses of conjunction and ellipsis (Pollard and Worth, 2015; Worth, 2014; Worth, 2016; Kubota and R. Levine, 2017; Kubota and R. Levine, 2012). For a morphological theory, this means that one is able to interface with that body of previous work, without needing to imagine what a syntactic analysis would look like. Especially given a goal of my work, which is that morphology should interface with independently motivated syntactic analyses, it is good to have them available.

A second strength of LCG is that it is formally well defined. Some of the work cited above was done in a sister theory, Hybrid Type-Logical Grammar (HTLCG) (Kubota, 2010; Kubota, 2014; Kubota and R. Levine, 2017; Kubota and R. Levine, 2012; Kubota and R. D. Levine, 2020), which explores slightly different formal premises, in particular, a weakening of the distinction between pheno and tecto, which at the very least is considered beneficial in simplifying the pheno. It is justifiable to count the strengths of that theory among the strengths of LCG because it was formally proven that HTLCG may be embedded in LCG (Needle, 2022) – though in a more advanced version of LCG than I share here. The fact that the theory is well defined enough to allow for proofs that it may embed another theory, or may be interdefinable with another theory, is a testament to the quality of its formal foundations. If one were to give a formal treatment of the morphological domain of Distributed Morphology (DM), for instance, they would need to be able to provide a formal treatment of a Minimalism-like syntax. This has never been done over the course of decades of work on the theory. The closest to some success are the Minimalist Grammars (Collins and Stabler, 2016), an extension of CG to Minimalism. This is so different from actual practice in the field that it accommodates DM analyses no better than LCG would. This leave DM in a kind of pseudo-formal space with no mathematically describable foundation, where *mathematically describable* is the equivalent of saying *coherent* or *makes sense*. This may sound like harsh criticism, but it is important to call out the consequences of such issues. It does not preclude that analyses in DM can be formalized, but it is alarming, given the long history of the framework that no one has demonstrated that it is so. Imagine another discipline carrying on in such a manner, where there is no mathematical foundation for chemistry, or the reasoning used in psychological experiments. There would be a serious question of whether the claims meant anything at all. The reasoning and analyses in DM deserve better foundations so that such questions need not be raised. If one wishes to have a morphological theory that provably makes sense, they need a syntactic theory that provably makes sense.

A third strength, and the most important to my work, is that LCG has a highly advanced interface to semantics. The basic qualities of this have been described above. Beyond this, LCG has proven to be a productive vehicle for semantic research. This can be seen in the respectable body of work in interfacing with innovative theories of static semantics (Pollard, 2015; Plummer and Pollard, 2012), and dynamic semantics (Needle, 2024; Martin and Pollard, 2014; Martin, 2013). This has resulted in research in areas of semantics not limited to correlative comparitives (Smith, 2010), supplements and implicatures (Martin, 2013; Martin, 2016), and focus constructions within the context of questions and answers (Yasavul, 2013; Yasavul, 2017).

I take the stance that not all morphology is syntax driven, not even all inflection-like syntax. Despite the popularity of syntactocentric realizationalism, I feel that this is far from an extreme point of view. There are a lot of questions about how this actually works, especially from a realizational perspective. By interfacing with a syntactic theory that emphasizes a well defined interface to meaning, the morphological theory inherits that interface. Especially when augmented with an interface to lexical semantics, the system is able to model how non-compositional morphological patterns interact with compositional meaning. For this reason I can say, that even though one would be able to interface my morphological framework with another syntactic theory, unless one were to choose a syntactic system with very similar properties, the value in the overall theory would decrease. LCG, and its sister HTLCG, are the standard at this point.

5.3 The Type of Semantics

A key mechanism in my ability to define realization over lexical meanings is an abstraction over semantic types provided by Needle (2024). It allows for an embedding of all semantic types under a single type. This is useful for form-sign relations and sign paradigm entries, because it allows for sign paradigm entries to have a uniform type, despite the fact that semantic expression may have a variety of types, such as e, t, $e \to t$, $(e \to t) \to t$, etc. I will not provide all of the details of the abstraction, since it is largely irrelevant to a work in morphology. Needle proves a number of properties of the semantics, which are very important. For my purposes, I simply need to demonstrate the encoding of the types of semantics and discuss the means of converting between an encoding of a semantic type and an actual semantic type.

Needle's semantic theory is based on Agnositic Hyperintensional Semantics (AHS) (Plummer and Pollard, 2012; Pollard, 2015). Please see those works if one is interested in AHS. All one needs to know here is that instead of using t, as is common in natural language semantics, I will be using a type p. So an intransitive verb will have a type $e \rightarrow p$, rather than $e \rightarrow t$.

To provide an embedding we need to begin with an encoding of semantic types, something which largely describes the syntax of the types. This is easily done using an inductive type, which is called $stat\_term$, following Needle. The name is an abbreviation of "static semantic term," in contrast to a term from dynamic semantics.

As can be seen in (XXVI), a $stat\_term$ has three constructors. The first two, ent and prp, correspond to e and p. The last is for encoding the types of functions. The function constructor takes two $stat\_term$ arguments, which will correspond to the types of the function's input, and output.

When looking at (XXVI), think of it as if one were stating the syntax of semantic types. Anything with a type of $stat\_term$ corresponds to some well-formed semantic type.

$$\begin{aligned} &\vdash stat\_term : Set := \\ &\quad |ent \\ &\quad |prp \\ &\quad |func : stat\_term \rightarrow stat\_term \rightarrow stat\_term \end{aligned} \tag{XXVI}$$

Due to the fact that $stat\_term$ is in Set, it cannot serve as the type of other types. All of ent, prp, and $func$, are terms. There can be no x, such that $x : ent$.

The semantic types these terms correspond to are defined through a number of axioms, given in (XXVII). I haven't included all of the details of what needs to be defined for an actual semantic theory in (XXVII). I only want to give the reader an idea. One can see that e and p are defined as having type Set, which allows them to have inhabitants of their own, such as $truth$ and $falsity$, which correspond to boolean truth values. The unary 'not' is defined, as is 'and'. The items in (XXVII) can be extended to provide a complete theory of a natural language semantics, with types defined as actual types, unlike the terms used for encodings in $stat\_term$.

$$\begin{aligned}
&\vdash e : Set \\
&\vdash p : Set \\
&\vdash truth : p \\
&\vdash falsity : p \qquad \text{(XXVII)} \\
&\vdash p_not : p \rightarrow p \\
&\vdash p_and : p \rightarrow p \rightarrow p \\
&\ldots
\end{aligned}$$

There isn't any clear way to define a general type over the semantic types, in (XXVII). Without some means of embedding types, a sign paradigm entry, such as $\langle \lambda s.s \bullet \mathrm{walk}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, walk \rangle$, has a type $str \rightarrow str \times tecto \rightarrow tecto \times e \rightarrow p$. Another entry, such as $\langle \mathrm{I}, \mathsf{Speaker}, speaker \rangle$, for the first person pronoun, would have $string \times tecto \times e$. My goal is to provide a means of constructing sign paradigm entries and verifying their validity. This is difficult if the types of sign paradigm entries are highly variable. I take shortcuts with the pheno and tecto, which can be seen in Appendix A, but the semantics needs more attention, since it is key to proving a system that allows for semantic realization.

Ideally, I would be able to use something like the $stat\_term$ directly. Then, assuming that I've defined the pheno and tecto types, which I'll call ϕ and τ, I could define a sign paradigm entry as having the type $\phi \times \tau \times stat\_term$. Unfortunately, I need both types and terms in the semantics. What I'd like to do then, is to have a family of types that can be encoded as $stat_term$s. As a first step, following Needle, I provide a function that converts a $stat\_term$ to a semantic type. I can do this under the Set type, which may lead one to ask, why not define the sign paradigm entry type as $\phi \times \tau \times Set$? The reason for this is that it would allow *everything* in Set, not just the semantic types.

The function for converting a $stat\_term$ to a Set type, is called Sns. It is a recursive function that in the simple cases replaces ent with e and prp with p. For $func$, it recurses on the arguments a and b, placing them on either side of the function constructor.

$$\begin{aligned} &\vdash Sns : stat\_term \to Set := \\ &\quad \lambda s.case\ s\ of \\ &\qquad ent \Rightarrow e \\ &\qquad prp \Rightarrow p \\ &\qquad func\ a\ b \Rightarrow (Sns\ a) \to (Sns\ b) \end{aligned} \tag{XXVIII}$$

What exists at this point are actual semantic types, which may have inhabitants, allowing for the definition of semantic constants and functions. There is also a type that describes the syntax of these types, but cannot have inhabitants, $stat\_term$. Finally, there is a function that can take a $stat\_term$ and return a semantic type under Set. Unfortunately, the semantic types remain undifferentiated within Set. It would be good to have something like a predicate that allowed one to say that the types they were interested in were just that sub-portion of Set that corresponds to outputs of Sns.

The mechanism to do this is called a dependent sum, or Σ-type. There is a $\Sigma(A)(P : A \to Prop) : Type$, and a $\Sigma(A)(P : A \to Type) : Type$. The first argument is a type A. The second argument is a type P parameterized by a term of type A. An inhabitant is a pair (based on the definition of projection functions). The first element is a term x of type A and the second element is an inhabitant of Px. The definition of Σ-types is inductive in standard CiC.

$$\begin{aligned} &\vdash \Sigma(A : Type)(P : A \to Prop) : Type := \\ &\quad |exist\ \forall x : A.Px \to \Sigma\ A\ P \end{aligned} \tag{XXIX}$$

Note that the $exist$ clause defines the type of a function. The second argument must have type Px. If an inhabitant of Px is supplied, $\Sigma\ A\ P$ is inhabited.

The projection functions use $exist$ to recover the term of type A (π_1), and an object corresponding to the inhabitant of Px (π_2). I use pair notation, instead of $exist$.

An example of the $Prop$ version of Σ is a type for the pair of even natural numbers, and proofs that they are evenly divided by 2. The type is constructed, like so $\Sigma(\mathbb{N})(\lambda x.x\ rem\ 2 = 0)$, where rem is a remainder operator. Let's say we have an axiom $\vdash \alpha : 2\ rem\ 2 = 0$, then we can say $(2, \alpha) : \Sigma(\mathbb{N})(x\ rem\ 2 = 0)$. Using an axiom is a nice shortcut, but the usual case involves providing proof through computation, which is verbose. Note, that 7 cannot be in the first projection of an inhabitant of $\Sigma(\mathbb{N})(x\ rem\ 2 = 0)$ because there is no means of constructing a proof of $7\ rem\ 2 = 0$.

The inhabitant pair is within the universe $Type$, for reasons described in §5.1. In case the reader is wondering why it is $Type$ and not $Type_1$, according to convention, the subscript on $Type$ is left off.

Returning to our core concern, semantic types, the idea is to use Sns in place of $x\ rem\ 2 = 0$. The Sns function has a type $stat\_term \rightarrow Set$. The non-$Prop$ version of Σ is $\Sigma(A)(P : A \rightarrow Type)$, where $Type$ is one of $Type_n$ large enough for the function P. The type $\Sigma(stat\_term)(Sns)$ is the type of all pairs of $stat_term$s, x, and an inhabitant of $Sns\ x$. To make the type less verbose, I alias it as $sense$.

$$\vdash sense := \Sigma(stat\_term)(Sns) \tag{XXX}$$

This is the type for the semantics of a sign paradigm entry.

An example of providing an inhabitant of $sense$ follows. Let us say we have defined a predicate $\vdash\ walk : e \rightarrow p$. First, we can supply a $stat\_term$, $func\ ent\ prp$, which results in $\Sigma(func\ ent\ prp)(Sns\ func\ ent\ prp)$. Since $Sns\ func\ ent\ prp = e \rightarrow p$, and $walk : e \rightarrow p$, we may supply $walk$ as an inhabitant, resulting in a pair $(func\ ent\ prp, walk) : sense$. This has taken some explanation, but the result is that given any $stat\_term$, which when provided as an argument to Sns, provides a semantic type, which has an inhabitant, there is a single type, $sense$. The upshot, is that it is possible to define a type $\langle \phi, \tau, sense \rangle$. Using $stat_term$s, it is still possible to provide constraints on the semantic types of the input and output of such functions that are finer grained than any $sense$. I make heavy use of $sense$ in defining form-sign relations in the next chapter.

Chapter 6

WP and Lexical Realization

In this chapter, I provide the details of a formal framework for the development of realizational, Word and Paradigm (WP) theoretical analyses of morphological systems, as well as the outline of a grammatical fragment of Wao Terero, focusing on the lexical suffix system, which uses the framework. I begin with a statement of what I believe a grammatical fragment and a theoretical framework must provide, so that the reader understands my goals. I then provide a high level overview of the theory architecture, which I compare to previous work in morphological theory. This includes a fairly detailed comparison to Paradigm Linking Theory (PLT) (Stump, 2016), a version of Paradigm Function Morphology (PFM) (Stump, 2001) that models the morphology-syntax interface using multiple tiers of paradigms. I further develop the concepts and formal attributes of the framework within the practical example of a formal fragment of Wao Terero. In addition to describing the basic morphotactics of lexical suffixes, and the mechanism used to map meanings to forms, the interface to syntax is described. I also provide an analysis of some verbal inflection, so that it can be compared to the PFM-like analysis provided in §4.2.4.

A more fleshed out fragment is provided in Appendix A, which includes Coq source code. I expect the average reader to be more interested in the *how* and *why* of the model than the formulas and listed information needed for a full fragment.

6.1 What is a Formal Grammatical Fragment?

A formal grammatical fragment is a type of grammatical analysis. A grammatical analysis is a theory. An analysis is an explanation for patterns found in linguistic data, which takes the form of a predictive model. An example of a predictive model in the linguistic context would be a characteristic function for defining the set of grammatical phrases from the set of random permutations of words. Another kind of predictive model may predict the entailments of sentences by proposing the semantics of lexical entries for tokens within grammatical phrases. There are also stochastic models and models with other goals, but the point is that the model provides testable hypotheses by providing predictions. In this chapter the goal is to describe a model that predicts a subset of well-formed words, phrases resulting from the composition of those words, and entailments those sentences have.

In a scientific theory there is also an implied, or explicit set of methods that link a model to an explanation. For instance, when a model divides a set of potential phrases into two, the behavior of the model may be linked to the explanatory concept that one portion of the division is grammatical and the other ungrammatical. The linkage between the explanatory concept, and the math of the model is an empirical method. In the case of the two sets of sentences, and the concept of grammaticality, the most common method for providing a methodological linkage is some variation of the grammaticality judgment. Other methods may link other formal mechanisms of a model to other concepts of an explanation.

The explanation portion of a theory is the part that we're really trying to get at, in a sense. A sound explanation links the math and measurement to some notion of understanding, which is the goal. Explanations are the concepts and philosophy of the theory. In this work, notions such as realization, paradigms, free versus bound, and word-forms are all components of an explanation for the patterns found in the co-variation of form and meaning of words.

A formal fragment is a specific type of analysis. The fragment methodology seeks to emphasize the model. This emphasis is not expected to be at the expense of explanation and empirical methods, but a fragment should provide a valid piece of logic that stands on its own, in the way that a valid computer program exists even without receiving input, or having its output interpreted by an end user. This is not the default in formal linguistic theory. An analysis, in general, may be vague when it comes to the model. It is not uncommon that an analysis is largely a set of diagrams and explanatory prose, possibly with a formula or two, which communicate how a formal model *might* be. If one reads Halle and Marantz (1993), for instance, in the discussion of Potawatomi, there are trees and movement arrows and some example words that are purported to result from a model. Yet, how one might go through an actual derivation step by step is vague, and left as an exercise for the reader. Neither is the concept of *tree*, *movement*, or many of the other core concepts ever defined rigorously in any cited work. According to the practice of the authors, it is only the job of the theorist to provide the outline, or the idea, because the emphasis is not on the model, but the philosophical explanation. This is problematic, because if everyone reading the work is filling in the blanks in their own mind, the message of the work becomes dependent on the subjective experience of the reader. Even if some reader arrives at some valid model in their mind, it is not because it was ever in the original work. Worse, many readers may simply trust that the author has it all worked out, which rests on a problematic appeal to authority. In order to avoid embracing an anti-scientific practice, these issues should be avoided. One way to do this is to be diligent in defining a valid model, at least as a proof of concept. Whatever its scope, the model should be like a machine that will arrive at the same outcome for the same inputs no matter who is observing it. To do this, all the axioms and rules of the system should be included.

Due to the fact that all axioms and rules of the system must be included, or at least referenced when they come from the underlying logic or framework, there is always a lot of listing that must accompany a formal fragment. Some readers may not be accustomed to this. It is known that a lexical theory of language must list lexical entries and lexical rules. Even a formal fragment for a non-lexical theory, which does not have lexical entries, like Distributed Morphology (DM), would necessarily contain a long listing of language specific (or Universal Grammar (UG) inherited) rules and $\sqrt{}$root properties. There are certain limits to parsimony when even a fairly small number of words and grammatical constructions are considered. I expose more duct work than some readers may be accustomed to, though I pack quite a bit of it in Appendix A to allow readers to focus on key attributes of the fragment.

6.2 What is a Framework?

An analysis, or fragment, is a theory, but the analysis is almost always performed within the context of an existing framework, which is sometimes called a theory. It is not a mistake to call a framework a theory. The word theory is simply overloaded. The idea here is that there are more precise ways to refer to frameworks than to loosely call them theories. The formal rules that define Head Driven Phrase Structure Grammar (HPSG) or DM are not really theories in the same sense that an analysis is a theory. The basic rules correspond to key ideas and formal mechanisms, but without actual linguistic data they cannot predict or express anything. It is also the case that HPSG and DM allow for multiple analyses, multiple theories, of the same data. That is why it is better to call them frameworks. According to this view, a framework is the formal analog of a scientific research programme[1] (Lakatos, 1970). A research programme is a collection of a hard-core of concepts adopted by a family of theories that are not easily falsifiable in and of themselves. The concepts may not be falsifiable for a number of reasons. It may be the case that the programme deals with something that is difficult to measure, or define, such as UG. It may also be a design choice, such as viewing grammar as a system of composing signs, or a hierarchy of constructions. Frameworks are instantiations of a programme in whole or in part. Analyses within a framework that stand the test of time support the insights of the research programme that the framework expresses. Rarely is a research programme falsified when a particular analysis is falsified, or some data is shown to be incompatible.

Frameworks are more concrete than programmes, since they are specific formal proposals within a research programme. By being more concrete, they are more susceptible to falsification. Yet, frameworks are not often designed with falsification in mind. Additionally, specific formal proposals for HPSG or DM often change over time, indicating that the formalism has a programme-like fluidity so long as certain core tenets are adhered to. For instance, HPSG had a proposal for how to handle inflection that involved hierarchies that described generalities across signs (Pollard and Sag, 1987). Information-Based Morphology (IbM) is a relatively new formal subsystem for describing inflection, which is considered compatible with HPSG. One can conceivably use either sub-system and still be working within HPSG.

With this fluidity in mind, it is important to understand that the framework presented here is explicitly a work in progress. I began by working from PFM-like assumptions and have been gradually shaping the formalism toward a more explicit WP system. The programmes that I am working under are various, and include realizationalism, WP, categorial grammar (CG), and others. Many of the core aspects of the system are well-developed, which includes those aspects of the system that are essential to the goals of this thesis. The most important of these is providing a realizational treatment of concrete meanings. There are certain

[1] I use the British spelling as is common when discussing the philosophical concept.

aspects of the system that are more ad hoc. I note areas where I feel progress needs to be made when they are relevant. The difficulty in formal work is that one cannot wave their hands at incompleteness. Something must fill the gap, even if temporary.

6.3 Framework Architecture

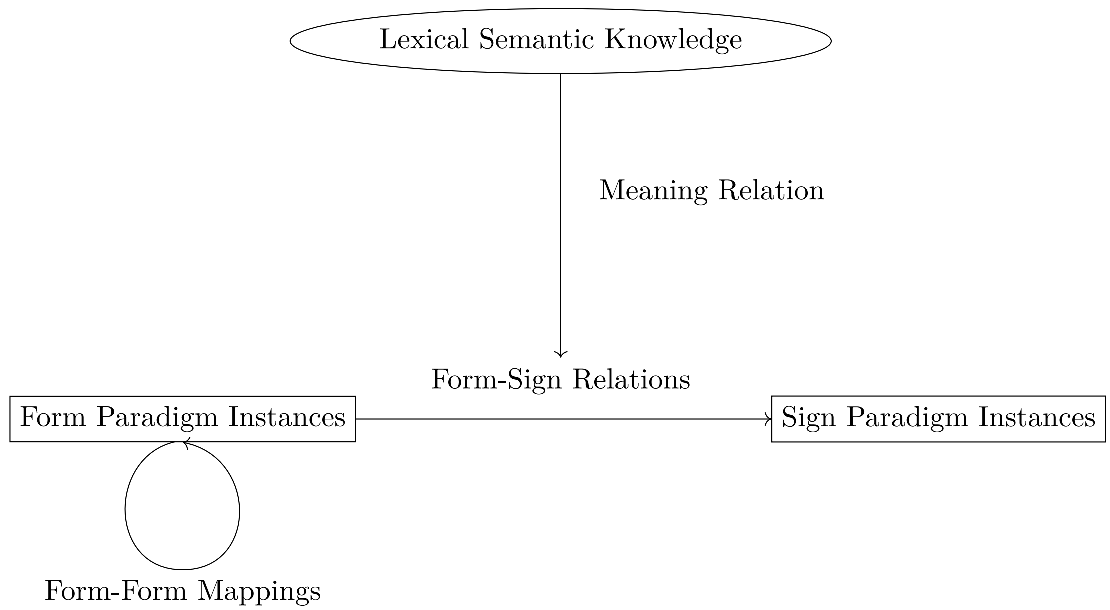


Figure 6.1: An overview of the architecture of the formal theory.

The basic architecture of the framework can be seen in Figure 6.1. A core attribute of the system is that it relates instances within two kinds of paradigms (actually taxonomies of paradigms). This dual paradigm approach is an important feature of the framework, which is consistent with separationism. All realizational theories are fundamentally separationist, but the term separationist is overloaded. Fundamentally, it describes a non-compositional relationship between morphological forms and meanings. One aspect of this is the lack of a functional relationship between forms and meanings, which manifests in morphs failing to compose in the same manner as phrasal units. I have also described another aspect of the concept, which is that *kinds* of morphological forms, such as roots versus affixes, do not rigidly correspond to kinds of meanings, specifically concrete versus abstract. Morphology can't be said to be a system of lexical roots that compose with functional affixes. The flexibility of √roots between morphological root and affix is an example of such separation. A third kind of separationism is that features, or categories, of the morphological level are distinct, to some degree, from the features and categories of the syntactic level. This is what multi-paradigm theories express.

6.3.1 Multi-paradigm Structure in Feature Theories

In feature-based theories, the structure of paradigms are defined by feature structures. This was described in detail in Chapter 4.1. The implication is that if one proposes separationism of morpho-grammatical categories, there will be morphological paradigms with one structure and syntactic paradigms with another. Such multi-paradigm theories are found in Stump (2016) and Sadler and Spencer (2001). In the case of Sadler and Spencer (2001), there are disjoint sets of features in both paradigms and rules that relate them. These are called s-features, which are required by phrases, and m-features, which regulate the morphological structure of words. An example of a pure m-feature is an inflection class feature. No features are shared in the system but it may be that the relation between features is trivial. As an example, one might imagine that in a language with no inflection classes that features PERSON and NUMBER would be divided into PERSON$_m$, NUMBER$_m$ and PERSON$_s$, NUMBER$_s$, where there is a bijective mapping between feature structures of the m and s-kinds. This can be seen in Figure 6.2. In a language with inflection classes, there may be an additional inflection class feature, which would not map to an s-feature. This would require that many m-feature structures map to an s-feature structure, as seen in Figure 6.3. The relation from m-feature values to s-feature values is a partial function in that case, and the inverse is a non-surjective relation. It is also possible that a particular language has no person or number feature at the morphological level, making them pure s-features. This would result in something like the inverse of Figure 6.3, where there would be many s-feature structures per m-feature structures. Of course, there may be both pure m and s-features in the same system.

Stump (2016) takes a slightly different approach from Sadler and Spencer (2001) in PLT. PLT proposes three paradigms. The first is called the *content* paradigm. It consists of a pair, $\langle L, \sigma \rangle$ of a lexeme, L, and morphosyntactic feature values, σ. This is the content that is expected to be at syntactic nodes, consistent with the Stump (2001) version of PFM, and much of the discussion in §4.1. The features σ in the content paradigm are the equivalent of the s-features in Sadler and Spencer (2001).

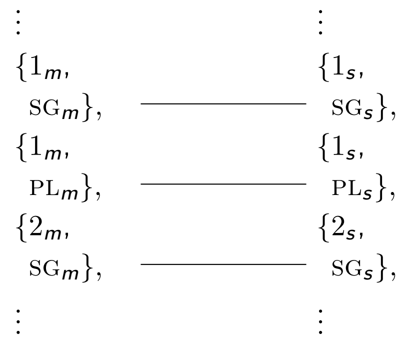


Figure 6.2: The bijective feature mapping that is possible between m-features and s-features in Sadler and Spencer (2001).

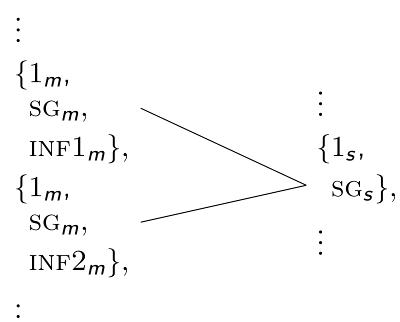


Figure 6.3: Inflection class m-features in Sadler and Spencer (2001) result in more than one morphological cell relation to a syntactic cell.

The second paradigm is the *form* paradigm. The PLT notion of form paradigm and my own are distinct. There is never a point in my system where whole word-forms are not represented within paradigm instances. The PLT form paradigm consists of pairs of stems, Z, and a possibly rewritten feature structure τ. Members of the content paradigm are mapped to the PLT form paradigm with the relation $Corr$. This is defined in terms of two sub-relations, $Stem$ and pm. The idea is that $Stem(L, \sigma)$ pairs $\langle L, \sigma \rangle$ with one or more stems Z. The relation $pm(\sigma)$ pairs σ and τ, where τ is a rewritten feature structure. I say *rewritten* because the notion is that, by convention, a τ intersects with its σ pair, and is not a completely independent feature structure – though nothing precludes the latter. There may be features in σ not found in τ, which for Sadler and Spencer (2001) would have been considered pure s-features. There may be also be features in τ not found in σ, which for Sadler and Spencer (2001) would have been pure m-features.

The form paradigm serves as the input to the *realized* paradigm. In that paradigm a word-form is paired with τ. The word-forms are computed using rule-blocks, which behave nearly the same as in the version of PFM I previously described.

There is an incomplete feature separation in PLT, since σ and τ will generally intersect, or even be equal. This is to say that in PLT, features are shared between paradigms. This is arguably less redundant, given the many trivial relations implied by Sadler and Spencer (2001). The system is like Sadler and Spencer (2001) in that some non-shared features exist. In comparison to Figure 6.3, one can more or less just erase the '$_m$' and '$_s$' subscripts to schematize the pm relation.

6.3.2 The Role of the Paradigms

I share with feature-based multi-tiered paradigm theories the idea that there are morphology-oriented paradigms, and syntax-oriented paradigms. I also share with these theories the idea that there is category relevant information that is not shared between the paradigms. Objects of the morphology-oriented paradigm and the syntax-oriented paradigm do not contain any sub-objects of the same type. The representation of objects in the two domains is entirely distinct. The taxonomy that delimits paradigms, the form classes, share a schema between both levels. A morphology-oriented object has its paradigm instance identified by a stem object. This is also the case for syntax-oriented objects. Despite this, the relationship to objects and the taxonomic schema is not the same, nor are the objects the same. For this reason, the morphology-oriented objects and

	SINGULAR	PLURAL
1	walk	walk
2	walk	walk
3	walks	walk

Table 6.1: A fully specified sub-paradigm of the English verb WALK.

syntax-oriented objects do not share a taxonomy of paradigms. See Appendix C for more information.

Whether paradigmatic separation in my framework is entirely distinct from the architectural assumptions of PLT requires some detailed comparison. The roles and justifications for the paradigms in my framework are often much different, and require a different kind of evidence for their justification. A comparison with structures, goals, and evidence required for PLT paradigms, and paradigms in my own system follows. This provides an explanation of my system through comparison with PLT, and seeks to partially answer the question of the extent to which the theories are convergent.

6.3.2.1 A conceptual mismatch

The notion of tiered paradigms in feature theories provides some additional room for morphological independence. Yet, the assumption is that the features seen in syntax are tightly coupled with morphology-level equivalents. For instance, consider the paradigms for WALK provided in §4.2.4.1. The realized paradigm for WALK in PLT would look roughly like the fully specified paradigm, repeated here in Table 6.1. The difference would be that there is something like a *weak* feature value, which would have been added by *pm*, to indicate the inflection class (Stump, 2016, p. 150).

The paradigm in Table 6.1 does not match well with either my concept of form paradigm or sign paradigm. I discuss some of its flaws in §4.2.4.1, §5.2, and elsewhere. To briefly recap, my conclusion is that it is based on sometimes spurious agreement information, and encodes contrasts seen only with the English copula, which is not a canonical verb. The shape is not syntax driven, since the syntactic theory that is generally assumed by PFM (and PLT), HPSG, does not require the full specification of the features used to define the paradigms on verbal nodes (Pollard and Sag, 1994) – at least, not in the Cartesian product sense of PFM. Therefore, there is evidence of a mismatch between theories. PLT predicts paradigms that have no conceptual basis in my approach. Neither is the realized, content, or form paradigm easily formulated, since without features, I cannot calculate the Cartesian product of features. Despite this, the systems are comparable. Below I go deeper into a comparison of the content, form, and realized paradigms with my own work.

6.3.2.2 The PLT content paradigm and my sign paradigm

In this section, I focus on the syntax oriented paradigms of the two frameworks, the PLT content paradigm and my sign paradigm. A paradigm instance of the syntax-oriented content paradigm would look like Table 6.1, but would have all of the cells filled with the lexeme WALK. My own notion of a syntax-oriented paradigm is structured according to different criteria. It is a natural paradigm based on measurement, and the terse description of the distribution of words within phrases. Instances consist of lexical entries of Linear Categorial Grammar (LCG), which are referred to as sign paradigm entries within their paradigmatic context.

An instance of the syntax-facing sign paradigm looks like (I) for intransitive WALK. The information found in sign paradigm entries was explained in §5.2.[2] The content of the paradigm instance follows directly from the analyses of syntacticians working in the LCG framework. Their motivation is independent of morphology-specific concerns. Each entry in (I) represents information that is necessary to a descriptive theory of English grammar. That is to say, that even if one were to chose a different formalism or theory, the meaning of the entries must be encoded, since the claims are measurable, utilizing standard syntactic diagnostics, such as replacement, constituency tests, grammaticality judgments, etc. That an intransitive English verb requires some analog of these six sign paradigm entries for an adequate description goes back, at least, to work in Generalized Phrase Structure Grammar (GPSG) (Gazdar et al., 1985).

[2]Note that Nom_{3s}, as described in §5.2, are unanalyzable symbols within the syntactic context. Additionally, remember that tecto terms are ordered, which is why they don't need case or agreement features.

$$\langle \lambda s.s \bullet \text{walk}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, \mathit{walk} \rangle \tag{Ia}$$

$$\langle \lambda s.s \bullet \text{walks}, \mathsf{Nom}_{3s} \multimap \mathsf{Fin}, \mathit{walk} \rangle \tag{Ib}$$

$$\langle \lambda s.s \bullet \text{walked}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{walk}) \rangle \tag{Ic}$$

$$\langle \text{walk}, \mathsf{PRO} \multimap \mathsf{Bse}, \mathit{walk} \rangle \tag{Id}$$

$$\langle \text{walking}, \mathsf{PRO} \multimap \mathsf{Prp}, \mathit{walk} \rangle \tag{Ie}$$

$$\langle \text{walked}, \mathsf{PRO} \multimap \mathsf{Psp}, \mathit{walk} \rangle \tag{If}$$

When I say that the paradigm is necessary, this may seem like a contradiction if PLT is able to encode inflectional analyses without reference to it. The truth is, that if PLT is assuming HPSG, a paradigm like (I) does exist in analyses of English, only in an HPSG formulation, as described in Pollard and Sag (1987). It is simply that PLT ignores it. I think that is a mistake.

The lexical entry patterns one sees in a syntactic analyses are one of the most natural paradigms one may observe in natural language. They arise by listing facts about the distributions of syntactic categories in a parsimonious fashion. Even theories without lexical entries have taxonomies of grammatical patterns that serve the same role. Equivalent patterns of valency and subcategorization are invoked in DM analyses. That a syntactic category manifests a finite number of forms, and a finite number of distributional patterns is fundamental to linguistic formal analysis. Across theories, syntacticians are able to agree on the facts that lexical entries encode in their debates on more exotic phenomena, such as the correct analyses of ellipsis. The extent to which this is possible is born out in works such as Kubota and R. Levine (2017), where coherent comparisons of analyses of ellipsis are made across CG (Hybrid Type-Logical Grammar (HTLCG) in particular), HPSG, and Minimalism. This is representative of convergence across analyses and diverse theories and formalisms.

There are some additional differences between sign paradigms and content paradigms go beyond their fundamental difference. First, the string support appears in sign paradigm instances, there is no “formless” state in the inter-paradigm derivation. Additionally, given the lack of lexeme object to identify paradigm members, the sign paradigm is not explicitly grouped in the same manner as a content paradigm. Its organi-

3 Singular	Elsewhere
walks	walk

Table 6.2: A counter proposal to full specification.

zation must be inferred from other information in the system, such as form classes and stems. This does not rule out suppletion, which I briefly discuss in Chapter 7.

6.3.2.3 The PLT Realizational Paradigm

The PLT realized paradigm is like my form paradigm in the sense that there are purely morphological features evident and full word-forms, rather than abstract stems or lexemes. There the similarities end.

The form paradigm of this work is a natural paradigm, based on form contrasts and shared stems. Unlike a realized paradigm, there is no syncretism evident. Syncretism is evident in sign paradigm instances, due to shared string supports. This is in contrast to the form paradigm. In practice, form paradigms are largely form contrastive. There is form ambiguity in form paradigms, a point I return to later, but it is irrelevant at the interface. To make an analogy between form paradigms and the realized paradigm, if my framework was translated to a feature theory, the form paradigm for intransitive verbs would look like the form-contrastive paradigm from §4.2.4.1, repeated in Table 6.2, using WALK as an example.

Syncretism is handled differently from PLT. Most syncretism falls out of the form-sign relations represented by the arrow between paradigms in Figure 6.1. There are two types of syncretism that exist in PLT that are not handled at the interface. First, unstipulated syncretism does not exist. Second, syncretism due to syntactic co-occurrence patterns is a relationship between syntactic categories, and is not morphological. It is handled by the LCG trace axiom schema, and the tecto term order. For instance, the fact that English *you*, unlike *I* and *me*, are identical in both the nominative and accusative is represented by the tecto term order. Otherwise, I expect syncretism to be an interface phenomenon that relates form-contrasts in the form paradigms to syntactic patterns encoded in the sign paradigms.

There is no paradigm like Table 6.2 in PLT. Instead, the concept that *walks* is specific to the third singular, and *walk* is the elsewhere case is captured by rules within rule blocks. If such a paradigm existed in the system, the rule blocks would be redundant. Therefore, the morphology facing, realized paradigm in PLT must look more like Table 6.1 if a claim is to be made that Pāṇinian rule ordering is explanatory.

$$\langle \mathsf{walk_m}, \text{walk} \rangle \tag{IIa}$$

$$\langle \mathsf{s_m} :: \mathsf{walk_m}, \text{walks} \rangle \tag{IIb}$$

$$\langle \mathsf{d_m} :: \mathsf{walk_m}, \text{walked} \rangle \tag{IIc}$$

$$\langle \mathsf{ing_m} :: \mathsf{walk_m}, \text{walking} \rangle \tag{IId}$$

An example form paradigm instance for WALK can be seen in (II). The example is more complete than Table 6.2, as it contains the present participle and past (participle). The first element of each pair in the paradigm instance is the category, which I call an M-cat. The second element can be thought of as a string, though it is actually a set of instructions for calculating a string. An important requirement of the form paradigm is that all form paradigm entries correspond to free word-forms. A form paradigm entry that contains a bound form, alone, is not compatible with the WP programme. The second projection of the form paradigm entry, the form, is mapped by a function η to the pheno string support of a sign paradigm entry. Therefore, the LCG string support is also non-bound content.

One can see from (II) that category names are form oriented. The categories could be named something else. I use a form-oriented naming convention to avoid an association with meaning or syntax, and because there is a close correspondence between morphological category and morphological form. One could replace $\langle \mathsf{d_m} :: \mathsf{walk_m}, \text{walked} \rangle$ with $\langle \mathsf{past_m} :: \mathsf{walk_m}, \text{walked} \rangle$ if they so wished.

The ‘m’ subscripts are part of the naming convention. They avoid naming conflicts. It may be useful to have other things named ‘walk’ or ‘d’ that are not part of M-cat names.

The symbol :: in the form paradigm entries joins portions of complex categories. Morphological systems accumulate category information through combination, in contrast to syntactic systems, which eliminate category information through combination. In the normal case, the addition of an affix results in the addition of category information, for instance $\langle \mathsf{walk_m}, \text{walk} \rangle$ goes to $\langle \mathsf{d_m} :: \mathsf{walk_m}, \text{walked} \rangle$, the opposite of the result of modus ponens in LCG, where an NP combined with $NP \multimap Fin$ results in Fin.

The M-cats primitives, such as $\mathsf{d_m}$, are not defined to be decomposable. They are not morphemes because they are not explicitly paired with either form or meaning. Neither are complex categories, such as

$\mathsf{d_m} :: \mathsf{walk_m}$, feature-like. One may think of them as a parse object, an abstract representation of the structure of a word-form. They are correlated with meaning and syntactic category only indirectly through a realizational calculation – a proof using form-sign relations. Structures like $\mathsf{d_m} :: \mathsf{walk_m}$ are categories because they determine the rules that apply to form paradigm entries. There are also abstract M-cats that do not occur in form paradigm entries, which provide abstractions over morphotactic and interface patterns.

In line with the concept of separationism expressed by multi-paradigm theories, the M-cats are purely morphological categories that do not occur in a phrasal context. They are different than m-features or τ. For instance, there is an entirely different system of categories to delineate form classes. M-cats are about the syntagmatic dimension of morphotactics. They also provide abstract categories that may be interpreted at the interface. Since allomorphy and inflection class-like phenomena is aligned with the paradigmatic dimension, it is encoded in a separate system.

This is a point of comparison worth expanding on. In PLT, pm allows for the addition of inflection class information that does not occur in syntax, but it does not provide an explanatory theory of inflection classes in the way that Network Morphology (NM) (Brown and Hippisley, 2012) does, where inheritance hierarchies describe patterns found across inflection classes. The generalizations expressed by such hierarchies provide valuable explanations that morphological inflection class feature value does not. For this reason, I provide a separate system of form classes, which captures similar relationships to NM, the hierarchies of HPSG (Pollard and Sag, 1987), as well as HPSG precursors (Flickinger, Pollard, and Wasow, 1985; Flickinger, 1987), and later work in construction grammar (Koenig and Jurafsky, 1994) and CG (McConville, 2006). This is detailed in §6.8.

If one is curious how the listing in (II) is grouped within a collection, a form paradigm instance is identified by a stem and a form class. There is no overt type, or data structure representation of (II), such as a set, list or attribute-value matrix. Paradigm instances may be reasoned over in complex ways (see §6.9), which do not require explicit data structure containers. That being said, data structures could be defined from framework primitives. In Appendix A, I provide a predicate in_{mp}, which indicates whether a form paradigm entry is a member of a paradigm instance. The predicate may serve as a characteristic function for both paradigm instances and paradigms.

6.3.2.4 Summary of Comparison to PLT

Through the comparison to PLT, the reader should have gained some knowledge of what the paradigms in my framework descriptively capture. The sign paradigm is a natural paradigm, which describes the finite distributional qualities of words. A content paradigm in PLT is a collection of inflectional categories, which are also called cells. I differ from PLT in the representation of those categories, and more fundamentally, in the standard of evidence for the categories. The form paradigm is also a natural paradigm. It describes the contrasting forms of a stem. There is no PLT equivalent, except that the realizational paradigm contains fully realized word-forms and purely morphological information.

6.4 Some Example Wao Tededo Paradigm Instances

In the interest of providing some familiar data for theory comparison, I have been using instances of a simple English language paradigm. Example (sub-)paradigm instances for Wao Terero are more relevant to this work.

6.4.1 A Wao Terero Sign paradigm Instance

The paradigms I present in this work require some speculation on the sign paradigm side. This is because a thorough syntactic analysis has not been performed. Peeke (1968) made important headway, but not in a manner that I believe can be easily translated from its tagmemic (K. L. Pike, 1958) origins. It is also the case that a comprehensive treatment of syntax is outside the scope of this work, even if I were able to supply it. For these reasons, I make a number of simplifications, as well as reasonable assumptions. All correspond to easily testable hypotheses.

One simplification is to use only an SOV pattern. SOV is a common Wao Terero word order. Peeke (1968) reported it to be default. Therefore, by focusing only on the SOV order, I will be producing a good foundation for later elaboration.

Another simplification is that agreement between subjects and verbs is grammatical. As stated in §3.2.1, there is reason to believe that mismatches are possible between verbal person inflection and the person of arguments, in the sense that the most specific inflection does not always match the person. For instance,

-*kã*, '3.H', being compatible with female subjects, despite the possibility of the more specific -*dã*, '3.F'. This does not necessarily mean that the agreement patterns are not grammatical. It may simply be that the grammatical constraint is not very specific, allowing variation. This could be modeled using the LCG trace axiom schema, and an order on tecto terms. The extent to which semantics comes into play is an open question. In this work, I'll assume a strict one-to-one correspondence between person inflection and the person of a pronominal subject, which does not allow any mismatch. I see it as a good starting point for later elaboration.

I also constrain the analysis to cases where items such as demonstratives and adjectives either occur alone, or directly preceding a nominal. There are other simplifications that don't merit specific mention, so long as the reader has the correct expectations concerning the completeness of the analysis.

The semantics will also be simple, and relatively opaque. For instance, I provide functions like future and past when tense needs to be represented, but I do not provide any information about what the functions do. This should not be surprising since such details would involve detailed analysis, and a considerable detour from the core topics of the thesis.

Given these simplifying assumptions, I provide some sample sign paradigm entries in (III) for intransitive GO, which in the intransitive use considered, has the meaning 'leave'. The '$\circ$' is function composition.

$$\langle \lambda s.s \bullet \mathrm{go}, \mathsf{Nom_t} \multimap \mathsf{Fin}, \mathit{leave} \rangle \tag{IIIa}$$

$$\langle \lambda s.s \bullet \mathrm{gobo}, \mathsf{Nom_{1.sg}} \multimap \mathsf{Fin}, \mathit{leave} \rangle \tag{IIIb}$$

$$\langle \lambda s.s \bullet \mathrm{gok\tilde{a}}, \mathsf{Nom_{3.h}} \multimap \mathsf{Fin}, \mathit{leave} \rangle \tag{IIIc}$$

$$\langle \lambda s.s \bullet \mathrm{gotabo}, \mathsf{Nom_{1.sg}} \multimap \mathsf{Fin}, \mathit{past} \circ \mathit{leave} \rangle \tag{IIId}$$

$$\langle \lambda s.s \bullet \mathrm{gokek\tilde{a}}, \mathsf{Nom_{3.h}} \multimap \mathsf{Fin}, \mathit{future} \circ \mathit{leave} \rangle \tag{IIIe}$$

The paradigm instance is quite large so only a few samples are provided. I'm leaving out a number of tenses of interest, and only providing finite entries for a few persons. For the tectos, I use a gloss-like name for the categories associated with various persons. One can find the glosses the tectos correspond to in Appendix E. The exception is $\mathsf{Nom_t}$, where t is a mnemonic for "thing". I use it in place of INANIM, for inanimate.

The type of the entries is $\phi \times \tau \times \mathit{sense}$, which is aliased struct_{sp}. The types of sign paradigm entry are described in §5.2 and §5.3. Not everything with the type is a valid sign paradigm entry. For this reason, a constructive predicate, $SE_{sp} : \mathit{struct}_{sp} \rightarrow \mathit{Prop}$, is defined. The predicate SE_{sp} is satisfied by valid struct_{sp}. It is defined inductively. Every means of proving SE_{sp} corresponds to a clause in its inductive definition. The clauses are referred to as constructors. One means of proving SE_{sp} is to use axiom-like constructors, such as $\mathit{bareFinGo} : SE_{sp} \langle \lambda s.s \bullet \text{go}, \mathsf{Nom_t} \multimap \mathsf{Fin}, \mathit{leave} \rangle$. The symbol $\mathit{bareFinGo}$ is an inhabitant, but can also be thought of as the name of the axiom-like constructor. The constructor states that $SE_{sp} \langle \lambda s.s \bullet \text{go}, \mathsf{Nom_t} \multimap \mathsf{Fin}, \mathit{leave} \rangle$ is proven. In general, a form-sign relation will fill an SE_{sp} clause, which takes a form paradigm entry as input and outputs a sign paradigm entry.

Sign paradigms instances do not exist in explicit data structures. There is a predication in_{sp}, which is used to determine whether a sign paradigm entry is a member of a paradigm instance, or paradigm. I do not use it in the analysis. One may find it in Appendix A.

As a point of comparison with a PFM-like paradigm instance, the information in a sign paradigm instance is more articulated. This has a number of advantages. Notably, for the verb in (III), the semantics does not specify anything concerning person. This is due to the assumption that Wao Terero verbs have purely grammatical agreement with subjects. The sign paradigm entry $\langle \lambda s.s \bullet \text{gobo}, \mathsf{Nom_{1.sg}} \multimap \mathsf{Fin}, \mathit{leave} \rangle$ has only the predicate leave, rather than something like $\lambda x.\mathit{leave}\, x \wedge \mathit{spearker}\, x$, $\mathit{leave} \circ \mathit{speaker}$, or otherwise. Where, in this example $\mathit{speaker}$ would be a predicate or function for speaker meanings. In other words, the semantics contributes nothing to the person meaning, not even a restriction. The person *meaning* is provided by normal semantic composition.

In a PFM-like paradigm, in order to get the proper realized form for something like *gobo*, there must be a feature structure that includes all of the person information. This structure does not differentiate between meaning and grammatical category. In §4.2.4, Table 4.10, repeated here as Table 6.3, one can see that person and number features are included. According to a theory like PFM these syntactic features are the inflectional meaning of the verb, as discussed in §4.2.3. Therefore, If person feature are properties of both the noun and the verb, the analysis is contrary to compositionality. There is nothing in the formal representation, nor the definition of the Hypothesis of Paradigm-Based Inflectional Semantics (Stump, 2001, p. 248) to differentiate

NUMBER	singular, plural, dual
PERSON	1, 2, 3
INCLUSIVE	inclusive, exclusive
SENTIENT	sentient, non-sentient
GENDER	feminine, neutral
MATERNAL	maternal, non-maternal

Table 6.3: The features to be used in the Wao Terero analysis.

a qualitative difference in the contribution of the features in different contexts.

Another issue is that due to the conflation of syntactic and semantic information, there is no trivial way to produce infelicitous, but grammatical, phrases in PFM that involve features, because meanings and syntactic categorizations are conflated.

There are other issues in the conflation of meaning and syntactic category in such a theory. In (III), the tectos for the past and future tenses are the same as for the present, per grammatical person. For instance $1.\mathsf{sg} \multimap \mathsf{Fin}$ is the tecto for both the past and present first person. I don't know with absolute certainty whether dependent clauses show grammatical concord with the tense of an independent clauses in some cases, but I have seen no evidence that they do. For that reason, there is no reason to propose a $\mathsf{Fin}_{\mathsf{pst}}$ as well as a $\mathsf{Fin}_{\mathsf{fut}}$. In (III), the tense is a property of the semantics. There must be *syntactic* evidence for a difference in syntactic category. This division may seem obvious, but in PFM-style paradigm tables, one *must* claim that something like tense is a syntactic distinction in order to allow for a tensed form to be licensed. There is no ability to appeal to a parallel semantics. Feature theories often conflate syntactic and semantic issues by pushing meaning into the syntax, due to a commitment to the notion that syntax must contain all of the information necessary for a phrase's interpretation and resulting phonological form.

Sign paradigms and their sign paradigm entries make useful distinctions that when not made, result in predictions about the nature of syntax and semantics that are poorly motivated.

6.4.2 An Example form paradigm Instance for Wao Terero

Some form paradigm entries are listed in (IV). Given what I have already said about (II), these should not look remarkable. A detail that I did not elaborate earlier is that the "string" portion of the entry is actually a list of processes. These result in a string when they are applied in succession. I present the second element of the pair as a string in some cases in order to make it clear what the entry represents. What I have been calling the string has the type $list\ process_{pr}$, where $process_{pr}$ is defined as $(string \rightarrow string) \rightarrow string \rightarrow string$. This means that "gobõda" in (IVj) is actually a list of three functions, da_{pr} :: b_{pr} :: go_{pr} :: nil. The nil, the empty list, is needed due to the formal definition of lists. I often leave off nil, since it is understood. The recurrence of :: for the process function list, which was previously seen for M-cats, is because :: is the general list constructor. Strings are produced from the process list by a function $apply_{pr}$, which is invoked at the interface. The goal is to have categories and processes accumulate in tandem as the result of form-form mappings. This eases the definition of lateral form-form mappings, which replace the heads of both the process and category lists. The head of a list is the left-most item, the last item appended. The $(string \rightarrow string) \rightarrow string \rightarrow string$ type states that a process takes a function from strings to strings and returns a function from strings to strings. The fact that processes are functions from string functions to string functions is a powerful and unique aspect of the system, which can be leveraged for elegant analyses of morphotactic phenomena. More is said below concerning processes and rules, when form-form mapping are described.

$$\langle \mathsf{go_m}, \text{go} \rangle \tag{IVa}$$

$$\langle \mathsf{bo_m} :: \mathsf{go_m}, \text{gobo} \rangle \tag{IVb}$$

$$\langle \mathsf{b\tilde{o}_m} :: \mathsf{go_m}, \text{gob\tilde{o}} \rangle \tag{IVc}$$

$$\langle \mathsf{bi_m} :: \mathsf{go_m}, \text{gobi} \rangle \tag{IVd}$$

$$\langle \mathsf{b\tilde{\imath}_m} :: \mathsf{go_m}, \text{gob\tilde{\imath}} \rangle \tag{IVe}$$

$$\langle \mathsf{da_m} :: \mathsf{go_m}, \text{goda} \rangle \tag{IVf}$$

$$\langle \mathsf{d\tilde{a}_m} :: \mathsf{go_m}, \text{god\tilde{a}} \rangle \tag{IVg}$$

$$\langle \mathsf{k\tilde{a}_m} :: \mathsf{go_m}, \text{gok\tilde{a}} \rangle \tag{IVh}$$

$$\langle \mathsf{da_m} :: \mathsf{b\tilde{o}_m} :: \mathsf{go_m}, \text{gob\tilde{o}da} \rangle \tag{IVi}$$

$$\langle \mathsf{di_m} :: \mathsf{b\tilde{o}_m} :: \mathsf{go_m}, \text{gob\tilde{o}di} \rangle \tag{IVj}$$

$$\langle \mathsf{da_m} :: \mathsf{b\tilde{\imath}_m} :: \mathsf{go_m}, \text{gob\tilde{\imath}da} \rangle \tag{IVk}$$

$$\langle \mathsf{di_m} :: \mathsf{b\tilde{\imath}_m} :: \mathsf{go_m}, \text{gob\tilde{\imath}di} \rangle \tag{IVl}$$

$$\langle \mathsf{di_m} :: \mathsf{d\tilde{a}_m} :: \mathsf{go_m}, \text{god\tilde{a}di} \rangle \tag{IVm}$$

The type of the form paradigm entry as a whole is $list\ m \times list\ process_{pr}$, a type aliased as $struct_{mp}$, where m is the type of M-cat primitives. Notably, some lists of m are M-cats, but some are not. All valid M-cats satisfy the predicate $M_m : list\ m \rightarrow Prop$. The proof is supplied by two mechanisms. The first is by axiom, such as stating $\vdash goM : M_m(go_m :: nil)$, where goM is the axiom inhabitant required in Calculus of Inductive Constructions (CiC). The second is that M-cat validity may be proven as the result of a form-form mapping rule, which are discussed later in this chapter.

Similar to lists of m in comparison to M-cats, the $struct_{mp}$ may or may not be a valid form paradigm entry. For this reason there is a constructive predicate $FE_{mp} : struct_{mp} \rightarrow Prop$, which functions in a similar manner to SE_{sp}, discussed above. Like SE_{sp}, FE_{mp} is defined inductively. It has axiom-like constructors, as well as form-form mapping constructors.

6.5 Example Relations

The types of the paradigm entries have been described. It is now time to describe how entries are related. In many ways, the system is within the tradition of formal grammar where the model consists of a characteristic function, which divides non-grammatical permutations of words from the grammatical. That is to say, that for a set of strings of words {“go be they”, “you”, “they be cats I”, “they are cats”,...}, the function divides the set and picks out only those that are grammatical. Since this is a constructive theory, it is more accurate to say that the theory provides the means to construct the set of valid objects. In the course of describing the paradigm entries, I introduced the constructive predicates, SE_{sp} and FE_{mp}. In some sense, providing proofs of these represents the goal of the model.

The syntactic theory LCG deals with grammatical phrases. It does so by combining lexical entries, which are sign paradigm entries. Yet, if the lexical entries are invalid, LCG cannot rule them out, and ungrammatical sentences may be predicted. To ensure that this does not happen, one could, conceivably, manually list all valid lexical entries of a language – though it is doubtful. Such a listing is not necessary. There are patterns in lexical entries. These reduce the amount of manual listing needed. For instance, there is uniformity in the distributions and forms of English weak intransitive verbs. Such patterns can be leveraged to provide greater lexical parsimony. One can view this as compressing the listing of lexical entries. The rules used to do so constitute a description of the morpho-lexical system. There is an established tradition of attempting to compress, or generate, lexical entries, with notable examples in unification grammars (Flickinger, Pollard, and Wasow, 1985; Flickinger, 1987; Pollard and Sag, 1987; Koenig and Jurafsky, 1994), but also in the CG tradition (McConville, 2006). The notion is that if word-like things are the tokens of syntax, describing the co-variation of form and grammatical-semantic meaning of word-like things, morphology, allows for the syntactic system to expand its token vocabulary, with less boilerplate listing. In so doing, one provides a scientific theory of something worthy of study in its own right, morphology, which may have attributes that are orthogonal to the optimal compression of lexical information. Such orthogonal attributes of a system include classic “mismatches”, such as syncretism and allomorphy.

The validation of lexical entries used in LCG is achieved through the SE_{sp} predicate. The form-sign relations provide the means of proving a valid entry satisfies SE_{sp}. The form-sign relations describe generalities across sign paradigm instances. The input to form-sign relations are valid form paradigm entries.

Validation of form paradigm entries is achieved through the FE_{mp} predicate.

Some form paradigm entries are validated using axiom-like constructors, which amounts to being listed. The notion of listing is often used in systems where the lexicon is the store of unpredictable information. The notion of least predictable is not entirely congruous with how axiom-like constructors may be used in the model. See §6.9 for examples of unique ways that listed information may be used in this framework.

Generally, morphological analyses assume roots or non-analyzable stems as a starting point for derivations. This is consistent with how I use axiom-like constructors. They tend to be the *simplest*, or *least analyzable* forms, which do not fall out of regular patterns.

Most form paradigm entries are validated by form-form mappings, which describe patterns across form paradigms. Given this, it is the form-form mappings and form-sign relations that describe the patterns that compress lexical information. They may also be considered the rules that *generate* lexical information.

6.5.1 Form-form mappings rules

There are currently two types of rule schema for form-form mappings, uncreatively named $rule1_{mp}$ and $rule2_{mp}$. This means that there are two types of form-form mappings. The first produces appending (not limited to affixing) form-form mappings rules. These result in the addition of category and process information to an existing form paradigm entry, for instance, $\langle \mathsf{go_m}, \mathrm{go_{pr}} \rangle$ to $\langle \mathsf{bo_m} :: \mathsf{go_m}, \mathrm{bo_{pr}} :: \mathrm{go_{pr}} \rangle$. The second produces lateral form-form mappings rules. These result in the dropping of the heads of category and process lists, with some new category and process appended to the tail, for instance $\langle \mathsf{bo_m} :: \mathsf{go_m}, \mathrm{bo_{pr}} :: \mathrm{go_{pr}} \rangle$ to $\langle \mathsf{bi_m} :: \mathsf{go_m}, \mathrm{bi_{pr}} :: \mathrm{go_{pr}} \rangle$. The reason that I place processes in lists is to allow for lateral rules, and other means of maintaining word-form to word-form relationships. It obviates the need to parse strings in an ad hoc manner, and simplifies process rules.

As mentioned previously, the process list is applied from the first element to the final element of the process list by a function $apply_{pr}$. The string function returned by the right-most element in the examples above, go_{pr}, is applied to an empty string first, providing the initial string element. Then, subsequent processes are applied. This is described in detail in §6.5.1.1.

Each rule schema is a function that returns a rule. Both have the same type, which can be seen in (V). It is complex so it will help to break it down. It also depends on some information that has not been covered, yet.

$$\forall(cat_m : list\ m)(\kappa : K)(new_m : list\ m)(new_{pr} : list\ process_{pr})$$
$$(\alpha : struct_{mp}).\pi_1\alpha \leq_m cat_m \to klass\ \pi_1\alpha \leq_k \kappa \to struct_{mp} \tag{V}$$

An important aspect of the type is that it describes a function. When defining form-form mappings, the analyst supplies an arguments to a function, a lambda term corresponding to (V). To define the rule, the analyst does not supply all of the arguments corresponding to the type. Only the arguments corresponding to:

$$\forall(cat_m : list\ m)(\kappa : K)(new_m : list\ m)(new_{pr} : list\ process_{pr})$$

When these arguments are provided, a rule is returned.

So called *currying* is assumed. This means that a multi-argument function is actually a series of single argument functions. For instance $addthree := \lambda\alpha\beta\gamma.\alpha + \beta + \gamma$, reduces such that $addthree\ 5$ returns the function $\lambda\beta\gamma.5 + \beta + \gamma$, etc. Given this, the function that is returned, when the schema is supplied the expected arguments is a form-form mapping function, which has a type as in (VI).

$$\forall(\alpha : struct_{mp}).\pi_1\alpha \leq_m cat_m \to klass\ \pi_1\alpha \leq_k \kappa \to struct_{mp} \tag{VI}$$

Remember that a $struct_{mp}$ is an alias for a form paradigm entry pair. The notation π_1 is the first projection function, which when given a pair, returns the first element, so $\pi_1(a, b) = a$. This means that $\pi_1\alpha$ is isolating the category information from the form paradigm entry. The $klass$ function takes form paradigm entry information and provides a form class, which is similar to an inflection class. Both M-cats and form paradigm entries have orders, relations, $\leq_m$ and $\leq_k$, respectively. All one needs to know about the orders at the moment is that they ensure that the input form paradigm entry has the correct category and class for the rule to apply. The output of the mapping function is a (possibly) new $struct_{mp}$, with some change to the category and process list. The check that input is a valid form paradigm entry, and the assertion that output is valid is provided through the inductive definition of FE_{mp}. The inductive definition of FE_{mp} is a way of formally stating, “These are the rules for producing valid form paradigm entries, and nothing else.”

One can read (VI) in technical English as: Given a form paradigm entry, which has an M-cat less than or equal to cat_m, and a form class less than or equal to κ, return a form paradigm entry.

The earlier portion of the type for a form-form mapping rule schema set up the conditions and described the new information that will be added to the form paradigm entry. This is isolated in (VII). It says that given some M-cat, named cat_m, a form class (which has type K), named κ, and new M-cat and process information, return a form-form mapping, where the rule only applies to a subset of form paradigm entries as restricted by cat_m and κ, as described above.

$$\forall(cat_m : list\ m)(\kappa : K)(new_m : list\ m)(new_{pr} : list\ process_{pr})... \tag{VII}$$

The type makes clear how the rule will be restricted. The form-form mapping term returned by the schema function is an expression that describes how the input form paradigm entry is altered. How the input is altered differs for the two schemata, where the content of $rule1_{mp}$ is in (VIII), $rule2_{mp}$ is in (IX). Both use a function $combine_{mp}$. The function works in a uniform manner for both schema types. The arguments to $combine_{mp}$ are a list of m to be added to the current M-cat, some portion of the current, input form paradigm entry M-cat, a new list of processes to be added, and some portion of the current, input processes. The only difference between the two types of rules is that the heads of the current input M-cat and process lists are dropped in (IX) using the $tail$ function, but not in (VIII). The $combine_{mp}$ function works by first checking to see if new_m intersects with the input M-cat information. If so, the input form paradigm entry is not altered. This means that successive applications of the same form-form mapping result in no change to the input form paradigm entry. If there is no intersection between new_m and $\pi_1\alpha$, then the new category and process information is appended to the old, and a new form paradigm entry is returned.

$$combine_{mp}\ new_m\ \pi_1\alpha\ new_{pr}\ \pi_2\alpha \tag{VIII}$$

$$combine_{mp}\ new_m\ (tail\ \pi_1\alpha)\ new_{pr}\ (tail\ \pi_2\alpha) \tag{IX}$$

An example of defining a form paradigm entry that would add first person inflectional information to a verb looks like (X). Remember that I omit nil on the lists. The $\mathsf{PERSONABLE_m}$ M-cat is an abstract M-cat. It is a super category for items that may be a base for person inflection. That the input item can take person marking, such as *-bo*, is indicated by the form class, person_k. Verbs and a number of other parts of speech are in this class. The naming may seem redundant, but there are distinct purposes. The form class person_k indicates that, according to class membership, there exist person marked form paradigm entries for the classed item. The M-cat $\mathsf{PERSONABLE}_\mathsf{m}$ specifies that the input form paradigm entry may serve as a stem for such affixation.

$$\mathit{rule}1_{mp}\ \mathsf{PERSONABLE}_\mathsf{m}\ \mathsf{person}_k\ \mathsf{bo}_\mathsf{m}\ \mathit{bo}_{pr} \tag{X}$$

It is the case that for $\langle \mathsf{go}_\mathsf{m}, \mathrm{go}_\mathrm{pr}\rangle$ that $\mathsf{go}_\mathsf{m} \leq_m \mathsf{PERSON}_\mathsf{m}$, and $\mathit{kass}\ \mathsf{go}_\mathsf{m} \leq_k \mathsf{person}_k$. This means that the rule defined by the schema can apply. In the analysis, below, the rule in (X) is named boMP. This means that one can provide boMP the argument $\langle \mathsf{go}_\mathsf{m}, \mathrm{go}_\mathrm{pr}\rangle$, and then the two proofs, first that $\mathsf{go}_\mathsf{m} \leq_m \mathsf{PERSONABLE}_\mathsf{m}$, and second that $\mathit{kass}\ \mathsf{go}_\mathsf{m} \leq_k \mathsf{person}_k$. The result of $\mathit{combine}_{mp}$ is $\langle \mathsf{bo}_\mathsf{m} :: \mathsf{go}_\mathsf{m}, \mathrm{bo}_\mathrm{pr} :: \mathrm{go}_\mathrm{pr}\rangle$. As a sub-clause of the inductive definition of FE_{mp}, one must prove that $\mathit{FE}_{mp}\ \langle \mathsf{go}_\mathsf{m}, \mathrm{go}_\mathrm{pr}\rangle$ before the rule may be applied.

All of these details are good to know. In general, I will use a simplified notation that does not reference the rule schemata. The rule schemata demonstrate that this is a truly formal system. Not just any rule is a valid rule. I find a natural deduction-like notation to be a little more readable. The definition of boMP can be seen in (XI). I use '$\diamond$' for the effect of $\mathit{combine}_{mp}$, alone. I use '$\triangleleft$' for the effect of applying tail before $\mathit{combine}_{mp}$. I reverse the order of how the list is written. This is because suffixing is common, and suffixes are to the right of stems in English writing conventions. I do not display the projection functions π_n. The klass function is omitted. In fact, I go somewhat further than (XI), and also omit $\mathit{FE}_{mp}\ \alpha$ and the FE_{mp} in the inference line. This can be seen in (XII). The notation is a bit loose, but so long as one knows that it is constrained by the underlying formalism, the simplicity is beneficial. There is enough information to map the notation in an unambiguous manner to a CiC formula.

$$\frac{\mathit{FE}_{mp}\ \alpha \qquad \alpha \leq_m \mathit{LSy}_m \qquad \alpha \leq_k \mathsf{person}_k}{\mathit{FE}_{mp}\ \langle \alpha \diamond \mathsf{bo}_\mathsf{m}, \alpha \diamond \mathrm{bo}_\mathrm{pr}\rangle}\ \text{boMP} \tag{XI}$$

$$\frac{\alpha \leq_m \mathit{LSy}_m \qquad \alpha \leq_k \mathsf{person}_k}{\langle \alpha \diamond \mathsf{bo}_\mathsf{m}, \alpha \diamond \mathrm{bo}_\mathrm{pr}\rangle}\ \text{boMP} \tag{XII}$$

6.5.1.1 A note on processes

Processes may seem mysterious at this point. I have said that a process type is $(\mathit{string} \rightarrow \mathit{string}) \rightarrow \mathit{string} \rightarrow \mathit{string}$. An example term of the type is an affixing process, as seen in (XIII) for the definition of da_{pr}. The '$+\!\!+$' is the string concatenation operator.

$$da_{pr} := \lambda p\ stem.p(stem +\!\!+ \text{"}da\text{"}) \tag{XIII}$$

Such processes are in a process list for all form paradigm entries. As stated above, there is an $apply_{pr}$ function. $apply_{pr}$ is a recursive function, which processes the list and outputs a string. It has a type of $list\ process_{pr} \rightarrow (\mathit{string} \rightarrow \mathit{string}) \rightarrow \mathit{string}$. The second argument, of type $\mathit{string} \rightarrow \mathit{string}$ is an accumulator. Accumulators are used within certain kinds of recursive functions, particularly *fold*s, where a variable is used to build up state after each recursive application of a function. The accumulator is initialized with id_{pr}, the identity function on strings. Given a list, such as $\mathrm{da_{pr}} :: \mathrm{b\tilde{o}_{pr}} :: \mathrm{Ke_{pr}} :: \mathrm{nil}$, where $\mathrm{Ke_{pr}}$ is the stem for 'to do', the $apply_{pr}$ function first applies $\mathrm{da_{pr}}$ to the accumulator id_{pr}. Given (XIII), this results in $\lambda stem.id_{pr}\ (stem +\!\!+ \text{"da"})$, which is the new accumulator value. Next, the similarly defined $\mathrm{b\tilde{o}_{pr}}$ is applied to the accumulator, resulting in $\lambda stem_2.(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))(stem_2 +\!\!+ \text{"bõ"})$. This continues for $\mathrm{Ke_{pr}}$, which, as an initial stem, is defined slightly differently. It ignores the $stem$ argument, as in $\lambda p\ stem_3.p\ \text{"ke"}$.

The end result, before nil is reached, is $\lambda stem_3.(\lambda stem_2.(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))(stem_2 +\!\!+ \text{"bõ"}))\text{"ke"}$. When nil is reached, the empty string 'ϵ' is provided as an argument to the accumulated functions. This reduces as follows:

$$\begin{aligned}
&(\lambda stem_3.(\lambda stem_2.(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))(stem_2 +\!\!+ \text{"bõ"}))\text{"ke"})\epsilon \\
&(\lambda stem_2.(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))(stem_2 +\!\!+ \text{"bõ"}))\text{"ke"} \\
&(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))\text{"kebõ"} \\
&id_{pr}\ \text{"kebõda"} \\
&\text{"kebõda"}
\end{aligned} \tag{XIV}$$

The reason for doing things this way is to have some structural parallelism between the M-cats and processes. This allows for the lateral rules, which are useful in systems where initial stems are commonly bound. Remember that form paradigm entries only correspond to free word-forms. In a language like Spanish, a verbal stem, like *habl-*, 'speak', cannot occur as a form paradigm entry. One must define an axiom for one of *hablo*, *habla*, or something similar. Lateral rules allow one to "trade" *o* for *a* in such cases, ensuring that there are no form paradigm entries that correspond to bound forms. The idea is to avoid error prone pattern matching on strings, to essentially work with a parsed object.

The fact that processes are functions to and from string functions also provides a number of useful properties. It results in a kind of lazy evaluation, which diminishes the notion of "before" and "after". This can be helpful for Wao Terero pronoun patterns. The first and second persons all have pronominal forms that end in a suffix *-tõ*, 'PRO', *bo-tõ*, *bi-tõ*, *bõ-di-tõ*, etc. The person marking is the same as on verbs and other items, but there is no stem. The identifiable morphological signal of the paradigm instance is *-tõ*, which must always occur as the final affix. None of *bo*, *bõ-di*, etc. can occur freely. Yet, the morphotactic system has no issue with this, while maintaining that form paradigm entries are free forms. To accomplish this, *-tõ* is treated like a stem. The definition of its process is $\lambda P\ stem_3.(P(stem_3)) +\!\!+ \text{"tõ"}$. This is in contrast to the process for *ke-*, $\lambda P\ stem_3.P\ \text{"ke"}$, which throws away the input stem and applies the input process to *ke-*. It is also in contrast to $\lambda P\ stem_2.P(stem_2 +\!\!+ \text{"bõ"})$, the affixation rule. What $\lambda P\ stem_3.(P(stem_3)) +\!\!+ \text{"tõ"}$ says is that *-tõ* should be added to the result of applying the input string-to-string function to the input string. Using this pattern, the string *bõditõ* can be represented as a processes list $di_{pr} :: b_{pr} :: t_{pr}$, which looks similar to a process list for a verb marked with the same person and number, e.g. $di_{pr} :: b_{pr} :: go_{pr}$

$$
\begin{aligned}
&(\lambda stem_3.((\lambda stem_2.(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))(stem_2 +\!\!+ \text{"bõ"}))stem_3) +\!\!+ \text{"tõ"})\epsilon \\
&(\lambda stem_2.(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))(stem_2 +\!\!+ \text{"bõ"}))\epsilon +\!\!+ \text{"tõ"} \\
&(\lambda stem_1.id_{pr}\ (stem_1 +\!\!+ \text{"da"}))\text{"bõ"} +\!\!+ \text{"to"} \\
&id_{pr}\ \text{"bõda"} +\!\!+ \text{"tõ"} \\
&\text{"bõdatõ"}
\end{aligned}
\tag{XV}
$$

The advantage of this is that the affix processes are identical, resulting in parsimony. There needs to be only one rule in the grammar for *-bõ*, whether for a verb, adjective, pronoun, etc. There is no need to posit some non-surfacing form like *bõda* in the grammar that exists prior to affixation by *-tõ*. There are no zero morphs, covert syntactic heads, or vacuous rules needed, despite a "null" root. The use of id_{pr} and ϵ, the empty string, are not associated with any grammatical information. As symbols they could be replaced by rules that eliminated them. The identity id_{pr} simply states: Begin composition with the head process. The ϵ states: Begin reduction with the first concrete string. They are fundamentally different than a phonologically null morpheme, or empty syntactic head, as there are no grammatical interpretations. Nor do they serve as phonologically null place holders for morphotactic categories. Such objects do not exist in the framework.

To elaborate on the last point, it is worth while to consider the phonologically null morphological information in templatic systems, such as IbM (Crysmann and Bonami, 2016), or PFM. One of the reasons that I did not adopt such a system is because both IbM and PFM only deal with inflection, so are non-starters for a system with broader goals. Yet, even within the core area of inflection, much of the success of templatic systems rely on the use of zero rules, which leave portions of the template empty. I do not understand what a zero rule corresponds to in nature. If such rules are necessary, I do not believe that the analyst is correctly describing the observable patterns of a morphotactic system. This is because they would be, by definition, adding something non-observable for no reason other than to protect the concept of template, which is not viable without such interventions. Of course, in a formal system there is always machinery that is used in the scaffolding in order to express a concept, which cannot be said to be natural. The $apply$ function, for instance, is not something I view as being natural. It is simply part of the computational machinery for expressing the relationship between a parse of a morphological form and the form itself. The difference between this and zero rules in theories such as IbM and PFM is that non-zero morphological processes are defined in the same manner as zeros, and interact with zeros in the system. For that reason, one would expect that the evidence for the zeros should rely on the same evidence as is used to justify the non-zeros. This is not the case, indicating a uniformity of status, without a uniformity of evidence.

Moving on to another aspect of processes, it is worth discussing their restrictions. Technically, the manipulation of strings by processes is restricted only by defining a functions of the correct type – in the sense of type theory. In principle this allows concatenation, deletion, replacement, and much more. This is likely too powerful. A sub-theory of processes should eventually provide restrictions, but I haven't adopted one, yet.

Formal theories of processes are found in Hoeksema and Janda (1988), Aksënova, Graf, and Moradi (2016), and Aboamer and Kracht (2019), which attempt to place limitations on string-to-string functions based on cross-linguistic evidence. I mention these works, in particular, because they are formally well-defined, and could easily be incorporated in my framework. Hoeksema and Janda (1988) defines a Sapir-like (Sapir, [1921] 2004) taxonomy of process types, which are formally defined, but constrained only by what has been empirically observed. This has limited power as an explanation, but avoids predicting impossible processes. Aksënova, Graf, and Moradi (2016) and Aboamer and Kracht (2019) identify morphological processes with classes of formal languages, which appear to have similar properties to those found in nature. Stating that there is a correlation between morphotactics and regular languages, for instance, could provide additional explanatory power if one could link regular languages to known psychological or biological phenomena. Until that is accomplished, there is only an imperfect correlation.

For the time being, I am content to have a powerful system, rather than add restrictions for restrictions sake. One can still evaluate an analysis on the basis of the naturalness or elegance of the process functions it relies on.

6.5.2 M-cat definition

I have described the basic type of an M-cat above. It is a list of m, which are named things like $\mathsf{PERSONABLE_m}$, $\mathsf{bo_m}$, or $\mathsf{k\tilde{a}_{2m}}$, all with a subscript 'm' to disambiguate them from other similarly named objects and relations of other types, such as a process bo_{pr}. Valid M-cats satisfy the constructive predicate $M_m : list\ m \rightarrow Prop$. This is done by axiom in some cases, but there is also a general rule that if $a : list\ m$ is in the first projection of $b : struct_{mp}$ and $FE_{mp}\ b$, then $M_m\ a$. That is to say that proven form paradigm entries have valid M-cats.

The type m is inductive. All this means is that there is a listing of inhabitants of m. For instance, $m : Set := \mathsf{LS_m}|\mathsf{bo_m}|\mathsf{k\tilde{a}_{1m}}|\mathsf{k\tilde{a}_{2m}}|...$, would be such a listing.

Not all M-cats are treated the same. Some number of them are singletons that are only defined by axiom. These are *abstract* M-cats. An example is $\mathsf{PERSONABLE_m} :: \mathsf{nil}$. In rules, these occur to the right of $\leq_m$, and are used to subcategorize *non-abstract* M-cats. The M-cat order is a partial order, which means that not every pair of M-cats are ordered with respect to another. See §5.1 for a reminder on some of the terminology for orders. The non-abstract M-cats are never on the right of the ordering relation $\leq_m$. They are always at the lowest tier. The characteristics of the abstract and non-abstract M-cats are by convention. One is free to experiment within the framework with different ways of arranging the order without extending the formal framework.

The relation $\leq_m$ is defined inductively, which establishes its basic properties. The inductive definition of $\leq_m$ also describes the order in concrete terms. A constructor exists for every pair of ordered elements. The constructors act as morphotactic descriptions. For instance, $\mathsf{NUMBERABLE_m}$ is an abstract M-cat for stems that may serve as input to form-form mappings that add dual or plural affixes. The M-cats that are ordered below $\mathsf{NUMBERABLE_m}$ are specified using the constructor in (XVI).

$$\mathit{numberable_m} : \forall\alpha.M_m\alpha \to \mathit{headIn}\,\alpha\,(\mathsf{b\tilde{o}_{2m}} :: \mathsf{b\tilde{\imath}_m}) \to \alpha \leq_m \mathsf{NUMBERABLE_m} \tag{XVI}$$

The list $\mathsf{b\tilde{o}_{2m}} :: \mathsf{b\tilde{\imath}_m}$ is not an M-cat, simply a list of m. The headIn predicate tests whether the head of α is in the list $\mathsf{b\tilde{o}_{2m}} :: \mathsf{b\tilde{\imath}_m}$. The definition in (XVI) says that if α is a valid M-cat, and the head of α is in the list provided, then α is ordered below $\mathsf{NUMBERABLE_m}$. Since the head of an M-cat is the last m added, the rule expresses that an item with a final suffix of *-bĩ* or (non-lexical suffix) *-bõ* may have a $\mathsf{NUMBERABLE_m}$ form-form mapping applied. Though, the form class is also relevant. This means that ordering rules often treat the input M-cat as a parse of the morphological form.

I invoke functions like headIn for the sake of grouping similar rules. One is free to individuate rules. Some may be uncomfortable with the lack of restriction implied by being able utilize any function one may chose in such definitions. I am restrained in practice. Below, I adopt notation that is equally expressive to my practice, but may prove more comfortable to linguists. It is a variation on the notation used in *Sound Pattern of English* (Chomsky and Halle, 1968). Using this I may express (XVI) as in (XVII). The notation communicates that the relation holds when the final M-cat is one of two options. All of the constructors that involve a non-abstract M-cat may be written in this way. In the future I may formally define a sub-language for defining rules over M-cats, but it was not a priority for this thesis.

$$\alpha \leq_m \mathsf{NUMBERABLE_m} /\ \{\mathsf{b\tilde{o}_{2m}}, \mathsf{b\tilde{\imath}_m}\}\ \# \tag{XVII}$$

As a final note on the grouping of $\mathsf{b\tilde{o}_{2m}}$ and $\mathsf{b\tilde{\imath}_m}$ in the constructor, a consequence is that an additional

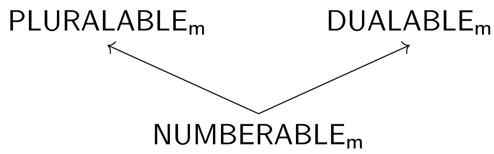


Figure 6.4: A relation defined using as a list of pairs represented as a directed graph diagram.

level of categorization takes shape. The M-cats $\mathsf{b\tilde{o}_{2m}}$ and $\mathsf{b\tilde{\imath}_{m}}$ correspond to a set of signals in the morphotactic space, which are distinct from others, such as $\mathsf{bo_m}$ and $\mathsf{bi_m}$. This is descriptively significant, whatever the formal theory, and nicely communicated by the constructor.

The abstract M-cats are ordered relative to one another differently. For these, I provide a listing of pairs. If the list is defined as $(\alpha, \beta) :: (\beta, \gamma)...$, then $\alpha \leq_m \beta$ and $\beta \leq_m \gamma$. For instance, form-form mappings do not directly reference $\mathsf{NUMBERABLE_m}$. Instead, there are two other abstract M-cats, $\mathsf{PLURALBLE_m}$ and $\mathsf{DUALABLE_m}$ both of which are the second element of a pair with $\mathsf{NUMBERABLE_m}$. There is one rule for adding the plural *-di* and another for the dual *-da*, which reference these ancestor M-cats. The plural version can be seen in (XVIII). This will apply to both items that are only able to take a plural affix, such as those that end in *-dã*, '3.F', as well as those that may also receive the dual affix, the $\mathsf{NUMBERABLE_m}$ items. The listing of pairs can be diagrammed as a directed graph, as seen in Figure 6.4. In order to save the reader from reading lists, I will provide such graphs in place of listings when possible.

$$\frac{\alpha \leq_m \mathsf{PLURALABLE_m} \qquad \alpha \leq_k \mathsf{person}_k}{\langle \alpha \diamond \mathsf{di_m}, \alpha \diamond \mathrm{di_{pr}} \rangle}\ \text{diMP} \qquad \text{(XVIII)}$$

Some of this may seem fairly abstract at this point, but concrete examples are provided below.

As a note on typographic conventions, I place abstract M-cats in all caps. Non-abstract M-cats that correspond to roots start with a capital letter. Otherwise, M-cats are lower case.

6.5.2.1 M-cat Identity

One may have noticed the numeric subscript on $\mathsf{b\tilde{o}_{2m}}$, above. This may naturally have lead to the question of whether there is a $\mathsf{b\tilde{o}_{1m}}$. This is the case. Both of the M-cats correspond to the same process $\mathrm{b\tilde{o}_{pr}}$. The item subscripted with '1' corresponds to the lexical suffix, 'LS.seed'. The other corresponds to an affix used for person marking. This leads to the question of when two affixes are the same or different.

Neither differing sounds nor meanings of an affix require distinct *m*s, necessarily. In general, different processes correspond to different *m*s. Different *m*s would generally be used in the case where a prefix and suffix are homophonous. Likewise, suffixes that sound different tend to distribute differently, and often require distinct *m*s. Despite this, there may be cases when the same *m* is used for different processes. For instance, suppletion such as *went* and *go* could be modeled using the same *m*, $\mathsf{go_m}$. There may also be cases of allomorphy or variable morphotactics (Crysmann and Bonami, 2016) that merit similar approaches. As this is a realizational theory, many meanings and syntactic categories can be associated with the same M-cat, and thus their component *m*s. This allows for syncretism and polysemy. Although the theory is not post-syntactic, the relationship between form and meaning remains similar.

For non-stems, the cases where the same process is associated with distinct *m*s are essentially identical to cases where PFM would place identical rules of exponence in separate rule blocks. That *-kã* should be considered two homophones, rather than a single suffix was discussed in §3.3.4.16. Given this status, the homophones are provided distinct *m*s, $\mathsf{kã_{1m}}$ and $\mathsf{kã_{2m}}$, even though they correspond to the same process.

For stems, I split *m*s according to form classes. I am more conservative with stem splitting than is the norm in PFM (Bonami and Stump, 2016). Yet, my reasoning is not always entirely distinct. In this work, my choices are somewhat heuristic. There are some stems such as *ĩ* that could be said to span multiple form classes. The stem is used for the copula, demonstratives, pronouns, accusative marking, and possibly other functions. Rather than consider *ĩ* to be the only member of some super-form class, which subsumes much of verbal, demonstrative, adjectival, and more exotic morphological classes, I treated it as a number of distinct stems. I leave it as an open question whether this is the best description. The determination is based on distributional criteria. The copula distributes like a verb, the demonstrative like a demonstrative, etc. I also treat homophonous nominal and verbal stems as distinct if there is no evidence of systematic derivation, or ambiguous categories. For instance, the stems *kẽ* are used for both 'manioc' words, where the stem is bound, and 'eat' or 'cut', where it is free. It is not impossible to provide an analysis that unifies such stems in the system, but I have not seen any motivation to do so within the context of this thesis.

6.5.2.1.1 Repetition Avoidance

There is more to the story of why multiple M-cats per process are used in the system.

The string *kẽ-kã*, with the stem *kẽ* for 'eat' or 'cut', is ambiguous. The *-kã* affix may be a lexical suffix, or person marking. The two annotations corresponding to *kẽ-kã* are 'eat-LS.body', 'It eats meat.', and 'eat-3.H', 'He eats.' Interpreted as a lexical suffix, it is allowed that there may also be a form *kẽ-kã-kã*, 'He eats flesh.' I wish to allow *kẽ-kã-kã*, 'eat-LS.body-3.H', but not a spurious *kẽ-kã-kã-kã* or *kẽ-kã-kã-kã-kã*. Keeping such repetition from occurring is not so difficult to solve in the verbal domain, and does not even require distinct ms. If only a single m is used for both the lexical suffix and person marking, an issue arises. By allowing *kẽ-kã-kã* for a verb, one would also allow *di·ka-ka*, a repetition of lexical suffixes for an inanimate noun. Inanimate nouns have no strict limits on the number of lexical suffixes, so long as they do not repeat. This is not exactly solved by using two ms, unless one can keep ms from repeating. The function $combine_{mp}$ specifies that the effect of applying a form-form mapping to a form paradigm entry more than once results in no change in M-cat or processes. It does this by ensuring that an m can only be appended to an M-cat once. Due to the definition of $combine_{mp}$, an affixation rule that adds kã_{1m} to an M-cat will not apply a second kã_{1m} in subsequent applications of the rule. The person marking is associated with a distinct m, kã_{2m}, which may apply to a stem with kã_{1m} because it passes the $combine_{mp}$ M-cat intersection check.

Note that none of this is relevant to processes such a reduplication, or even expressive repeating, which I consider to be distinct process types.

The strategy for avoiding repetition has two negative consequences. First, the multiple ms create situations where form paradigms are not strictly form-contrastive. For *kẽ-kã* there are two categories, kã_{2m} :: kẽ_{m} and kã_{1m} :: kẽ_{m}. Being strictly form-contrastive implies that a form never has more than one category. Given the current morphotactic scheme, being strictly form contrastive was viewed as difficult to enforce for other reasons, as well. For instance, there are languages with prefixing and suffixing, where a prefix may be added before or after a suffix. This may also result in distinct categories for the same item. For instance, un :: able_{m} :: lock_{m} versus able_{m} :: un :: lock_{m}, for *unlockable*. In my theory, the different interpretations of *unlockable* are not due to morphotactic scope, but realizational interpretations. For this reason, having two M-cats is redundant. I wish to maintain contrastive forms in form paradigm instances. There are ways to enforce this. Yet, by doing so, I would be erasing something that is true. As will be seen later, the theory celebrates the notion of multiple routes to a proof. It is not just an artifact of the theory that creates M-cat ambiguities. Linguists perceive more that one parse of linguistic forms such as *kẽkã* or *unlockable*. There truly is more than one parse, and they truly are ambiguous. I have no desire to erase a truth to preserve a truth. In order to preserve the notion that forms contrast with forms, but also that more than more parse may correspond to a form, I introduce a notion of form paradigm entry equivalence, which I discuss below.

Second, $\mathit{combine}_{mp}$ may be empirically inadequate. There are dialects of Quechua with two basic plural exponents *-kuna* and *-s*. These can occur by themselves, but may also co-occur in the following patterns *-s-kuna*, *kuna-s*, and *-s-kuna-s* (Muysken, 2002). There is no reason to say that the *-s*, which may sometimes occur twice, is not simply the same morph with the same category occurring more than once. Additionally, in some dialects of Quechuan suffixes like *-yuk*, 'owner', or *-chi*, CAUS, may occur more than once in a word (Muysken, 1981). There is little justification for supplying distinct ms in such cases. Therefore the $\mathit{combine}_{mp}$ function is too restrictive in some cases, and requires revision.

6.5.2.1.2 Form Equivalence

As noted above, I wish to both preserve the notion that forms contrast, while at the same time acknowledging that a form may correspond to multiple parses. To accomplish this, the framework has a concept of form paradigm entry equivalence, '$\equiv_{mp}$', which I invoke at the interface. It is defined such that for two form paradigm entry, α and β, $\alpha \equiv_{mp} \beta$ if $\mathit{apply}_{pr}\ \pi_2\alpha = \mathit{apply}_{pr}\ \pi_2\beta$. For example, $\langle \mathsf{k\tilde{a}}_{1m} :: \mathsf{K\tilde{e}}_{1m}, \mathrm{k\tilde{a}}_{\mathrm{pr}} :: \mathrm{k\tilde{e}}_{\mathrm{pr}}\rangle \equiv_{mp} \langle \mathsf{k\tilde{a}}_{2m} :: \mathsf{K\tilde{e}}_{1m}, \mathrm{k\tilde{a}}_{\mathrm{pr}} :: \mathrm{k\tilde{e}}_{\mathrm{pr}}\rangle$ because $\mathit{apply}_{pr}\ \mathrm{k\tilde{a}}_{\mathrm{pr}} :: \mathrm{k\tilde{e}}_{\mathrm{pr}} = \mathit{apply}_{pr}\ \mathrm{k\tilde{a}}_{\mathrm{pr}} :: \mathrm{k\tilde{e}}_{\mathrm{pr}}$. Remember, apply_{pr} requires string identity, not process list identity. Therefore, two form paradigm entries have an equivalence up to string identity. The effect is that when two things sound the same, they share properties. It doesn't erase distinctions It allows that English /ɹos/ can be identified with either the past tense of *rise*, or the flower until it is locked into syntactic context.

Given this identity, a form-sign relation may replace the M-cat of the input form paradigm entry, so long as the processes resolve to the same string. Any form-form mapping conditions on the input M-cats must still be met by the M-cat of a replacement form paradigm entry. This means that given *kẽ-kã* with the person marking *-kã*, one may successfully use it as input to a sign paradigm entry, which requires a lexical suffix interpretation of *-kã*. The sign paradigm entry allows for any M-cats of string identical form paradigm entries to be used. Additionally, imagine a language with many prefixes and suffixes, where one might be able to prove many parses that have no affect on interpretation. One can chose one as canonical, and avoid redundantly relating syntactic categories and meanings to every parse.

6.5.3 Form Class Definition

At the morphological level, the M-cat hierarchy is used to express the syntagmatic dimension of morphotactics. In the example above, it is used to say that stems ending in *-bõ* correspond to the category of stems that may have either dual or plural affixes following. The notion of form class is to cover the paradigmatic dimension, which in many languages would be concerned with allomorphy. Such classes do not only delineate conjugations from conjugations, they also delineate conjugations from declensions. That is to say, that they delineate lexical categories into groups that share certain syntagmatic patterns from those that exhibit other syntagmatic patterns. These often closely correspond to syntactic parts of speech, but rarely, if ever, to my knowledge, in a strict fashion. There may be several form classes per syntactic part of speech. There may be classes that are shared. For instance, in Wao Terero, both adjectives and verbs take person marking, but only verbs take tense and other inflection. They share a person_k, form class, but there is a distinct verb_k class, for verbs.

As an example, the form-form mapping in (XIX) was discussed above. That form-form mapping applies to both adjectives and verbs. In contrast, ignoring the M-cat information, for now, (XX), which affixes the 'DECL' suffix, is restricted only to verb-like items. In that rule I also use the M-cat, $\mathsf{CLOSEABLE}_\mathsf{m}$, which I discuss further below.

$$\frac{\alpha \leq_m \mathit{PERSONABLE}_m \qquad \alpha \leq_k \mathsf{person}_k}{\langle \alpha \diamond \mathsf{bo}_\mathsf{m}, \alpha \diamond \mathrm{bo}_{\mathrm{pr}} \rangle}\ \text{boMP} \tag{XIX}$$

$$\frac{\alpha \leq_m \mathit{CLOSEABLE}_m \qquad \alpha \leq_k \mathsf{verb}_k}{\langle \alpha \diamond \mathsf{pa}_{2\mathsf{m}}, \alpha \diamond \mathrm{pa}_{\mathrm{pr}} \rangle}\ \text{paMP} \tag{XX}$$

In order to make (XIX) apply to verb_k, as well as other form classes, the person_k form class must be ordered above verb_k, which is to say $\mathsf{verb}_k \leq_k \mathsf{person}_k$.

Utilizing hierarchies to describe the relationship between form classes is an established idea. Similar hierarchies are used in a multitude of theories (Flickinger, Pollard, and Wasow, 1985; Flickinger, 1987; Pollard and Sag, 1987; Koenig and Jurafsky, 1994; Riehemann, 1998; McConville, 2006; Brown and Hippisley, 2012; Guzmán Naranjo, 2019) for similar purposes.

An important aspect of the concept of M-cat and form class in this work is that they are not viewed as completely orthogonal. A form class spans over and delimits certain syntagmatic patterns. In the M-cat order, form classes are directly referenced to delineate certain patterns. For instance, an adjective stem may receive either lexical suffix or person marking, so only the root is either a $PERSONABLE_m$ or $LSABLE_m$ category, which allows such suffixation. For classes of verbs that allow a lexical suffix, one lexical suffix may precede person affixes. For that reason, both the root and the lexical suffixed stem are $\mathsf{PERSONABLE_m}$. The difference in ordering, which results in distinct syntagmatic patterns, is parameterized by form classes.

It is also the case that syntagmatic patterns signal form class membership. In some languages, a derivational morph may signal a denominal verb, and a subsequent difference in form class from the nominal stem. For this reason, the form class system explicitly appeals to M-cat information when determining the class associated with a form paradigm entry. It does so with the $klass$ function.

All form classes are of type $K : Set$, which is inductively defined. The inductive definition is essentially a listing of class names. They take no arguments, nor is the ordering expressed by the inductive definition. We've already seen examples, such as $\mathsf{verb}_k : K$, and $\mathsf{person}_k : K$. The typical subscripting strategy is used to disambiguate their names from similarly named items in the system. Their order, $\leq_k$, is defined in a much simpler fashion than, $\leq_m$. Remember from above that abstract M-cats have their order defined using a list of pairs. This is the case for all form classes.

Returning to the $klass$ function, $klass$ examines an M-cat structure, and matches particular patterns. For instance, if the pattern $ka_{1m} :: Di_m :: nil$ is encountered, the function returns inanim_k, for inanimate noun. The $klass$ function is descriptive of the criteria used to determine class membership. I will use a similar notation for $klass$ as I do for $\leq_m$ constructors, which looks like phonological rule notation.

Due to the fact that form classes hierarchies are not as exotic as M-cats, I wait until the analysis proper to flesh out the system.

6.5.4 Form-sign relation definition

At its simplest, a form-sign relation rule is a function that takes the pair provided by a form paradigm entry, as well as a meaning associated with its M-cat, and outputs a sign paradigm entry. Like the form-form mappings, the form-sign relations are defined in terms of rule schema. There is only one, called $rule_{sp}$. Like the form-form mappings, it can be viewed as two parts. There are the arguments used to define the rule and the arguments that the rule takes when applied.

$$\begin{aligned}
&\forall(cat_m : list\ m)\\
&(\kappa : K)\\
&(P : \phi \to \phi)\\
&(T : list\ m \to \tau)\\
&(s_1 s_2 : stat_term)\\
&(Q : Sns\ s_1 \to Sns\ s_2)
\end{aligned} \tag{XXI}$$

The type above corresponds to the arguments needed to define a rule. The first two arguments should look familiar, $cat_m : list\ m$ and $\kappa : K$. The applicability of a form-sign relation remains dependent on the input form paradigm entry meeting category and class constraints.

The following argument $P : \phi \to \phi$, is a function from a pheno to a pheno. The function $\eta \circ apply_{pr}$ is embedded in the definition of rule construction (see Appendix A), so no reference is made to a non-pheno string type.[3] This results in a pheno term that serves as a string support. The input form paradigm entry only supplies a string support. The mapping rule must output some useful concrete syntax, which embeds the string support. The string support is input to the function supplied as an argument, which will result in some pheno function or string. For instance, for an intransitive use of *go*, $P := \lambda st.t \bullet s$, may be the supplied function, which would be provided the argument η go when the rules is applied in order to produce $\lambda t.t \bullet \eta$ go.

[3]The type $\phi \to \phi$ may seem odd if someone is familiar with the pheno theory. The "pheno" in use here is not the pheno as introduced in Chapter 5, though it will have no impact on the discussion of the morphological theory. Pheno types are complex. A pheno term can have a type such as $s \to s \to s \to s$, or simply s. This is not compatible with reducing sign paradigm entries structures to a single type. For a complete treatment, the pheno needs to be embedded under a Σ-type, just as is done in the semantics. This was done for the semantics because the linkage between morphological forms and lexical semantics is vital to the core concerns of this work. The pheno is not important to the same extent. Dealing with the Σ-type for semantics involves a lot of overhead, not just for its definition, but in adding complexity to form-sign relations. I felt that adding this complexity to the system for the pheno was a distraction.

Rather than implement the pheno, I provide a stopgap pheno type. It is an inductive type ϕ, with a number of constructors. For this reason, η is a type constructor for ϕ, which explains the type provided in (XXI). The type of η is $string \to \phi$.

The stopgap implementation can be seen in Appendix A.

The argument $T : list\ m \rightarrow \tau$ is a function from an M-cat to a tecto term. The requirements of a syntactic category are often orthogonal. Ts are used to provide a general tecto type, which may express some case-like restrictions. This type is then parameterized by the input $list\ m$, which allows for basic co-occurrence information to be added. T can also be used for other types of tecto manipulations. It may also ignore its input, such as $\lambda t.\mathsf{N}$. See Appendix B for other examples of how it may be used. The results of T are opaque to the syntax, which uses tecto term orders to describe relationships between tectos.

The way that Ts are used is somewhat the opposite of conventional feature theories, which utilize concepts such as person and number within the morphology. They inject a portion of a parse into the syntax. For instance, given a T function, $\lambda t.\mathsf{Nom\ t} \multimap \mathsf{Fin}$, and an M-cat, $\mathsf{di_m} :: \mathsf{b\tilde{\imath}_m} :: \mathsf{go_m}$, for *gobĩdi* 'go-2-PL', 'You all leave.', I provide a function $personFilter$, which strips out all non-person related ms, and expands the T definition to be the composition of $\lambda t.Nom\ t \multimap Fin$ and $personFilter$, $(\lambda t.\mathsf{Nom\ t} \multimap \mathsf{Fin}) \circ personFilter$. When $\mathsf{di_m} :: \mathsf{b\tilde{\imath}_m} :: \mathsf{go_m}$ is provided as an argument, it results in $\mathsf{Nom}\ (\mathsf{di_m} :: \mathsf{b\tilde{\imath}_m}) \multimap \mathsf{Fin}$. Likewise, a noun phrase may have a similar type to $\mathsf{Nom}\ (\mathsf{di_m} :: \mathsf{b\tilde{\imath}_m})$, due to the form of the adjective, demonstrative, etc. To emphasize the opaque nature of the category annotation, and to provide readers with something a bit more intuitive to look at, I alias the annotations, such that the result of T, above, is presented as $\mathsf{Nom_{2.pl}}$.

$$(s_1 s_2 : stat\_term) \tag{XXII}$$

$$(Q : Sns\ s_1 \rightarrow Sns\ s_2) \tag{XXIII}$$

The rest of the type deals with the semantics. I've repeated that portion above. Two $stat_term$s are supplied. The first is expected to correspond to a lexical semantic expression that is associated with the category of the form paradigm entry. The second corresponds to the type of the semantic expression used in syntactic combination. The idea is that some conventional inflectional meanings may be associated with the output sign paradigm entry. These meanings will correspond to what people usually mean by grammatical meanings. For instance, if a grammatical category is strongly correlated with semantic past tense, it may make sense to provide a function $Q := \lambda(R : e \rightarrow p)(x : e).Past(Rx)$, or something similar. Then s_1 would be the $stat\_term$, $s_1 := func\ ent\ prp$, and s_2 would be $s_2 := func\ ent\ prp$, as well. The expectation would be that some intransitive verb meaning would be passed to the rule. It is therefore necessary to provide a function with the right type to *convert* that meaning to past tense. The final argument $Q : Sns\ s_1 \rightarrow Sns\ s_2$ could be $Q := \lambda(P : e \rightarrow p)(x : e).Past(Px)$, which accomplishes that need.

The above serves to define a form-sign relation rule. A rule once defined will require specific arguments in order to produce a sign paradigm entry. A form-sign relation is then the function below, with some variables provided from above.

$$\lambda(mp_1 : struct_{mp})(mp_2 : struct_{mp}) \tag{XXIV}$$

$$(\beta : Sns\ s_1) \tag{XXV}$$

$$(proof_{ev} : mp_1 \equiv_{mp} mp_2) \tag{XXVI}$$

$$(proof_m : (\pi_1 mp_2) \leq_m cat_m) \tag{XXVII}$$

$$(proof_k : klass(\pi_1 mp_2) \leq_k \kappa) \tag{XXVIII}$$

$$(proof_s : meaning(\Sigma\ Sns\ s_1\ \beta)(\pi_1\ mp_2)). \tag{XXIX}$$

$$(P(\eta(apply_{pr}(\pi_2\ mp_2)\ id_{pr})), T\ \pi_1\ mp_1, \Sigma\ Sns\ s_2\ (Q\ \beta)) \tag{XXX}$$

A rule takes two form paradigm entries, which are verified to be FE_{mp}, indirectly, in the course of supplying proofs to the rule. Two entries are provided to allow for a proof of their equivalence, which is on line (XXVII). The second entry, mp_2 is then used throughout, and mp_1 is ignored. This allows one to provide proofs within the system of the consequences of equivalence. It also means that if there are two or more ambiguous form paradigm entries, that one can define the meaning rules according to the M-cat pattern of only one. For a reminder on the justification of form paradigm entry equivalence, see §6.5.2.1.2. I would expect that in the usual case, one would simply provide the same form paradigm entries twice, allowing for a proof of equivalence by reflexivity.

The rule also takes a meaning argument of type $Sns\ s_1$ (XXVI), which serves as input to the $meaning$ relation, and the function Q, from above.

The second and third proofs, $proof_m$ and $proof_k$ (XXVIII,XXIX), should be familiar from form-form mappings.

The last proof, $proof_s$ (XXX) involves the $meaning$ relation, which is the means of performing the lexical semantic portion of the realizational calculation. The $meaning$ relation is a non-functional relationship between meanings and M-cats. The type that the meaning must have is supplied by s_1. By also supplying β and the M-cat of the form paradigm entry, one attempts to provide an inhabitant of the $meaning$ relation. One can think of it as simply matching a valid meaning to a valid category. I discuss this key relation in detail below.

The final line is the template for the sign paradigm entry.

$$(P(\eta(apply_{pr}(\pi_2\ mp_2)\ id)), T\ \pi_1\ mp_2, \Sigma\ Sns\ s_2\ (Q\ \beta)) \qquad \text{(XXXI)}$$

The first portion corresponds to the pheno. At its center, $apply_{pr}$ is used to produce a string. This string is injected into the type ϕ using η. The function P, provided when defining the rule, takes the string as input, constructing an expression of the concrete syntax. The second portion corresponds to the tecto. The function T, discussed above, is provided the input M-cat. The final portion is of type $sense$, a Σ-type. The first member of the pair is s_2, though it is not explicitly provided, since the second member of the pair takes the first member as an argument, $Sns\ s_2$, leaving no ambiguity as to its type. An inhabitant of $Sns\ s_2$ is provided, which is Q applied to β.

Technically, this is not a sign paradigm entry, only a $struct_{sp}$. The inductive definition of SE_{sp}, defines such rules as constructors, which result in valid sign paradigm entry. For instance, a rule for adjectives is defined below:

$$adjmodSP: \quad \text{(XXXII)}$$

$$\forall mp_1\ mp_2\ \beta\ proof_{ev}\ proof_m\ proof_k\ proof_s. \quad \text{(XXXIII)}$$

$$SE_{sp}((rule_{sp}\ \mathsf{INF_m}\ \mathsf{adj}_k \quad \text{(XXXIV)}$$

$$(\lambda st, s \bullet t) \quad \text{(XXXV)}$$

$$(\lambda \mathsf{t.N\ t} \multimap \mathsf{Adj\ t}) \circ \mathsf{filterPerson} \quad \text{(XXXVI)}$$

$$(func(func\ ent\ prp)(func\ ent\ prp)) \quad \text{(XXXVII)}$$

$$(func(func\ ent\ prp)(func\ ent\ prp)) \quad \text{(XXXVIII)}$$

$$(\lambda x.x)) \quad \text{(XXXIX)}$$

$$mp_1\ mp_2\ \beta\ proof_{ev}\ proof_m\ proof_k\ proof_s) \quad \text{(XL)}$$

The first line is the constructor name. It is a mnemonic for adjectival modifier. The second line declares the arguments that the constructor takes. The third line indicates that the resulting structure is an inhabitant of the SE_{sp} predicate. It also provides the beginning to the $rule_{sp}$ application to its arguments. The first of these are the M-cat and class. This expresses that the input M-cat must be an $\mathsf{INF_m}$, which is a kind of catch all, since for adjectives every state, from being a singleton, root, to having any further inflectional information will result in a valid M-cat for this rule. The mnemonic for $\mathsf{INF_m}$ is *inflected*. Of course, all form paradigm entries correspond to some free form, which is why it is a bit of a catch all. For finite verbs, it is a bit more specific, in that $\mathsf{INF_m}$ corresponds to verbs that are sufficiently inflected to unambiguously serve as finite verbs, which involves an affix such as *-pa*, 'DECL'. The fourth line provides a function for constructing the pheno. It corresponds to P, above. The string for the adjectival element will replace s, indicating that the adjective precedes its argument, t. The fifth line provides the function that constructs the tecto. I do not use an English-like $\mathsf{N} \multimap \mathsf{N}$ tecto. I have seen no evidence that more than one adjective may modify a single nominal, so there is an Adj category to block further adjectival modification. This does not eliminate the possibility that some other modifier can take an Adj argument, but adjective modifiers are defined to take N arguments, not Adj. It is important to remember that Adj is an N-like category. The naming is not intended to imply that it is an independent adjective. Notably, the t will parameterize the categories, to ensure that co-occurrence constraints are both met when applying the adjective and passed on in the resulting adjectival category, so that any further modification must follow suit. Following this, there are two lines specifying the semantic types of β and the resulting semantics of the sign paradigm entry. The obvious Q value, given the identity of the two $stat_term$s, is the identity function, given here as $\lambda x.x$. Finally, the arguments are provided to the rule, on the final line.

An example application will look like (XLI). In place of proofs, I simply supply the types. I also shorten the type of the $meaning$ relation since the Σ-type is verbose.

$$adjmodSP\ \langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathrm{y\tilde{e}d\tilde{e}_{pr}} \rangle\ \langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathrm{y\tilde{e}d\tilde{e}_{pr}} \rangle\ \ big \tag{XLI}$$

$$\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathrm{y\tilde{e}d\tilde{e}_{pr}} \rangle \equiv_{mp} \langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathrm{y\tilde{e}d\tilde{e}_{pr}} \rangle \tag{XLII}$$

$$\mathsf{Y\tilde{e}d\tilde{e}_m} \leq_m \mathsf{INF_m} \tag{XLIII}$$

$$klass\ Yd_m \leq_k \mathsf{adj}_k \tag{XLIV}$$

$$meaning(big\ \mathsf{Y\tilde{e}d\tilde{e}_m}) \tag{XLV}$$

The application of the constructor results in the sign paradigm entry seen in (XLVI). Note that the type for big is $(e \to p) \to e \to p$, as in English.

$$\vdash \lambda \mathrm{t.y\tilde{e}d\tilde{e}} \bullet \mathrm{t}; \mathsf{N_t} \multimap \mathsf{Adj_t}; big \tag{XLVI}$$

The '$_t$' is the alias for nil, which I consider a mnemonic for *thing*. Given another sign paradigm entry, such as (XLVII).

$$\vdash \mathrm{daa}; \mathsf{N_t}; thorn \tag{XLVII}$$

The syntax can combine the two via modus ponens, resulting in (XLVIII).

$$\vdash \mathrm{y\tilde{e}d\tilde{e}} \bullet \mathrm{daa}; \mathsf{Adj_t}; big\ thorn \tag{XLVIII}$$

As with the form-form mappings, it is also possible to provide a simplified notation, once the formal definition has been provided.

$$\frac{\alpha \leq_m \mathit{INF}_m \qquad \alpha \leq_k \mathsf{adj}_k \qquad \mathit{meaning}(\beta, \alpha)}{\langle (\lambda \mathrm{st}, \mathrm{s} \bullet \mathrm{t})\alpha, (\lambda \mathsf{t}.\mathsf{Nt} \multimap \mathsf{Adjt})\alpha, (\lambda x.x)\beta \rangle} \text{adjSP} \tag{XLIX}$$

In (XLIX), I leave out a lot of what can be filled in by a reader once they know how the rules work. I abuse the α variable to be either projection of the form paradigm entry. I don't include the equivalence information, since that will not be relevant in the normal case. I leave the apply_{pr} and η applications as understood. Since I utilize the same function to parameterized tectos throughout in this work, I leave out that detail as well. I don't write in the Σ-types, either. I leave what the rule expresses. For some inflected adjective, with a meaning β, construct a sign paradigm entry as so. Providing quasi-formal notation of this kind is harmless, so long as there is a verifiable formal analog.

6.5.5 The *meaning* relation

The meaning relation provides the lexical semantics to the form-sign relations. It is very simple, once one has established the *sense* type, discussed in §5.3. It is nothing more than a non-functional relation between morphological form categories and meanings.

$$\vdash \mathit{meaning} : \mathit{sense} \rightarrow \mathit{list}\ m \rightarrow \mathit{Prop} \tag{L}$$

The claim here is that there may be many meanings per form, and many forms per meaning. In other words, it is a realizational interface. What one puts on the other side of such an interface, in order to calculate forms, is largely up to them, so long as they are able to provide inhabitants of *sense*. I will be working with basic listings of meanings, and matching them to patterns in M-cats. As a theory of lexical semantics, it is extremely basic. A more ambitious theory of lexical semantics may be used in its place. Perhaps one would be interested in interfacing with a theory of compositional distributional semantics (Coecke, Sadrzadadeh, and Clark, 2010), or a theory in the Generative Lexicon family (Pustejovsky, 1991; Luo, 2012a; Asher, 2011; Chatzikyriakidis and Luo, 2017), with some minor tweaking, it should be possible to do so. I believe something more sophisticated is needed to capture the full range of gradience and analogy found in the lexical space. This is not to downplay the significance of *meaning* as it stands. To date, I am aware of no other formally well-defined system of lexical semantic realization, though its necessity has been noted elsewhere (Harley, 2014). By removing the distinction between $\surd$roots (or lexemes) and feature structures, realization, whether lexical or grammatical is simply a matter of matching an abstract parse object to meanings and categories at other levels of grammar.

The meaning relation is incredibly simple. It is inductively defined, with a clause per semantic type.

$$
\begin{aligned}
&\mathit{meaning} : \mathit{sense} \rightarrow \mathit{list}\ m \rightarrow \mathit{Prop} := \\
&\quad |e_s : \forall \mathit{cat}_m\ (\alpha : e).\mathit{meaning}_e\ \alpha\ \mathit{cat}_m \rightarrow \mathit{meaning}\ (\Sigma \mathit{Sns}\ \mathit{ent}\ \alpha)\ \mathit{cat}_m \\
&\quad |\mathit{in}_s : \forall \mathit{cat}_m\ (\alpha : e \rightarrow p).\mathit{meaning}_{e\rightarrow p}\ \alpha\ \mathit{cat}_m \rightarrow \\
&\qquad \mathit{meaning}\ (\Sigma \mathit{Sns}\ ((\mathit{func}\ \mathit{ent}\ \mathit{prp})\ \alpha)\ \mathit{cat}_m \\
&\quad |\mathit{tr}_s : \forall \mathit{cat}_m\ (\alpha : e \rightarrow e \rightarrow p).\mathit{meaning}_{e\rightarrow e\rightarrow p}\ \alpha\ \mathit{cat}_m \rightarrow \\
&\qquad \mathit{meaning}\ (\Sigma \mathit{Sns}\ (\mathit{func}\ \mathit{ent}\ (\mathit{func}\ \mathit{ent}\ \mathit{prp}))\ \alpha)\ \mathit{cat}_m \\
&\quad |\mathit{adj}_s : \forall \mathit{cat}_m\ (\alpha : (e \rightarrow p) \rightarrow e \rightarrow p).\mathit{meaning}_{(e\rightarrow p)\rightarrow e\rightarrow p}\ \alpha\ \mathit{cat}_m \rightarrow \\
&\qquad \mathit{meaning}(\Sigma \mathit{Sns}\ (\mathit{func}\ (\mathit{func}\ \mathit{ent}\ \mathit{prp})\ (\mathit{func}\ \mathit{ent}\ \mathit{prp}))\ \alpha)\ \mathit{cat}_m \\
&\quad |\mathit{indef}_s : \forall \mathit{cat}_m\ (\alpha : (e \rightarrow p) \rightarrow p).\mathit{meaning}_{(e\rightarrow p)\rightarrow p}\ \alpha\ \mathit{cat}_m \rightarrow \\
&\qquad \mathit{meaning}\ (\mathit{Sns}(\mathit{func}\ (\mathit{func}\ \mathit{ent}\ \mathit{prp})\ \mathit{prp})\ \alpha)\ \mathit{cat}_m \\
&\quad |\mathit{quant}_s : \forall \mathit{cat}_m\ (\alpha : (e \rightarrow p) \rightarrow (e \rightarrow p) \rightarrow p). \\
&\qquad \mathit{meaning}_{(e\rightarrow p)\rightarrow (e\rightarrow p)\rightarrow p}\ \alpha\ \mathit{cat}_m \rightarrow \\
&\qquad \mathit{meaning}(\Sigma \mathit{Sns}\ (\mathit{func}\ (\mathit{func}\ \mathit{ent}\ \mathit{prp})\ (\mathit{func}\ (\mathit{func}\ \mathit{ent}\ \mathit{prp})\ \mathit{prp}))\ \alpha)\ \mathit{cat}_m
\end{aligned}
\tag{LI}
$$

Each clause is a constructor, which for an M-cat and a semantic term α of the given type, constructs an instance of $\mathit{meaning}$. Each clause uses a separately defined $\mathit{meaning}_x$ relation, where x is a semantic type matching the type of α. The semantic term provided by the respective $\mathit{meaning}_x$ is used to construct $\mathit{meaning}$.

The behavior of adjectives is fairly uniform. This makes the adj_s clause an easy example, prior to the analyses of other parts of speech.

$$meaning_{(e\to p)\to e\to p} : (e \to p) \to e \to p) \to list\ m \to Prop :=$$
$$|big_s : \forall cat_m.cat_m = \text{Yẽdẽ} \to meaning_{(e\to p)\to e\to p}\ big\ cat_m$$
$$|tall_s : \forall cat_m.cat_m = \text{Yẽdẽ} \to meaning_{(e\to p)\to e\to p}\ tall\ cat_m$$
$$\ldots \tag{LII}$$
$$|adjls_s : \forall cat_m\ \alpha\ \beta.cat_m \leq_m \text{LS}_\text{m} \to LSmeaning\ alpha\ (hd\ cat_m) \to$$
$$meaning_{(e\to p)\to e\to p}\ \beta\ (tail\ cat_m) \to$$
$$meaning_{(e\to p)\to e\to p}\ (intersect\ \beta\ \alpha)\ cat_m$$

The first line of the definition of $meaning_{(e\to p)\to e\to p}$ declares the type. Following this, there are two clauses for polysemous root meanings of *yẽdẽ*. The semantic terms big and $tall$ were previously defined. This makes clear that listing is involved for unpredictable meanings. It also demonstrates that patterns within the M-cat are matched to determine rule applicability. When supplying the analysis of nouns, we'll see examples of larger patterns being matched, such as ka_{1m} :: Di_m, which will have a non-compositional meaning resolving to only the 'rock' lexical suffix meaning associated with ka_{1m}. Either big or $tall$ may be proven for the *yẽdẽ* meaning. It depends on the choices made during a proof. Either are available. For an item like DRAW, in English, one would expect a more expansive and diverse listing to express its polysemy.

The final clause demonstrates some of the power of the approach, which goes beyond simple listing. Adjectives have a fairly predictable relationship to the meanings of lexical suffixes. The classifier use enforces semantic concord with the adjective's argument. Additionally, any lexical suffix appears to be usable with an adjective, though the result may be infelicitous in some cases. Here we're concerned with well-formedness, rather than felicity. An adjective may have only one lexical suffix, which is enforced by form-form mappings. Therefore, for a clause that deals with adjectival meanings, when affixed by a lexical suffix, the first test is whether the item contains a lexical suffix, $cat_m \leq_m \text{LS}_\text{m}$. The inequality specifically requires that a lexical suffix is the final suffix. If this can be demonstrated, the $LSmeaning$ relation can be used. The relation pairs a meaning, here α, with an m, the head of the input M-cat. I give all lexical suffixes the basic type of $e \to p$, like nominals.

It is expected that the lexical suffix meaning will combine with an adjectival meaning. Therefore, the tail of the M-cat is provided to $meaning_{(e\to p)\to e\to p}$, which makes the relation recursive. This allows for a proof of an adjectival meaning β.

Given both α and β, the meanings need to be combined. At the time of writing, I treat them as having intersecting meanings in all cases. Determining whether this is correct requires further work. I do not necessarily feel that it is the best way to express their discourse sensitivity, but such questions must be left for future work. The *intersect* function is fairly simple $\lambda adj\ ls\ n\ x.(adj\ n\ x)\ and\ (ls\ x)\ and\ (n\ x)$. It takes and adjectival meaning and a lexical suffix meaning. It returns a function of an adjectival type that takes a nominal meaning and an argument of type e. Given the adjectival meaning *big*, and the lexical suffix meaning *rock*, the adjectival meaning is $\lambda n\ x.(big\ n\ x)\ and\ (rock\ x)\ and\ (n\ x)$. From this, the various polysemous meanings of both the root, and the suffix may be provided. I consider it fairly brute force, but it is a good starting point.

More examples of the use of the meaning relation are provided in the analysis below.

6.5.5.1 A note on felicity

The above discussion of the architecture of the framework describes most of the machinery needed for realization relative to compositional semantics and syntax – even if I have not said how, yet. It does not include any information about discourse. Discourse errors result in infelicity. Predicting infelicity judgments therefore requires a model of discourse. Providing such a model is outside the scope of this work. Though, some effort has been made to interface the system to a theory of dynamic semantics. Needle (2024) provides a compositional dynamic semantics that interfaces with LCG. In principle, this provides a basic interface for free, since the dynamic semantics neither needs to look into the morphology, nor does the morphology need to play any role in semantic composition. The idea would be that the morpho-lexicon simply needs to generate signs, which may combine in either a felicitous or infelicitous manner. Needle's dynamic semantics could take the static semantic terms in the sign and lift them into the dynamic space.

One issue is that it may not be so easy to separate concerns. I believe that some of what I treat with simple static semantics, or the lack of any semantics, may be better treated using some appeal to pragmatics. I treat verbal agreement morphology as entirely grammatical. Yet, it does seem to be true that the use of the feminine, for instance, on a verb will carry some information. If that information is to be held in the static semantics, in a normal fashion, it may produce erroneous static semantics. If the information is not in the

a-	The bound stem for 'plant'
a	The verb 'see'
ã	The verb 'say'
bẽye-	The bound stem for 'fruit'
daa	The noun 'thorn'
dãta	The verb 'hurt'
di-	The bound stem for 'stone'
e-	The bound stem for 'water'
kẽ-	The bound stem for 'manioc'
kẽ	The verbs 'eat' and 'cut'
ĩ	The copula
ĩ	A multipurpose pronominal
okiye	The noun 'woman'
peẽ-	The bound stem for 'plantain'
te-	The bound stem for 'chonta palm'
-to	The pseudo-stem for some pronouns
yẽdẽ	The adjective 'big'
wi-	The bound stem for 'canoe'

Table 6.4: The stems covered by the analysis.

static semantics, then a lifting operation into a dynamic setting will not have the information available. For this reason, something more is needed, and I do not believe the interface is trivial. Despite this, I think it is entirely doable without substantially disrupting the interfaces as currently defined.

6.6 What the analysis covers

The sample of the morphological system of Wao Terero I plan to cover is provided in Tables 6.4 through 6.6. What I call the *stems* in Table 6.4 are heavy on picky nominal forms. I call them "picky" because they exhibit non-productive restrictions on which lexical suffixes may occur – at least immediately after the stem. The *di-* stem only allows a following *-ka*, for instance. Following the first affix *-ka*, I assume *dika*, as well as all other nouns, may potentially take additional lexical suffixes in a process of coinage, which is hypothesized to have a lexical semantic, rather than a morphotactic restriction. The stems *wi-*, *peẽ-*, *te-*, *e-*, *a-*, *bẽye-*, and *kẽ-*, in their nominal uses, all have some similar restrictions to *di-*.

I include two free nominals, one animate *okiye*, which takes no lexical suffixes, but may take person marking. The other is *daa*, an inanimate, which allows lexical affixes, but not person marking.

There are two verbs that take lexical suffixes, *dãta*, 'hurt', and *kẽ*. Both are restricted to body-part

bõ	LS.seed
dẽ	LS.food
ka	LS.stone
kã	LS.body
pa	LS.board
pẽ	LS.liquid
po	LS.canoe
ta	LS.shell
wẽ	LS.plant
yabo	LS.leaf$_1$

Table 6.5: The lexical affixes covered by the analysis.

suffixes. The rest of the verbs, *ĩ*, *ã*, and *a*, do not take lexical suffixes. I do not know the patterning of verbs that take non-body-part lexical suffixes well enough to feel like I could provide a reasonable analysis. I expect verbs to be much more complex than what I provide in the analysis. Some of the choices for verbs were made due to their stems being homophonous with other items in the analysis. My thought is that since I have the processes defined, there is no harm including them. Strictly speaking, I need only one verb that hosts lexical suffixes, and one that does not to demonstrate the pattern.

For this reason, I include only one adjective. The adjectival pattern is so similar to the demonstrative pattern that I did not include the demonstrative.

The pseudo-stem *-to* and the pronominal *ĩ* are included, partially to show off some qualities of the morphotactics, but also to allow for more options in example sentences.

This may seem like a small sample, but it can be extended without difficulty. Adding another adjective, for instance, is largely a matter of providing the information about the root. The patterns are already in place, just not the specific item.

As already alluded to, verbs require the most detail for a correct description. I discuss this below, but the gist is that verbs are a diverse class. Nouns can be picky, in the sense I discussed above, but it is largely a matter of listing their details correctly. Verbs introduce theoretical and descriptive questions that are only partially understood in Wao Terero.

The lexical affixes chosen in Table 6.5 allow me to get into the weeds of less productive patterns without small patterns distracting from the big picture. One of my goals is to demonstrate how small patterns, in aggregate, feed productive patterns. I provide an analysis of *-dẽ*, which assumes that the ‘food’ and ‘abdomen’

bi	2
bĩ	2.MO (2 with DU or PL)
bo	1
bõ	1.INCL (1 with DU or PL)
da	DU
dã	3.F
di	PL
kã	3.H
ke	FUT
ke	LIM
pa	DECL
ta	PST
te	GER

Table 6.6: The inflectional affixes covered by the analysis.

meanings are polysemous. This may be wrong, but as discussed in §3.3.3 and §3.3.4.7, more data is needed to make a determination. It would be interesting to investigate (the also possibly wrong) hypothesis that *-dẽ* is two lexical suffixes, one of which is confined to nominal uses, and the other with classificatory uses. It would demonstrate that even if it is logical to posit that small patterns, such as *di·ka*, 'stone', feed reasoning about classificatory uses of lexical suffixes, it need not necessarily be the case. There may be small nominal lexical suffix patterns that are not associated with classificatory uses, even if there is a homophonous affix that is classificatory. Unfortunately, the alternative analysis of *-dẽ* was not explored due to time constraints. Such an analysis is trivial. It involves providing two M-cat primitives to represent the homophones, and adding a nominal inflection class associated with forms that have the 'food' meaning, which is not referenced by form classes of verbs, adjectives, etc.

In Table 6.6 the majority of inflectional affixes are associated with person and number marking. I include some tense marking, though only for the morphotactic complexity that they add. I have nothing to say about them in terms of meaning. The *-pa* affix is essential to providing a treatment of verbal inflection. The *-te* affix is needed for accusative marking of sentient objects. It also provides a limited version of overabundance, since it is optional in those cases.

Clearly, the above will not result in anything near a comprehensive fragment. The goal is to provide a clear path forward to a more complete fragment.

I begin with a section that describes the morphotactic patterns in the system that provides the definitions

root	te			
x	(ke)	(person)		(pa)
x	ta	first		(pa)
x		(non-first)	ta	pa

Table 6.7: The basic verbal morphotactic pattern, as treated here. The parenthesis indicate optional items in a row. The x is replaced with ‘root’ for non-lexical suffix verbs, and ‘root (ls)’, for lexical suffix verbs.

of the M-cats, form classes, and form-form mappings that are used. I then move on to the form-sign relation and *meaning* relation. As a final step, I show another means of proving form paradigm entries, which utilizes paradigmatic inference. I also discuss the definition of paradigms. Following the analysis, I perform some further theory comparison and discuss the implications of the analysis.

6.7 Morphotactic analysis

I begin with a discussion of the basic verbal facts. These can be seen in Table 6.7. The ‘x’ is different depending on whether the verb takes a lexical suffix or not. If it takes a lexical suffix, ‘x’ is ‘root (ls)’, where ‘ls’ is lexical suffix. If it does not, ‘x’ is simply ‘root’. The parenthesis indicate that something is optional. The ‘x’ difference indicates that there are two established classes.

The *-te*, ‘GER’, suffix blocks further affixation for verbs. This is also true of *-pa*, ‘DECL’, but *-te* occurs only at the root.

The most general pattern is in the second row of Table 6.7. A root (or root (ls)), may be followed by a tense affix, with optional person marking, and an optional *-pa* affix. The *-ta*, ‘PST’, affix disrupts this, as it cannot occur word finally. It may follow non-first person marking, which obligates the use of *-pa*, or it must precede first person marking. This means that questions, which cannot have *-pa*, are not explicitly marked for past tense by *-ta* for non-first persons. Due to the *-ta*, there is a split in the categories used for person marking rules.

The patterns for other parts of speech are provided in Table 6.8. An animate noun may take person marking. An inanimate noun may take one or more lexical suffixes. An adjective may take either person marking, or one lexical suffix. Participant pronouns are pronouns that take first or second person marking, including plural and dual, which end in *-to*. The general pronouns are used for all persons when forming

Animate Noun	root	(person)	
Inanimate Noun	root	ls*	
Adjective	root	(ls|person)	
Participant Pronoun		participant	to
General Pronoun	ĩ	(person)	(te)

Table 6.8: The patterns found across non-verbal parts of speech. The parenthesis indicate optionality. The '|' is used to indicate exclusive disjunction. The '*' means zero or more of the preceding.

verb_k	General verbs (x = root).
verbls_k	Verbs that take lexical suffixes (x = root (ls)).
anim_k	Animate noun
inanim_k	Inanimate nouns
adj_k	Adjectives
prt_k	Participant pronouns
pro_k	General pronoun

Table 6.9: A listing with a short description of form classes that roughly align with part of speech.

an accusative argument, and for non-participant subjects. The divide between participant and other persons requires a divide in class and category.

Based on these seven patterns, there must be seven form classes. These are named in Table 6.9.

There are also a number of morphotactic states, which correspond to M-cats, that are shared when looking at Tables 6.7 and 6.8. First, there are roots in almost all cases, except for the participant pronouns. In one case, there is only one instance of a root, the general pronoun, which is why I provided it as *ĩ* 'PRO' in Table 6.8. For the participant pronoun, I will be treating the final *-to* 'PRO' affix as a root-like element. This is largely due to the obligatory status of the affix *-to*. It serves as the identifier of the part of speech. If it were not obligatory, the person marking, itself, could serve as the root. Unlike √roots, the roots referred to here are simply the unanalyzable core of a word-form. The category has no special status, except that things that have the M-cat ROOT_m, are singletons. So, singletons, categorized as ROOT_m are an important shared category across classes.

Most parts of speech allow the full range of person marking. The M-cat for this is $\mathsf{PERSONABLE}_\mathsf{m}$. For adjectives, animate nouns, and general pronouns only the root may receive person marking. For verb_k, the root and *-ke* tensed categories are also $\mathsf{PERSONABLE}_\mathsf{m}$. For verbls_k, the root, the *-ke* tensed, and the lexical

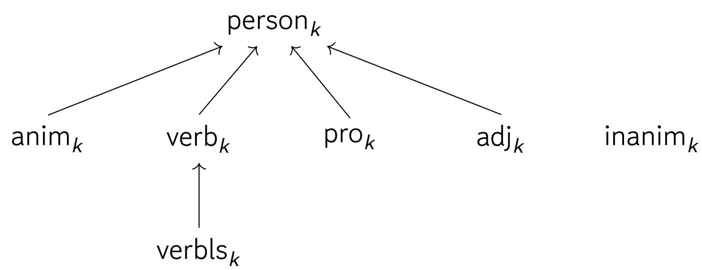


Figure 6.5: An initial form class diagram, which shows the relations among various parts of speech to the class $person_k$.

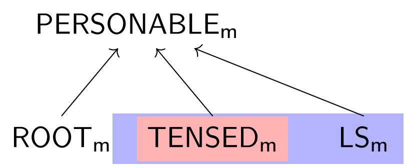


Figure 6.6: An initial M-cat diagram, which shows the relations among various M-cats relative to PERSONABLE$_m$. The area surrounded by only pink applies when $verb_k$. The area surrounded by light blue applies when $verbls_k$.

suffix affixed categories may receive person marking. All of these share the quality that the root may receive person marking, so we may say, without qualification, $\mathsf{ROOT_m} \leq_m \mathsf{PERSONABLE_m}$. We don't need to worry about specifying that this doesn't apply to inanimate nouns and participant pronouns. This is because there is a class of items that take full person marking, person_k, where $\{\mathsf{adj}_k, \mathsf{anim}_k, \mathsf{pro}_k, \mathsf{verb}_k, \mathsf{verbls}_k\} \leq_k \mathsf{person}_k$, which excludes prt_k or inanim_k. So person marking rules will not apply for those parts of speech corresponding to prt_k or inanim_k. The applicability of person marking diverges beyond $\mathsf{ROOT_m} \leq_m \mathsf{PERSONABLE_m}$. Other items are parameterized to have the $\mathsf{PERSONABLE_m}$ category based on class. The *-ke* tensed category is $\mathsf{TENSED_m}$. Given this category it is defined that $\forall\alpha.klass\ \alpha \leq_k \mathsf{verb}_k \rightarrow \mathsf{TENSED_m} \leq_m \mathsf{PERSONABLE_m}$. The verbls_k class inherits this property by stating $\mathsf{verbls}_k \leq_k \mathsf{verb}_k$. The category of having a stem ending in a lexical suffix is $\mathsf{LS_m}$. Given this, it is possible to provide a final ordering rule, $\forall\alpha.klass\ \alpha \leq_k \mathsf{verbls}_k \rightarrow \mathsf{LS_m} \leq_m \mathsf{PERSONABLE_m}$. The form class and M-cat relationships under discussion can be seen in Figures 6.5 and 6.6.

Moving forward to another common pattern, three of the form classes take lexical suffixes. The adjectives and lexical suffix verbs may take a single suffix, immediately following a root. The inanimate nouns may take one or more lexical suffixes following a root. For this reason, we can say that $\mathsf{ROOT_m} \leq_m \mathsf{LSABLE_m}$, where $\mathsf{LSABLE_m}$ is the category of items that are able to accept a lexical suffix. This can be done without qualifications. The form class hierarchy will ensure that only the proper items receive the affixes. For the inanimate nouns, it must also be specified that $\forall\alpha.\mathit{klass}\ \alpha \leq_k \mathsf{inanim}_k \rightarrow \mathsf{LS_m} \leq_m \mathsf{LSABLE_m}$, which allows for a limitless number of lexical suffixes to be affixed. These will not surface in the syntax unless some meaning can be matched to them. There is also a limit in that each suffix may only apply once. This may seem like over-generation to some, but I would contend that coinage often follows lexical patterns of word-forms. Until these word-forms take on meaning and occur in syntax, they are simply potential words. The alternative, limiting the model only to what has been observed, offers no interesting room for predictions. Another analyst is free to take another route. My choice to embrace generality in this case is not due to a limitation of the system. One can be very fine grained.

For instance, there are stems that take only one lexical suffix, such as *di-*, *a-*, *peẽ-*, *e-*, and *wi-*. Each of these must have a form class specific to their limitations. I name these dika_k, $\mathsf{aw\tilde{e}}_k$, $\mathsf{pe\tilde{e}d\tilde{e}}_k$, $\mathsf{ep\tilde{e}}_k$, and wipo_k. There are some stems that take only two lexical affixes, *bẽye-*, *kẽ-*, and *te-*. Both *kẽ-* and *te-*, with form classes $\mathsf{k\tilde{e}d\tilde{e}}_k$ and $\mathsf{tep\tilde{e}}_k$, respectively, take the *-we* suffix, like *a-*. Therefore, $\{\mathsf{k\tilde{e}d\tilde{e}}_k, \mathsf{tep\tilde{e}}_k\} \leq_k \mathsf{aw\tilde{e}}_k$. The stem *bẽye-* takes the affix *-ka*, so its form class, $\mathsf{b\tilde{e}yeb\tilde{o}}_k$ is ordered under dika_k. The stem also takes the *-bõ* affix but since there is no more specific item in the fragment, it does not inherit this behavior. The rules for affixing *-ka*, *-wẽ*, *-dẽ*, *-pẽ*, and *-po*, reference the most specific class. Larger classes are born of examples from smaller classes.

Adjectives may take all classificatory lexical suffixes, which are the only kind of lexical suffix in this analysis. Inanimate nouns are picky, but beyond the affixes that may attach directly to a root, I place no further constraints. One might easily do so. The verbs under consideration only take affixes with body-part meanings associated with them. Therefore $\mathsf{body}_k \leq_k \{\mathsf{b\tilde{e}yeb\tilde{o}}_k, \mathsf{wipo}_k\}$. The general class for any lexical suffix is thing_k. I state that $\{\mathsf{adj}_k, \mathsf{inanim}_k\} \leq_k \mathsf{thing}_k \leq_k \mathsf{body}_k$. The verbs are limited to body-parts, so they do not inherit from thing_k, but directly from body_k. These classes only limit *which* affixes may apply. For an inanimate noun, the order is essentially free. This can be constrained by fine grained class parameters in the M-cat order.

One can see that fine grained distinctions are not only possible, but do not have the character of *excep-*

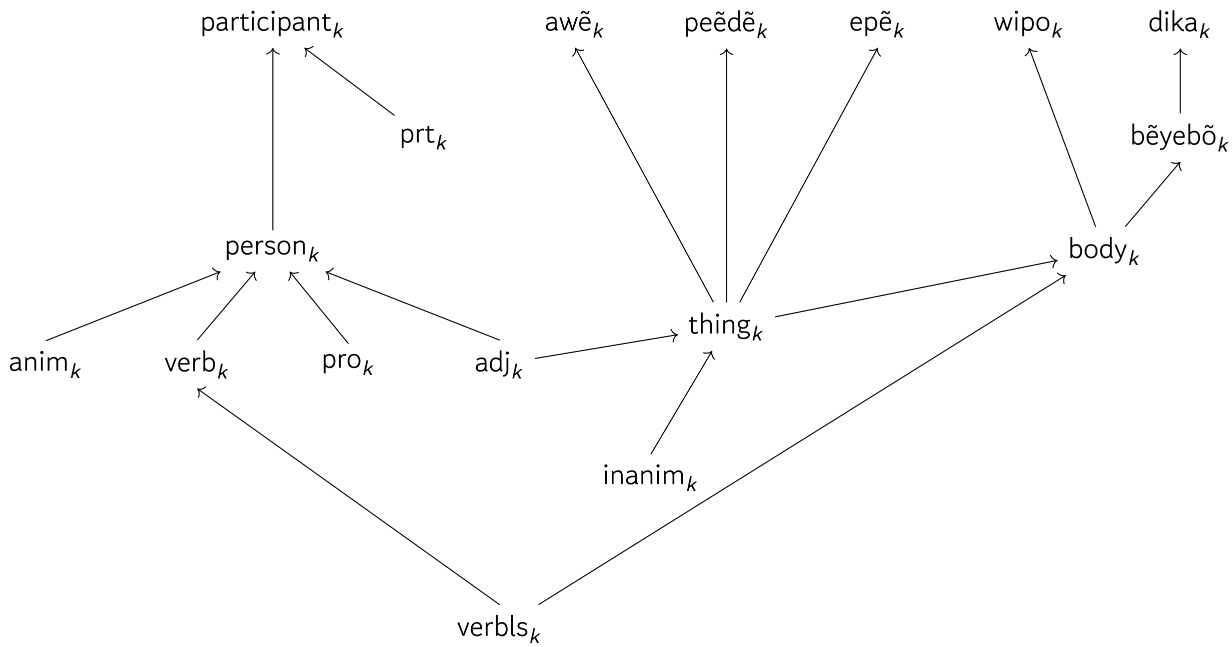


Figure 6.7: An elaborated form class diagram, which shows the relations among various parts of speech to the class person$_k$, as well as lexical suffix relations.

tions. This is in contrast to truly exceptional items in other languages. That BE has suppletive forms, such as *is* and *are*, has a different character. The small patterns of Wao Terero lexical suffixes seem to feed generalities. The BE example appears to stand alone, with nothing that can be generalized from it. The manner in which Wao Terero nouns fit a pattern of the specific feeding the general in the lexical suffix space suggests that there may be some deeper explanation, likely diachronic.

There is a tendency for meanings to get "locked in" by nominal constructions. Not only are there items that take only one or two affixes, there are items where only a single meaning of the affix is available. In some cases, these are constraints of stem. The affix *-bõ*, LS.seed, may mean 'fruit', 'seed', or 'eye', but on a stem like *tewẽ*, 'chonta palm', it will be interpreted as the palm fruit.

The current form class diagram can be seen in Figure 6.7. In addition to what has already been covered, I added a class participant$_k$, which is for first and second person inflection. The class prt$_k$ and person$_k$ inherit from this. There are a number of useful form classes that I introduce at the interface to generalize some rules, which are not in the diagram, yet.

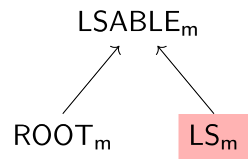


Figure 6.8: An M-cat diagram, which shows the relations among various M-cats relative to LSABLE$_m$. The area surrounded by pink applies when inanim$_k$.

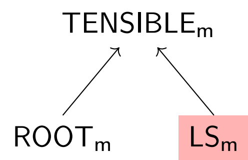


Figure 6.9: An M-cat diagram, which shows the relations among various M-cats relative to TENSIBLE$_m$. The area surrounded by pink applies when verbls$_k$.

6.7.1 Person marking

Given the form classes above, there are further M-cat restrictions on whether certain affixes occur, and the order they occur in. The initial M-cat ordering was provided in Figures 6.6 and 6.8. To allow for the *-ke* tense to follow a root or lexical suffix, there needs to be a TENSIBLE$_m$ M-cat, which can be seen in Figure 6.9. It should be emphasized that there are affixes in competition with *-ke*. The order is more general than it may seem given only *-ke* in this example.

With the TENSIBLE$_m$ category established and the patterns for lexical suffixes essentially covered, the person patterns must be considered. These have interactions with affixes like *-ta, -to*, and *-te*. I will begin with the patterns within prt$_k$, which inherits only from the participant$_k$ pattern.

boto	1
bõto	1.INCL
bõdato	1.DU
bõdito	1.PL
bito	2
bĩto	2.MO
bĩdato	2.DU
bĩdito	1.PL

Table 6.10: The various forms of the participant pronoun.

The pattern for the pronoun type can be seen in Table 6.10. Only the first and second persons are represented. There is always an ending *-to*. Since the person marking is normally suffixed to a stem, it looks like a null stem. The recognizability of patterns based on a stem is important to this theory. Despite this, it is the function that the stem serves that is important, rather than some correspondence to a particular formal pattern. Clearly the pattern in Table 6.10 is recognizable as a formal grouping due to the shared *-to* affix and the highly salient positioning of the person marking word initially, which happens no place else. I have not developed a system that allows for recognizing contrasts of the latter kind, which involves some notion of morphotactic comparison, though it is likely an important cue. It is possible to treat the *-to* as a pseudo-stem.

In the normal case, at least one member of a form paradigm is defined by an axiom-like constructor. I choose *bito* for this purpose, providing a constructor that species $FE_{mp}\,\langle \mathsf{bi_m} :: \mathsf{to_m}, \mathrm{bi_{pr}} :: \mathrm{to_{pr}}\rangle$. Remember from §6.5.1.1, that to_{pr} is defined as $\lambda P\ s.(P\ s) +\!\!+ \text{“to”}$. The bi_{pr} process is defined as a normal affixing process. Since the initial affix is added by axiom, I don't bother with additional form classes.

The question is, how do I go from *boto* to *bito*? In §6.5.1.1, I defined $\mathrm{b\tilde{o}}_{pr}$ as being added using the $rule_1$ schema. Instead of this, I will define the rules using a $rule_2$ schema. This will also be the case for *-bo* and *-bĩ*. Only *-bi* uses the $rule_1$ schema.

I define two ordering rules, introducing the abstract M-cats $\mathsf{FIRST_m}$ and $\mathsf{SECOND_m}$.

$$\alpha \leq_m \mathsf{FIRST_m}/bo_m\# \tag{LIII}$$

$$\alpha \leq_m \mathsf{SECOND_m}/bi_m\# \tag{LIV}$$

(LV)

These describe the condition of ending in *-bi* or *-bo*.

As will be seen below, I don't care very much about the number of proof steps needed for certain items. This is because each intermediary proof is necessary for the definition of a paradigm member, and the goal is the definition of the paradigm instance as a whole. There is also a descriptive truth that emerges in my approach, which is that the second person plural marking is morphotactically more informative than other person marking. This is explained below.

Given the ordering rule above, the following form-form mapping may be defined.

$$\frac{\alpha \leq_m \mathsf{SECOND_m} \qquad \alpha \leq_k \mathsf{participant}_k}{\langle \alpha \lhd \mathsf{bo_m}, \alpha \lhd \mathrm{bo_{pr}} \rangle} \text{boMP} \tag{LVI}$$

$$\frac{\alpha \leq_m \mathsf{SECOND_m} \qquad \alpha \leq_k \mathsf{participant}_k}{\langle \alpha \lhd \mathsf{b\tilde{\imath}_m}, \alpha \lhd \mathsf{bo_{pr}} \rangle} \text{bĩMP} \tag{LVII}$$

$$\frac{\alpha \leq_m \mathsf{FIRST_m} \qquad \alpha \leq_k \mathsf{participant}_k}{\langle \alpha \lhd \mathsf{b\tilde{o}_m}, \alpha \lhd \mathsf{b\tilde{o}_{pr}} \rangle} \text{bõMP} \tag{LVIII}$$

$$\frac{\alpha \leq_m \mathsf{PERSONABLE_m} \qquad \alpha \leq_k \mathsf{participant}_k}{\langle \alpha \diamond \mathsf{bi_m}, \alpha \lhd \mathsf{bi_{pr}} \rangle} \text{biMP} \tag{LIX}$$

Given that $FE_{mp}\,\langle \mathsf{bi_m} :: \mathsf{to_m}, \mathrm{bi_{pr}} :: \mathrm{to_{pr}} \rangle$ has already been defined, there is no need to apply biMP for the paradigm. This is not ad hoc, since the start state, so to speak, for defining the paradigm instance usually involves one complete entry. Since there is only one paradigm of the prt_k kind, the obvious choice will be the most informative member. It will be seen why $FE_{mp}\,\langle \mathsf{bi_m} :: \mathsf{to_m}, \mathrm{bi_{pr}} :: \mathrm{to_{pr}} \rangle$ is most informative as the analysis progresses.

From $FE_{mp}\,\langle \mathsf{bi_m} :: \mathsf{to_m}, \mathrm{bi_{pr}} :: \mathrm{to_{pr}} \rangle$, it can be seen that boMP may apply. The $\lhd$ indicates that the $\mathsf{bi_m}$ and bi_{pr} will be popped from the list, and the *-bo* equivalents applied. This means that $FE_{mp}\,\langle \mathsf{bo_m} :: \mathsf{to_m}, \mathsf{bo_{pr}} :: \mathsf{to_{pr}} \rangle$ is true.

This leaves the plural and dual affixes. As was discussed in §6.5.2, the rules are defined largely in terms of preceding concrete M-cats, except for the third person dual, which applies for the category $\mathsf{PERSONABLE_m}$, as well. To allow the affixes to apply for prt_k, as well as in other contexts, the rules reference the $\mathsf{participant}_k$ class. The two number relevant ordering rules that make reference to concrete M-cats are below. The ordering rules above fit into the M-cat hierarchy as seen in Figure 6.10.

$$\alpha \leq_m \mathsf{NUMBERABLE_m} / \{\mathsf{b\tilde{o}_{2m}}, \mathsf{b\tilde{\imath}_m}\}\,\# \tag{LX}$$

$$\alpha \leq_m \mathsf{PLURALABLE_m} / \mathsf{d\tilde{a}_m}\# \tag{LXI}$$

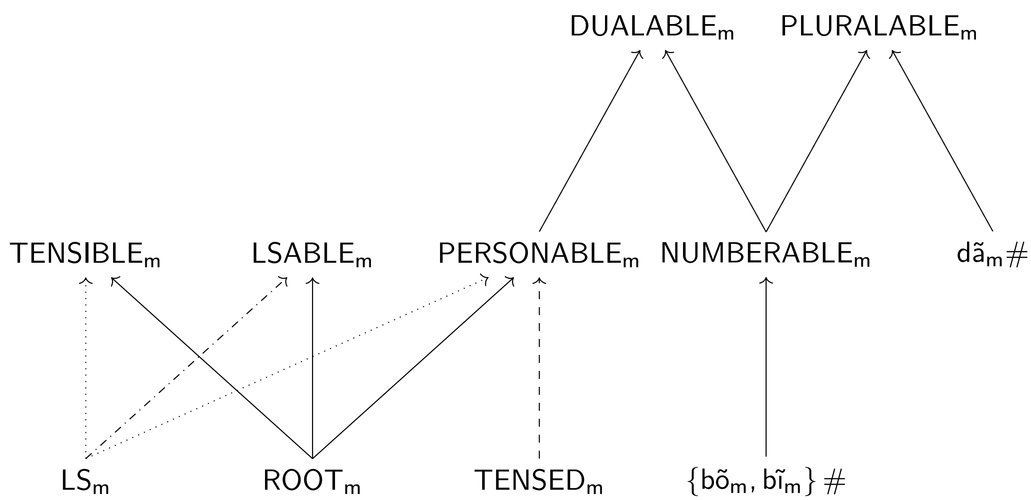


Figure 6.10: An elaborated M-cat diagram. Dashed lines apply when verb_k. Dotted lines apply when verbls_k. Dash dotted lines apply when inanim_k.

Given the above, it is possible to step through a derivation. The target will be *bõdito*.

$$\dfrac{\dfrac{\dfrac{FE_{mp}\,\langle \mathsf{bi_m} :: \mathsf{to_m}, \mathrm{bi}_{\mathrm{pr}} :: \mathrm{to}_{\mathrm{pr}}\rangle \quad \begin{array}{l} \mathsf{bi_m} :: \mathsf{to_m} \leq_m \mathsf{SECOND_m} \\ klass\ \mathsf{bi_m} :: \mathsf{to_m} \leq_k \mathsf{participant}_k \end{array}}{FE_{mp}\,\langle \mathsf{bo_m} :: \mathsf{to_m}, \mathrm{bo}_{\mathrm{pr}} :: \mathrm{to}_{\mathrm{pr}}\rangle}\ \text{boMP} \quad \begin{array}{l} \mathsf{bo_m} :: \mathsf{to_m} \leq_m \mathsf{FIRST_m} \\ klass\ \mathsf{bo_m} :: \mathsf{to_m} \leq_k \mathsf{participant}_k \end{array}}{FE_{mp}\,\langle \mathsf{b\tilde{o}_m} :: \mathsf{to_m}, \mathrm{b\tilde{o}}_{\mathrm{pr}} :: \mathrm{to}_{\mathrm{pr}}\rangle}\ \text{bõMP} \quad \begin{array}{l} \mathsf{b\tilde{o}_m} :: \mathsf{to_m} \leq_m \mathsf{PLURALABLE_m} \\ klass\ \mathsf{b\tilde{o}_m} :: \mathsf{to_m} \leq_k \mathsf{participant}_k \end{array}}{FE_{mp}\,\langle \mathsf{di_m} :: \mathsf{b\tilde{o}_m} :: \mathsf{to_m}, \mathrm{di}_{\mathrm{pr}} :: \mathrm{b\tilde{o}}_{\mathrm{pr}} :: \mathrm{to}_{\mathrm{pr}}\rangle}\ \text{diMP} \qquad \text{(LXII)}$$

In (LXII), I stack the conditions for the rule to save horizontal space. I begin with $FE_{mp}\,\langle \mathsf{bi_m} :: \mathsf{to_m}, \mathrm{bi}_{\mathrm{pr}} :: \mathrm{to}_{\mathrm{pr}}\rangle$, which passes the category and class conditions for boMP. This proves $FE_{mp}\,\langle \mathsf{bo_m} :: \mathsf{to_m}, \mathrm{bo}_{\mathrm{pr}} :: \mathrm{to}_{\mathrm{pr}}\rangle$, which allows for a proof of $FE_{mp}\,\langle \mathsf{b\tilde{o}_m} :: \mathsf{to_m}, \mathrm{b\tilde{o}}_{\mathrm{pr}} :: \mathrm{to}_{\mathrm{pr}}\rangle$. Since this ends in $\mathsf{b\tilde{o}_m}$, ordered under $NUMERABLE$, the plural rule applies.

Something to note about this derivation is that there are no hidden layers. Each step along the way

go-ta-bo-pa	go-PST-1-DECL
go-ta-bõ-pa	go-PST-1.INCL-DECL
go-ta-bõ-di-pa	go-PST-1-PL-DECL
go-ta-bõ-da-pa	go-PST-1-DU-DECL
go-bi-ta-pa	go-2-PST-DECL
go-bĩ-ta-pa	go-2.MO-PST-DECL
go-bĩ-di-ta-pa	go-2-PL-PST-DECL
go-bĩ-da-ta-pa	go-2-DU-PST-DECL
go-da-ta-pa	go-DU-PST-DECL
go-kã-ta-pa	go-3.H-PST-DECL
go-dã-ta-pa	go-3.F-PST-DECL
go-dã-di-ta-pa	go-3.H-PL-PST-DECL
go-ta-pa	go-PST-DECL

Table 6.11: The past tense in Wao Terero distributes differently depending on person. For the first person, in gray, the *-ta* affix precedes person marking. For all other persons, the affix follows person marking, and cannot occur word finally.

produced a valid word-form. They all resolve to a concrete free form (string support) in Wao Terero syntax.

Continuing on to an interesting pattern in Wao Terero, the Wao Terero past tense has a challenging distribution. The pattern can be seen in Table 6.11. The past tense in Wao Terero distributes differently depending on person. For the first person, in gray, the *-ta* affix precedes person marking. For all other persons, the affix follows person marking, and cannot occur word finally.

This pattern requires two additional rules. They are remarkable in that they add more than one M-cat and process at the same time.

$$\frac{\alpha \leq_m \mathsf{TENSIBLE_m} \qquad \alpha \leq_k \mathsf{verb}_k}{\langle \alpha \diamond \mathsf{bo_m} :: \mathsf{ta_m}, \alpha \diamond \mathsf{bo}_{\mathrm{pr}} :: \mathsf{ta}_{\mathrm{pr}} \rangle} \text{ taboMP} \tag{LXIII}$$

The first rule relies only on classes and categories that have already have been defined. The addition of non-first person affixes in taboMP is blocked due the fact that first replaces second person, and not vice versa. Additionally, the other persons simply cannot apply, because the stem is not $\mathsf{PERSONABLE_m}$ after there is person marking. No *-pa* is needed because *-ta* is not word final. There is no reason why multiple M-cats and processes cannot be added at the same time. Having more than one M-cat occur per process is reasonable for fused morphs that are allomorphs of two distinct morphs. Adding more than one process per M-cat is relevant to multiple exponence. In (LXIII), the rule exposes a certain truth. The *-ta* only occurs prior to person marking for the first person.

$$\frac{\alpha \leq_m \mathsf{TAABLE_m} \qquad \alpha \leq_k \mathsf{verb}_k}{\langle \alpha \diamond \mathsf{bi_m}, \alpha \diamond \mathsf{bi_{pr}} \rangle} \text{ tapaMP} \tag{LXIV}$$

For the second rule, in (LXIV), there are two conditions that can block the application of *-tapa* – other than *-pa* already having been affixed. The first is the occurrence of the first person. The second is competition among tenses, assuming that *-ta* is incompatible with earlier tense marking. Given the two conditions, a fairly specific category constraint is needed.

$$\alpha \leq_m \mathsf{TAABLE_m}/\neg(\mathit{first} \wedge \mathit{tense}) \tag{LXV}$$

The negative and the logical symbols in the ordering rule are intended to make it easy to read, but negation is not actually needed. There is a list of all M-cats, from this list, one may define sub-lists of the first persons and the relevant tenses. The rule actually references such a list, where the first persons and tenses have been removed. $\mathsf{TAABLE_m}$ is not seem the most natural looking category, but it reflects a morphotactic reality.

One further point. There are affixes in competition with *-pa*. I define these using lateral rules. There would not need to be a tapaMP-like rule for each affix.

To return to the point about the second person being an informative final affix, the affix tells us that the preceding stem contains no *-ta*. Therefore, any participant affix can replace *-bi*. The occurrence of *-bo* does not entailment a lack of *-ta*. For that reason, only *-bi* utilizes the rule_1 schema. Going from second to first is always safe, the opposite is not.

A final pattern of interest is the *-te* affix. This affix is incompatible with person marking on verbs. On the *ĩ* stem, it may occur with person marking, which is used to form pronouns and accusative marking. The details of how to deal with the affix are not difficult. Since *-te* occurs with both verbs and pronouns, an additional super class is needed for these. The distribution of *-te* is unique. I am not aware of anything it is in simple competition with. For that reason, and due to a lack of creativity, I'll call the new class te_k. I have not seen *-te* co-occur with a lexical suffix, so I'll assume it attaches only to roots.

$$\frac{\alpha \leq_m \mathsf{ROOT_m} \qquad \alpha \leq_k \mathsf{te}_k}{\langle \alpha \diamond \mathsf{te_m}, \alpha \diamond \mathsf{te_{pr}} \rangle} \text{ teMP} \tag{LXVI}$$

The te_{pr} process is defined like to_{pr}, assuring that it will always be affixed after subsequent, standard affixing processes. The only additional rule needed is to ensure that person marking may be added after *-te* for items of the pro_k class.

$$\alpha \leq_m \mathsf{PERSONABLE_m}/\mathsf{pro}_k, \mathsf{te_m}\# \tag{LXVII}$$

This sums up person marking. The affixation of *-kã* and *-dã* is unremarkable. One can see Appendix A for details.

6.7.2 Some example derivations involving lexical suffixes

In comparison to the person marking, the morphotactics of lexical suffixes is straight forward. They always follow a root, or do not occur at all. In this section, I run through some examples, some of which will be used later in the analysis.

The adjective *yẽdẽ* has its root entry defined by axiom $\langle \mathsf{Yẽdẽ_m}, \mathsf{yẽdẽ}_{\mathrm{pr}} \rangle$. To apply the *-ka*, LS.stone affix, the kaMP rule is applied.

$$\frac{FE_{mp}\langle \mathsf{Yẽdẽ_m}, \mathsf{yẽdẽ}_{\mathrm{pr}}\rangle \quad \boxed{\begin{array}{l}\mathsf{Yẽdẽ_m} \leq_m \mathsf{LSABLE_m} \\ klass\ \mathsf{Yẽdẽ_m} \leq_k \mathsf{dika}_k\end{array}}}{FE_{mp}\langle \mathsf{ka_m} :: \mathsf{Yẽdẽ_m}, \mathsf{ka}_{\mathrm{pr}} :: \mathsf{yẽdẽ}_{\mathrm{pr}}\rangle}\ \text{kaMP} \tag{LXVIII}$$

In the next example, kã1MP is used to derive ‘eat meat’. Then the identical person marking is added with kã2MP. Finally, *-pa* is added. This introduces a category I have not discussed. The category $\mathsf{CLOSEABLE_m}$ is perhaps not well named. It should not be confused with similar terms in other works. It is an ad hoc choice. It describes a form paradigm entry that may serve as input to a form-form mapping that applies a suffix that closes off the word to further affixation. It implies some group of affixes that have this quality, which would include *-pa*, ‘DECL’, and *-ke*, ‘LIM’. The affixes of this kind are dependent on the class. For a verb_k, all M-cats that have names that end in “-ABLE” are ordered below $\mathsf{CLOSEABLE_m}$. Basically, there is no M-cat that occurs in a verbal form paradigm entry where a *-pa* affixing form-form mapping may not be applied, except where *-pa* (or an affix in competition) has already been applied. A similar situation holds for other parts of speech with respect to the limitive *-ke*.

$$\dfrac{\dfrac{\dfrac{FE_{mp}\,\langle \mathsf{K\tilde{e}_{1m}}, \mathrm{k\tilde{e}_{pr}} \rangle \quad \boxed{\begin{array}{l} \mathsf{K\tilde{e}_{1m}} \leq_m \mathsf{LSABLE_m} \\ klass\ \mathsf{K\tilde{e}_{1m}} \leq_k \mathsf{body}_k \end{array}}}{FE_{mp}\,\langle \mathsf{k\tilde{a}_{1m}} :: \mathsf{K\tilde{e}_{1m}}, \mathrm{k\tilde{a}_{pr}} :: \mathrm{k\tilde{e}_{pr}} \rangle \quad \boxed{\begin{array}{l} \mathsf{k\tilde{a}_{1m}} :: \mathsf{K\tilde{e}_{1m}} \leq_m \mathsf{PERSONABLE_m} \\ klass\ \mathsf{k\tilde{a}_{1m}} :: \mathsf{K\tilde{e}_{1m}} \leq_k \mathsf{person}_k \end{array}}}\ \text{kã1MP}}{FE_{mp}\,\langle \mathsf{k\tilde{a}_{2m}} :: \mathsf{k\tilde{a}_{1m}} :: \mathsf{K\tilde{e}_{1m}}, \mathrm{k\tilde{a}_{pr}} :: \mathrm{k\tilde{a}_{pr}} :: \mathrm{k\tilde{e}_{pr}} \rangle \quad \boxed{\begin{array}{l} \mathsf{k\tilde{a}_{2m}} :: \mathsf{k\tilde{a}_{1m}} :: \mathsf{K\tilde{e}_{1m}} \leq_m \mathsf{CLOSEABLE_m} \\ klass\ \mathsf{k\tilde{a}_{2m}} :: \mathsf{k\tilde{a}_{1m}} :: \mathsf{K\tilde{e}_{1m}} \leq_k \mathsf{verb}_k \end{array}}}\ \text{kã2MP}}{FE_{mp}\,\left\langle \mathsf{pa_{2m}} :: \mathsf{k\tilde{a}_{2m}} :: \mathsf{k\tilde{a}_{1m}} :: \mathsf{K\tilde{e}_{1m}}, \mathrm{pa_{pr}} :: \mathrm{k\tilde{a}_{pr}} :: \mathrm{k\tilde{a}_{pr}} :: \mathrm{k\tilde{e}_{pr}} \right\rangle}\ \text{pa2MP} \qquad \text{(LXIX)}$$

After the ordering rules and form-form mapping have been defined, specific derivations require no cleverness.

For a bound noun stem, a non-bound state must be defined by axiom. The form *kẽwẽ*, ‘manioc (plant)’, is one such. Below I provide an example of *kẽwẽyabo*, ‘manioc leaf’.

$$\dfrac{FE_{mp}\,\langle \mathsf{w\tilde{e}_m} :: \mathsf{K\tilde{e}_{2m}}, \mathrm{w\tilde{e}_{pr}} :: \mathrm{k\tilde{e}_{pr}} \rangle \quad \boxed{\begin{array}{l} \mathsf{w\tilde{e}_m} :: \mathsf{K\tilde{e}_{2m}} \leq_m \mathsf{LSABLE_m} \\ klass\ \mathsf{w\tilde{e}_m} :: \mathsf{K\tilde{e}_{2m}} \leq_k \mathsf{thing}_k \end{array}}}{FE_{mp}\,\left\langle \mathsf{yabo_m} :: \mathsf{w\tilde{e}_m} :: \mathsf{K\tilde{e}_{2m}}, \mathrm{yabo_{pr}} :: \mathrm{w\tilde{e}_{pr}} :: \mathrm{k\tilde{e}_{pr}} \right\rangle}\ \text{yaboMP} \qquad \text{(LXX)}$$

The last demonstrates adding a lexical suffix to a stem that already has a lexical suffix.

This concludes the section on morphotactics. For those who are accustomed to justifying every morphotactic move based on some form of universal, or typological, feature structure, morphotactics in the system have little to offer. The names I have chosen for classes and categories are language specific in most cases. Yet, I have also suggested categories that would be useful for language comparison, and haven’t been pedantic in avoiding common terms such as “first person”. There are categories in the morphological space that correspond to grammatical animacy, sentience, and male-female gender, which may be of interest to comparative studies.

What I have demonstrated is the power of some fairly old ideas in a new context. Inflection class hierarchies have been used with similar aims in the past. Dividing stems up into categorized pieces is also an old idea. The level of attention paid to the formal definitions of these categories is unusual. A unique aspect of the system is that it borrows both these ideas while being word-form-based.

Leaving morphotactics, the next section deals with form-sign relations.

6.8 The Wao Terero sign paradigm

The portion of the sign paradigm that I define is fairly simple, though it covers a lot of ground. There is a focus on the fact that the definite status of nominal items are not signaled by a compositional device. I assume that nominals have variants that, in addition to a basic noun, $e \rightarrow p$, meaning, behave as though they have been modified by items such as *some* or *the* in English. When adjectives occur without nominals, they are treated similarly, but additionally have semantics equivalent to taking a *one* argument in English. Adjectives also behave like covert definite and indefinite modifiers/quantifiers for nouns. This is because the noun phrase with an adjective may be definite or indefinite, yet the expectation is that an adjective takes a simple nominal argument.

There exist proposals for syntactic encodings of definiteness (Armoskaite and Gillon, 2013) in other languages. I have seen no evidence that there is any rigid word-order distinction in Wao Terero relevant to definiteness.

Wao Terero nouns and verbs do not have much case or valence morphology. For this reason, the M-cat order is less important at the interface than it might be in another language. Unlike, Kichwa, where causative, reflexive, reciprocal, and other valence relevant affixes distribute uniformly across the verbal class, Wao Terero verbs, despite having some valence changing morphology, are relatively ambiguous. For that reason, it is enough to know, in many cases, that something is a nominal or verbal element, and then provide the transitive-intransitive distinction largely based on whether the item has a transitive or intransitive lexical semantic interpretation. For that reason, all of the form-sign relations reference the $\mathsf{INF_m}$ category, a kind of catch all. In order to limit the scope of the analysis, I only consider *-pa* affixed verbs. I do not deal with multi-verb expressions. Expanding the analysis in some areas would be trivial, yet there needs to be some limit.

Given that many nominal expressions follow general syntactic patterns, whether an argumentless adjective, or an inanimate or animate noun, I use a general $\mathsf{nominal}_k$ in form-sign relations that subsumes those classes. This is the first example of a sign-oriented form classes, which may raise question of whether *form* class is the best name for the paradigmatic schema. Form classes provide the paradigm schema for both form paradigms and sign paradigms. To call them inflection classes would be wrong, since that is only a subset of what they classify. Additionally, they really are about form patterns. At both the level of form and the level of signs, a paradigm instance is identified by a stem and a form class.[4]

I define a number of form-sign relation for the fragment, sufficient to provide a reasonably sized grammar. The first is the adjectival rule, which was already introduced, repeated below.

$$\frac{\alpha \leq_m \mathit{INF}_m \qquad \alpha \leq_k \mathsf{adj}_k \qquad \mathit{meaning}(\beta, \alpha)}{\vdash (\lambda \mathrm{st}, \mathrm{s} \bullet \mathrm{t})\alpha; (\lambda \mathsf{t}.\mathsf{N}_\mathsf{t} \multimap \mathsf{Adj}_\mathsf{t})\alpha; (\lambda x.x)\beta} \text{ adjSP} \tag{LXXI}$$

The rule was described in §6.5.4. I do not provide any additional details here.

An example application appears below, for *yẽdẽ*.

$$\frac{\mathsf{Y\tilde{e}d\tilde{e}_m} \leq_m \mathit{INF}_m \qquad \mathsf{adj}_k \leq_k \mathsf{adj}_k \qquad \mathit{meaning}(\mathit{big}, \mathsf{Y\tilde{e}d\tilde{e}_m})}{\vdash (\lambda \mathrm{t}, \mathrm{y\tilde{e}d\tilde{e}} \bullet \mathrm{t})\alpha; (\mathsf{N}_\mathsf{t} \multimap \mathsf{Adj}_\mathsf{t})\alpha; \mathit{big}} \text{ adjSP} \tag{LXXII}$$

The application of these rules requires no chaining, as was the case for form-form mappings. There are no rules from sign paradigm entry to sign paradigm entry.

In addition to adjSP, there are other adjective specific rules. There are definite and indefinite forming rules that take a noun argument.

$$\frac{\alpha \leq_m \mathit{INF}_m \qquad \alpha \leq_k \mathsf{adj}_k \qquad \mathit{meaning}(\beta, \alpha)}{\vdash (\lambda \mathrm{st}, \mathrm{s} \bullet \mathrm{t})\alpha; (\lambda \mathsf{t}.\mathsf{N}_\mathsf{t} \multimap \mathsf{Nom}_\mathsf{t})\alpha; (\lambda x.x)\beta} \text{ adj-n-defSP} \tag{LXXIII}$$

The rule adj-n-defSP for a definite argument taking adjective is very similar adjSP, given that most of the differences in meaning are handled by $\mathit{meaning}$.

Below, I provide an example application with *yẽdẽbidi*, the second person plural form of the adjective.

[4]In the future, in order to accommodate suppletion, a sign paradigm instance will be identified by a list of m that corresponds to a stem, providing slightly more abstraction.

$$\frac{\begin{array}{l}\mathsf{di_m :: b\tilde{ı}_m :: Y\tilde{e}d\tilde{e}_m} \leq_m INF_m \\ \mathsf{adj}_k \leq_k \mathsf{adj}_k \\ meaning(\lambda nx.\iota(tall(n\,x) \wedge n\,x), \mathsf{di_m :: b\tilde{ı}_m :: Y\tilde{e}d\tilde{e}_m})\end{array}}{\vdash (\lambda \mathrm{t}, \mathrm{y\tilde{e}d\tilde{e}b\tilde{ı}di} \bullet \mathrm{t}); (\mathsf{N_{2.pl}} \multimap \mathsf{Nom_{2.pl}}); (\lambda nx.\iota(tall(n\,x) \wedge n\,x)} \text{ adj-n-defSP} \qquad \text{(LXXIV)}$$

Since the adjective takes a nominal argument, I consider it an agreement target and there is no explicit person meaning. The tecto does require a syntactic second person plural argument, due to parameterization using the function T.

In cases where person marking should carry some meaning, I treat the person meaning as a property. The nominal rule is provided below.

$$\frac{\alpha \leq_m INF_m \qquad \alpha \leq_k \mathsf{nominal}_k \qquad meaning(\beta, \alpha)}{\vdash (\lambda \mathrm{t.t})\alpha; (\lambda \mathrm{t.Nt})\alpha; (\lambda x.x)\beta} \text{ nSP} \qquad \text{(LXXV)}$$

If nSP is provided with a form paradigm entry with the person marking for second plural, the result encodes a meaning as well as a category.

$$\frac{\begin{array}{l}\mathsf{di_m :: b\tilde{ı}_m :: Okiye_m} \leq_m INF_m \\ \mathsf{anim}_k \leq_k \mathsf{nominal}_k \\ meaning(\lambda x.addressees \Rightarrow woman\,x, \mathsf{di_m :: b\tilde{ı}_m :: Okiye_m})\end{array}}{\vdash \mathrm{okiyeb\tilde{ı}di}; \mathsf{N_{2.pl}}; \lambda x.addressees\,x \Rightarrow woman\,x} \text{ nSP} \qquad \text{(LXXVI)}$$

The result of (LXXVI) may combine with the result of (LXXIV).

Below I derive a phrase that means ‘You tall women eat the big one (meat)’, which sounds a bit odd in English, but can be grammatically formed in Wao Terero. The parenthesis indicate that the ‘meat’ meaning is supplied by a lexical suffix. There is a choice concerning the placement of the lexical suffix. It may occur on the adjective, verb or both, since the concord is not be grammatical. Two pieces that are needed to supply a poof of a phrase with the intended meaning have been provided so far. In the interest of brevity, I won’t go through the trouble of introducing the form-sign relation before using them in the proof, from here on out. One can use Appendix A as a reference. The proof of the verb needed is below.

$$\frac{\begin{array}{l}\mathsf{di_m} :: \mathsf{b\tilde{\imath}_m} :: \mathsf{k\tilde{a}_m} :: \mathsf{K\tilde{e}_{1m}} \leq_m INF_m \\ \mathsf{verbls}_k \leq_k \mathsf{verb}_k \\ meaning(\beta, \mathsf{di_m} :: \mathsf{b\tilde{\imath}_m} :: \mathsf{k\tilde{a}_m} :: \mathsf{K\tilde{e}_{1m}})\end{array}}{\vdash \lambda \mathrm{tu.t} \bullet \mathrm{u} \bullet \mathrm{k\tilde{e}k\tilde{a}b\tilde{\imath}di}; \mathsf{Nom_{2.pl}} \multimap \mathsf{Acc} \multimap \mathsf{Fin}; \lambda xy.eat\ x\ y \Rightarrow meat\ y} \text{transSP} \tag{LXXVII}$$

If the semantics look a little unsophisticated, above, remember that I am explicitly simplifying. The point to note is that the $meat$ information is included without saturating an argument.

Now that all the sign paradigm entries needed are defined, it is useful to take a look at the $meaning$. I repeat a portion of the $meaning$ definition below.

$$\begin{array}{l} meaning : sense \to list\ m \to Prop := \\ \quad |e_s : \forall cat_m\ (\alpha : e).meaning_e\ \alpha\ cat_m \to meaning\ (\Sigma Sns\ ent\ \alpha)\ cat_m \\ \quad |in_s : \forall cat_m\ (\alpha : e \to p).meaning_{e\to p}\ \alpha\ cat_m \to \\ \qquad meaning\ (\Sigma Sns\ ((func\ ent\ prp)\ \alpha)\ cat_m \\ \quad |tr_s : \forall cat_m\ (\alpha : e \to e \to p).meaning_{e\to e\to p}\ \alpha\ cat_m \to \\ \qquad meaning\ (\Sigma Sns\ (func\ ent\ (func\ ent\ prp))\ \alpha)\ cat_m \\ \ldots \end{array} \tag{LXXVIII}$$

As a reminder, to construct a meaning matching an input M-cat, one performs a sub-proof for a particular meaning. We're not really trying to build a meaning. We're hypothesizing a meaning, and then seeing if we can prove it. Obviously, for the 'eat' verb, above, the route we'll want to take is to prove a transitive, using tr_s. The application of the rule is so: $tr_s(\mathsf{di_m} :: \mathsf{b\tilde{\imath}_m} :: \mathsf{k\tilde{a}_m} :: \mathsf{K\tilde{e}_{1m}}, \lambda xy.eat\ x\ y \Rightarrow meat\ y)$. This requires a proof of $meaning_{e\to e\to p}\ \alpha\ cat_m$. In the definition of $meaning_{e\to e\to p}$ there is a constructor specific to verbls_k.

$$bodyverb\_trans_s : \forall cat_m\ \alpha\ \beta. \tag{LXXIX}$$

$$klasscat_m \leq_k \mathsf{verbls}_k \rightarrow \tag{LXXX}$$

$$bodylsmeaning_s\ \alpha(hd(stripverb_m\ cat_m)) \rightarrow \tag{LXXXI}$$

$$meaning_{e\rightarrow e\rightarrow p}\ \beta\ (tail\ (stripverb_m\ cat_m)) \rightarrow \tag{LXXXII}$$

$$meaning_{e\rightarrow e\rightarrow p}\ (\lambda xy.\beta\ x\ y \Rightarrow \alpha y)\ cat_m \tag{LXXXIII}$$

On line (LXXX) there are some bindings for the category and two meanings, the lexical suffix meaning, α, and the general verbal meaning, β. On the second line there is a check for class. On line (LXXXII) there is a relation specific to the lexical suffix meanings related to body parts. There is a general $lsmeaning_s$ relation, which is articulated into sections for different types of meanings that are specific to particular lexical patterns. These sections, such as $bodylsmeaning_s$, may be appealed to directly outside of the main $lsmeaning_s$ relation when only more specific meanings should be considered. Appealing to $bodylsmeaning_s$ will result in affixes like *-ka* being associated only with heads, *-po* only with hands, and *-bõ* only with eyes, for instance. By appealing to $lsmeaning_s$, these body-part meanings, plus 'rock', 'canoe', and 'seed' may be associated with the three affixes, respectively, in addition to other possible meanings. The sections under $lsmeaning_s$ are less a theory of lexical semantics than a necessary division of meanings due to quirks of Wao Terero's grammar and lexicon. Some verbs only take body part meanings so body-part meanings have to be divided from other meanings. Returning to the proof, within the form-meaning pairs in $bodylsmeaning_s$, a match is attempted between the lexical suffix and a meaning. The hd function should be familiar by now. There is also a $stripverb_m$ function, which removes all inflectional information. This is more or less an implementation detail. Within the $meaning$ relation, as elsewhere, the M-cat is treated like a parse object that may be examined. The use of $stripverb_m$ is used to isolate the lexical suffix. The clauses in $bodylsmeaning_s$, and the related lexical suffix relations all have a fairly simple format, which can be seen below.

$$\ldots \tag{LXXXIV}$$

$$|ka_{head} : \forall m.m = \mathsf{ka_{1m}} \rightarrow bodylsmeaning\ head\ m \tag{LXXXV}$$

$$|k_{body} : \forall m.m = \mathsf{ka_{1m}} \rightarrow bodylsmeaning\ body\ m \tag{LXXXVI}$$

$$|k_{meat} : \forall m.m = \mathsf{ka_{1m}} \rightarrow bodylsmeaning\ meat\ m \tag{LXXXVII}$$

$$\ldots \tag{LXXXVIII}$$

It is more or less, a lookup table. For this portion, I could have provided a list of pairs and simply searched the list, but I decided to use a more elaborate pattern in order to maintain consistency with the other meaning relations.

$$meaning_{e \rightarrow e \rightarrow p}\ \beta\ (tail\ (stripverb_m\ cat_m)) \rightarrow$$

Returning to the discussion of $meaning_{e \rightarrow e \rightarrow p}$ proper, on line (LXXXIII), repeated above, the portion of the M-cat preceding the lexical suffix, essentially the stem, is matched against a meaning β. This is done by proving another instance of the same relation, $meaning_{e \rightarrow e \rightarrow p}$, in a recursive manner. In this case, β and the M-cat will match a non-lexical suffix clause for $\mathsf{K\tilde{e}_{1m}}$, which may be 'cut' or 'eat'. The clause is provided below.

$$eat\_trans : \forall cat_m.cat_m = \mathsf{K\tilde{e}_{1m}} \rightarrow meaning_{e \rightarrow e \rightarrow p}\ eat\ cat$$

This means that the value of α is $meat : e \rightarrow p$ and that of β, $eat : e \rightarrow e \rightarrow p$. The final step is to combine these meanings, and declare an instance of the relation, which can be seen below.

$$meaning_{e \rightarrow e \rightarrow p}\ (\lambda xy.\beta\ x\ y \Rightarrow \alpha y)\ cat_m$$

The relationship between the two meanings is somewhat ad hoc. In reality, there may be many relation types. Yet, it provides an articulation of concerns, which can be elaborated.

That I can prove the definite sign paradigm entry for *yẽdẽkã*, does not require a detailed walk through.

In order to expedite the proof, I will present its lexical entry as Acc, with the knowledge that this would either require an Acc specific clause for SE_{sp} or some application of the trace axiom schema and the order on tectos. Below are the lexical entries for syntactic combination that have been proven. I have aliased the semantic terms, which can be verified against the full forms by the reader.

$$\vdash \text{okiyebĩdi}; \mathsf{N}_{2.pl}; \mathit{you.women} \tag{LXXXIX}$$

$$\vdash \lambda t.\text{yẽdẽbĩdi} \bullet t; \mathsf{N}_{2.pl} \multimap \mathsf{Nom}_{2.pl}; \mathit{tall} \tag{XC}$$

$$\vdash \text{yẽdẽkã}; \mathsf{Acc}; \mathit{big.meat} \tag{XCI}$$

$$\vdash \lambda st.s \bullet t \bullet \text{kẽkãbĩdi}; \mathsf{Nom}_{2.pl} \multimap \mathsf{Acc} \multimap \mathsf{Fin}; \mathit{eat.meat} \tag{XCII}$$

Given these entries, the adjectival modifier may be applied to *okiyebĩdi*, via modus ponens.

$$\frac{\vdash \lambda t.\text{yẽdẽbĩdi} \bullet t; \mathsf{N}_{2.pl} \multimap \mathsf{Nom}_{2.pl}; \mathit{tall} \qquad \vdash \text{okiyebĩdi}; \mathsf{N}_{2.pl}; \mathit{you.women}}{\vdash \text{yẽdẽbĩdi} \bullet \text{okiyebĩdi}; \mathsf{Nom}_{2.pl}; \mathit{tall}(\mathit{you.women})}\ \text{MP} \tag{XCIII}$$

This provides the first argument to the verb.

$$\frac{\begin{array}{ll} \vdash \lambda st.s \bullet t \bullet \text{kẽkãbĩdi}; & \vdash \text{yẽdẽbĩdi} \bullet \text{okiyebĩdi}; \\ \mathsf{Nom}_{2.pl} \multimap \mathsf{Acc} \multimap \mathsf{Fin}; & \mathsf{Nom}_{2.pl}; \\ \mathit{tall}(\mathit{you.women}) & \mathit{eat.meat} \end{array}}{\vdash \lambda t.\text{yẽdẽbĩdi} \bullet \text{okiyebĩdi} \bullet t \bullet \text{kẽkãbĩdi}; \mathsf{Acc} \multimap \mathsf{Fin}; \mathit{eat.meat}(\mathit{tall}(\mathit{you.women}))}\ \text{MP} \tag{XCIV}$$

The final step adds the object.

$$\cfrac{\begin{array}{ll} \vdash \lambda \text{t.yẽdẽbĩdi} \bullet \text{okiyebĩdi} \bullet \text{t} \bullet \text{kẽkãbĩdi}; & \vdash \text{yẽdẽkã}; \\ \mathsf{Acc} \multimap \mathsf{Fin}; & \mathsf{Acc}; \\ \mathit{eat.meat}(\mathit{tall}(\mathit{you.women})) & \mathit{big.meat} \end{array}}{\vdash \text{yẽdẽbĩdi} \bullet \text{okiyebĩdi} \bullet \text{yẽdẽkã} \bullet \text{kẽkãbĩdi}; \mathsf{Fin}; \mathit{eat.meat}(\mathit{tall}(\mathit{you.women}))(\mathit{big.meat})} \text{MP} \tag{XCV}$$

This provides a sample of the system. There are more interesting sentences that I could prove, but the end goal was to demonstrate the morphological system. The other form-sign relation are not significantly different.

One further example that is of interest, though it is very hypothetical, is how multiple lexical suffixes behave on an inanimate noun. Plant vocabulary is fairly productive, and initially convinced me that an arbitrary limit on the number of lexical suffixes that can occur on nouns would be premature. I talk about plant vocabulary in §3.3.5.1.

I have already discussed how there is a kind of recursion in the morphotactics because an LS stem for an inanimate noun, which already has a lexical suffix, is also categorized as LSABLE. Again, I consider the meaning system to be the arbiter of whether a noun with multiple lexical suffixes exists in the syntax. I have been conservative with the nouns I recognize in the *meaning* relation. I only have a few items listed. Yet, what if I wished to express a productive system, where transparent meanings are associated with nouns that contain many lexical suffixes? As an exploration, I designated a number of stems as plant_k, a subclass of inanim_k.

$$\mathit{plant\_inanim}_s : \forall \mathit{cat}_m\ \alpha\ \beta. \tag{XCVI}$$

$$\mathit{klasscat}_m \leq_k \mathsf{plant}_k \rightarrow \tag{XCVII}$$

$$\mathit{plantlsmeaning}\ \alpha\ (\mathit{hd}\ \mathit{cat}_m) \rightarrow \tag{XCVIII}$$

$$\mathit{inmeaning}\ \beta\ (\mathit{tail}\ \mathit{cat}_m) \rightarrow \tag{XCIX}$$

$$\mathit{inmeaning}\ (\mathit{combine}_s\ \alpha\ \beta)\mathit{cat}_m \tag{C}$$

The above is a sub-clause of $inmeaning$, for items of type $e \to p$. The reference to $plantlsmeaning$ is similar to $bodylsmeaning$. The recursive reference to $inmeaning$ should also be familiar. What is slightly different is that the meanings of the suffix and the stem are combined using the function $combine_s$, which is fairly unsophisticated and simply applies a composing function to the arguments. I'll replace it with an infix c, below. As the system is defined, the plant_k status of the stem does not change as affixes are added, one can recursively build meaning on to the base. For the hypothetical $yabo_m :: w_m :: ka_{1m} :: Daa_m$, one could potentially end up with a meaning similar to 'thorn fruit tree leaf', for the leaf of a tree with thorny fruit. With the first pass through $plant\_inanim_s$, there is a meaning $leaf$, which may be combined with some stem ($leaf\ c\ inmeaning\ stem$). The second round of $inmeaning$ results in finding the meaning $tree$ for $\mathsf{w\tilde{e}_m}$, resulting in ($leaf\ c\ tree\ c\ inmeaning\ stem$). Following this there is the fruit meaning ($leaf\ c\ tree\ c\ fruit\ c\ inmeaning\ stem$). Finally, $\mathsf{Daa_m}$ matches against a different clause, specific to the $thorn$ meaning, resulting in ($leaf\ c\ tree\ c\ fruit\ c\ thorn$). Does this mean 'thorn fruit tree leaf'? Not really, there isn't any real interpretation offered by $combine_s$. Yet, as a proof of concept it serves a purpose. If the relationship where consistent between stem and affix, the $combine_s$ function could be replaced with that relationship. On the other hand, if there are very specific relationships allowed, one may add information to the $meaning$ relation to match particular patterns and apply specific, rather than general relations. The versatility is there. The plant_k pattern may be too general for any language, but the tools exist, even in this highly imperfect lexical semantics, to explore complex and semi-recursive morphological meanings. This is why the realizational infrastructure has the power to relate the forms and meanings of inflectional, derivational, and lexical content in a uniform manner.

In the next section, I present additional WP functionality. I describe rules that allow for one paradigm instance to be instantiated using another paradigm instance of the same class as an example. The biconditional rules that were promised in the introduction are described. These ingredients make the framework a truly WP theory. What has been presented, so far, is largely the scaffolding for these qualities.

6.9 WP Properties

There are a number of things which a WP theory should provide. First, there should be a notion of biconditional relationships between members of an instance of a particular paradigm. Second, the existence of a form and category in one paradigm instance should imply the existence of a similar form and category in another instance of the same paradigm. The form should differ only in its stem between the two instances, but otherwise have the same category and form. Further abstraction should allow for inferences even when allomorphy is involved, but I do not supply such mechanisms here. The first condition is conceptually dependent on the second. In order to demonstrate biconditionality, reference is made to more than one paradigm instance. This may be done reflexively for paradigms with a single instance, but one cannot prove anything that has not already been proven by axiom-like constructors or form-form mappings. For this reason, prt_k may be shown to have biconditionality among its form paradigm entries, though proofs using biconditionality are redundant. The second condition also implies that a paradigm instance may have all of its form paradigm entries constructed by comparison to other paradigm instances.

Before continuing, I wish to provide some orientation to what follows. First, there is no data structure defined for paradigm instances. Inferential reasoning relies on relational form class patterns. A paradigm instance is identified by a stem and a form class. Reasoning about the instance precedes from there. I do discuss equivalence classes that exist within and between paradigm instances. Equivalence classes have a graph interpretation, but I do not provide such an interpretation. I only go so far as to discuss proofs of equivalence classes using rules within defined within the system. I see no benefit to providing a data structure interpretation of paradigm instances at this time.

A second point is that reasoning about paradigm instances precedes from existent form paradigm entries. There may be rules that could apply to form paradigm entry with particular properties, but if such a form paradigm entry cannot be proven from existent form paradigm entry, those rules cannot apply. For instance, for participant pronouns I proposed that the *-bi* form should be supplied by axiom, allowing for all other forms to be proven from that base. If the *-bi* form axiom were to not exist, the paradigm instance would not be provable. It may be the case that some paradigms instances can only be partially proven due to similar sparseness. Even with sparseness of proven form paradigm entries one can prove biconditionality among some limited number of form paradigm entries, but empirically it may be clear that sparseness has resulted in poor coverage. This is not an issue with formal validity, it is due to a theoretical stance, where actual examples are needed in order for reasoning to occur.

One may ask why there would ever be sparseness. One reason is that sparse paradigms allow for theoret-

yẽdẽ ‘big’	giitã ‘small’	wẽtabõ ‘dark’	dãbẽta ‘light’
$\langle \textsf{Yẽdẽ}_{\textsf{m}}, \text{yẽdẽ}_{\text{pr}} \rangle$			
	$\langle \textsf{wẽ}_{\textsf{m}} :: \textsf{Giitã}_{\textsf{m}}, \text{wẽ}_{\text{pr}} :: \text{giitã}_{\text{pr}} \rangle$		
		$\langle \textsf{bi}_{\textsf{m}} :: \textsf{wẽtabõ}_{\textsf{m}}, \text{bi}_{\text{pr}} :: \text{wẽtabõ}_{\text{pr}} \rangle$	

Table 6.12: Four partial instances of the adj_k form class.

ical thought experiments about paradigmatic reasoning. Another is that paradigm instance sparseness is an empirical fact. There are no natural corpora where there are attested instances of each form of some word. Therefore, a central issue in WP theories is the so-called cell filling problem.

A third point is that when one chooses a stem to identify a paradigm instance, that stem need not be the simplest stem. For instance, for *kẽ*, ‘eat’, one could choose the root, a form with a lexical suffix, *kẽ-po*, or a form with person marking, as well, *kẽ-po-bĩ*. One is free to reason about sub-instances and their relationships to other sub-instances, rather than the whole of the paradigm instance. A full instance is informally understood to corresponding to all of the form paradigm entry of the same form class that share the same minimal stem category, which may or may not correspond to a root, since derivation may result in form class changes.

A final point is that a paradigm for a stem is identified by the form class returned by $klass$. It is not identified using a form class greater than that form class in the hierarchy. The base form class are expected to have instances where the forms differ only in their stems, which may or may not be analyzable. If this is not the case for some reason, it is considered an error in the analysis.

To demonstrate the rules that allow for inferential reasoning within the system, I will provide a simple example using the adj_k class. I confine the example to three M-cat and four adjectives. These can be seen in Table 6.12. Omitting the FE_{mp}, the first three columns have only one form paradigm entry defined. The last column is empty. The goal in this section will be to fill all of the empty spaces in Table 6.12 without using $rule_1$ or $rule_2$ based form-form mapping.

To do so, a rule is introduced, a constructor, which is a clause under the inductive definition of FE_{mp}, which I call $interchangeMP$. Its definition is in (CI).

$$
\begin{aligned}
&\mathit{interchangeMP} : \\
&\qquad \forall\, (\mathit{inStem}\ \mathit{outStem}\ \mathit{inWord}\ \mathit{outWord} : \mathit{struct}_{mp})(\kappa : K). \\
&\qquad \mathit{FE}_{mp}\ \mathit{inWord} \\
&\qquad \to \kappa = \mathit{klass}\ (\pi_1\ \mathit{inStem}) \\
&\qquad \to \kappa = \mathit{klass}\ (\pi_1\ \mathit{outStem}) \\
&\qquad \to \kappa = \mathit{klass}\ (\pi_1\ \mathit{inWord}) \\
&\qquad \to \mathit{isStem}_{mp}\ \mathit{inStem}\ \mathit{inWord} \\
&\qquad \to \mathit{isStem}_{mp}\ \mathit{outStem}\ \mathit{outWord} \\
&\qquad \to \mathit{remove}_{mp}\ \mathit{inStem}\ \mathit{inWord} = \mathit{remove}_{mp}\ \mathit{outStem}\ \mathit{outWord} \\
&\qquad \mathit{FE}_{mp}\ \mathit{outWord}
\end{aligned} \tag{CI}
$$

The definition is straight forward. Two stems are provided from different paradigm instances. Then an existing word is provided from the paradigm instance of the inStem. Finally, a target $\mathit{outWord}$, which may not have been otherwise proven, is supplied for the paradigm instance of the $\mathit{outStem}$. A form class is also provided.

On the following line, only a proof that the inWord is a form paradigm entry is required. No tests are performed on the stems. The reason for this is that paradigmatic patterns may be recognized for stems of different sizes. Any word-form in a paradigm instance serves as a form class identifier, since there must be some klass value associated with its M-cat. In addition, many non-FE_{mp} struct_{mp}s have a $\mathit{list}\ m$ that resolves to a form class. In some cases it is necessary to use these for paradigmatic inferences. This is due to the existence of bound stems. For instance, Di_{m} is a singleton that can be identified as a member of dika_k, but it is a bound root, and so cannot serve as an M-cat.

After ensuring that inWord is a valid form paradigm entry, checks are required to ensure that the stems and inWord are all of the required form class. The predicate isStem_{mp} checks that inStem is actually a stem of inWord, and that $\mathit{outStem}$ is actually a stem for $\mathit{outWord}$. Finally, the function remove_{mp} strips out the stems, and an equal length of process information. The inStem is removed from inWord and $\mathit{outStem}$ from $\mathit{outWord}$. The results of the removals are compared to ensure that no more or no less than the difference between the $inStem$ and $inWord$ is proven to be applicable to $outStem$ to achieve $outWord$. If these conditions are met, $outWord$ is a form paradigm entry.

The basic idea is that one starts from a premise that some target word-form exists in some paradigm instance, which corresponds to a source form paradigm entry in another instance. The source and target differ only in their respective stems. The stems of the source and target word-forms may or may not be form paradigm entries, but they are believed to contain, at least, the non-identical portions of the source and target. The stems must be identifiable as being of the same form class as the source and target, the inference cannot be made across paradigm boundaries. The stems may actually be of different size, which allows stems formed by complex derivation to be compared with unanalyzable stems. The core notion is that if one stripped out the stem of the source, and replaced it with the stem of the target, the result would be the target. This metaphorical exchange of stems is why I named the constructor $interchangeMP$.

Given this, one can see how Table 6.12 may be filled. One may state both of the following, and provide the necessary proofs:

$$\begin{aligned} &interchangeMP \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \text{y\~ed\~e}_{\text{pr}} \right\rangle \left\langle \mathsf{Giit\tilde{a}_m}, \text{giit\~a}_{\text{pr}} \right\rangle \\ &\qquad\qquad\qquad \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \text{y\~ed\~e}_{\text{pr}} \right\rangle \left\langle \mathsf{Giit\tilde{a}_m}, \text{giit\~a}_{\text{pr}} \right\rangle \ \mathsf{adj}_k \\ &interchangeMP \left\langle \mathsf{Giit\tilde{a}_m}, \text{giit\~a}_{\text{pr}} \right\rangle \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \text{y\~ed\~e}_{\text{pr}} \right\rangle \\ &\qquad\qquad\qquad \left\langle \mathsf{w\tilde{e}_m} :: \mathsf{Giit\tilde{a}_m}, \text{w\~e}_{\text{pr}} :: \text{giit\~a}_{\text{pr}} \right\rangle \left\langle \mathsf{w\tilde{e}_m} :: \mathsf{Y\tilde{e}d\tilde{e}_m}, \text{w\~e}_{\text{pr}} :: \text{y\~ed\~e}_{\text{pr}} \right\rangle \ \mathsf{adj}_k \end{aligned} \tag{CII}$$

Notice how this leverages the category representation and the process abstraction system. What is possible to do now is define an equivalence relationship between form paradigm entries of different instances of the same form class. This allows for statements such as the following.[5]

$$FE_{mp} \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \text{y\~ed\~e}_{\text{pr}} \right\rangle \sim_{\mathsf{adj}_k} FE_{mp} \left\langle \mathsf{Giit\tilde{a}_m}, \text{giit\~a}_{\text{pr}} \right\rangle \tag{CIII}$$

The '$\sim_x$' can be read as iff, similar to, or congruent to. Since '$\sim_x$' only holds within a form class, its form class is indicated with a subscript. The equivalence relation is not defined using axiom-like rules in system,

[5]I do not actually define an operator $\sim_x$. It is only presented as an example. That the relation can be proved is sufficient.

since it may be proven. In order to be an equivalence relation, three conditions must be met, reflexivity, transitivity, and symmetry. Reflexivity is demonstrated by examples such as the following:

$$\mathit{interchangeMP}\ \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle$$

This can be generalized to $\forall\alpha.\mathit{FE_{mp}}\ \alpha \rightarrow \mathit{interchangeMP}\ \alpha\ \alpha\ \alpha\ \alpha$. Transitivity is clearly the case for *interchangeMP*. Symmetry requires that $\forall\alpha\ \beta.\alpha \sim \beta \wedge \beta \sim \alpha$. Symmetry is also clearly true of *interchangeMP*. For single instance paradigms, symmetry and transitivity require that the *outWord* form paradigm entry is the *isWord*, which has already been proven by other means.

An issue with *interchangeMP* is that it assumes an analysis where *m*s and processes are in a one-to-one relationship. In order to generalize the rule to other analyses the structure of $\mathit{struct_{mp}}$ must either change to explicitly group *m*s and processes, or process information will need to exist in the input of *interchangeMP*. That being said, in this context, *interchangeMP* serves the purpose for which it was designed, and can easily be adapted or generalized for other analyses.

Given that an equivalence relation exists between paradigm instances, the goal is to leverage that relation to provide biconditionality within a paradigm instance. Considering Table 6.12, there is now a horizontal equivalence relation, but we also want a vertical equivalence relation. The import of the relation is that if one has proved $\mathit{FE_{mp}} \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle$ they can prove $\mathit{FE_{mp}} \left\langle \mathsf{we_m} :: \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{w\tilde{e}_{pr}} :: \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle$. If one has proved $\mathit{FE_{mp}} \left\langle \mathsf{we_m} :: \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{w\tilde{e}_{pr}} :: \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle$, then they can prove $\mathit{FE_{mp}} \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle$.

This equivalence relation can partially be proven using combinations of *interchangeMP* and form-form mappings, though any contribution of form-form mappings will be ad hoc, and specific to an analysis. Reflexivity can be demonstrated using *interchangeMP*, as above. As an example of how form-form mappings may be used in the relation, rules such as *wMP*, which appends *-wẽ*, demonstrate that $\mathit{FE_{mp}} \left\langle \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle$ implies $\mathit{FE_{mp}} \left\langle \mathsf{we_m} :: \mathsf{Y\tilde{e}d\tilde{e}_m}, \mathsf{w\tilde{e}_{pr}} :: \mathsf{y\tilde{e}d\tilde{e}_{pr}} \right\rangle$, but there is no rule to provide the reverse implication. A more general solution is provided by the constructor, *analogyMP*. Its definition can be seen in (CIV.[6]

[6] The line Proof steps of: *interchangeMP inStem medStem inWord medWord*$_1$ κ is not valid. I abuse the notation here to give the reader an idea of the relationship between the two rules. See Appendix A for the actual definition.

$$
\begin{aligned}
& \mathit{analogyMP} : \\
& \qquad \forall\, (\mathit{inStem}\ \mathit{medStem}\ \mathit{inWord}\ \mathit{medWord}_1\ \mathit{medWord}_2\ \mathit{outWord} : \mathit{struct}_{mp})(\kappa : K). \\
& \qquad \mathit{FE}_{mp}\ \mathit{medWord}_2 \\
& \qquad \to \kappa = \mathit{klass}\ (\pi_1\ \mathit{medWord}_2) \\
& \qquad \to \mathit{isStem}_{mp}\ \mathit{medStem}\ \mathit{medWord}_2 \\
& \qquad \to \mathit{isStem}_{mp}\ \mathit{inStem}\ \mathit{outWord} \\
& \qquad \to \text{(Proof steps of: } \mathit{interchangeMP}\ \mathit{inStem}\ \mathit{medStem}\ \mathit{inWord}\ \mathit{medWord}_1\ \kappa) \\
& \qquad \to \mathit{remove}_{mp}\ \mathit{medStem}\ \mathit{medWord}_2 = \mathit{remove}_{mp}\ \mathit{inStem}\ \mathit{outWord} \\
& \qquad \mathit{FE}_{mp}\ \mathit{outWord}
\end{aligned}
\tag{CIV}
$$

The argument list is increased in comparison to *interchangeMP*. The *inStem* and *medStem* are similar to the *inStem* and *outStem* arguments of *interchangeMP*. The *inWord* is a proven form paradigm entry, with a stem of *inStem*. The argument $\mathit{medWord}_1$ is a potentially unproven form paradigm entry in a paradigm instance of forms with *medStem* stems. The argument $\mathit{medWord}_2$ is a proven form paradigm entry from the same paradigm instance. The *outWord* is an unproven form paradigm entry from the instance of *inWord*. The idea is that, given the *inWord*, we can prove that there is some similar $\mathit{medWord}_1$ in another instance. If that instance already has a proven member, $\mathit{medWord}_2$, with the same stem as $\mathit{medWord}_1$, then there is an *outWord* similar to $\mathit{medWord}_2$, with the same stem and paradigm instance as *inWord*. Basically, the *inWord* to $\mathit{medWord}_1$ relationship establishes the shared paradigm. If the paradigm is the same, then something that is true in one instance should be true in the other. Given the previous description of *interchangeMP*, the rest of the constructor definition should be clear.

Given *analogyMP*, the intra-instance symmetric and transitive qualities of an equivalence relationship may be trivially proven. What *analogyMP* expresses is something fundamental to WP theory, which is that intra-instance inferences depend on comparison to other paradigm instances. With *analogyMP* and *interchangeMP*, one could potentially do away with form-form mappings and instead provide example form paradigm entries sufficient to fill all gaps. The fact that both kinds of reasoning are represented in the theory is because I believe that both are relevant to morphological systems. I believe that rule based coinage is possible using form-form mappings, such that one can add categories to a paradigm in a rule governed manner. Once the coinage is established, it may stand as an example for other instances of the same paradigm.

A number of theories assume something like the equivalence relationships I have described above (Singh and Starosta, 2003; Bonami and Strnadová, 2019). Yet, I have never seen a means of providing proof. The relation is generally simply taken as basic. It appears that two things are necessary for such a relation. One is a specific class over which the relation holds. If one can't say that everything in the class follows the same formal pattern, it is a non-starter. The second is that one must explicitly appeal to other instances of a class, at least if sparseness within instances is assumed, which it generally is in stochastic studies of paradigms (Ackerman and Malouf, 2013), though not in semi-formal treatments, where paradigm relations are defined discretely (Singh and Starosta, 2003).

An innovation of my approach compared to others is that it handles lexical and derivational patterns in exactly the same way as inflectional pattern. Other theories tend to assume that one can know all of the cells, or categories, ahead of time. This simply isn't true, except for very carefully chosen samples. Positing that a paradigm structure's bounds are already known is not at all necessary in this framework. Intra-paradigm instance proofs rely on inter-paradigm instance proofs, not on some top down knowledge of the whole. The total size, in terms of categories, of any paradigm is bounded by the number of M-cats defined in the system. This is due to $combine_{mp}$ failing to allow repetition of ms. Yet, in some cases, such as a fully specified lexical suffix system, the number of permutations allowed could be quite large. Even if the Wao Terero system proves to be formally more constrained for the inanim_k paradigm than my analysis entails, there are other productive systems in the world's languages where, even if the number of possible word-forms appears finite, the bounds of the system are not obvious, and may grow according to incremental rules, where by incremental, I mean that for every process, some category information is added. This does not mean that stochastic measurements of paradigms that rely on the simplifying assumption of easily bounded paradigms are incompatible with my system. To perform such measurements one would need to chose some bounding heuristics, but it is possible.

All of the relations above concern form paradigms. The theory of sign paradigms is less developed at this time. It is dependent on information used in the morphological domain, which for the most part works in grouping sign paradigm entries. This doesn't make it obvious how the theory handles suppletion. It was not an important issue in this thesis given that there is no suppletion in Wao Terero. In Chapter 7, I review some options for suppletion that are available within the framework, though they may require minor extensions.

For sign paradigm instances, I provide a limited grouping mechanism, which shows that a sign paradigm entry is an instance of a class that is identified by a stem and form class. One does so by demonstrating that one can utilize a form paradigm entry with the stem and form class to construct a rule that constructs a sign paradigm entry equal to the target sign paradigm entry. Unlike the rules above, the notion of membership is expanded to include form classes greater than what is returned by $klass$. This is because the form classes relevant to signs are generally more abstract. No data structure for sign paradigm instances is automatically provided. There is only a means of demonstrating membership in an instance. See the predicate in_{sp} in Appendix A for the implementation details.

Chapter 7

Conclusion

In this concluding chapter, I begin by briefly discussing issues relevant to the sign paradigm and suppletion. I then consider whether my framework is truly realizational, given that the architecture is different from previous approaches. I also discuss some claims that have been made that words don't exist, which would be problematic for my theory. I conclude with a discussion of what was achieved with the analysis, and the extent to which the framework allowed me to resolve issues presented in the introduction.

7.1 Lexemes and the sign paradigm

An area of incompleteness in this work is the articulation of sign paradigms. The only organizational predicate, in_{sp}, makes an appeal to the form paradigm and form class used to construct a sign paradigm entry. This is actually fairly adequate. In Paradigm Function Morphology (PFM), the usual case is to have one stem per lexeme.

One issue that some might have with an appeal to form-oriented categories is that homophones such as *bore*, which can mean 'to make disinterested' or 'to make a hole', would not have their paradigm instances differentiated, unless some syntax oriented form class or M-cat were to be defined. Differences in form class would be justified if there were demonstrable differences in syntactic distributions, which would affect the kinds of sign paradigm entries needed in each case. If the two items are syntactically identical, except in their lexical semantics, it may not be necessary to differentiate them in a syntactic sense. The lexical meanings are already differentiated semantically, and sign paradigm entries are not semantic paradigm entries.

Therefore, semantics should not delineate sign paradigm instances. Items that are considered polysemous such as *draw* and *run* also have unrelated meanings, so it may be legitimate to consider the semantics to be of little importance in a syntactic paradigm. That is to say, the tecto and the pheno should define the paradigm, allowing the semantics to vary. Given that sign paradigms are *syntactic* paradigms, that makes sense to me. In fact, it brings some order to sign paradigm instances that is needed in any case. Given that polysemy can offer many interpretations for items like *run*, one wouldn't want to say that these are paradigm defining.

Another issue with references to form-oriented categories involves suppletion. I didn't focus on suppletion in this work because Wao Terero doesn't have suppletive items, neither does it exhibit much allomorphy. For that reason, I adopted a simplifying assumption that stems are uniform in sign paradigms instances. Despite this, grouping suppletive items in some way is not difficult, though it necessarily deviates from form-orientation. Some suppletion can be treated as allomorphy, requiring some elaboration of the M-cat and form class system, but otherwise, not necessarily distinct from allomorphy in inflection. One analysis for some types of allomorphy involves the same m for distinct processes, with the form class managing the division. This may be a means of handling verbal Spanish allomorphy in cases such as the imperfect past tense, where the AR verbs have a base of *-aba* and ER and IR verbs have a base of *-ía*, but otherwise have the same pattern of person and number morphology. Another is to utilize both distinct m and form classes, but to use the M-cat hierarchy to express the shared category of certain items. For instance, English *hit* is of a form class that does not take a past tense affix, while *love* is. The M-cat hierarchy can express that something like $\mathsf{hit_m}$ and $\mathsf{d_m}$ :: love share a $\mathsf{PAST_m}$ category. Both strategies may be applied to suppletion. Essentially, ms are form-*oriented* but not form-*detemined*.

In some cases, simple axiom-like rules are sufficient for suppletion. One could add a type constructor to prove $\langle \mathsf{d_m} :: \mathsf{go}, \mathrm{went_{pr}} \rangle$ for English *went*. The predicate in_{sp}, which is used to prove whether pairs of sign paradigm entry are in the same paradigm, makes reference to form paradigm instances, but it could be altered to make reference to only the M-cats, rather both projections of $struct_{mp}$, which includes process information. This would be a minimal change allowing for special cases like *went*. I don't necessarily advocate the approach, since I have not tested its consequences for an English fragment, but it is an example of how the system can accommodate suppletion in a trivial manner.

Despite the fact that there are options for handling suppletion, I do not wish to commit to a uniform approach to suppletion. This is because suppletion is not a uniform phenomenon that should have only one analysis. I would rather do more than accommodate the diversity. Instead, I would like to provide an articulated theory that respects the diversity.

7.2 What does it mean to be realizational, really?

There is a case to be made that the theory I presented in this work is not realizational. By adopting the term *realization* in this work, I force a distinction between post-syntactic realization, and what one might call generative realization, or form-oriented realization. This distinction has not been made previously in the literature. I may be hijacking the term *realization* to mean something similar but different from previous usage. Despite this, I believe it is useful to call my approach realizational.

Generative form-oriented realization is explicitly contrary to the definition of realization provided by post-syntactic authors. I have not encountered a self-described realizational theory in the literature that is not explicitly post-syntactic. For instance, Stump (2001, p. 2) states "According to realizational theories [...] a word's association with a particular set of morphosyntactic properties licenses the introduction of those properties' inflectional exponents[.]" Zwicky (1985a, p. 372) is perhaps the earliest users of the term in a precise sense, and says that realization rules "refer to principles describing when and how morphosyntactic features are realized as morphological processes. These principles belong in a morphological component, which follows a syntactic component[.]" Aronoff (1994, p. 13) specifically endorses this definition, which he also equates with the *morphological spelling component* of Beard (1995), an important precursor to Distributed Morphology (DM). Carstairs-McCarthy (1987, p. 6) also describes realization as being relative to morphosyntactic features. Anderson (1992, p. 41) is explicit about relative ordering of syntax prior to morphology, and the importance of morphosyntactic features. He does not invoke the term realization, but his works are clearly within the realizational cannon. I have never encountered a work that invokes the term that does not emphasize that realization means post-syntactic. There also is no story in these theories of how morphology might sidestep syntax and be realized relative to semantic information in the compositional, lexical, sociological or discourse domains. Such information, to the extent that it is taken into account, must be recoverable from information encoded directly in the syntax as a feature, lexeme or $\sqrt{}$root. The question then is, what room is there to have a distinct definition of realization given this well established tradition?

Perhaps, it would be clearer to say that realizational theories are just those theories that have adopted the term historically, and which behave essentially the same. Yet, as discussed in Chapter 6, particularly in Paradigm Linking Theory (PLT), with its multiple tiers of paradigms, some realizational theories are shifting toward a less syntacto-centric position. In PLT there is no longer a function from morphosyntactic features on syntactic nodes to a realized paradigm. Between the form and content paradigms of PLT, there is a *relation*. This means that there is no functional directionality between the paradigms. The feature set σ is no more primary than τ. As defined, one cannot say that one is more dependent on the other. For that reason, despite some continued machinery that relates to the legacy Stump (2001) PFM, PLT could be considered to be something other than post-syntactic. I suspect that the trend away from syntacto-centrism will grow. There are empirical reasons for its abandonment in PLT that one cannot avoid, such as overabundance. For that reason, my framework will likely look increasingly less exotic in time.

7.3 The existence of words

Haspelmath (2017) details reasons why a notion of morphosyntactic word is flawed. The basic premise is that the concept of morphosyntactic word, which functions as an opaque syntactic unit, may not be discovered using a reliable standard set of cross-linguistic diagnostics. I begin this section by discussing how the theoretical framing of some arguments in the work make engagement with their substance complex. I discuss the alternative to a grammar of word tokens and phrases that Haspelmath argues for, and ask whether it offers any clear advantages. I then introduce Haspelmath's notion of the *criterion of independent occurrence* and the question diagnostic, which despite the author's claims to the contrary, is an excellent means of delineating free word-forms and bound content. I compare the question test to diagnostics for bound content and phrasal constituents, concluding that these have their own issues as objective cross-linguistic measures.

I do not engage with the entirety of Haspelmath (2017). It requires deconstruction and examination of many frames imposed by the way questions are posed in the literature. As one example of framing, Haspelmath discusses arguments against the notion of anaphoric islands (Postal, 1969), where items within a word cannot serve as antecedents to anaphora, nor may they serve as anaphora to outside antecedents. In my discussion of anaphoric properties of classifier constructions in §3.3.6, I discussed how negation diagnostics reveal that even single morph items, such as English pronouns, can be revealed to have multiple meaning associations, some of which may be canceled under negation, while others are not. Under a specific notion of the relationship between syntax and semantics, such a multiplicity of meaning would require that the pronoun must be syntactically elaborated in order to both encode referential properties and descriptive content. If this were done, English pronouns could all be said to have some node within the whole that represented the anaphoric nature of the pronoun, which would be distinct from the node of descriptive content. If such nodes are associated with morphemes, as in Déchaine and Wiltschko (2002), anaphoric islandhood is violated. In Linear Categorial Grammar (LCG), an English pronoun can be treated as atomic in the tecto, while encoding the complexities of its meaning in the semantics, alone. This would preserve anaphoric islandhood in a trivial manner, since there are no morpheme nodes. These are not the only two theoretical options, of course. It takes a lot of work to untie assumptions in the area of lexical integrity and word-hood. Argumentation often only applies in a theory internal manner. For this reason, engaging with all of Haspelmath's argumentation requires a work of large scope.

Haspelmath's goal is to argue that the treatment of morphological forms and syntactic phrases should be collapsed. He claims that both morphology and syntax are a system of "sign combinations" (Haspelmath, 2017, p. 33). This is an ambitious conjecture. Even in DM, which largely collapses syntactic and morphological combination, the system is not strictly sign-combining. Not all bound content is a √root. In fact, with late insertion models (Harley, 2014), sign-like relationships of form and meaning for √roots are only occasional accidents. He makes claims such as "It is now very widely recognized that many complex words are semantically compositional in exactly the same way as phrases and clauses" (Haspelmath, 2017, p. 36). My understanding is that semantic transparency of morphological forms remains an open area of research, where considerations such as token frequency and storage are relevant to a greater extent than they are in the phrasal semantics literature (Johnson, Elsner, and Andrea D. Sims, 2023; Bybee, 1985; Baayan, 1993). One interesting study is Dąbrowska (2014), who demonstrated that implicit morphological knowledge allows English speakers to guess the meanings of unfamiliar words above chance. Yet, this does not mean that speakers of English can throw out their dictionaries. Even though speakers were given multiple choices from which to pick a definition of an unfamiliar word, thereby increasing the likelihood of being correct beyond what one might expect if they were forced to provide their own definition or paraphrase, they were only able to guess a correct meaning of an unfamiliar word 45% of the time on average. The study is illustrative. The results seem far different from what one would expect for the comprehension of phrases that have never been encountered before. No results are found in DuckDuckGo, the search engine, for *the orange computer chip*, indicating that it is a rare phrase, yet I would expect a very high rate of understanding from people who share my language and culture. Creating a truly parallel study comparing performance at guessing the meanings of phrases versus words would not be trivial, but thinking about what it would entail to do so suggests that a word isn't just a short phrase. For one, it is odd to even talk about phrases that one knows or does not know. There is some difference in our assumptions concerning storage, at the least.

To deepen the parallel, Haspelmath points to phrasal idioms, which have opaque meanings. Yet, a big difference between reasoning about unfamiliar words and many syntactic idioms is that something like *kick the bucket*, can be interpreted as striking a pail with one's foot, which is perfectly legitimate. There are idioms that have no compositional interpretation. Authors such as Culicover and Jackendoff (2005), Goldberg (1995), and Fillmore, Kay, and O'Connor (1988) have pointed out many interesting examples where sentence meanings are not the sum of their parts. But there are some very good reasons to believe that reasoning about word meanings is different from how one reasons about phrasal compositions. Whatever the difficulties in word measurement the author exposes, the proposed alternative of treating both morphology and syntax as a single system of composing signs is not obviously better. I also believe that Haspelmath overstates the measurement issue.

The concept of word-form is not immeasurable cross-linguistically, even according to Haspelmath's own account. A key quote is found below.

> [T]he criterion of independent occurrence is helpful at least in one regard: If an element can occur independently (i.e. as a complete utterance all by itself), it clearly cannot be an affix but must be minimally a morphosyntactic word. (Haspelmath, 2017, p. 40)

The criterion of independent occurrence essentially states that words are free forms, which can occur alone. The means suggested for isolating non-bound content by the author involved question answers. There are known issues for such a diagnostic. As Haspelmath points out, answers to questions may not isolate a morphosyntactic word. There is ambiguity due to compounds, as well as additional material in an answer, such as an article. He also points out other issues. None of these negate the utility of the diagnostic.

As a simple example of a question diagnostic, one may point at a flower and ask *What is this?* The resulting answer may be something like *a morning glory*, *a blue flower*, or *It's a flower, obviously.* Clearly, such answers are not guaranteed to isolate word-forms. But there is a guarantee, which is very powerful. Bound items will never occur alone in the answer. This means that as part of a first pass in an iterative process of word-form discovery one can isolate a set of non-bound content, some of which may be word-forms, simply by asking run-of-the-mill questions. Further iterations are needed to further isolate minimal free forms, but the process is straight forward. One starts with general questions and progresses to more targeted questions, which require greater knowledge of the language in question. For instance, finding that *blue* is a word may require learning to ask about colors. For instance, *What color is this flower?* One can also ask whether a specific question answer is acceptable. For instance, *If I asked what color this flower is, could someone answer 'blue'?* This process of seeking out word-forms by iterative questioning is a sound practice.

It is worth comparing the question diagnostic to the discovery procedures for other grammatical units.

Consider performing constituency tests cross-linguistically. Question answering also works for gaining some notion of constituency. Words are constituents, as are other non-bound content. The issue is that one can't be sure what kinds of constituents they are getting in answers, and simply drilling down to the smallest units that can serve as an answer isn't the goal. Constituency tests need to impose some additional sentence structure around the point of interest to see where a piece fits in the syntactic system. For this reason, constituency probes tend to be highly language specific. One may need to have a very good idea of what a grammar is like before beginning to make headway. As an example, in English, some constituents may be recognized using a cleft construction, such as *It was the clown who jumped on the mattress.*, which isolates *the clown* within a specific phrasal context. There may or may not be such a construction available in another language, and it may take some work to find out if it does. Constituency tests tend to require batteries, since one kind of test cannot isolate every kind of constituent. The cleft construction cannot isolate verb phrases, * *It was jumped on the box which ....* Various types of ellipsis constructions also complicate the concept of phrase and constituent. Constituency behaves in radically different ways across languages, so much so that there are debates concerning whether some languages have word order. It is also the case that diagnostics that have been relied on in the literature have later proven to be far less reliable than previously assumed when closely examined. For instance Houser (2010) raises major concerns for the English *do so* diagnostic for verb phrases, demonstrating that even in English more work needs to be done to provide reliable diagnostics. Therefore, the diagnostic foundations for phrases look at least as shaky, if not more, than for word-forms.

Likewise, it is not the case that isolating units smaller than a word-form is free of complications, or more amenable to cross-linguistic uniformity of diagnostics, especially if word-forms are not considered legitimate units of the grammar. Consider the diagnostics necessary to determine whether some (possible) aggregate of lexical suffixes were multiple suffixes or only a single suffix in Chapter 3. Looking at verbal usage, as I did there, was an extremely language specific criteria, which I only thought to employ after considerable experience with the system. Diagnostics for affixes presuppose a word-form for minimal pair comparison. I suppose that the minimal pairs for affix identification could be complete phrases. That may create some difficulties when investigating affixes that act as co-signals with word order, such as English passive morphology. The phrase *The pie was eaten.* does not pair with *The pie was eat.* The issue may not be great for the English passive, since the affix also occurs in other contexts, but one could imagine an unfamiliar language where such issues make it difficult to find key sentence-sized minimal pairs.

The point is, Haspelmath describes a means for word-form discovery that is at least as sound as the discovery procedures for phrasal constituents or bound content. The only thing one must be aware of when using question-like diagnostics to isolate word-forms is that one must continue their exploration beyond an initial naive probe, and augment their findings with additional diagnostics.

Haspelmath's ultimate argument for dismissing independent occurrence is the following.

> The criterion of independent occurrence does not have real practical value, it is not surprising that it is hardly used in recent wordhood controversies. It seems that it is mentioned in discussions of wordhood primarily because of Bloomfield's (and Hockett's) authority. (Haspelmath, 2017, p. 40)

I find the notion that modern linguists unquestionably accept the works of Hockett and Bloomfield odd. Even then, it is not a strong argument, since one must assess the quality of the authority's arguments in order to counter them, rather than simply suggesting that those who repeat them are doing so uncritically. Likewise, the popularity of an idea in the current literature is not a measure of its merit.

Haspelmath's substantive argument against probing for independent occurrence is that one may get question answers that are not single word-forms. This is true, but it does not demonstrate an appreciation of the iterative process of probing that one must always do in a fieldwork setting.

7.4 Goals achieved in analysis

The primary goal of the analysis was to provide a realizational treatment of lexical content in a manner consistent with functional, grammatical content. This might imply that the goal was to reduce representations and perform realization relative to a single kind of source. In fact, I diversified the relationships that morphs and word-forms may have to meaning and category in comparison to post-syntactic theories. For instance, the correspondence between form and meaning for a single person marker, such as *-bo*, is not homogeneous. The same form may correspond to a semantic expression, or a purely grammatical category, depending on its role as agreement or as a primary signal.

Lexical suffixes are licensed based on semantic criteria, which requires that their conventional semantic options are listed in a lexical semantics. Feature-based competition is ill-suited to such a task. An advantage of the approach is that it also allows for a well-articulated means of realizing items associated with lexical content, in general. The DM late insertion literature (Acquaviva, 2009; Harley, 2014) recognizes the need for such a mechanism, but there have been no detailed formal proposals for doing so. I justify the need for realizational treatments of lexical suffixes because they may be associated with multiple meanings. The interpretation of *-ka*, *-po*, and *-wẽ* depend on context, especially when they occur on an adjectives or demonstratives. The converse is true as well, the affixes are infelicitous when used without a licensing context. Representing this relationship is an innovative contribution of this work, and its primary goal.

Of course, it was not simply a question of representing lexical realization. There was also a goal of doing so according to Word and Paradigm (WP) principles. The form-form mappings provide a word-form-based theory. The basic diagnostic notion of minimal pairs is baked into these rules. The fundamental discovery procedures of morphology involve the recognition of patterns based on the relationship between minimal pairs of freely occurring forms. This is captured in an unambiguous manner in my framework. There is a scientific principle that underlies the framework design. The formal objects and mechanisms of the framework should encode the discovery procedure. Designing a grammatical theory where the objects have been disconnected from fundamental evidence is an error. I briefly described the Potawatomi analyses of Halle and Marantz (1993), Anderson (1992), and Stump (2001) in Chapter 4. Analysts with nothing in common but a corpus of an unknown language and a translation should produce convergent analyses given minimal pair analysis. Somehow, in the Potawatomi case, three authors working in extremely similar post-syntactic theories came up with highly divergent feature proposals, despite all of them sharing a source that is a masterwork of descriptive linguistics. There is something wrong there, a disconnect between representation and measurement.

My proposal is to provide representations that encode the evidence used to justify them. This is seen in categorial grammars (CGs), in the sense that replaceability is encoded in the notion of category. It is seen in formal semantics in the sense that entailments must match intuitions. The same is true of distributional semantics, where vectors encode the environments of tokens.

I seek something similar in the morphological domain. We know that a suffix exists because it signals a contrast between minimal pairs. Therefore, suffixes relate pairs. The nature of a suffix is that relation. Utilizing form-form mappings as the basis of an analysis forces an engagement of the model with evidence. One must express the chain of minimal pair comparisons that would be used to justify a morphological analysis.

The form classes used in my system allow for the particular and idiosyncratic to form part of the general systematicity of the morphological grammar. Small patterns such as *dika*, or *tewẽ* and *tepẽ*, do not exist in isolation. Their lack of productivity is not stated as a restriction using form classes. Instead, their small patterns inform more general patterns through the form class hierarchy. This does not cause their productivity to be overstated. The *-ka* on the ending of *dika* is related to the suffix usage, but given that *di* never occurs alone or with another ending, it would be odd to propose that *dika* is the result of affixation. Due to the fact that some forms may be stated to exist by axiom, there isn't any reason to treat *dika* as the result of some productive process. It may be associated with a category shared by *-ka* affixed items despite the fact that it is essentially listed. Theories such as DM, in particular, have difficulty expressing the relationship between unproductive instances of a signal and productive instances of a signal, without treating all cases as fully productive (Macaulay and Salmons, 2017). It is an unnecessaryly strident approach to treat *di* as the same as a productive base like *yẽdẽ*, 'big', just because *dika* exists and is related to *-ka*.

Another accomplishment of the form class system is that it improves on previous formal mechanisms for expressing similar ideas. Inheritance systems, such as those used in Network Morphology (NM), simply are not well suited to natural linguistic patterns. I have mentioned issues such as overabundance in §4.3.3.3, but neither is it obvious how they cope with other morphological phenomena, such as defectiveness (Andrea D Sims, 2015). The major innovation of form classes are to define the class relationships as a scaffold of hierarchical relationships that rules reference, rather than conflating the building of grammatical objects with the hierarchy definition. Though I am not the only one to recognize that form class-like hierarchies should be referenced as external structures (Guzmán Naranjo, 2019; McConville, 2006), I believe that I have developed the idea further than previously seen in the literature.

Realization relative to lexical meaning is awkward, at the very least, in other morphological theories. There is no reason why it had to be so. Realization relative to lexical meanings is fairly straight forward. Conventional meaning associations are expressed through a listing of meanings that match patterns found in a morphological form, represented by an M-cat-abstraction in my framework. Basically, when a word-form sounds a particular way, it may be associated with certain lexical meanings. There isn't a much more clear way of expressing this than I have presented. To show that *yẽdẽ-ka*, 'big-CLF.stone', may be associated with stones, one simply examines whether such an option exists in what is essentially a lookup table. Just as in a print dictionary, one looks up a form and often has a number of definitions listed. There is no reason why those options need to compete. One can simply pick out the meaning needed for evidence. To show that *yẽdẽ-ka* may be associated with heads, one simply chooses that meaning instead of the stone one. One may be interested in a more sophisticated theory of meaning relationships within a lexical semantic system, but that is beside the point. There is nothing complicated at all about matching lexical meanings to forms in a manner consistent with realizational principles. It is trivial, but it rests on some foundational choices that are not adopted in other theories.

Constructive proofs, where the goal is to provide an example object, simplify many aspects of modeling morphological phenomena. I do not need to worry about the fact that a gerund, or the accusative marker, may or may not have a *-te* ending. In some realizational theories, two distinct *ĩ* stems are needed to explain why one can signal the accusative for a *okiye* as either *okiye ĩdãte*, *okiye ĩdã* (among others) (Stump, 2016). Rule block competition would always rule out one option, otherwise. In my framework one can simply prove the existence of either option. There is no tension between the two. The two form options do not interfere with one another. This is all due to a simple idea. The lexicon lists the rules that explain the possibilities. One may perform a proof that demonstrates a particular possibility, but it does not mean that it is the only option.

I believe that I have gone further than any previous discrete formal theory in modeling inferences that can be made within a WP system. The system provides more than one route to proving that a form exists. There is the shared stem route, where form-form mappings that share a stem are related. This allows one to prove that if *yẽdẽ* exists, then *yẽdẽ-po*. Likewise, if *yẽdẽ-bi* exists, so does *yẽdẽ-bo*.

There is also the route where two word-forms that have different stems that belong to the same form class can be proven to be related by *interchangeMP*. If there is a *yẽdẽ-po*, then there is a *giita-po*, given that both follow the adjectival pattern.

This is not a system that requires that all forms or form categories of a paradigm are known ahead of time. In feature theories, there are a finite number of features, with a finite number of values that are multiplied together to define a paradigm. The scheme expresses the conceit that one already knows all morphosyntactic categories of a language. I doubt this has ever actually been the case for any language. Theories that focus on the question of how forms are inferred across paradigms, the so-called cell-filling problem, due to the reification of the feature-based paradigm (Blevins, Milin, and Ramscar, 2017), recognize sparseness of known forms across paradigms, but they do not consider paradigms with unknown boundaries. Incrementalist theories, such as Item and Arrangement (IA) or Item and Process (IP), which build words according to general rules, express how word-forms may continue to develop from some base in an open-ended manner, but have not engaged with the question of how speakers make inferences within a shared class of word-forms – at least, not to my knowledge. The interaction of form-form mappings and rules like *interchangeMP* means that in my framework one may not know all cells (so to speak) of a paradigm, yet still make inferences across paradigm instances for those cells that are known. One could build up a series of ever more complex words in a language like Cochabamba Quechua or Turkish from a base α using only form-form mappings. So long as another base β is of the same class, all forms associated with β could be proven by reference to α, rather than any form-form mappings. Notably, one could use no form-form mappings, and instead base every proposed form on a documented form, which is encoded by axiom in the framework, and use the *interchangeMP* rule to fill in blanks. I view this last property as what WP theories, such as Singh and Starosta (2003), have been promising for some time, but discrete formal models have not materialized. In addition to allowing greater realizational coverage of morphological phenomena by including lexical information in licensing conditions, the framework provides an example of viable discrete formal theory of an inferential WP system.

Bibliography

Aboamer, Yousuf and Marcus Kracht (2019). “A Purely Surface-Oriented Approach to Handling Arabic Morphology”. In: *Formal Grammar. FG 2019*. Ed. by R. Bernardi, G. Kobele, and S. Pogodalla. Vol. 11668. Lecture Notes in Computer Science. Springer, pp. 1–17. DOI: 10.1007/978-3-662-59648-7_1.

Ackerman, Farrell and Robert Malouf (2013). “Morphological Organization: The Low Conditional Entropy Conjecture”. In: *Language* 89.3, pp. 429–464.

Acquaviva, Paolo (2008). *Lexical Plurals: A Morphosemantic Approach*. Oxford University Press.

— (2009). “Roots and Lexicality in Distributed Morphology”. In: *Fifth York-Essex Morphology Meeting (YEMM), 9th February and 10th February 2008*. University of York Department of Language and Linguistic Science.

Adelaar, Willem and Pieter Muysken (2004). *The Languages of the Andes*. Cambridge: Cambridge University Press.

Aikhenvald, Alexandra Y. (1994). “Classifiers in Tariana”. In: *Anthropological Linguistics* 36.4, pp. 407–465.

— (2000). *Classifiers. A Typology of Noun Categorization Devices*. OUP Oxford.

Ajdukiewicz, Kasimir (1935). “Die Syntaktische Konnexität”. In: *Studia Philosophica* 1, pp. 1–27.

Aksënova, Alëna, Thomas Graf, and Sedigheh Moradi (2016). “Morphotactics as Tier-Based Strictly Local Dependencies”. In: *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pp. 121–130. URL: https://www.aclweb.org/anthology/W/W16/W16-2019.pdf.

Allan, Keith (1977). “Classifiers”. In: *Language* 53.2, pp. 283–310.

Almor, Amit (1999). “Noun-Phrase Anaphora and Focus. The Information Load Hypothesis”. In: *Psychological Review* 106.4, pp. 748–765.

Anderson, Stephen R (1992). *A-Morphous Morphology.* Cambridge University Press.

Andrews, Avery D (1990). “Unification and Morphological Blocking”. In: *Natural Language & Linguistic Theory* 8, pp. 507–557.

Armoskaite, Solveiga and Carrie Gillon (2013). “50 Shades of Definiteness”. In: *Proceedings of the 2013 Annual Conference of the Canadian Linguistics Association.*

Aronoff, Mark (1994). *Morphology by Itself. Stems and Inflection Classes.* MIT Press.

Ascoop, Kristen and Torsten Leuschner (2006). “Affixoidhungrig? Skitbra! Comparing Affixoids in German and Swedish”. In: *Language Typology and Universals* 59.3, pp. 241–252. DOI: 10.1524/stuf.2006.59.3.241.

Asher, Nicholas (2011). *Lexical Meaning in Context. A Web of Words.* Cambridge University Press.

Baayan, Harald (1993). “On frequency, transparency and productivity”. In: *Yearbook of Morphology 1992.* Ed. by G. Booij and J. van Marle. Yearbook of Morphology.

Bar-Hillel, Yehoshua (1953). “A Quasi-Arithmetical Notation for Syntactic Description”. In: *Language* 29.1, pp. 47–58.

Barnes, Janet (1990). “Classifiers in Tuyuca. Studies in Lowland South American Languages”. In: *Amazonian Linguistics. Studies in Lowland South American Languages.* Ed. by Doris L. Payne. University of Texas Press, pp. 273–292.

Beard, Robert (1995). *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation.* Suny Press.

Becker, Thomas (1993). “Back-formation, cross-formation, and ‘bracketing paradoxes’ in paradigmatic morphology”. In: *Yearbook of Morphology 1993.* Ed. by Geert Booij and Jaap van Marle, pp. 1–26.

Bekki, Daisuke (2012). “Dependent Type Semantics. An Introduction”. In: *Logic and Interactive RAtionality (LIRA). Yearbook 2012.* Ed. by Zoé Christoff, Paolo Galeazzi, Nina Gierasimczuk, Alexandru Marcoci, and Sonja Smets. Vol. 1.

Berlin, Brent (1968). *Tzeltal Numeral Classifiers. A Study in Ethnographic Semantics.* Mouton.

Biedny, Jerome, Matthew Burner, Andrea Cudworth, and Monica Macaulay (2021). "Classifier Medials across Algonquian. A First Look". In: *International Jounal of American Linguistics* 87.1, pp. 1–47.

Bischoff, Shannon T. (2011). "Lexical affixes, incorporation, and conflation. The case of Coeur D'Alene". In: *Studia Linguistica.* A Journal of General Linguistics 65.1, pp. 1–31.

Blevins, James P. (1995). "Syncretism and Paradigmatic Opposition". In: *Linguistics and Philosophy* 18, pp. 113–152.

— (2006). "Word-Based Morphology". In: *Journal of Linguistics* 42.3, pp. 531–573.

Blevins, James P., Peter Milin, and Michael Ramscar (2017). "The Zipfian Paradigm Cell Filling Problem". In: *Perspectives on Morphological Oragnization. Data and Analyses.* Ed. by Ferenc Kiefer, James Blevins, and Huba Bartos. Vol. 10. Empirical Approaches to Linguistic Theory. Brill, pp. 139–156. to appear.

Bloomfield, Leonard (1934). *Plains Cree Texts.* American Ethnological Society Publications. AMS Press. ISBN: 978-0-404-58166-4. URL: http://books.google.ca/books?id=_yAmAQAAMAAJ.

— [1933] (1984). *Language.* University of Chicago Press.

Bobaljik, Jonathan David (2002). "Syncretism without Paradigms. Remarks on Williams 1981, 1994". In: *Yearbook of Morphology 2001.* Ed. by Geert Booij and Jaap Van Marle, pp. 53–85.

— (2017). "Distributed Morphology". In: *Oxford Research Encyclopedia of Linguistics.* DOI: 10.1093/acrefore/9780199384655.013.131.

Bonami, Olivier (2015). "Periphrasis as Collocation". In: *Morphology* 25, pp. 63–110.

Bonami, Olivier and Jana Strnadová (2019). "Paradigm Structure and Predictability in Derivational Morphology". In: *Morphology* 28.2, pp. 167–197.

Bonami, Olivier and Gregory Stump (2016). "What Is Paradigm Function Morphology?" In: *The Cambridge Handbook of Morphology.* Ed. by Andrew Hippisley and Gregory Stump. Cambridge University Press, pp. 449–481.

Braitenberg, Valentino, Detlef Heck, and Fahad Sultan (1997). "The Detection and Generation of Sequeunces as a Key to Cerebellar Function: Experiments and Theory". In: *Behavioral and Brain Sciences* 20, pp. 229–277.

Braitenberg, Valentino and Almut Schüz (2013). *Anatomy of the Cortex: Statistics and Geometry.* Studies of Brain Function 18. Springer Science and Business Media.

Bresnan, Joan and Sam A. Mchombo (1987). "Topic, Pronoun, and Agreement in Chicheŵa". In: *Language* 63.4, pp. 741–782.

Brown, Dunstan and Andrew Hippisley (2012). *Network Morphology. A Defaults-Based Theory of Word Structure.* Cambridge University Press.

Burling, Robins (1965). "How to Choose a Burmese Numeral Classifier". In: *Context and Meaning in Cultural Anthropology.* Ed. by M.E. Spiro. Free Press, pp. 243–264.

Butler, Nancy E. and Elizabeth M. Ekdahl (1979). *Aprenda Terêna.* Vol. 1. Summer Institute of Linguistics. URL: https://archive.org/details/rosettaproject_ter_book-2.

Butt, Miriam (2014). "The Light Verb Jungle. Still Hacking Away". In: *Complex Predicates. Cross-Linguistic Perspectives on Event Structure.* Ed. by Mengistu Amberber, Brett Baker, and Mark Harvey. Cambridge University Press, pp. 48–78.

Bybee, Joan L (1985). *Morphology.* John Benjamins Publishing.

Camacho Rios, Gladys, Simeon Floyd, and Félix Julca Guerrero (2024). "¿Cuántas lenguas quechuas hay? Una estimación del número de lenguas quechuas". In: *Lexis* 48.1, pp. 34–77. DOI: 10.18800/lexis.202401.002.

Campbell, Donald and Donald Fiske (1959). "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix". In: *Psychological Bulletin* 56.2, pp. 81–105.

Carstairs-McCarthy, Andrew (1987). *Allomorphy in Inflection.* London: Croom Helm.

Chatzikyriakidis, Stergios and Zhaohui Luo (2013). "Adjectives in a Modern Type-Theoretical Setting". In: *Formal Grammar*, pp. 159–174.

— (2017). "Adjectival and Adverbial Modification. The View from Modern Type Theories". In: *Journal of Logic, Language and Information* 26.1, pp. 45–88.

Chomsky, Noam (1957). *Syntactic Structures.* Mouton, The Hague.

— (1995). *The Minimalist Program.* MIT Press.

Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English.* New York: Harper & Row.

Cinque, Guglielmo and Luigi Rizzi (2004). “The Cartography of Syntactic Structures”. In: *The Structure of CP and IP*. Vol. 2, pp. 3–15.

Coecke, Bob, Mehrnoosh Sadrzadadeh, and Stephen Clark (2010). “Mathematical foundations for a compositional distributional model”. In: *CoRR* abs/1003.3494. URL: http://arxiv.org/abs/1003.4394.

Colby, B. N., Olga Akhmanova, Ethel M. Albert, E. Pendleton Banks, Wallace L. Chafe, Harold C. Conklin, J. L. Fischer, Willem A. Grootaers, Dell Hymes, Paul Kay, Roger M. Keesing, Edward A. Kennard, J. Knobloch, F. G. Lounsbury, Louise E. Sweet, G. L. Trager, Francis Lee Utley, and Roger W. Wescott (1966). “Ethnographic Semantics. A Preliminary Survey [and Comments and Replies]”. In: *Current Anthropology* 7.1, pp. 3–32.

Cole, Peter (1985). *Imbabura Quechua*.

Collins, Chris and Edward Stabler (Mar. 2016). “A Formalization of Minimalist Syntax”. In: *Syntax* 19.1, pp. 43–78. ISSN: 1368-0005, 1467-9612. DOI: 10.1111/synt.12117. URL: https://onlinelibrary.wiley.com/doi/10.1111/synt.12117 (visited on 01/07/2024).

Cooper, Richard P and Tim Shallice (2006). “Hierarchical Schemas and Goals in the Control of Sequential Behavior.” In: *Psychological Review* 113.4, pp. 887–916.

Copot, Maria, Sara Court, Noah Diewald, Stephanie Antetomaso, and Micha Elsner (2022). “A Word-and-Paradigm Workflow for Fieldwork Annotation”. In: *Proceedings of the 5th Workshop on Computational Methods for Endangered Languages (ComputEL-5)*. Dublin, Ireland: Association for Computational Linguistics, pp. 159–169. DOI: 10.18653/v1/2022.computel-1.20. URL: https://aclanthology.org/2022.computel-1.20.

Coquand, Thierry and Gérard Huet (1988). “The Calculus of Constructions”. In: *Information and Computation* 76.2, pp. 95–120. ISSN: 0890-5401. DOI: 10.1016/0890-5401(88)90005-3.

Coquand, Thierry and Christine Paulin-Mohring (1990). “Inductively Defined Types”. In: *Proceedings of Colog-88*. Ed. by Per Martin-Löf and Grigori Mints. Vol. 417. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 50–66. DOI: 10.1007/3-540-52335-9_47.

Corbett, Greville G. (1991). *Gender*. Cambridge University Press.

— (2003). *Agreement*. Cambridge University Press.

Corbett, Greville G. (2005). “The Canonical Approach in Typology”. In: *Linguistic Diversity and Language Theories*. Ed. by Frajzyngier Zygmunt, Adam Hodges, and David S Rood. John Benjamins, pp. 25–49.

Corbett, Greville G. and Sebastian Fedden (2016). “Canonical gender”. In: *Journal of Linguistics* 52.3, pp. 495–531. DOI: 10.1017/S0022226715000195.

Corbett, Greville G. and Norman Fraser (1993). “Network Morphology. a DATR Account of Russian Nominal Inflection”. In: *Journal of Linguistics* 29, pp. 113–142.

Corbett, Greville G. and Norman M. Fraser (2000). “Gender assignment”. In: *Systems of Nominal Classification*. Ed. by Gunter Senft. Cambridge University Press.

Cowper, Elizabeth and Daniel Currie Hall (2022). “Morphosemantic Features in Universal Grammar. What We Can Learn from Marshallese Pronouns and Demonstratives”. In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.3, pp. 242–266. DOI: 10.1017/cnj.2022.25.

Crysmann, Berthold and Olivier Bonami (2016). “Variable Morphotactics in Information-Based Morphology”. In: *Journal of Linguistics* 52, pp. 311–374.

Culicover, Peter and Ray Jackendoff (2005). *Simpler Syntax*. Oxford University Press.

Curry, Haskell B (1961). “Some Logical Aspects of Grammatical Structure”. In: *Structure of Language and Its Mathematical Aspects*. Vol. 12. American Mathematical Society, pp. 56–68.

Dąbrowska, Ewa (2001). “Learning a Morphological System without a Default. The Polish Genetive”. In: *Journal of Child Language* 28.3, pp. 545–574.

— (2014). “Implicit Lexical Knowledge”. In: *Linguistics* 52.1, pp. 205–223.

De Saussure, Ferdinand (2011). *Course in General Linguistics*. Columbia University Press.

Déchaine, Rose-Marie and Martina Wiltschko (2002). “Decomposing pronouns”. In: *Linguistic Inquiry* 33.3, pp. 409–442. DOI: 10.1162/002438902760168554.

Denny, J. Peter (1976). “What Are Noun Classifiers Good For?” In: *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society*, pp. 122–132.

— (1981). “Algonquian word structure from the point of view of logical semantics”. In: *Montreal Working Papers in Linguistics*. Ed. by Lynn Drapeau. Vol. 16, pp. 21–29.

Denny, J. Peter (1983). “Micmac semantics. Medials for noun classes”. In: *Actes du quatorzième Congrès des Algonquinistes*. Ed. by William Cowan. Carlton University. Ottawa, pp. 363–368.

Derbyshire, Desmond and Doris L. Payne (1990). “Noun Classification Systems of Amazonian Languages”. In: *Amazonian Linguistics*.

Di Sciullo, Anna-Maria and Edwin Williams (1987). *On the Definition of the Word*. Cambridge, MA: MIT Press.

Dickinson, Connie, Uboye Gaba, Oswando Nenquimo, Nenquihui Nemo, Mery Nenquihui, and Casey High, collab. (2013). *Documentation of Wao Terero*. URL: `https://elar.soas.ac.uk/Collection/MPI1029685`.

Dixon, R. M. W. (1986). “Noun Classes and Noun Classification in Typological Perspective”. In: *Noun Classes and Categorization. Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*. Ed. by Colette G. Craig. John Benjamins, pp. 105–112. DOI: `10.1075/tsl.7`.

Downing, Pamela (1986). “The Anaphoric Use of Classifiers in Japanese”. In: *Noun Classes and Categorization. Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*. Ed. by Colette G. Craig. John Benjamins, pp. 345–375. DOI: `10.1075/tsl.7`.

Drapeau, Lynn and Renée Lambert-Brétière (2011). “Verbal Classifiers in Innu”. In: *Anthropological Linguistics* 53.4, pp. 293–322.

Dryer, Matthew S (1997). “Passive vs Indefinite Actor Construction in Plains Cree”. In: *Twenty-Sixth Algonquian Conference*. Winnipeg: University of Manitoba.

Embick, David (2000). “Features, Syntax, and Categories in the Latin Perfect”. In: *Linguistic Inquiry*, pp. 185–230.

Embick, David and Morris Halle (2005). “On the Status of *stems* in Morphological Theory”. In: *Romance Languages and Linguistic Theory 2003*. Ed. by Twan Geerts, Ivo van Ginneken, and Haike Jacobs. John Benjamins, pp. 59–88.

Fawcett, Alexia Zandra (2012). “Documenting Language, Culture, and Cognition. Language and Space among the Waorani”. BA thesis. Bryn Mawr College.

Fawcett, Alexia Zandra (2018). “Ideophone Integration and Expressiveness in Wao Terero”. MA thesis. University of California – Santa Barbara.

— (2023). “Wao Terero”. An International Handbook. In: *Volume 2 Language Isolates II. Kanoé to Yurakaré*. Ed. by Patience Epps and Lev Michael. Vol. 2. Amazonian Languages. De Gruyter Mouton, pp. 1191–1241.

Fedden, Sebastian and Greville G. Corbett (2017). “Gender and Classifiers in Concurrent Systems. Refining the Typology of Nominal Classification”. In: *Glossa-an International Journal of Linguistics. A journal of general linguistics* 2.1, pp. 1–47.

Fiddler, Michael (2011). “Bound Nominal Roots in Waorani”.

Fillmore, Charles, Paul Kay, and Catherine O’Connor (1988). “Regularity and Idiomaticity in Grammatical Constructions. The Case of *let alone*”. In: *Language* 64, pp. 501–538.

Fitzgerald, Colleen M. (2000). “A Reanalysis of Bidirectionality in Auca”. In: *Proceedings of the Twenty-Eight Western Conference on Linguistics (WECOL99).* 11, pp. 106–118.

Fleck, David (2006). “Body-part prefixes in Matses. Derivation or noun incorporation?” In: *International journal of American linguistics* 72.1, pp. 59–96. DOI: 10.1086/505279.

Flickinger, Daniel (1987). “Lexical Rules in the Hierarchical Lexicon”. PhD thesis. Stanford University.

Flickinger, Daniel, Carl Pollard, and Thomas Wasow (1985). “Structure-Sharing in Lexical Representation”. In: *Proceedings of the 23rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pp. 262–267.

Gardiner, Mary (2003). “Identifying and Resolving One-Anaphora”. Bachelor of science. Macquarie University.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag (1985). *Generalized Phrase Structure Grammar*. Harvard University Press.

Gerdts, Donna B. (2003). “The Morphosyntax of Halkomelem Lexical Suffixes”. In: *Internaltional Journal of American Linguistics* 69.4, pp. 345–356.

Gilbert, Gaëtan, Jesper Cockx, Matthieu Sozeau, and Nicolas Tabareau (2019). “Definitional proof-irrelevance without K”. In: *Proceedings of the ACM on Programming Languages*. Vol. 3. POPL. New York, NY, USA: Association for Computing Machinery. DOI: 10.1145/3291648.

Girard, Jean-Yves (1987). “Linear Logic”. In: *Theoretical computer science* 50.1, pp. 1–101.

Goldberg, Adele (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Goldsmith, John (2001). “Unsupervised Learning of the Morphology of a Natural Language”. In: *Computational Linguistics* 27.2, pp. 153–198.

Gomez-Imbert, Elsa (2007). “Tukanoan Nominal Classification”. In: *Language Endangerment and Endangered Languages. Linguistic and Anthropological Studies with a Special Emphasis on the Language and Cultures of the Andean-Amazonian Border Area*, pp. 401–428.

Greenberg, Joseph, ed. (1963). *Universals of Language*. MIT Press.

Grice, Herbert P. (1975). “Logic and Conversation”. In: *Speech Acts*. Brill.

Grinevald, Colette (2000). “A Morphosyntactic Typology of Classifiers”. In: *Systems of Nominal Classification*. Ed. by Gunter Senft. Cambridge University Press, pp. 50–92.

Grinevald, Colette and Frank Seifart (2004). “Noun Classes in African and Amazonian Languages. Towards a Comparison”. In: *Linguistic Typology* 8.2, pp. 243–285.

Grupen, Roderic A. (2023). *The Developmental Organization of Robot Behavior*. MIT Press.

Gualapuro, Santiago David Gualapuro, Jordan Lacher, Sean Burke, Conor Snoek, Michelle Garcia-Vega, Megan Bontogon, and Evan Lloyd (2018). *Kichwa English Shimiyuk Kamu Dictionary*. Universidad de San Francisco de Quito Press.

Gũnther, Christine (2011). “Noun ellipsis in English. Adjectival modifiers and the role of context”. In: *English Language & Linguistics* 15.2, pp. 279–301. DOI: 10.1017/S1360674311000049.

Guzmán Naranjo, Matías (2019). “Analogy in Formal Grammar”. PhD thesis. Universität Leipzig.

Hale, Ken (1983). “Warlpiri and the Grammar of Non-Configurational Languages”. In: *Natural Language & Linguistic Theory* 1.1, pp. 5–47.

Halle, Morris and Alec Marantz (1993). "Distributed Morphology and the Pieces of Inflection". In: *The View from Building 20*. Ed. by Kenneth L. Hale and Samuel J. Keyser. Cambridge, MA: MIT Press, pp. 111–176.

Harel, David (1987). "Statecharts: A Visual Formalism for Complex Systems". In: *Science of computer programming* 8.3, pp. 231–274.

Harley, Heidi (2008). "When Is Syncretism More than a Syncretism?. Impoverishment, Metasyncretism and Underspecification". In: *Phi Theory. Phi-Features across Modules and Interfaces*. Ed. by Daniel Harbour, David Adger, and Susana Béjar. Oxford University Press, pp. 251–294.

— (2012). "Compounding in Distributed Morphology". In: *The Oxford Handbook of Compounding*. Ed. by Rochelle Lieber and Pavol Štekauer. Oxford University Press, pp. 129–145. DOI: 10.1093/oxfordhb/9780199695720.013.0007.

— (2014). "On the Identity of Roots". In: *Theoretical Linguistics* 40.3/4, pp. 225–276.

Harris, Alice C. (2000). "Where in the Word Is the Udi Clitic?" In: *Language* 76.3, pp. 593–616.

Harris, Zellig S. (1942). "Morpheme Alternants in Linguistic Analysis". In: *Language* 18.3, pp. 169–180.

— (1951). *Methods in Structural Linguistics*. University of Chicago Press.

Haspelmath, Martin (2017). "The Indeterminacy of Word Segmentation and the Nature of Morphology and Syntax". In: *Folia Linguistica* 51.s1000, pp. 31–80. DOI: doi:10.1515/flin-2017-1005. URL: https://doi.org/10.1515/flin-2017-1005 (visited on 12/28/2023).

— (2019). *How Comparative Concepts and Descriptive Linguistic Categories Are Different*. Ed. by Daniël van Olmen, Tanja Mortelmans, and Frank Brisard. Vol. 324. Trends in Linguistics. Studies and Monographs [TiLSM]. De Gruyter Mouton, pp. 83–114. DOI: 10.1515/9783110607963.

Haugen, Jason D. and Daniel Siddiqi (2013). "Roots and the Derivation". In: *Linguistic Inquiry* 44.3, pp. 493–517.

Hay, Jennifer (2001). "Lexical Frequency in Morphology: Is Everything Relative?" In: *Linguistics* 39.6, pp. 1041–1070.

Heim, Irene (1983). "File Change Semantics and the Familiarity Theory of Definiteness". In: *Meaning, Use, and Interpretation of Language*. Ed. by Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow. Walter de Gruyter, pp. 164–189.

Heim, Irene and Angelika Kratzer (1998). *Semantics in Generative Grammar*. Blackwell Publishing.

Hinkson, Mercedes Q. (1999). "Salishan lexical suffixes. A study in the conceptualization of space". Simon Fraser University.

Hockett, Charles F. (1947). "Problems of Morphemic Analysis". In: *Language* 23.4, pp. 321–343.

— (1948). "Potawatomi 2. Derivation, Personal Prefixes, and Nouns". In: *International Journal of American Linguistics* 14.2, pp. 63–73.

— (1954). "Two Models of Grammatical Description". In: *Word-journal of The International Linguistic Association* 10.2–3, pp. 210–234. DOI: 10.1080/00437956.1954.11659524.

Hoeksema, Jack and Richard D. Janda (1988). "Implications of Process-Morphology for Categorial Grammar". In: *Categorial Grammars and Natural Language Structures*. Ed. by Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler.

Hornstein, Norbert, Jairo Nunes, and Kleanthes K. Grohmann (2005). *Understanding Minimalism*. Cambridge University Press.

Houser, Michael J. (2010). "The Sysntax and Semantics of *So So* Anaphora". phd. University of California, Berkeley.

Jackendoff, Ray (2002). *Foundations of Language*. Oxford University Press.

Jelinek, Eloise (1984). "Empty Categories, Case, and Configurationality". In: *Natural Language & Linguistic Theory* 2.1, pp. 39–76.

Johns, Alana (2017). "Noun Incorporation". In: *The Whiley Blackwell Companion to Syntax, Second Edition*. Ed. by M Everaert and H C Riemsdijk. DOI: 10.1002/9781118358733.wbsyncom093.

Johnson, Martha Booker, Micha Elsner, and Andrea D. Sims (2023). "High frequency derived words have low semantic transparency mostly only if they are polysemous". International Symposium of Morphology 2023. URL: https://ismo2023.ovh/fichiers/slides/3-johnsonsimselsner.pdf.

Jones, William (1904). “Some Principles of Algonquian Word-Formation”. In: *American Anthropologist* 6.3, pp. 369–411.

Kalashnikova, Marina, Aimee Oliveri, and Karen Mattock (2018). “Acceptance of Lexical Overlap by Monolingual and Bilingual Toddlers”. In: *International Journal of Bilingualism* 23.6, pp. 1517–1530.

Katz, Jerrold J. (1981). *Language and Other Abstract Objects*. Roman and Littlefield.

Katz, Jerrold J. and Jerry A. Fodor (1963). “The Structure of a Semantic Theory”. In: *Language* 39.2, pp. 170–210.

Kelley, Patricia M. (1988). “Issues for Literacy Materials Development in a Monolingual Amazonian Culture : The Waodani of Ecuador”. In: University of British Columbia. DOI: 10.14288/1.0054664. URL: https://doi.library.ubc.ca/10.14288/1.0054664 (visited on 01/14/2024).

Key, Harold (1967). *Morphology of Cayuvava*. Mouton.

Khullar, Payal, Arghya Bhattacharya, and Manish Shrivastava (2020). “Finding the Right One and Resolving It”. In: *Proceedings of the 24th Conference on Computational Natural Language Learning*. Association for Computational Linguistics, pp. 132–141.

Kinkade, M. Dale (1963). “Phonology and Morphology of Upper Chehalis: II”. In: *International Journal of American Linguistics* 29.4, pp. 345–356.

— (1998). “Origins of Salishan Lexical Suffixes”. In: *Papers for the 33rd International Conference on Salish and Neighboring Languages*, pp. 266–295.

Koenig, Jean-Pierre and Daniel Jurafsky (1994). “Type Underspecification and On-Line Type Construction in the Lexicon”. In: *West Coast Conference on Formal Linguistics*. Vol. 13. CSLI Publications/SLA, pp. 270–285.

Krasnoukhova, Olga (2012). “The Noun Phrase in the Languages of South America”. Universite de Liege.

Kubota, Yusuke (2010). “(In) Flexibility of Constituency in Japanese in Multi-Modal Categorial Grammar with Structured Phonology”. PhD thesis. The Ohio State University.

Kubota, Yusuke (2014). "The Logic of Complex Predicates". In: *Natural Language & Linguistic Theory* 32.4, pp. 1145–1204.

Kubota, Yusuke and Robert Levine (2012). "Gapping as Like-Category Coordination". In: *International Conference on Logical Aspects of Computational Linguistics*, pp. 135–150.

— (2017). "Pseudogapping as Pseudo-VP Ellipsis". In: *Linguistic Inquiry* 48, pp. 213–257.

Kubota, Yusuke and Robert D. Levine (2020). *Type-Logical Syntax*. MIT Press.

Kubota, Yusuke, Koji Mineshima, Robert Levine, and Daisuke Bekki (2019). "Underspecification and Interpretive Parallelism in Dependent Type Semantics". In: *Proceedings of the IWCS 2019 Workshop on Computing Semantics with Types, Frames and Related Structures*, pp. 1–9. URL: https://www.asc.ohio-state.edu/levine.1/publications/Gothenburg2019.pdf.

Lakatos, Imre (1970). "Falsification and the Methodology of Scientific Research Programmes". In: *Criticism and the Growth of Knowledge*. Ed. by Imre Lakatos and Alan Musgrave. Cambridge University Press, pp. 91–195.

Lakoff, George (2008). *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. University of Chicago Press.

Lambek, Joachim (1958). "The Mathematics of Sentence Structure". In: *American Mathematical Monthly* 65.3, pp. 154–170.

— (1997). "Type Grammar Revisited". In: *International Conference on Logical Aspects of Computational Linguistics*. Springer, pp. 1–27.

— (2001). "Type Grammars as Pregroups". In: *Grammars* 4.1, pp. 21–39. ISSN: 1572-848X. DOI: 10.1023/A:1011444711686. URL: http://dx.doi.org/10.1023/A:1011444711686.

Lashley, Karl Spencer (1951). "The Problem of Serial Order in Behavior". In: *Cerebral Mechanisms in Behavior*, pp. 112–136.

Leipzig Glossing Rules (2015). *Leipzig Glossing Rules. Conventions for Interlinear Morpheme-by-Morpheme Glosses*. Ed. by Department of Linguistics Max Planck Institute for Evolutionary Anthropology. Version 2015-05-31. URL: https://www.eva.mpg.de/lingua/resources/glossing-rules.php (visited on 11/09/2022).

Lewis, David (1970). “General Semantics”. In: *Synthese. An International Journal for Epistemology, Methodology and Philosophy of Science* 22, pp. 18–67.

Luo, Zhaohui (1990). “An Extended Calculus of Constructions”. PhD thesis. University of Edinburgh.

— (2012a). “Common Nouns as Types”. In: *International Conference on Logical Aspects of Computational Linguistics*. Ed. by Denis Béchet and Alexander Dikovsky. Springer.

— (2012b). “Formal Semantics in Modern Type Theories with Coercive Subtyping”. In: *Linguistics and Philosophy* 35, pp. 491–513.

— (2012c). “Notes on Universes in Type Theory”. URL: `http://www.cs.rhul.ac.uk/home/zhaohui/universes.pdf`.

Luo, Zhaohui, Sergei Soloviev, and Tao Xue (2013). “Coercive Subtyping. Theory and Implementation”. In: *Information and Computation* 233, pp. 18–42.

Macaulay, Monica and Joseph Salmons (2017). “Synchrony and Diachrony in Menominee Derivational Morphology”. In: *Morphology* 27, pp. 179–215.

Marantz, Alec (1996). ““Cat” as a Phrasal Idiom. Consequences of Late Insertion in Distributed Morphology”. URL: `https://www.academia.edu/17275409/Cat_as_a_phrasal_idiom`. draft.

— (1997). “No Escape from Syntax. Don’t Try Morphological Analysis in the Privacy of Your Own Lexicon”. In: *University of Pennsylvania Working Papers in Linguistics* 4.2, pp. 201–225.

Martin, Scott (2013). “The Dynamics of Sense and Implicature”. PhD thesis. The Ohio State University.

— (2015). “A Unidimensional Syntax-Semantics Interface for Supplements”. In: *Proceedings for the ESSLLI 2015 Workshop on Empirical Advances in Categorial Grammar (CG 2015)*, pp. 82–106.

— (2016). “Supplemental Update”. In: *Semantics and Pragmantics* 9.5, pp. 1–61. DOI: `10.3765/sp.9.5`.

Martin, Scott and Carl Pollard (2014). “A Dynamic Categorial Grammar”. In: *Formal Grammar*. Springer, pp. 138–154.

Martin-Löf, Per (1984). *Intuitionistic Type Theory*. Naples: Bibliopolis.

McConville, Mark (2006). “Inheritance and the CCG Lexicon”. In: *11th Conference of the European Chapter of the Association for Computational Linguistics*. URL: http://aclweb.org/anthology/E06-1001.

Mihaliček, Vedrana (2012). “Serbo-Croatian Word Order: A Logical Approach”. PhD thesis. The Ohio State University.

Mihaliček, Vedrana and Carl Pollard (2012). “Distinguishing Phenogrammar from Tectogrammar Simplifies the Analysis of Interrogatives”. In: *Formal Grammar*. Ed. by Philippe de Groote and Mark-Jan Nederhof. Springer, pp. 130–145.

Miller, Marion (1999). “Desano Grammar”. Summer Institute of Linguistics.

Mithun, Marianne (1984). “The Evolution of Noun Incorporation”. In: *Language* 60, pp. 847–893.

— (1986). “The Convergence of Noun Classification Systems”. In: *Noun Classes and Categorization. Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*. Ed. by Colette G. Craig. Vol. 7. Typological Studies in Language. John Benjamins, pp. 379–395. DOI: 10.1075/tsl.7.

— (1997). “Lexical affixes and morphological typology”. In: *Essays on Language Function and Language Type. Dedicated to T. Givón*. Ed. by Joan Bybee, John Haiman, and Sandra A. Thompson. John Benjamins, pp. 357–371.

Moura, Leonardo de and Sebastian Ullrich (2021). “The Lean 4 Theorem Prover and Programming Language”. In: *Automated Deduction – CADE 28*. Ed. by André Platzer and Geoff Sutcliffe. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 625–635. ISBN: 978-3-030-79876-5. DOI: 10.1007/978-3-030-79876-5_37.

Muysken, Pieter (1981). “Quechua Word Structure”. In: *Binding and Filtering*. Ed. by Frank Heny. The MIT Press Cambridge, MA, pp. 279–326.

— (2002). “La Categoría Del Plural En El Quechua Boliviano”. In: *La Romania Americana. Procesos Lingüísticos En Situaciones de Contacto*. Ed. by Norma Díaz, Ralph Ludwig, and Stefan Pfänder, pp. 209–217.

— (2008). *Functional Categories*. Cambridge Studies in Linguistics 117. Cambridge University Press.

Needle, Jordan (2022). "Embedding HTLCG into LCG ". In: *Journal of Logic, Language and Information* 31, pp. 677–721. DOI: 10.1007/s10849-022-09388-5.

— (2024). "Relational Dynamic Categorial Grammar". PhD. Ohio State University.

Nerbonne, John, Masaya Lida, and William Ladusaw (1989). "Running on Empty. Null Heads in Head-Driven Grammar". In: *Proceedings of the 8th West Coast Conference of Formal Linguistics*. Vol. 8, pp. 276–288.

Nerbonne, John and Tony Mullen (2000). "Null-Headed Nominals in German and English". In: *Computational Linguistics in the Netherlands 1998. Selected Papers from the Ninth CLIN Meeting*. Leiden, The Netherlands: Brill, pp. 125–146. ISBN: 978-90-04-48889-2. DOI: 10.1163/9789004488892_012. URL: https://brill.com/view/book/9789004488892/B9789004488892_s012.xml.

Nevins, Andrew (2011). "Marked Targets versus Marked Triggers and Impoverishment of the Dual". In: *Linguistic Inquiry* 42.3, pp. 413–444.

Oehrle, Richard T (1994). "Term-Labeled Categorial Type Systems". In: *Linguistics and Philosophy* 17.6, pp. 633–678.

Panagiotidis, Phoevos (Feb. 2024). *Against Semantic Features: The View from Derivational Affixes*. URL: https://lingbuzz.net/lingbuzz/007291 (visited on 02/21/2024). preprint.

Parrochia, Daniel (Aug. 2025). "Classification". In: *The Internet Encyclopedia of Philosphy*. URL: https://iep.utm.edu/.

Passer, Matthias B. (2016). "(What) do verb classifiers classify?" In: *Lingua* 174, pp. 16–44. DOI: 10.1016/j.lingua.2015.12.003.

Payne, Doris L. (1986). "Noun Classification in Yagua". In: *Noun Classes and Categorization. Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*. Ed. by Colette G. Craig. John Benjamins, pp. 113–131. DOI: 10.1075/tsl.7.

— (1987). "Noun Classification in the Western Amazon". In: *Language Sciences* 9.1, pp. 22–44.

Payne, Thomas (1997). *Describing Morphosyntax*. Cambridge University Press.

Peeke, M. Catherine (1968). "Preliminary Grammar of Auca". PhD thesis. Indiana University.

Peeke, M. Catherine (1979). *El Idioma Huao. Grammática Pedagógica. Tomo 1.* Ed. by Summer Institute of Linguistics. Cuadernos Etnolingüísticos 3.

— (1991). "Negación En Huarani". In: *Estudios Gramaticales En Napo Quichua y Huarani.* Ed. by Carolyn Orr, Stephen H. Levinson, and M. Catherine Peeke. Cuadernos Etnolingüísticos 16. Instituto Lingüístico de Verano.

Pike, Evelyn G. and Rachel Saint, eds. (1988). *Workpapers Concerning Waorani Discourse Features.* Summer Institute of Linguistics.

Pike, Kenneth L. (1958). "On Tagmemes, Née Gramemes". In: *International Journal of American Linguistics* 24.4, pp. 273–278.

— (1964). "Stress Trains in Auca". In: *In Honour of Daniel Jones*, pp. 425–431.

Plummer, Andrew and Carl Pollard (2012). "Agnostic Possible Worlds Semantics". In: *Logical Aspects of Computational Linguistics*. Springer, pp. 201–212.

Pollard, Carl (2015). "Agnostic Hyperintensional Semantics". In: *Synthese. An International Journal for Epistemology, Methodology and Philosophy of Science*, pp. 1–28.

Pollard, Carl and Ivan Sag (1987). *Information-Based Syntax and Semantics*. Vol. 13.

— (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Pollard, Carl and Chris Worth (2015). "Coordination in Linear Categorial Grammar with Phenogrammatical Subtyping". In: *Proceedings for the ESSLLI 2015 Workshop on Empirical Advances in Categorial Grammar (CG 2015)*, pp. 162–182.

Postal, Paul (1969). "Anaphoric Islands". In: *Chicago Linguistics Society* 5, pp. 205–239.

Pullum, Geoffrey K and Hans-Jorg Tiede (2010). "Inessential Features and Expressive Power of Descriptive Metalanguages". In: *Features: Perspectives on a Key Notion in Linguistics.* Ed. by Greville G Corbett and Anna Kibort. Oxford University Press, pp. 271–292.

Pustejovsky, James (1991). "The Generative Lexicon". In: *Computational Linguistics* 17.4, pp. 409–441.

Ranta, Aarne (1994). *Type-Theoretical Grammar*. Oxford University Press.

Riehemann, Susanne Z. (1998). "Type-Based Derivational Morphology". In: *The Journal of Comparative German Linguistics* 2, pp. 49–77. DOI: 10.1023/A:1009746617055.

Ritter, Elizabeth and Sara Thomas Rosen (2010). “Animacy in Blackfoot: Implications for Event Structure and Clause Structure”. In: *Syntax, Lexical Semantics and Event Structure*. Ed. by Malka Rappaport-Hovav, Edit Doron, and Ivy Sichel. Oxford University Press.

Roberts, Craige (2003). “Uniqueness in Definite Noun Phrases”. In: *Linguistics and philosophy* 26.3, pp. 287–350.

— (2012). “Information Structure in Discourse. Towards an Integrated Formal Theory of Pragmatics”. In: *Semantics & Pragmatics* 5.6, pp. 1–69. DOI: 10.3765/sp.5.6.

Robins, Robert H. (1959). “In Defence of WP”. In: *Transactions of the Philological Society* 58.1, pp. 116–144.

Rose, Françoise and An Van linden (2022). “The derivational use of classifiers in Western Amazonia”. In: *Binominal lexemes in cross-linguistic perspective. Towards a typology of complex lexemes*. Ed. by Steve Pepper, Francesca Masini, and Simone Mattiola. Empirical Approaches to Language Typology 62. De Gruyter Mouton, pp. 237–276. DOI: 10.1515/9783110673494-008.

Rosen, Sara T. (1998). “Two Types of Noun Incorporation. A Lexical Analysis”. In: *Language* 65, pp. 294–317.

Sadler, Louisa and Andrew Spencer (2001). “Syntax as an Exponent of Morphological Features”. In: *Yearbook of Morphology 2000*. YOMO (Yearbook of Morphology). Springer, pp. 71–96.

Saint, Rachel and Kenneth L. Pike (1962). “Auca Phonemics”. In: *Studies in Ecuadorian Indian Languages 1*. Ed. by Benjamin Franklin Elson. Summer Institute of Linguistics Publications in Linguistics and Related Fields 7. Norman, Summer Institute of Linguistics of the University of Oklahoma, pp. 2–30.

Sapir, Edward (1911). “The Problem of Noun Incorporation in American Languages”. In: *American Anthropologist* 13.2, pp. 250–282.

— [1921] (2004). *Language*. Dover Publications.

Seifart, Frank (2002). *El Sistema de Classificatión Nominal Del Miraña*. CCELA and Universidad de los Andes.

— (2004). “Nominal Classification in Miraña, a Witotoan Language of Columbia”. In: *Language Typology and Universals* 57.2/3. Ed. by Alexandra Y. Aikhenvald, pp. 228–246.

Seifart, Frank (2007). “The Prehistory of Nominal Classification in Witotoan Languages”. In: *International Journal of American Linguistics* 73.4, pp. 411–445.

— (2009). “Multidimensional Typology and Miraña Class Markers”. In: *New challenges in typology. Transcending the borders and refining the distinctions*, pp. 365–385.

Seifart, Frank and Doris L. Payne (2007). “Nominal Classification in the North West Amazon. Issues in Areal Diffusion and Typological Characterization”. In: *International Journal of American Linguistics* 73.4, pp. 381–387.

Simons, Gary F. and Charles D. Fennig, eds. (2017). *Ethnologue. Languages of the World*. Version online. URL: http://www.ethnologue.com.

Sims, Andrea D (2015). *Inflectional Defectiveness*. Vol. 148. Cambridge University Press.

Sims, Andrea D, Adam Ussishkin, Jeff Parker, and Samantha Wray (2022). “At the Intersection of Cognitive Processes and Linguistic Diversity”. In: *Morphological Diversity and Linguistic Cognition*. Ed. by Andrea D Sims, Adam Ussishkin, Jeff Parker, and Samantha Wray. Cambridge University Press, pp. 1–28. DOI: 10.1017/9781108807951.

“Explorations in Seamless Morphology” (2003). In: *Explorations in Seamless Morphology*. Ed. by Rajendra Singh and Stanely Starosta. Sage Publications.

Slavin, Tanya (2012). “The syntax and semantics of stem composition in Ojicree”. University of Toronto.

Smith, E. Allyn (2010). “Correlational Comparison in English”. PhD thesis. The Ohio State University.

Steedman, Mark (1996). *Surface Structure and Interpretation*. 30. MIT Press.

Stump, Gregory (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge University Press.

— (2016). *Inflectional Paradigms. Content and Form at the Syntax-morphology Interface*. Cambridge University Press.

Sundholm, Göran (1989). “Constructive Generalized Quantifiers”. In: *Synthese. An International Journal for Epistemology, Methodology and Philosophy of Science* 79.1, pp. 1–12.

The Coq Development Team (2019). *The Coq Proof Assistant, Version 8.9.0.* DOI: 10.5281/zenodo.2554024. URL: https://doi.org/10.5281/zenodo.2554024.

Thornton, Anna M. (2012). "Reduction and Maintenance of Overabundance. A Case Study on Italian Verb Paradigms". In: *Word Structure* 5.2, pp. 183–207.

Valenzuela, Pilar M. (2016). "Classifiers in Shilwilu (Kawapanan) in Northwestern Amazonian Perspective". In: *Anthropological Linguistics* 58.4, pp. 333–380.

Van Benthem, Johan (1988). "The Semantics of Variety in Categorial Grammar". In: *Categorial Grammar*, pp. 37–55.

Van der Walt, Paul and Wouter Swierstra (2013). "Engineering Proof by Reflection in Agda". In: *Implementation and Application of Functional Languages.* Ed. by Ralf Hinze. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 157–173. ISBN: 978-3-642-41582-1. DOI: 10.1007/978-3-642-41582-1_10.

Wadler, Philip (1991). "Is There a Use for Linear Logic?" In: *ACM SIGPLAN Notices* 26.9, pp. 255–273.

White, John R. and Alejandro Monteros Altamirano (2024). "Genes in the chagra. Crop and crop wild relative landesque capital among Amazonian Runa in Ecuador". In: *Journal of Ethnobiology* 44.2, pp. 172–187. DOI: 10.1177/02780771241248905.

Wiltschko, Martina (2009). "√Root Incorporation. Evidence from Lexical Suffixes in Halkomelem Salish". In: *Lingua. International review of general linguistics. Revue internationale de linguistique générale* 119.2, pp. 199–223.

Wolfart, H Christoph (1973). "Plains Cree: A Grammatical Study". In: *Transactions of the American Philosophical Society*, pp. 1–90.

Wolvengrey, Arok (2001). *Cree Words.* Vol. 2. University of Regina Press.

Worth, Chris (2014). "The Phenogrammar of Coordination". In: *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS).* Association for Computational Linguistics, pp. 28–36.

— (2016). "English Coordination in Linear Categorial Grammar". PhD thesis. The Ohio State University.

Yasavul, Murat (2013). "Two Kinds of Focus Constructions in K'iche'". In: *Semantics and Linguistic Theory*. Vol. 23, pp. 611–632.

— (2017). "Questions and Answers in K'iche"'. PhD thesis. The Ohio State University.

Yeti Caiga, Cawetipe (2012). *Gramática Waodani Tededo Para Nivel Básico*. 1st ed. Sabiduría Amazónica 9. Cuenca, Ecuador: Universidad de Cuenca. ISBN: 978-9978-14-000-0.

Yeti Caiga, Cawetipe and Daniel Quimontari Tocari Ahua (2012). *Sabiduría de La Cultura Waodani de La Amazonía Ecuatoriana*. 1st ed. Sabiduría Amazónica 3. Cuenca, Ecuador: Universidad de Cuenca. ISBN: 978-9978-14-000-0.

Zwicky, Arnold M (1985a). "Clitics and Particles". In: *Language* 61.2, pp. 283–305.

— (1985b). "How to Describe Inflection". In: *Annual Meeting of the Berkeley Linguistics Society*. Vol. 11, pp. 372–386.

Appendix A

COQ Code for Grammatical Fragment

This appendix contains the coq code used to define the formal fragment.[1] I remove distracting details, such as pulling in libraries, and other necessary aspects of using the theorem prover. Interspersed with code examples there is discussion, which is intended to help a reader unfamiliar with coq understand the code.[2] This is not supposed to be read independently of Chapter 6. I assume the reader is able to reference the analysis.

I begin with the definition of `m`, the primitive elements of M-cats, that is to say, the things that M-cats are lists of. It is an inductive listing, where `m` is a type in `Set`. Remember that not all `list m` are M-cats. There is a predicate `Mₘ : (list m) → Prop` for valid M-cats. Only these lists are M-cats. Remember also that there are concrete and abstract M-cats, all of length one. I call length one M-cats singletons.

The first portion of the inductive type features the *m*s that correspond to roots. Numeric subscripts are used to distinguish homophonous stems. The convention is to use an initial capital letter in the names of root `m`s. Note that the root `m` do not distinguish free or bound status. There isn't really a special categorization for bound, since only a list of `m`s such that `Mₘ` is proven for the list have categories. Apart from abstract M-cats, any categorized form must correspond to something free.

The `m` subscript occurs on `m` related data structures and functions.

```
Inductive m : Set :=
```

[1]The typographic conventions are different in this Appendix due to the fact that references are being made to code.

[2]The full source code can be found at: https://github.com/noahdiewald/morphology-examples/blob/main/Wao2.v

```
| A_1m (* The stem of the noun 'plant'. *)
| A_2m (* The stem of the noun 'to see'. *)
| Ã_m (* The stem of the verb 'to say'. *)
| Ado_m (* The stem of 'same' and the numeral 'one'. *)
| Bẽye_m (* The bound stem of the noun 'fruit'. *)
| Daa_m (* The stem of 'thorn'. *)
| Di_m (* The bound stem of the noun 'stone'. *)
| Dãta_m (* The stem of the verb 'to ache'. *)
| E_m (* The water stem. *)
| Ĩ_1m (* The stem of the verb 'to be'. *)
| Ĩ_2m (* The stem of pronominal elements. *)
| Ke_m (* The stem of the verb 'to do'. *)
| Kẽ_1m (* The stem of the verb 'to eat' or 'to cut'. *)
| Kẽ_2m (* The stem of the manioc noun. *)
| Okiye_m (* The stem for 'woman'. *)
| Peẽ_m (* The bound stem for the noun 'plantain'. *)
| Te_m (* The bound stem for 'chonta' and 'chicha'. *)
| Wi_m (* The bound stem of the noun 'canoe'. *)
| Yẽdẽ_m (* The stem of the adjective 'big'. *)
```

I call $\texttt{to}_m$ a quasi-stem because it identifies the short first and second person pronouns.

```
(* quasi-stems *)
| to_m (* The pronoun suffix. *)
```

Here the lexical suffix *m* are listed. They may have numerical subscripts if there are other homophonous affixes in the system.

```
(* Lexical suffixes *)
| bõ_1m (* The 'seed' lexical suffix. *)
| dẽ_m (* The 'food' lexical suffix. *)
```

| ka$_{1m}$ (* The ’fruit’ lexical suffix. *)

| kã$_{1m}$ (* The ’body’ lexical suffix. *)

| pa$_{1m}$ (* The ’board’ lexical suffix. *)

| pẽ$_{m}$ (* The ’liquid’ lexical suffix. *)

| po$_{m}$ (* The ’hand’ lexical suffix. *)

| ta$_{1m}$ (* The ’shell’ lexical suffix. *)

| wẽ$_{m}$ (* The ’plant’ lexical suffix. *)

| yabo$_{m}$ (* The ’leaf’ lexical suffix. *)

Inflectional affixes are listed below.

(* Tense affixes *)

| ke$_{1m}$ (* The future tense suffix. *)

| ta$_{2m}$ (* The past tense suffix. *)

(* Person and number affixes *)

| bi$_{m}$ (* The second person singular suffix. *)

| bĩ$_{m}$ (* The non-singular second person suffix. *)

| bo$_{m}$ (* The first person singular suffix. *)

| bõ$_{2m}$ (* The non-singular first person suffix. *)

| da$_{m}$ (* The dual suffix. *)

| dã$_{m}$ (* The feminen suffix. *)

| di$_{m}$ (* The plural person suffix. *)

| kã$_{2m}$ (* The third person sentient suffix. *)

(* Final verbal inflection *)

| pa$_{2m}$ (* The declarative suffix. *)

| te$_{m}$ (* The gerundive suffix. *)

(* Final nominal inflection. *)

| ke$_{2m}$ (* The limitive suffix. *)

| ka$_{2m}$ (* The instrumental suffix. *)

The m of abstract M-cats are listed below. The final period indicates the end of the listing. I define a $\mathtt{none}_m$ to avoid using what is called the Maybe or option monad to define what would otherwise be partial functions.

```
(* Abstract m categories. *)
| CLOSEABLEm (* A stem that may take an affix such as -ke or -pa, *)
             (* which do not allow further affixation. *)
| DUALABLEm (* A stem for the dual. *)
| FIRSTm (* A stem with the first singular. *)
| GERABLEm (* A stem for the gerund. *)
| INFm (* A stem for inflection. *)
| LSm (* Ending in a lexical suffix. *)
| LSABLEm (* A stem for lexical suffixes. *)
| NONFIRSTm (* Not a first person stem. *)
| NUMBERABLEm (* A stem for number suffixes. *)
| PERSONm (* Ending in a person suffix. *)
| PERSONABLEm (* A stem for person affixation. *)
| PLURALABLEm (* A stem for the plural suffix. *)
| ROOTm (* A singleton. *)
| SAFEFIRSTm (* The -bi affix can replace -bo but not if -ta *)
             (* precedes -bo. The stem has -bo but not -tabo. *)
| SECONDm (* A stem for the second singular. *)
(* For instances when a meaningless m is needed. *)
| nonem.
```

Below I define boolean equality of m. The $+$ is the disjoint sum constructor. Both equality = and inequality $\neq$ of m are of type Prop, for instance, $\mathtt{LS}_m$ = $\mathtt{LS}_m$: Prop. Working under Prop involves manual proofs in cases where one may prefer an automatic calculation. For instance, providing a manual calculation that something is a member of a list can be time consuming. If one is only examining the formalism, it isn't an issue, but if one wishes to actually use the definitions in this document within Coq, the definitions below are convenient time savers. For simple inductive types like m, it is not difficult to provide a conversion from proofs of equality to boolean values. Booleans are in the universe of Set and behave according the classical rules of boolean logic. The first step to doing the conversion is declaring a type m_dec, which takes two m and returns a disjoint sum type. The sum type in Coq, used here, is called sumbool, and takes two Prop arguments. It is defined inductively, such that inleft (_:A) or inright (_:B), where A and B are the two Prop arguments. So, a proof of A results in the construction of the type using inleft. Note that curly braces are used around the (in)equalities. This means that these are implicit arguments. Their witnesses may be inferred from the definition of the inductive type m. For this reason, a proof of m_dec may be supplied directly with the decide equality tactic. The result of this is to say, that m may be equal or not equal but not both. They will always be one or the other. The if _ then _ else _ function is defined such that the first clause is returned on inleft, while the second, else clause is returned on inright. This provides the conversion to boolean values. The result is that functions that utilize boolean equality are now compatible with m.

```
Definition m_dec : ∀ α β : m, {α = β} + {α ≠ β}.
Proof. decide equality. Defined.
Definition eqₘ α β := if m_dec α β then true else false.
```

M_m is a predicate on list of ms that indicates an M-cat is well-formed. The reason for doing this is that not all permutations of all sublists of m correspond to grammatical categories of form paradigm entries.

```
Axiom Mₘ : (list m) → Prop.
```

As a reminder of where this is heading, form paradigm entries are pairs of lists of ms and processes, called $struct_{mp}$s, which are well formed according to the predicate $FE_{mp} : struct_{mp} \rightarrow Prop$. Form paradigm entries belong to form paradigm instances. There is no literal data structure that groups form paradigm entries. A form paradigm instance is identified by a form class and a stem, which is a $struct_{mp}$ who's pairs match some stem of any length of a form paradigm entry. Both the form paradigm entry and the stem must have the same form class. The stem need not be a form paradigm entry. The $klass$ function maps lists of ms to form classes. These lists do not need to be M-cats. The predicate in_{mp}, at the bottom of the code examples, indicates form paradigm membership.

Returning to the discussion of M-cats, according to convention, which may depend on the needs of a particular analysis or the preferred style of the theorist, some m are only indirectly associated with morphological process lists, and only occur as singletons. I call the M-cats that correspond to particular morphological processes *concrete*. Other categories are used to provide super categories for these categories. I call these categories *abstract* categories, and it is usefult to be able to refer to a list of them.

The abstract M-cats come in two flavors. One describes how a stem is, such as LS_m, used when the stem ends in a lexical suffix. The other describes the stems potential for further affixation, such as LSABLE_m, a stem that may be affixed with an lexical suffix.

```
Definition abstractsₘ : list m :=
  [ CLOSEABLEₘ ;
    DUALABLEₘ ;
    FIRSTₘ ;
    GERABLEₘ ;
    INFₘ ;
    LSₘ ;
    LSABLEₘ ;
    NONFIRSTₘ ;
    NUMBERABLEₘ ;
    PERSONₘ ;
    PERSONABLEₘ ;
    PLURALABLEₘ ;
    ROOTₘ ;
    SAFEFIRSTₘ ;
    SECONDₘ
  ].
```

There are other m that are usefully grouped to make the rules that define the partial order on M more succinct.

`lssₘ` is a list of lexical suffixes. Items are described in the comments on the definition of m, above.

```
Definition lss_m : list m :=
  [ bõ_1_m ;
    dẽ_m ;
    ka_1_m ;
    kã_1_m ;
    pa_1_m ;
    po_m ;
    pẽ_m ;
    ta_1_m ;
    wẽ_m ;
    yabo_m
  ].
```

nonfirsts$_m$ are affixes used for non-first person and number marking.

```
Definition nonfirsts_m : list m :=
  [ bi_m ;
    bĩ_m ;
    da_m ;
    dã_m ;
    di_m ;
    kã_2_m
  ].
```

persons$_m$ are all person marking.

```
Definition persons_m : list m := bo_m :: bõ_2_m :: nonfirsts_m.
```

Classes of roots that may be useful to refer to are included in the lists below.

```
Definition oneroots_m : list m := [ Ado_m ].
```

```
Definition verbrootsₘ : list m := [ A₂ₘ ; Ãₘ ; Keₘ ].

Definition verblsrootssₘ : list m := [ Dãtaₘ ; Kẽ₁ₘ ].

Definition adjrootsₘ : list m := [ Yẽdẽₘ ].

Definition inanimrootsₘ : list m :=
[ A₁ₘ ; Daaₘ ; Diₘ ; Eₘ ; Kẽ₂ₘ ; Peẽₘ ; Teₘ ; Wiₘ ].

Definition animrootsₘ : list m := [ Okiyeₘ ].
```

K are names of form classes, similar in concept to inflection classes. The uppercase kappa K is a mnemonic for /klæs/. Like m the k subscript is used for names of types, functions and relations associated with K. Variables of type K are written as κ or κ_n. The none_k class is for the null case of a list of categories. It has no theoretical meaning. It is used so that the $klass$ function below can be defined as a total function, rather than using $Maybe$/$option$ monads.

```
Inductive K : Set :=
| adjₖ (* Adjective-like items have person marking and lexical
  suffixes in competition. *)
| animₖ (* Animate noun-like items have person marking but no lexical
  suffixes. *)
| awẽₖ (* The noun awẽ ends in -wẽ, only. *)
| bẽyebõₖ (* -bõ and -ka lexical suffix *)
| bodyₖ (* Some items take only body-part affixes. *)
| deverbₖ (* Items that take -te *)
| dikaₖ (* The noun dika ends in -ka, only. *)
| epẽₖ (* The noun epẽ ends in -pẽ, only. *)
| inanimₖ (* Inanimate noun-like items may have more than one LS, but
  no person marking. *)
```

```
| kẽdẽₖ (* The stem for manioc has a -we, and -dẽ ending but nothing
  else. *)
| oneₖ (* The class of adoke, $one’, and ado, $same’. *)
| õdõbõₖ (* -bõ lexical suffix *)
| õyaboₖ (* -yabo lexical suffix *)
| nominalₖ (* Things that may serve as NPs *)
| participantₖ (* The class that gets first and second person *)
| personₖ (* Items that take person marking. *)
| peẽdẽₖ (* The noun peẽdẽ ends only in -dẽ. *)
| plantₖ (* Compatible with plant meanings. *)
| proₖ (* The ĩ pronouns *)
| singlelsₖ (* This covers adjectives and LS taking verbs. *)
| tepẽₖ (* The noun tepẽ and tewẽ, which have only two endings. *)
| thingₖ (* Items that take any LS. *)
| verbbodyₖ (* Verbs that take only body LSs *)
| verblsₖ (* Verbs that take LSs. *)
| verbₖ (* General verbs, which may not take LSs. *)
| wipoₖ (* The noun wipo ends only in -po. *)
| noneₖ. (* The default class. *)
```

See above for the definition of eq_m, to understand the definitions below.

```
Definition K_dec : ∀ α β : K, {α = β} + {α ≠ β}.
Proof. decide equality. Defined.
Definition eqₖ α β := if K_dec α β then true else false.
```

Below is the list of rules for the form class order. Items on the left of the pairs are ordered below those on the right. These rules are referred to in the definition of the $\leq_k$ order.

```
Definition le_rulesₖ : list (K * K) :=
  [ (bodyₖ, dikaₖ) ; (* Items with the LS -ka may be body part LSs. *)
```

```
(body_k, wipo_k) ; (* Items with the LS -po may be body part LSs. *)
(tepẽ_k, epẽ_k) ; (* The te- stem may end in -pẽ. *)
(tepẽ_k, awẽ_k) ; (* The te- stem may end in -wẽ. *)
(kẽdẽ_k, peẽdẽ_k) ; (* The kẽ- stem may end in -dẽ. *)
(kẽdẽ_k, awẽ_k) ; (* The kẽ- stem may end in -wẽ. *)
(bẽyebõ_k, õdõbõ_k) ;
(bẽyebõ_k, dika_k) ;
(body_k, bẽyebõ_k) ;
(thing_k, õyabo_k) ;
(thing_k, kẽdẽ_k) ; (* The general ls-taking class includes -wẽ and -dẽ *)
(thing_k, tepẽ_k) ; (* The general ls-taking class includes -wẽ and -pẽ *)
(thing_k, body_k) ; (* The general ls-taking class includes body affixes *)
(person_k, participant_k) ; (* Some items only take first person or second person *)
(anim_k, person_k) ; (* Animate nouns may take person marking *)
(adj_k, person_k) ; (* Adjectives may take person marking *)
(verb_k, person_k) ; (* Verbs may take person marking *)
(one_k, person_k) ; (* The numeral one may take person marking *)
(pro_k, person_k) ;
(verbls_k, verb_k) ; (* Verbs that take lexical suffixes are verbs *)
(verbbody_k, verbls_k) ; (* Verbs that take body LS only are LS taking verbs *)
(verb_k, deverb_k) ; (* Deverb is for deverbal affixes, like the gerund. *)
```

```
    (proₖ, deverbₖ) ; (* The copula is split off from normal verbs. *)
    (verbbodyₖ, bodyₖ) ; (* Verbs that take body LS *)
    (adjₖ, thingₖ) ; (* Adjectives can take any lexical suffix *)
    (oneₖ, thingₖ) ; (* The numeral one may take any lexical suffix *)
    (inanimₖ, thingₖ) ; (* Inanimate nominals may take any lexical
    suffix *)
    (plantₖ, inanimₖ) ; (* some inanimate things have plant meanings *)
    (verblsₖ, singlelsₖ) ; (* Verbs only take one LS *)
    (adjₖ, singlelsₖ) ; (* Adjectives only take one LS *)
    (adjₖ, nominalₖ) ;
    (animₖ, nominalₖ) ;
    (inanimₖ, nominalₖ) ;
    (participantₖ, nominalₖ) ;
    (proₖ, nominalₖ)
  ].
```

The definition of the form class order is below. The normal properties of an order are defined. There is also a reference to the rules above, which list explicitly ordered `K` instances. The `Infix` line provides and alias, so that $\leq_k$ can be used as an infix, rather than `leₖ`, which occurs before its arguments.

```
Inductive leₖ : K → K → Prop :=
| reflₖ : ∀ α, leₖ α α
| transₖ : ∀ α β γ, leₖ α β → leₖ β γ → leₖ α γ
| rulesₖ : ∀ α β, inrulesₖ α β le_rulesₖ = true → leₖ α β.

Axiom antisymₖ : ∀ α β, leₖ α β → leₖ β α → α = β.

Infix ”≤ₖ” := leₖ (at level 60, right associativity).
```

The *klass* function returns the ‘formclass’ for a stem. It is a recursive function, which assumes that the important stem information is early in the list. If it doesn't match anything, it returns none_k, which by convention is never used in rules. A subset of `list m` terms may not correspond to a M_m. For instance, the item may correspond to a bound root. For that reason, the definition of $klass$ takes $listm$, rather than M_m.

I haven't explained pattern matching much in the code above. Most of the items below match a singleton value, such as `[Bẽyeₘ]`. Some match a specific sequence of more than one m, such as `[wẽₘ ; Teₘ]`. Sometimes, especially for bound stems that take only one affix, one doesn't need to specify which affix one in concerned with, and a _ is used as a wildcard, for instance, `[_ ; Diₘ]`.

```
Fixpoint klass (α : list m) : K :=
  match α with
  | [] => noneₖ
  | [A₁ₘ] => awẽₖ
  | [A₂ₘ] => verbₖ
  | [Ãₘ] => verbₖ
  | [Adoₘ] => oneₖ
  | [Bẽyeₘ] => bẽyebõₖ
  | [Daaₘ] => plantₖ
  | [Diₘ] => dikaₖ
  | [Dãtaₘ] => verbbodyₖ
  | [Eₘ] => epẽₖ
  | [Ĩ₁ₘ] => verbₖ
  | [Ĩ₂ₘ] => proₖ
  | [Keₘ] => verbₖ
  | [Kẽ₁ₘ] => verbbodyₖ
  | [Kẽ₂ₘ] => kẽdẽₖ
  | [Okiyeₘ] => animₖ
  | [Peẽₘ] => peẽdẽₖ
  | [Teₘ] => tepẽₖ
  | [toₘ] => participantₖ
```

```
  | [Wiₘ] => wipoₖ
  | [Yẽdẽₘ] => adjₖ
  | [_ ; A₁ₘ] => plantₖ
  | [_ ; Bẽyeₘ] => plantₖ
  | [_ ; Diₘ] => inanimₖ
  | [_ ; Eₘ] => inanimₖ
  | [_ ; Kẽ₂ₘ] => plantₖ
  | [_ ; Peẽₘ] => plantₖ
  | [wẽₘ ; Teₘ] => plantₖ
  | [pẽₘ ; Teₘ] => plantₖ
  | [_ ; Wiₘ] => inanimₖ
  | _ :: t => klass t
  end.
```

I define some helper predicates for determining super categories of M_m. The in_m function looks to see whether some m is a member of a list. The $inabb_m$ function is a helper function to a helper function, called $inab_m$, which is short for "in abstract". There is a list of pairs of abstract M-cats. The $inabb_m$ portion does nothing more than read through an input list and attempt to match a pair of singleton lists of m. The main $inab_m$ returns false for bad input, such as empty lists, before calling $inabb_m$. The $hdIn_m$ function takes two lists and determines if the head of the first list is a member of the second list. The $isNonFirst_m$ function references a list of m that correspond to person and number marking, which excludes the first persons, which I have proved above the function definition. It takes a list of m as input and matches on the first two heads. If the head is in the $nonfirsts_m$ list, and the second element is not bõ$_m$, which precedes number marking, then true is returned. Otherwise false is returned.

```
Fixpoint inₘ (α : m) (l : list m) : bool :=
  match l with
  | [] => false
  | x :: t => if eqₘ x α then true else inₘ α t
  end.
```

```
Fixpoint inabbₘ (α : m) (β : m) (l : list (list m * list m)) : bool :=
      match l with
      | [] => false
      | ([x], [y]) :: t =>
          match (eqₘ α x), (eqₘ β y) with
          | true, true => true
          | _, _ => inabbₘ α β t
          end
      | _ :: t => inabbₘ α β t
      end.


Definition inabₘ (α : list m) (β : list m) : bool :=
    match α, β with
    | [], _ => false
    | _, [] => false
    | _ :: _ :: _, _ => false
    | _, _ :: _ :: _ => false
    | [x], [y] => inabbₘ x y abstract_leₘ_rulesₘ
    end.

Definition hdIn (l₁ : list m) (l₂ : list m) : bool :=
  match l₁ with
  | [] => false
  | x :: _ => inₘ x l₂
  end.
```

```
Definition nonfirstsₘ : list m :=
  [ biₘ ;
    bĩₘ ;
    daₘ ;
    dãₘ ;
    diₘ ;
    kã₂ₘ
  ].

Definition isNonFirstₘ (l₁ : list m) : bool :=
  match l₁ with
  | [] => false
  | [x] => false
  | x :: y :: _ => andb (inₘ x nonfirstsₘ) (negb (eqₘ y bõ₂ₘ))
  end.
```

Relationships between abstract M_m, which are the same for all form classes, are defined in a manner similar to the form class order.

```
Definition abstract_leₘ_rulesₘ : (list (list m * list m)) :=
  [ ([ ROOTₘ ], [ PERSONABLEₘ ]) ;
    ([ PERSONABLEₘ ], [ DUALABLEₘ ]) ; (* Although not all roots take
    person affixes, most do and those that don't will be of  the wrong
    form class for this ordering to matter. *)
    ([ NUMBERABLEₘ ], [ DUALABLEₘ]) ;
    ([ NUMBERABLEₘ ], [ PLURALABLEₘ ]) ;
    ([ PERSONABLEₘ ], [ CLOSEABLEₘ ]) ;
    ([ PERSONₘ ], [ CLOSEABLEₘ ])
  ].
```

The below states that all singeltons of abstract *m* are M_m.

```
Axiom Mₘ_abstractsₘ : ∀ (α : m), inₘ α abstractsₘ = true → Mₘ [α].
```

A partial order is defined over *m* lists, where only M_m instances are ordered. The basic properties are defined in the inductive type and a following axiom of antisymetry. The meat of the definition are the more specific constructors that define the language specific category order.

```
Inductive leₘ : (list m) → (list m) → Prop :=
| reflₘ : ∀ α, Mₘ α → leₘ α α
| transₘ : ∀ α β γ, Mₘ α → Mₘ β → Mₘ γ → leₘ α β → leₘ β γ → leₘ α γ
(* All non-abstract singleton M-cats are ≤ₘ [ ROOTₘ ]. *)
| rootₘ : ∀ α, Mₘ α → length α = 1 → ~ hdIn α abstractsₘ = true →
  leₘ α [ROOTₘ]
(* When the last process was a lexical suffix process, the category is
a subcategory of the M-cat [LSₘ]. *)
| lsₘ : ∀ α, Mₘ α → hdIn α lssₘ = true → leₘ α [LSₘ]
(* When the last $m$ corresponds to person marking, the category is a
subcategory of the M-cat [PERSONₘ]. *)
| prₘ : ∀ α, Mₘ α → hdIn α personsₘ = true → leₘ α [PERSONₘ]
(* bõ₂ₘ and bĩₘ correspond to stem endings where person number affixes
may be added. Either plural or dual may be added to such
stems. Only plural may be added to a stem with dãₘ. See below. *)
| numerableₘ : ∀ α, Mₘ α → hdIn α [bõ₂ₘ;bĩₘ] = true →
  leₘ α [NUMBERABLEₘ]
(* The placement of the past tense morph depends on whether or not the
person marking on a verb is first person or otherwise. *)
| nonfirstₘ : ∀ α, Mₘ α → isNonFirstₘ α = true → leₘ α [NONFIRSTₘ]
| firstₘ : ∀ α, Mₘ α → hd noneₘ α = boₘ → leₘ α [FIRSTₘ]
| safefirstₘ : ∀ α, Mₘ α → hd noneₘ α = boₘ →
```

```
hd noneₘ (tail α) ≠ ta₂ₘ → leₘ α [SAFEFIRSTₘ]
| secondₘ : ∀ α, Mₘ α → hd noneₘ α = biₘ → leₘ α [SECONDₘ]
(* The affix dãₘ may be followed by plural marking. *)
| dãplₘ : ∀ α, Mₘ α → hdIn α [dãₘ] = true → leₘ α [PLURALABLEₘ]
(* The relation between abstract M-cats is listed separately. *)
| abstract_leₘ : ∀ α β, inabₘ α β = true → leₘ α β
(* The M-cats which describe morphotactics are parameterized based on class.
 *)
(* Some verbs, demonstratives, and adjectives allow a single LS *)
| singleₘ : ∀ α, Mₘ α → klass α ≤ₖ singlelsₖ → leₘ [ROOTₘ] [LSABLEₘ]
(* LSs and person marking don't compete on verbs. *)
| verblsₘ : ∀ α, Mₘ α → klass α ≤ₖ verblsₖ → leₘ [LSₘ] [PERSONABLEₘ]
(* Non-finite verb-like things may take a gerund directly on the root *)
| verbgerₘ : ∀ α, Mₘ α → klass α ≤ₖ deverbₖ → leₘ [ROOTₘ] [GERABLEₘ]
(* The "copula" may take the gerund after person marking *)
| accₘ : ∀ α, Mₘ α → klass α ≤ₖ proₖ → leₘ [PERSONₘ] [GERABLEₘ]
(* Inanimate nouns may take any number of LS *)
| inanimₘ : ∀ α, Mₘ α → klass α ≤ₖ inanimₖ → leₘ [LSₘ] [LSABLEₘ]
(* Two definitions of fully inflected *)
| nominfₘ : ∀ α, Mₘ α → klass α ≤ₖ nominalₖ → leₘ α [INFₘ]
| verbinfₘ : ∀ α, Mₘ α → klass α ≤ₖ verbₖ → hd noneₘ α = pa₂ₘ →
  leₘ α [INFₘ].

Axiom antisymₘ : ∀ α β : (list m), leₘ α β → leₘ β α → α = β.

Infix "≤ₘ" := leₘ (at level 60, right associativity).
```

Rather than eagerly building up strings, morphological rules build a list of processes that are applied at some point of evaluation, such as evaluating string equality. The $apply_{pr}$ function applies all of the processes in order to the empty string. Processes and process related functions have a *pr* subscript.

```
Definition processpr := (string → string) → (string → string).


Fixpoint applypr (processes : list processpr)
  (acc : string → string) : string :=
  match processes, acc with
  | nil, acc' => acc' ""
  | p :: ps, acc' => applypr ps (p acc')
  end.
```

Processes are not string to string functions but functions from string to string functions to string to string functions. Stems, notably, take an input string to string function and return a constant-like function, which disgards its input. I provide two helper functions for stems and suffixes below. Only to_{pr} has a significantly different definition.

```
Definition stempr (s : string) (p : string → string) : (string → string) :=
  λ (_ : string), p s.


Definition suffixpr (s : string) (p : string → string) : (string → string) :=
  λ (stem : string), p (stem ++ s).


Definition apr := stempr "a".


Definition ãpr := stempr "ã".


Definition adopr := stempr "ado".


Definition bẽyepr := stempr "bẽye".
```

```
Definition bi_pr := suffix_pr "bi".

Definition bĩ_pr := suffix_pr "bĩ".

Definition bo_pr := suffix_pr "bo".

Definition bõ_pr := suffix_pr "bõ".

Definition da_pr := suffix_pr "da".

Definition dã_pr := suffix_pr "dã".

Definition daa_pr := stem_pr "daa".

Definition dãta_pr := stem_pr "dãta".

Definition dẽ_pr := suffix_pr "dẽ".

Definition di1_pr := stem_pr "di".

Definition di2_pr := suffix_pr "di".

Definition e_pr := stem_pr "e".

Definition ĩ_pr := stem_pr "ĩ".

Definition ka_pr := suffix_pr "ka".
```

```
Definition kãₚᵣ := suffixₚᵣ "kã".

Definition kẽₚᵣ := stemₚᵣ "kẽ".

Definition keₚᵣ := stemₚᵣ "ke".

Definition okiyeₚᵣ := stemₚᵣ "okiye".

Definition paₚᵣ := suffixₚᵣ "pa".

Definition pẽₚᵣ := suffixₚᵣ "pẽ".

Definition peẽₚᵣ := stemₚᵣ "peẽ".

Definition poₚᵣ := suffixₚᵣ "po".

Definition taₚᵣ := suffixₚᵣ "ta".

Definition teₚᵣ := suffixₚᵣ "te".

Definition toₚᵣ (p : string → string) : (string → string) :=
  λ (s : string), (p s) ++ "to".

Definition wẽₚᵣ := suffixₚᵣ "wẽ".

Definition wiₚᵣ := suffixₚᵣ "wi".

Definition yaboₚᵣ := suffixₚᵣ "yabo".
```

```
Definition yẽdẽpr := stempr "yẽdẽ".
```

The id_{pr} function is used as input to the result of applying processes.

```
Definition idpr (s : string) : string := s.
```

The data structure used by the form paradigm is a pair of a category `list m` and a list of processes.

```
Definition structmp := (list m * list processpr).
```

The $combine_{mp}$ function, which will only append additional m and process information if there is no intersection of m is defined below.

```
Fixpoint intersectionm (α β : list m) : bool :=
  match α with
  | [] => false
  | x :: t => match inm x β with
              | true => true
              | false => intersectionm t β
              end
  end.

Definition combinemp (newm prevm : list m)
  (newpr prevpr : list processpr) : structmp :=
  match newm with
  | [] => (prevm, prevpr)
  | _ => match intersectionm newm prevm with
         | true => (prevm, prevpr)
         | false => (app newm prevm, app newpr prevpr)
         end
  end.
```

The following are helper functions for the $\texttt{interchange}_{mp}$ constructor below. The $isStem_{mp}$ predicate tests whether $stem_{mp}$ is a stem of $word_{mp}$. I'll explain this in detail so that it is clear what a stem means in this context. The first step reverses the m and process lists of the potential stem because manipulation and matching on lists is more convenient from head to tail. The stem information would be some tail of the m and process lists for $word_{mp}$. The lengths of the $stem_{mp}$ lists are used to take an equally sized portion of the reversals of the $word_{mp}$ lists. These sub-portions of the lists can then be directly compared with the stem lists to determine whether $stem_{mp}$ is an actual stem.

The $remove_{mp}$ function removes a stem from a word, leaving only what followed the stem.

```
Definition isStemₘₚ (stemₘₚ wordₘₚ : structₘₚ) : Prop :=
  let stemₘ := rev (fst stemₘₚ) in
  let stemₚᵣ := rev (snd stemₘₚ) in
  let lengthₘ := length stemₘ in
  let lengthₚᵣ := length stemₚᵣ in
  let wordₘ := firstn lengthₘ (rev (fst wordₘₚ)) in
  let wordₚᵣ := firstn lengthₚᵣ (rev (snd wordₘₚ)) in
  stemₘ = wordₘ ∧ stemₚᵣ = wordₚᵣ.

Definition removeₘₚ (stemₘₚ wordₘₚ : structₘₚ) : structₘₚ :=
  let stemₘ := fst stemₘₚ in
  let stemₚᵣ := snd stemₘₚ in
  let lₘ := length stemₘ in
  let lₚᵣ := length stemₚᵣ in
  let wordₘ := fst wordₘₚ in
  let wordₚᵣ := snd wordₘₚ in
  let newₘ := skipn lₘ (rev wordₘ) in
  let newₚᵣ := skipn lₚᵣ (rev wordₚᵣ) in
  (newₘ, newₚᵣ).
```

A form to form mapping rule has the following structure. There is an input compound category, a form class constraint, a new category that will be added to the compound category and a new process to be added to the process list.

The first rule schema for defining a form-form mapping is `rule1ₘₚ`. This rule takes the category that the rules will be constrained by, the class it will be constrained by, some new category information and new process information. It provides a function that takes an input form entry `struct`, and proofs that the entry matches the category and class conditions. It then returns a new entry where the new category and process information is appended to the existing information.

```
Definition rule1ₘₚ (catₘ : list m) (κ : K) (newₘ : list m)
  (newₚᵣ : list processₚᵣ) :=
  λ (α : structₘₚ)
    (proofₘ : fst α ≤ₘ catₘ)
    (proofₖ : klass (fst α) ≤ₖ κ),
    combineₘₚ newₘ (fst α) newₚᵣ (snd α).
```

The `rule2ₘₚ` schema is essentially the same as `rule1ₘₚ`, except that `combineₘₚ` takes the tails of the input category and process information, roughly swapping out the heads. This is why these rules are called lateral rules.

```
Definition rule2ₘₚ (catₘ : list m) (κ : K) (newₘ : list m)
  (newₚᵣ : list processₚᵣ) :=
  λ (α : structₘₚ)
    (proofₘ : fst α ≤ₘ catₘ)
    (proofₖ : klass (fst α) ≤ₖ κ),
    combineₘₚ newₘ (tail (fst α)) newₚᵣ (tail (snd α)).
```

The inductive definition of valid form paradigm entries is given as `FEₘₚ`. Some items have constructors that I have referred to as axioms, because they are simply given, such as `dikaMP`. There are also a fair number of invocations of `rule2ₘₚ` and `rule1ₘₚ`, which define rules. There is also `interchangeMP`.

```
Inductive FEₘₚ : structₘₚ → Prop :=
```

```
| interchangeMP : ∀ (s₁ s₂ s₃ s₄ : structₘₚ) (κ : K),FEₘₚ s₃ →
  κ = klass (fst s₁) → κ = klass (fst s₂) →
  κ = klass (fst s₃) → isStemₘₚ s₁ s₃ →
  s₄ = combineₘₚ (fst s₂) (fst (removeₘₚ s₁ s₃))
    (snd s₂) (snd (removeₘₚ s₁ s₃)) → FEₘₚ s₄
| analogyMP :
  ∀ (stem₁ stem₂ inWord medWord₁ medWord₂ target : structₘₚ) (κ : K),
    FEₘₚ medWord₂ →
    κ = klass (fst stem₁) →
    κ = klass (fst stem₂) →
    κ = klass (fst inWord) →
    κ = klass (fst medWord₂) →
    isStemₘₚ stem₁ inWord →
    isStemₘₚ stem₁ target →
    isStemₘₚ stem₂ medWord₁ →
    isStemₘₚ stem₂ medWord₂ →
    removeₘₚ stem₁ inWord = removeₘₚ stem₂ medWord₁ →
    removeₘₚ stem₂ medWord₂ = removeₘₚ stem₁ target →
    FEₘₚ target
| aMP : FEₘₚ ( [A₂ₘ], [aₚᵣ] )
| awẽMP : FEₘₚ ( [wẽₘ ; A₁ₘ], [wẽₚᵣ ; aₚᵣ] )
| ãMP : FEₘₚ ( [Ãₘ], [ãₚᵣ] )
| adoMP : FEₘₚ ( [Adoₘ], [adoₚᵣ] )
| bẽyekaMP : FEₘₚ ( [ka₁ₘ ; Bẽyeₘ], [kaₚᵣ ; bẽyeₚᵣ] )
| bẽyebõMP : FEₘₚ ( [bõ₁ₘ ; Bẽyeₘ], [bõₚᵣ ; bẽyeₚᵣ] )
| botoMP : FEₘₚ ( [boₘ ; toₘ], [boₚᵣ ; toₚᵣ] )
| daaMP : FEₘₚ ( [Daaₘ], [daaₚᵣ] )
| dikaMP : FEₘₚ ( [ka₁ₘ ; Diₘ], [kaₚᵣ ; di₁ₚᵣ] )
```

```
| dãtaMP : FEₘₚ ( [Dãtaₘ], [dãtaₚᵣ] )
| epẽMP : FEₘₚ ( [pẽₘ ; Eₘ], [pẽₚᵣ ; eₚᵣ] )
| ĩ₁MP : FEₘₚ ( [Ĩ₁ₘ], [ĩₚᵣ] )
| ĩ₂MP : FEₘₚ ( [Ĩ₂ₘ], [ĩₚᵣ] )
| keMP : FEₘₚ ( [Keₘ], [keₚᵣ] )
| kẽMP : FEₘₚ ( [Kẽ₁ₘ], [kẽₚᵣ] )
| kẽdẽMP : FEₘₚ ( [dẽₘ ; Kẽ₂ₘ], [dẽₚᵣ ; kẽₚᵣ] )
| kẽwẽMP : FEₘₚ ( [wẽₘ ; Kẽ₂ₘ], [wẽₚᵣ ; kẽₚᵣ] )
| okiyeMP : FEₘₚ ( [Okiyeₘ], [okiyeₚᵣ] )
| peẽdẽMP : FEₘₚ ( [dẽₘ ; Peẽₘ], [dẽₚᵣ ; peẽₚᵣ] )
| tepẽMP : FEₘₚ ( [pẽₘ ; Teₘ], [pẽₚᵣ ; teₚᵣ] )
| tewẽMP : FEₘₚ ( [wẽₘ ; Teₘ], [wẽₚᵣ ; teₚᵣ] )
| wipoMP : FEₘₚ ( [poₘ ; Wiₘ], [poₚᵣ ; wiₚᵣ] )
| yẽdẽMP : FEₘₚ ( [Yẽdẽₘ], [yẽdẽₚᵣ] )
| bõ₁MP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] õdõbõₖ [bõ₁ₘ] [bõₚᵣ])
                    α proofₘ proofₖ)
| dẽMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] peẽdẽₖ [dẽₘ] [dẽₚᵣ])
                    α proofₘ proofₖ)
| kaMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] dikaₖ [ka₁ₘ] [kaₚᵣ])
                    α proofₘ proofₖ)
| kã₁MP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] bodyₖ [kã₁ₘ] [kãₚᵣ])
                    α proofₘ proofₖ)
| paMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] thingₖ [pa₁ₘ] [paₚᵣ])
```

```
                    α proofₘ proofₖ)
| poMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] wipoₖ [poₘ] [poₚᵣ])
                    α proofₘ proofₖ)
| pẽMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] epẽₖ [pẽₘ] [pẽₚᵣ])
                    α proofₘ proofₖ)
| taMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] thingₖ [ta₁ₘ] [taₚᵣ])
                    α proofₘ proofₖ)
| wẽMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] awẽₖ [wẽₘ] [wẽₚᵣ])
                    α proofₘ proofₖ)
| yaboMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] õyaboₖ [yaboₘ] [yaboₚᵣ])
                    α proofₘ proofₖ)
| ke₁MP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [LSABLEₘ] verbₖ [ke₁ₘ] [keₚᵣ])
                    α proofₘ proofₖ)
| biMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [SAFEFIRSTₘ] participantₖ [biₘ] [biₚᵣ])
                    α proofₘ proofₖ)
| bĩMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule2ₘₚ [SECONDₘ] participantₖ [bĩₘ] [bĩₚᵣ])
                    α proofₘ proofₖ)
| boMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [PERSONABLEₘ] participantₖ [boₘ] [boₚᵣ])
                    α proofₘ proofₖ)
```

```
| bõ₂MP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule2ₘₚ [FIRSTₘ] participantₖ [bõ₂ₘ] [bõₚᵣ])
                    α proofₘ proofₖ)
| daMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [DUALABLEₘ] personₖ [daₘ] [daₚᵣ])
                    α proofₘ proofₖ)
| dãMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [PERSONABLEₘ] personₖ [dãₘ] [dãₚᵣ])
                    α proofₘ proofₖ)
| diMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [PLURALABLEₘ] personₖ [diₘ] [di₂ₚᵣ])
                    α proofₘ proofₖ)
| kã₂MP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [PERSONABLEₘ] personₖ [kã₂ₘ] [kãₚᵣ])
                    α proofₘ proofₖ)
| pa₂MP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [CLOSEABLEₘ] verbₖ [pa₂ₘ] [paₚᵣ])
                    α proofₘ proofₖ)
| teMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [GERABLEₘ] verbₖ [teₘ] [teₚᵣ])
                    α proofₘ proofₖ)
| tapaMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [NONFIRSTₘ] verbₖ [pa₂ₘ ; ta₂ₘ] [paₚᵣ ; taₚᵣ])
                    α proofₘ proofₖ)
| taboMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [PERSONABLEₘ] verbₖ [boₘ ; ta₂ₘ] [boₚᵣ ; taₚᵣ])
                    α proofₘ proofₖ).
```

Anything that is the category of a proveable form paradigm member has a validly named compound category.

```
Axiom Mₘ_are_FE_fst : ∀ (cat : list m) (α : structₘₚ),
  FEₘₚ α → cat = fst α → Mₘ cat.
```

Form paradigm string equivalence states that when two structures that have process lists that reduce to the same string, they are equivalent, so long as their compound categories are valid names. It may be that the string is produced by distinct process lists or that the compound category is different. The `str_equivₘₚ` constructor states that they are never the less equivalent, equal strings are the determining factor. If two structures are equivalent, compound categories may be swapped.

```
Inductive equivₘₚ : structₘₚ → structₘₚ → Prop :=
| reflₘₚ : ∀ α : structₘₚ, Mₘ (fst α) → equivₘₚ α α
| symₘₚ : ∀ α β : structₘₚ, Mₘ (fst α) → Mₘ (fst β) →
  equivₘₚ α β → equivₘₚ β α
| transₘₚ : ∀ α β γ, Mₘ (fst α) → Mₘ (fst β) → Mₘ (fst γ) →
  equivₘₚ α β → equivₘₚ β γ → equivₘₚ α γ
| str_equivₘₚ : ∀ α β : structₘₚ, Mₘ (fst α) → Mₘ (fst β) →
  applyₚᵣ (snd α) = applyₚᵣ (snd β) → equivₘₚ α β.

Infix ”≡ₘₚ” := equivₘₚ (at level 90).
```

Below is a highly simplified tectogrammatical type for LCG.

As far as I have been able to tell, all of what I have named Nom, N, and Adj take the same full range of person annotations. If this were found not to be the case, the inductive definition of τ need not be the place to add constraints. This can also be done in the definition of form-sign relations.

The tecto N ⊸ Adj is used instead of N ⊸ N because it does not appear to be possible to use multiple adjectives with a noun. Some category other that N is needed for that reason. That being said, Adj is a nominal category, not an adjectival category.

```
Inductive τ_anno : Set :=
```

```
| t1_du
| t1_inc
| t1_pl
| t1_sg
| t2_du
| t2_mo
| t2_pl
| t2_sg
| t3_du
| t3_f
| t3_h
| t3_pl
| t.

Inductive τ : Set :=
| Nom (α : τ_anno)
| Acc
| N (α : τ_anno)
| Adj (α : τ_anno)
| Fin
| infτ (α β : τ).

Infix ”-o” := infτ (at level 60, right associativity).
```

These are some helper functions to strip out non-person *m* from an M-cat.

```
Definition personFilterₘ : list m → list m :=
  filter (λ x,inₘ x personsₘ).

Definition nonfirstFilterₘ : list m → list m :=
```

```
filter (λ x,inₘ x personsₘ).
```

Below is a portion of Jordan Needle's formalization of Agnostic Hyper-intentional Semantics. Much of what makes the theory agnostic and hyper-intentional is not provided here. The goal is simply to embed the terms of the semantic theory below a single type. The idea is to provide an encoding of the types of the many types of the semantic theory under an inductive type `stat_term`, which is a single type within *Set*. The *ent*, *prp* and *func* constructors are simple terms of type `stat_term` and may not have inhabitants. `Sns` is a recursive function that returns a type within *Set* given a `stat_term` encoding. The `stat_term` is basically just a syntactic expression of the types of the semantic logic. `Sns` converts that syntactic representation into acutal types.

I start with a listing of a number of meanings.

Note that there is an axiom named `doo` below because *do* is a reserved word in Coq.

There are also a number of properties that begin with the word "both", below. These are used for dual number meanings. The property `both` is not intended to correspond to something like the English word *both*.

```
(** A number of types for meanings. *)

Axiom big : (e → prop) → e → prop.

Axiom tall : (e → prop) → e → prop.

Axiom head : e → prop.

Axiom rock : e → prop.

Axiom fruit : e → prop.

Axiom thorn : e → prop.
```

```
Axiom plant : e → prop.

Axiom pole : e → prop.

Axiom branch : e → prop.

Axiom hurt : e → prop.

Axiom say : e → prop → prop.

Axiom liquid : e → prop.

Axiom see : e → e → prop.

Axiom doo : e → prop → prop.

Axiom same : (e → prop) → e → prop.

Axiom cutt : e → e → prop.

Axiom eat : e → prop.

Axiom woman : e → prop.

Axiom plantain : e → prop.

Axiom chonta_palm : e → prop.
```

```
Axiom chicha : e → prop.

Axiom canoe : e → prop.

Axiom seed : e → prop.

Axiom food : e → prop.

Axiom body : e → prop.

Axiom meat : e → prop.

Axiom board : e → prop.

Axiom flat_thing : e → prop.

Axiom shell : e → prop.

Axiom paper : e → prop.

Axiom small_flat_thing : e → prop.

Axiom small_round_thing : e → prop.

Axiom round_thing : e → prop.

Axiom hand : e → prop.
```

```
Axiom cluster : e → prop.

Axiom river : e → prop.

Axiom leaf : e → prop.

Axiom past : e → prop.

Axiom future : e → prop.

Axiom onenoun : e → prop.

Axiom to_trans : (e → prop) → (e → e → prop).

Axiom ι : (e → prop) → e.

Axiom there_is : (e → prop) → prop.

Axiom thing : e → prop.

Axiom speaker : e → prop.

Axiom speakers : e → prop.

Axiom both_speakers : e → prop.

Axiom speakers_addressees : e → prop.
```

```
Axiom addressee : e → prop.

Axiom addressees : e → prop.

Axiom both_addressees : e → prop.

Axiom mother : e → e → prop.

Definition my_mother_addressee : e → prop :=
  λ x,(mother (ι speaker) x and addressee x).

Axiom both : e → prop.

Axiom feminine : e → prop.

Axiom person : e → prop.

Axiom people : e → prop.

Axiom some : (e → prop) → (e → prop) → prop.

Axiom all : (e → prop) → (e → prop) → prop.
```

$sense$ is a Sigma type, a dependent sum. $\Sigma(x : A), B(x)$ is the notation for the constructor in type theory. So for the coq express `{ s : stat_term & Sns s }`, the type corresponds to:

$$\Sigma(s : stat\_term), Sns(s)$$

s is a stat term and $Snss$ is a type *indexed* by s, belonging to a family of types, in this case e, $prop$, and the types of functions of things of type e and $prop$.

Usually, in the literature, the type is written $\Sigma(x : A), B(x)$, but given that $x : A$ is recoverable from the type of B, the type need only invoke the predicate, as below.

```
Definition sense := sigT Sns.


Definition adjsense := existT Sns (func (func ent prp) (func ent prp)).
```

The following are useful helpers for different meaning types.

```
Definition intranssense := existT Sns (func ent prp).


Definition transsense := existT Sns (func ent (func ent prp)).


Definition indefsense := existT Sns (func (func ent prp) prp).


Definition quantsense := existT Sns (func (func ent prp)
  (func (func ent prp) prp)).


Definition intersectls : ((e → prop) → e → prop) → (e → prop) →
  (e → prop) → e → prop :=
  λ adj ls n x,(adj n x) and (ls x) and (n x).


(* The following is used because I don’t want to deal with tense and
other details. *)


Definition stripverb_m : list m → list m :=
  filter
    (λ x,in_m x (A_2m :: Ã_m :: Dãta_m :: Ĩ_1m :: Ke_m :: Kẽ_1m :: lss_m)).
```

```
(* I am treating person information on nouns as grammatical, which is
an oversimplification. *)

Definition stripanim_m : list m → list m :=
  filter (λ x,in_m x [Okiye_m]).
```

covertard$_s$ and overtarg$_s$ are defined below because verbs that take body-part lexical suffixes may or may be exhibit overt transitivity. In fact, it is reasonable to suspect that verbs like *dãta*, ‘hurt’, may have ambiguous status. The basic idea is to provide a meaning suitable for verbs that don’t take object arguments, but do take a body-part suffixes. These correspond to covertard$_s$. I do not have a developed analysis for such meanings. The semantic role associated with the property may be various. I use a stopgap, which is to say that there has to be something of with a body-part property, but I don’t say how it relates to the subject.

A meaning is also needed when there is an explicit argument. That is what overtarg$_s$ is for. In that case the object should have the body-part property, which is what the meaning expresses.

The functions below are only used with body-part verbs.

```
Definition covertarg_s (α β : e → prop) : e → prop :=
  λ x,there_is α implies β x.

Definition overtarg_s (α : e → prop) (β : e → e → prop) :
e → e → prop :=
  λ x y,β x y and α y.
```

The below supplies the basic *m* parse to meaning logic. There is a lot of listed information, but it is all fairly simple logic.

```
Inductive bodylsmeaning : (e → prop) → m → Prop :=
| ka_1s_head : ∀ m, m = ka_1m → bodylsmeaning head m
| kã_1s_body : ∀ m, m = ka_1m → bodylsmeaning body m
```

```
| kã₁ₛmeat : ∀ m, m = ka₁ₘ → bodylsmeaning meat m
| poₛhand : ∀ m, m = poₘ → bodylsmeaning hand m.

Inductive plantlsmeaning : (e → prop) → m → Prop :=
| bõ₁ₛfruit : ∀ m, m = bõ₁ₘ → plantlsmeaning fruit m
| bõ₁ₛseed : ∀ m, m = bõ₁ₘ → plantlsmeaning seed m
| ka₁ₛfruit : ∀ m, m = ka₁ₘ → plantlsmeaning fruit m
| ka₁ₛseed : ∀ m, m = ka₁ₘ → plantlsmeaning seed m
| wẽₛplant : ∀ m, m = wẽₘ → plantlsmeaning plant m
| wẽₛbranch : ∀ m, m = wẽₘ → plantlsmeaning branch m
| yaboₛleaf : ∀ m, m = yaboₘ → plantlsmeaning leaf m.

Inductive lsmeaning : (e → prop) → m → Prop :=
| bodylsₛ : ∀ α m, bodylsmeaning α m → lsmeaning α m
| plantlsₛ : ∀ α m, plantlsmeaning α m → lsmeaning α m
| bõ₁ₛround : ∀ m, m = bõ₁ₘ → lsmeaning round_thing m
| bõ₁ₛsmall : ∀ m, m = bõ₁ₘ → lsmeaning small_round_thing m
| dẽₛfood : ∀ m, m = dẽₘ → lsmeaning food m
| ka₁ₛrock : ∀ m, m = ka₁ₘ → lsmeaning rock m
| pa₁ₛboard : ∀ m, m = pa₁ₘ → lsmeaning board m
| pa₁ₛflat : ∀ m, m = pa₁ₘ → lsmeaning flat_thing m
| poₛcanoe : ∀ m, m = poₘ → lsmeaning canoe m
| poₛcluster : ∀ m, m = poₘ → lsmeaning cluster m
| pẽₛliquid : ∀ m, m = pẽₘ → lsmeaning liquid m
| ta₁ₛshell : ∀ m, m = ta₁ₘ → lsmeaning shell m
| ta₁ₛpaper : ∀ m, m = ta₁ₘ → lsmeaning paper m
| wẽₛpole : ∀ m, m = wẽₘ → lsmeaning pole m.
```

```
Definition personmeaningₛ (catₘ : list m) : e → prop :=
  match (personFilterₘ catₘ) with
  | [ diₘ ; dãₘ ] => people
  | [ diₘ ; bõ₂ₘ ] => speakers
  | [ diₘ ; bĩₘ ] => addressees
  | [ daₘ ; bõ₂ₘ ] => both_speakers
  | [ daₘ ; bĩₘ ] => both_addressees
  | [ biₘ ] => addressee
  | [ bĩₘ ] => my_mother_addressee
  | [ boₘ ] => speaker
  | [ bõ₂ₘ ] => speakers_addressees
  | [ daₘ ] => both
  | [ dãₘ ] => feminine
  | [ kã₂ₘ ] => person
  | otherwise => thing
  end.

(* I am using a generic and opaque combine because the actual
determination of potential meaning relationships would require an
advanced study of lexical semantics *)

Axiom combineₛ : (e → prop) → (e → prop) → e → prop.

Inductive inmeaning : (e → prop) → list m → Prop :=
| plantyₛ : ∀ cat α, cat = [ wẽₘ ; A₁ₘ ] → lsmeaning α wẽₘ →
  inmeaning α cat
| fruit₁ₛ : ∀ cat, cat = [ ka₁ₘ ; Bẽyeₘ ] → inmeaning fruit cat
| fruit₂ₛ : ∀ cat, cat = [ bõ₁ₘ ; Bẽyeₘ ] → inmeaning fruit cat
```

```
| waterₛ : ∀ cat, cat = [ pẽₘ ; Eₘ ] → inmeaning liquid cat
| riverₛ : ∀ cat, cat = [ pẽₘ ; Eₘ ] → inmeaning river cat
| thornₛ : ∀ cat, cat = [ Daaₘ ] → inmeaning thorn cat
| womanₛ : ∀ cat, cat = [ Okiyeₘ ] → inmeaning woman cat
| hurtₛ : ∀ cat, stripverbₘ cat = [ Dãtaₘ ] → inmeaning hurt cat
| eat_intransₛ : ∀ cat, stripverbₘ cat = [ Kẽ₁ₘ ] → inmeaning eat cat
| plant_inanimₛ : ∀ cat α β, klass cat ≤ₖ plantₖ →
  plantlsmeaning α (hd noneₘ cat) → inmeaning β (tail cat) →
  inmeaning (combineₛ α β) cat
| bodyverbₛ : ∀ cat α β,klass cat ≤ₖ verbbodyₖ →
  bodylsmeaning α (hd noneₘ (stripverbₘ cat)) →
  inmeaning β (tail (stripverbₘ cat)) →
  inmeaning (covertargₛ α β) cat.

Inductive trmeaning : (e → e → prop) → list m → Prop :=
| seeₛ : ∀ cat, cat = [ A₂ₘ ] → trmeaning see cat
| cutₛ : ∀ cat, cat = [ Kẽ₁ₘ ] → trmeaning cutt cat
| eat_transₛ : ∀ cat, cat = [ Kẽ₁ₘ ] → trmeaning (to_trans eat) cat
| hurt_transₛ : ∀ cat, cat = [ Kẽ₁ₘ ] → trmeaning (to_trans hurt) cat
| bodyverb_transₛ : ∀ cat α β,klass cat ≤ₖ verbbodyₖ →
  bodylsmeaning α (hd noneₘ (stripverbₘ cat)) →
  trmeaning β (tail (stripverbₘ cat)) →
  trmeaning (overtargₛ α β) cat.

Inductive adjmeaning : ((e → prop) → e → prop) → list m → Prop :=
| bigₛ : ∀ cat, cat = [ Yẽdẽₘ ] → adjmeaning big cat
| tallₛ : ∀ cat, cat = [ Yẽdẽₘ ] → adjmeaning tall cat
| sameₛ : ∀ cat, cat = [ Adoₘ ] → adjmeaning same cat
```

```
| adjlsₛ : ∀ cat α β, cat ≤ₘ [ LSₘ ] → lsmeaning α (hd noneₘ cat) →
  adjmeaning β (tail cat) → adjmeaning (intersectls β α) cat.

Inductive emeaning : e → list m → Prop :=
| definite_adjₛ : ∀ cat (α : (e → prop) → e → prop),
  adjmeaning α cat → emeaning (ι (α onenoun)) cat
| definite_inanimₛ : ∀ cat (α : e → prop), klass cat ≤ₖ inanimₖ →
  inmeaning α cat → emeaning (ι α) cat
| definite_animₛ : ∀ cat (α : e → prop), klass cat ≤ₖ animₖ →
  inmeaning α (stripanimₘ cat) → emeaning (ι α) cat
| definite_participantₛ : ∀ cat, klass cat ≤ₖ participantₖ →
  emeaning (ι (personmeaningₛ (personFilterₘ cat))) cat
| definite_proₛ : ∀ cat, klass cat ≤ₖ proₖ →
  nonfirstFilterₘ cat ≠ nil →
  emeaning (ι (personmeaningₛ  (nonfirstFilterₘ cat))) cat.

Inductive quantmeaning : ((e → prop) → (e → prop) → prop) →
list m → Prop :=
| quant_adjₛ : ∀ cat (α : (e → prop) → e → prop),
  adjmeaning α cat →
  quantmeaning (λ (x : e → prop),some (α x)) cat.

Inductive indefmeaning : ((e → prop) → prop) → list m → Prop :=
| indef_adjₛ : ∀ cat (α : (e → prop) → e → prop),
  adjmeaning α cat → indefmeaning (some (α onenoun)) cat
| inder_inanimₛ : ∀ cat α, klass cat ≤ₖ inanimₖ →
  inmeaning α cat → indefmeaning (some α) cat
| inder_animₛ : ∀ cat α, klass cat ≤ₖ animₖ →
```

```
  inmeaning α cat → indefmeaning (some α) cat.

Inductive meaning : sense → list m → Prop :=
| eₛ : ∀ cat (α : e), emeaning α cat → meaning (existT Sns ent α) cat
| inₛ : ∀ cat (α : e → prop), inmeaning α cat →
  meaning (intranssense α) cat
| trₛ : ∀ cat (α : e → e → prop), trmeaning α cat →
  meaning (transsense α) cat
| adjₛ : ∀ cat (α : (e → prop) → e → prop), adjmeaning α cat →
  meaning (adjsense α) cat
| indefₛ : ∀ cat (α : (e → prop) → prop), indefmeaning α cat →
  meaning (indefsense α) cat
| quantₛ : ∀ cat (α : (e → prop) → (e → prop) → prop), quantmeaning α
  cat → meaning (quantsense α) cat.
```

Below a stopgap "pheno" type for LCG. The "pheno" in use here is not the pheno as introduced in Chapter 5, though it will have no impact on the discussion of the morphological theory. Pheno types are complex. A pheno term can have a type such as $s \to s \to s \to s$, or simply s. This is not compatible with reducing sign paradigm entries structures to a single type. For a complete treatment, the pheno needs to be embedded under a Σ-type, just as is done in the semantics. This was done for the semantics because the linkage between morphological forms and lexical semantics is vital to the core concerns of this work. The pheno is not important to the same extent. Dealing with the Σ-type for semantics involves a lot of overhead, not just for its definition, but in adding complexity to form-sign relations. I felt that adding this complexity to the system for the pheno was a distraction.

Rather than implement the pheno, I provide a stopgap pheno type. It is an inductive type ϕ, with a number of constructors. For this reason, η is a type constructor for ϕ, which explains the type provided in (XXI).

Since I don't actually use λ-terms, or real functions with the type, I provide dummy stand in variables, as though there was some binder to express relative positioning of arguments. These are the $x\phi, y\phi, z\phi$ constructors. This allows me to work through morphological proofs as though there was an actual pheno defined. Nothing in the morphological analysis requires pheno calculations, as is the case with the semantics.

Below the stopgap are the initial definitions needed for a Σ-type alternative. I do not develop it and incorporate it because it is irrelevant to the core goals of the thesis.

```
Inductive ϕ : Set :=
| ε
| η : string → ϕ
| xϕ
| yϕ
| zϕ
| fϕ (α β : ϕ)
| concatϕ (α β : ϕ).

Infix ”•” := concatϕ (at level 60, right associativity).
```

Below a pheno definition follow the pattern of the semantics. The type `ps` is not much different than the type ϕ I defined above. Functions of type `ps` are what raises issues, and necessitates and embedding. The type `pheno_term` plays a similar role to `stat_term`. One can specify what a type should be. The function Φ plays the role of `Sns`. Using these, plus some additional string property information, I can implement a real phenogrammar.

```
Inductive ps : Set :=
| pempty : ps
| peta : string → ps
| pconcat : ps → ps → ps.

Inductive pheno_term : Set :=
| pheno_string : pheno_term
| pheno_function : pheno_term -> pheno_term -> pheno_term.
```

```
Fixpoint Φ (p : pheno_term) : Set :=
  match p with
  | pheno_string   => ps
  | pheno_function a b  => (Φ a) -> (Φ b)
  end.
```

The type of sign paradigm entries is below.

```
Definition structₛₚ := (ϕ * τ * sense).
```

The below maps *m* to annotated tectos.

```
Definition τₘ (catₘ : list m) : τ_anno :=
  match (personFilterₘ catₘ) with
  | [ diₘ ; dãₘ ] => t3_pl
  | [ diₘ ; bõ₂ₘ ] => t1_pl
  | [ diₘ ; bĩₘ ] => t2_pl
  | [ daₘ ; bõ₂ₘ ] => t1_du
  | [ daₘ ; bĩₘ ] => t2_du
  | [ biₘ ] => t2_sg
  | [ bĩₘ ] => t2_mo
  | [ boₘ ] => t1_sg
  | [ bõ₂ₘ ] => t1_inc
  | [ daₘ ] => t3_du
  | [ dãₘ ] => t3_f
  | [ kã₂ₘ ] => t3_h
  | otherwise => t
  end.
```

Below is the rule schema for form to sign mappings.

```
Definition ruleₛₚ (catₘ : list m) (k : K) (P : ϕ → ϕ)
```

```
(T : list m → τ) (s₁ : stat_term) (s₂ : stat_term)
(Q : Sns s₁ → Sns s₂) :=
λ (mp₁ : structₘₚ)
  (mp₂ : structₘₚ)
  (β : Sns s₁)
  (proofₑᵥ : mp₁ ≡ₘₚ mp₂)
  (proofₘ : (fst mp₂) ≤ₘ catₘ)
  (proofₖ : klass (fst mp₂) ≤ₖ k)
  (proofₛ : meaning (existT Sns s₁ β) (fst mp₂)),
  (P (η( applyₚᵣ (snd mp₂) idₚᵣ)), T catₘ, existT Sns s₂ (Q β)).
```

The inductive definition of sign paradigm entries is given as SE_{sp}. It is entirely invocations of the rule schema above. I run through the arguments taken be the rule schema in §6.5.4.

```
Inductive SEₛₚ : structₛₚ → Prop :=
| adjSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] adjₖ (λ s,s•xϕ) ((λ t,N t -o Adj t) ∘ τₘ)
             (func (func ent prp) (func ent prp))
             (func (func ent prp) (func ent prp))
             (λ x,x))
            mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| n_def_nomSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] adjₖ (λ s,s•xϕ) ((λ t,N t -o Nom t) ∘ τₘ)
             (func (func ent prp) ent)
             (func (func ent prp) ent)
             (λ x,x))
            mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| def_nomSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] nominalₖ (λ s,s)
             ((λ t,Nom t) ∘ τₘ) ent ent (λ x,x))
```

```
          mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| n_quant_nomSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] adjₖ (λ s,s•xϕ)
           ((λ t,N t □ (Nom t -o Fin) -o Fin) ∘ τₘ)
           (func (func ent prp) (func (func ent prp) prp))
           (func (func ent prp) (func (func ent prp) prp))
           (λ x,x))
          mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| indef_nomSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] nominalₖ (λ s,s)
           ((λ t,(Nom t -o Fin) -o Fin) ∘ τₘ)
           (func (func ent prp) prp)
           (func (func ent prp) prp)
           (λ x,x))
          mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| nom_accSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] proₖ (λ s,s)
           ((λ t,Nom t -o Acc) ∘ τₘ)
           (func ent ent)
           (func ent ent)
           (λ x,x))
          mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| nounSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] nominalₖ (λ s,s)
           ((λ t,N t) ∘ τₘ)
           (func ent ent)
           (func ent ent)
           (λ x,x))
```

```
            mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| intransSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] verbₖ (λ s,s)
             ((λ t,Nom t -o Fin) ∘ τₘ)
             (func ent ent)
             (func ent ent)
             (λ x,x))
            mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
| transSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [INFₘ] verbₖ (λ s,s)
             ((λ t,Nom t -o Acc -o Fin) ∘ τₘ)
             (func ent ent)
             (func ent ent)
             (λ x,x))
            mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ).
```

Although they are not used or needed in order to perform inferential reasoning over paradigm instances, I supply predicates for indicating membership in form paradigm and sign paradigm instances below.

The definition of $\texttt{in}_{mp}$ is very simple. A form paradigm instance is identified by a stem and a form class. The instance identified in this manner may be a sub-instance. For instance, if the target word is *go-bõ-di*, 'go-1.INCL-PL', any of the stems *go*, *go-bõ* or *go-bõ-di* may successfully be applied with $\texttt{verb}_k$ as the form class. It is up to the individual using the predicate to determine the instance scope of interest and choose the stem they feel is appropriate for their needs. If needed, one could define a function that determines the smallest stem for which $\texttt{in}_{mp}$ would be true for a particular form paradigm entry and form class.

The definition of $\texttt{in}_{sp}$ is more complex, but the basic idea is simple. The idea is to test whether a target sign paradigm entry, called `sign` can be constructed from a form paradigm entry that is an instance of a form paradigm. Sign paradigm instances are also identified by a stem and a form class. After the first four arguments, the following should look familiar from clauses in the inductive definition of $\texttt{SE}_{sp}$. These arguments are used to define a rule in the first `let` clause. Then, in the second `let` clause a $\texttt{struct}_{sp}$ is defined with arguments that reference `word`. Following this, the same means of proving a form paradigm instance need to be proved, in addition to a proof of `SEₛₚ sign`. Finally, `sign` must be shown to be equal to the constructed `newsign`.

Clearly this is arduous. I have not had any reason to use these predicates, so far. They are presented for those who would like some means of proving membership in paradigm instances. The system could certainly be optimized to make the predicates require less work.

```
Axiom inₘₚ : ∀ (κ : K) (stem word : structₘₚ),
    FEₘₚ word
    → klass (fst stem) = klass (fst word)
    → klass (fst stem) ≤ₖ κ
    → isStemₘₚ stem word
    → Prop.

Definition inₛₚ
  (κ₁ : K)
  (stem word : structₘₚ)
  (sign : structₛₚ)
  (catₘ : list m)
  (k₂ : K)
  (P : ϕ → ϕ)
  (T : list m → τ)
  (s₁ : stat_term)
  (s₂ : stat_term)
  (Q : Sns s₁ → Sns s₂)
  (β : Sns s₁)
  (proofₑᵥ : word ≡ₘₚ word)
  (proofₘ : (fst word) ≤ₘ catₘ)
  (proofₖ : klass (fst word) ≤ₖ k)
```

```
(proofₛ : meaning (existT Sns s₁ β) (fst word))
: Prop :=
  let rule := ruleₛₚ catₘ k P T s₁ s₂ Q in
  let newsign := rule word word β proofₑᵥ proofₘ proofₖ proofₛ in
  (klass (fst stem) = klass (fst word)) ∧
  (klass (fst word) ≤ₖ κ) ∧
  (isStemₘₚ stem word) ∧
  (SEₛₚ sign) ∧
  newsign = sign.
```

Appendix B

Toy Analyses

B.1 Walk

During the course of the thesis, I used *walk* as an example due to its likely familiarity to the audience. I judged that basic concepts of the framework would be clarified by familiar data. Yet, I only took the example so far, and the Wao Terero data is sufficiently different from English inflection that it may be difficult to see how form-form mappings and form-sign relations would be defined for an English example.

In this section, I will only supply a very basic example, just enough for walk and other weak intransitive verbs. It will also be toy-like in some other ways. I will not map out an elaborated meaning relation, or a particularly elaborated M-cat hierarchy. The example will be sufficient to demonstrate that the framework can handle simple English data, in addition to Wao Terero.

In Chapter 6, I presented a form paradigm instance for *walk* on page 286, which is presented here as (I) with elaborated process information.

$$\langle \mathsf{Walk_m}, \mathrm{walk} \rangle \quad \text{(Ia)}$$

$$\langle \mathsf{s_m} :: \mathsf{Walk_m}, \mathrm{s_{pr}} :: \mathrm{walk_{pr}} \rangle \quad \text{(Ib)}$$

$$\langle \mathsf{d_m} :: \mathsf{Walk_m}, \mathrm{d_{pr}} :: \mathrm{walk_{pr}} \rangle \quad \text{(Ic)}$$

$$\langle \mathsf{ing_m} :: \mathsf{Walk_m}, \mathrm{ing_{pr}} :: \mathrm{walk_{pr}} \rangle \quad \text{(Id)}$$

The rules for constructing the form paradigm entry in (I) are defined as part of the inductive definition of FE_{mp}. The Coq definition for Wao Terero is discussed on page 410. Similar clauses for *walk* would appear as below:

```
Inductive FEₘₚ : structₘₚ → Prop :=
...
| walkMP : FEₘₚ ( [Walkₘ], [walkₚᵣ] )
...
| sMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [BASEₘ] weakₖ [sₘ] [sₚᵣ])
                     α proofₘ proofₖ)
| dMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [BASEₘ] weakₖ [dₘ] [dₚᵣ])
                     α proofₘ proofₖ)
| ingMP : ∀ α proofₘ proofₖ,
    FEₘₚ α → FEₘₚ ((rule1ₘₚ [BASEₘ] weakₖ [ingₘ] [ingₚᵣ])
                     α proofₘ proofₖ)
```

The $\mathsf{BASE_m}$ M-cat is for uninfected forms. The weak_k form class is for weak verbs. In reality, the form classes would need to be articulated to handle the full range of classes in English. The rules above can be translated to the semi-formal notation that I introduced in Chapter 6. An example for sMP is provided below.

$$\frac{\alpha \leq_m \mathsf{BASE_m} \qquad \alpha \leq_k \mathsf{weak}_k}{\langle \alpha \diamond \mathsf{s_m}, \alpha \diamond \mathrm{s_{pr}} \rangle}\ \mathrm{sMP} \tag{II}$$

As an aside, this rule could be elaborated to handle nominal plurals and the possessive (Saxon genitive). It is not really the third person singular, plural, or possessive until the form is licensed in a particular context. The M-cat is not associated with a meaning or syntactic category at this point. Further, if one decided to go this route $\langle \mathsf{s_m} :: \mathsf{Walk_m}, \mathrm{s_{pr}} :: \mathrm{walk_{pr}} \rangle$ could serve as a form paradigm entry for third person singular, nominal plural, and possessive *walks*. This is a choice that is available within the framework, but there are mechanisms for accommodating homophonous affixes in the system. One is not forced to unify the plural and third person singular. I believe that all English form paradigms contain one *-s* affixed form, no matter their part of speech, exploring this hypothesis is beyond the scope of this toy analysis.

Intransitive *walk* has six sign paradigm entries. They are listed in (III).

$$\langle \lambda s.s \bullet \text{walk}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, \mathit{walk} \rangle \tag{IIIa}$$

$$\langle \lambda s.s \bullet \text{walks}, \mathsf{Nom}_{3s} \multimap \mathsf{Fin}, \mathit{walk} \rangle \tag{IIIb}$$

$$\langle \lambda s.s \bullet \text{walked}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{walk}) \rangle \tag{IIIc}$$

$$\langle \text{walk}, \mathsf{PRO} \multimap \mathsf{Bse}, \mathit{walk} \rangle \tag{IIId}$$

$$\langle \text{walking}, \mathsf{PRO} \multimap \mathsf{Prp}, \mathit{walk} \rangle \tag{IIIe}$$

$$\langle \text{walked}, \mathsf{PRO} \multimap \mathsf{Psp}, \mathit{walk} \rangle \tag{IIIf}$$

The first two and last three can be collapsed. Three form-sign relations are needed.

The first two can use annotations. Annotations may be mistaken for features, but they are not necessary in opposition. One would need to add such an interpretation to the grammar. As I have used them, they are more about predictable forms recurring in meaningful ways in otherwise identical syntactic contexts. In this case, it is morphologically predictable, based on form, which items are $\mathsf{Nom}_{\neg 3s}$ and which are Nom_{3s}. They could be named $\mathsf{Nom}_{\text{no s}}$ and Nom_{s}.

In order to express this, I provide two annotations. In Coq, their definition looks like the following:

```
Inductive τ_anno : Set :=
| ts
| t
```

In more standard Linear Categorial Grammar (LCG) the first is $3s$ and the other is $\neg 3s$.

The basic tectos for the fragment are listed in Coq notation below:

```
Inductive τ : Set :=
| Nom (α : τ_anno)
| Fin
```

```
| PRO
| Bse
| Prp
| Psp
...
```

Remember that these cannot be deconstructed in LCG. They are necessarily reinterpreted as symbols in another logic. Technically, there should be more said about this mapping. Yet, the fragment does not extend to actually being able to perform proofs in LCG at this point. When that point is reached, the nature of these symbols in that logic will need to be explicit.

For Wao Terero, I used some tecto constructor helpers. Here I will use two. The first τ_m takes the head of the list of m and matches it against either ts, if it is $\mathsf{s_m}$, or t, otherwise. The second does not return an annotation, but a tecto term. Likewise, it takes the head of the list, but it maps $\mathsf{ing_m}$ to Psp, $\mathsf{d_m}$ to Prp, and anything else to Bse. I call it T_m.

The Coq code for the finite entries that take a nominal is below.

```
| finargSP : ∀ mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ,
    SEₛₚ ((ruleₛₚ [FINARGₘ] verbₖ (λ s,s•xϕ) ((λ t,Nom t -o Fin) ∘ τₘ)
            (func (func ent prp) (func ent prp))
            (func (func ent prp) (func ent prp))
            (λ x,x))
          mp₁ mp₂ β proofₑᵥ proofₘ proofₖ proofₛ)
```

In the pseudo-formal notation I have used, it looks like the following:

$$\frac{\alpha \leq_m FINARG_m \qquad \alpha \leq_k \mathsf{verb}_k \qquad meaning(\beta, \alpha)}{\langle(\lambda \mathrm{st.s} \bullet \mathrm{t})\alpha, (\lambda \mathsf{t.Nom_t} \multimap \mathsf{FIN}) \circ \tau_\mathsf{m}\alpha, (\lambda x.x)\beta\rangle} \text{finargSP} \qquad \text{(IV)}$$

The M-cat $\mathsf{FINARG_m}$ will match the bare forms and forms with *-s*. The input item must be in the verb_k class, which would subsume weak_k and other patterns.

In the pseudo-formal notation, the following produces the three entries for non-finite sign paradigm entries.

$$\frac{\alpha \leq_m \mathit{NONFIN}_m \qquad \alpha \leq_k \mathsf{verb}_k \qquad \mathit{meaning}(\beta, \alpha)}{\langle (\lambda \mathrm{s.s})\alpha, (\lambda \mathsf{t}.\mathsf{PRO} \multimap \mathsf{t}) \circ \mathsf{T_m}\alpha, (\lambda x.x)\beta \rangle} \text{nonfinSP} \tag{V}$$

The notable difference is that T_m has a type $\mathit{list}\ m \to \tau$, from lists of m to tectos, rather than some annotation type.

The rule for past tense is notably only in ignoring the input list of m.

$$\frac{\alpha \leq_m \mathit{PAST}_m \qquad \alpha \leq_k \mathsf{verb}_k \qquad \mathit{meaning}(\beta, \alpha)}{\langle (\lambda \mathrm{st.s} \bullet \mathrm{t})\alpha, (\lambda \mathsf{t}.\mathsf{Nom} \multimap \mathsf{Fin})\alpha, (\lambda x.\mathit{past}(x))\beta \rangle} \text{pstSP} \tag{VI}$$

The rule compression that I am performing may not seem to have natural motivations to all readers. It is certainly language specific at this time. Though, it may be that as the number of languages under analysis grows, patterns emerge. It is likely that agreement information will tend to be semi-orthogonal to categories associated with nouns and adjectives in many languages. It is also likely that the English patterns for non-finite verbs have analogs in other languages.

The use of verb_k in the form-sign relations, but weak_k in the form-form mappings, is suggestive of a pattern. The intent is that some verbs will be of a class with a different past tense form, for instance, placing it in another class relevant to formal patterns, but that there will be a certain uniformity to syntactic paradigm instances. One can look at all the form-sign relation that reference the same form class, and observe the structure of the paradigm associated with the class.

B.2 Hablar y Comer

As another toy analysis, I will present a simplified pattern that can be used with languages like Spanish to capture conjugation classes.

I will consider the data in Table B.1.

The point I wish to make about this data is very simple. It is similar to the point made above concerning verb_k versus weak_k. In order to capture that both *hablar*, 'to speak', and *comer*, 'to eat', have *-o, -a* and *-e* forms, there is a form-oriented generalization one can make. Yet, to express that the syntactic distributions of these forms diverge, one must make a seemingly contrary generalization.

The form paradigms instances for both words follow the pattern below, where X is the stem information.

	AR	ER
1 indicative	hablo	como
3 indicative	habla	come
3 subjunctive	hable	coma

Table B.1: Three forms from the Spanish ER and AR conjugations that demonstrates contrasts for the indicative and subjunctive third persons.

$$\langle \mathsf{o_m} :: \mathsf{X}, \mathrm{o_{pr}} :: \mathrm{X} \rangle \tag{VIIa}$$

$$\langle \mathsf{a_m} :: \mathsf{X}, \mathrm{a_{pr}} :: \mathrm{X} \rangle \tag{VIIb}$$

$$\langle \mathsf{e_m} :: \mathsf{X}, \mathrm{e_{pr}} :: \mathrm{X} \rangle \tag{VIIc}$$

Let us say that the *-o* forms are specified by axiom. In actual Spanish, this may not be the best choice. Since the forms are identical for this subset, the form class can be the same. I call it cverb_k, which is the class of forms common to verbs. Abstracting away from the stem the rule in (VIII) uses an M-cat, $\mathsf{O_m}$, for *-o* ending forms. Remember that $\triangleleft$ results in the replacement of the final element on the process and *m* lists.

$$\frac{\alpha \leq_m \mathsf{O_m} \qquad \alpha \leq_k \mathsf{cverb}_k}{\langle \alpha \triangleleft \mathsf{e_m}, \alpha \triangleleft \mathrm{e_{pr}} \rangle} \text{ eMP} \tag{VIII}$$

I do not want to get into the specifics of a Spanish sign paradigm instance. All we need to know is that there is some indicative third person rule, and a subjunctive third person rule. Both of these rules are going to reference a form class of verb_k. The subjunctive rule will reference an M-cat of $\mathsf{SUBJ_m}$, and the indicative, $\mathsf{IND_m}$. What makes this work is that in addition to *hablar* belonging to a form class cverb_k and verb_k, it belongs to ar_k, where $\mathsf{ar}_k \leq_k \mathsf{cverb}_k \leq_k \mathsf{verb}_k$. Likewise, *comer* is a similarly subsumed er_k. Due to the fact that the M-cat order allows for parameterization based on form class, an er_k must have an *-a* ending to be ordered below $\mathsf{SUBJ_m}$, and an ar_k must have an *-e* ending. Clearly, a full analysis would be more complex, but such patterns hold over subsets of the Spanish verbal paradigm.

Appendix C

The Definition of Paradigm Objects

Although there are significant similarities between the framework presented in this thesis and other Word and Paradigm (WP) frameworks, the differences are likely more apparent to the reader. This may lead to the question of whether my framework is dealing with the same concept of paradigm as those theories. In some cases, the answer is obviously “no”. I am not working with paradigms defined by the Cartesian multiplications of features, for instance. In other cases, I believe that what I am doing is not significantly different. I capture the same types of generalizations as theories that use inflection class hierarchies (Brown and Hippisley, 2012; Pollard and Sag, 1987) through the form class hierarchy.

C.1 Intensional Versus Extensional Classification

In this framework, I do not use inheritance. I use an alternative mechanism to provide morphological taxonomies. One may review section §4.3.3 for a reminder of its benefits. The idea is to build up a schema, which specifies a taxonomic structure, and then classify objects using the *klass* function. This is part of a general trend in the framework design, which emphasizes relations and functions, rather than complex objects. My work is not unique in this emphasis. I follow general design patterns used in type theory. One may compare my notion of *klass* with the concept of coercive subtyping and universe definitions (Luo, Soloviev, and Xue, 2013). There types, or universes, are similarly defined to have a schema or structure, which is not based on subdividing a set. A coercion operator is used to coerce (or classify) otherwise unorganized types within the schema. When the coercive approach is compared to subset-like subtyping and universes (Luo, 2012c), one can see similar benefits to my approach to classification over inheritance. More broadly, what I am doing is called an *intensional* approach to classification (Parrochia, 2025).

The more popular approach to classification in linguistics is the extensional approach. There, classification is a matter of dividing sets, which contain the objects to be classified. For this reason, despite the validity of intensional classification, there may be readers who do not accept my notion of "being a member of", if there is no set or explicit collection. This is a philosophical attitude that I clearly do not accept. Saying that a bear is a member of the family Ursidae, is compatible with both intensional and extensional models. Since "member" is a matter of definition, popularity is relevant. The usage in intensional classification is widely accepted (Parrochia, 2025).

C.2 Primary Objects and Relations

There have been questions from committee members as to whether my concept of paradigm is a secondary phenomenon, or even an epiphenomenon, in terms of explanatory role. Perhaps this is due to its non-extensional nature. A paradigm and its structure may *seem* more concrete if there are sets of members. Despite the intensional design, the notion of taxonomy and paradigm are primary phenomena in the framework.

Similar to inflection class hierarchy theories, the form class hierarchy is a fundamental object. It is not a secondary consequence of any other mechanism in the theory. The form classes and their order are defined using axioms. There is no means of generating them from more basic rules or objects. The implication is that the hierarchy must be empirically discovered. It is a central empirical claim.

Likewise, form classes are the most important category of constructive proofs. Even the other important category type, abstract M-cats, are parameterized by form classes. Even the notion of syntactic category could be considered secondary, since the form-sign relations reference classes such as verb_k in their role of defining what syntactic categories a verb may correspond to. There is no more basic classificatory unit in the theory, which is to be expected, because the taxonomy defines the paradigms of the WP theory.

The place of form paradigm entries and sign paradigm entries within the form class hierarchy is also primary phenomenon, in terms of explanatory role. The function klass matches on lists of m and assigns form paradigm entries their primary class. A primary class is the lowest class in the hierarchy that a form paradigm entry or sign paradigm entry is a member of. Sign paradigm entries indirectly utilize *klass* for their classification. Importantly, the definition of *klass* is listed. The patterns cannot be derived from some other system of rules. They must be empirically discovered, and are thus one of the claims of an analysis. The *klass* function, though it utilizes stem patterns in its logic, classifies objects that correspond to free words, which is consistent with a WP theory.

C.3 Definition of Taxonomies

The qualities I described above are fundamental to a taxonomic system. The definition of taxonomy that I assume requires four things.

1. Objects that represent taxonomic classes.
2. A relation between classes.
3. Objects that represent instances.
4. A means of classifying instances.

The first two of these requirements constitute a schema. In my framework, the first requirement is supplied by form class objects, which name classes. In this framework, classes are also referred to as paradigms. This means that a form class equals a paradigm. The form classes are related by a partial order. This constitutes the taxonomic schema, the global structure.

There are two taxonomies mapped to the schema. This is because there are two types of objects, form paradigm entries and sign paradigm entries. A taxonomy is not separable from its instance objects. As an example, one could define a schema for describing the makeup of faculty, staff, and students at a college or university. Mapping the individuals at Ohio State University to the schema would result in a different taxonomy than mapping individuals at Madison Area Technical College.

At first, I was uncomfortable with using a shared schema for both form and sign objects. Yet, the fact that a sign paradigm entry belongs to a particular conjugation, in addition to being a transitive verb, is relevant at both the form and sign levels, in a descriptive sense. What is important is that sign paradigm entries do not have form classes as part of their data structure. Membership in a conjugation must involve a demonstration that the sign paradigm entry is constructed according to a morphological pattern that is consistent with the conjugation.

C.3.1 Form Classification

The last requirement of a taxonomy is a means of classifying instances. In order to determine the primary class of a form paradigm entry, a simple application of *klass* on its M-cat is required. This is all that is required to define the form taxonomy.

There are two reasons for an additional classifying relation. First, it may be useful to know whether a form paradigm entry is a member of ancestor form classes. Second, in terms of scientific description, it is good to state what it means to be an instance of a form class.

For those reasons, I supply a predicate in_{mp}. I have reproduced the coq code below.

```
inₘₚ : ∀ (κ : K) (stem word : structₘₚ),
    FEₘₚ word
    → klass (fst stem) = klass (fst word)
    → klass (fst stem) ≤ₖ κ
    → isStemₘₚ stem word
    → Prop.
```

The predicate takes an identifier for a paradigm instance. An identifier for a paradigm instance is a form class and a stem. A stem is represented by a $struct_{mp}$, a pair of a list of m and a list of processes. I will call the identifying form class, κ, the *target*. The predicate also takes a second $struct_{mp}$, which is the word-form to be tested. If both the stem and form paradigm entry have the same form class, and that class is a subclass of the target, so long as the stem is a stem of the word, the predicate is satisfied. This describes what it means to be part of form paradigm instance in the theory.

Essentially, instances share stems and classes.

C.3.2 Sign Classification

Sign paradigm entries do not expose M-cats, so *klass* cannot be used directly. For the sake of a complete taxonomic definition, and in order to describe what it means for a sign paradigm entry to be a member of a paradigm instance, I provide the predicate in_{sp}. The definition in coq is verbose because it requires a form-sign relation definition to complete the proof. In order to focus on the core idea, I provide a pseudo-formal definition below.

$$
\begin{aligned}
& in_{sp} \\
& \quad (\text{target} : K) \\
& \quad (\text{stem word} : struct_{mp}) \\
& \quad (\text{sign} : struct_{sp}) \\
& \quad (\text{rule} : \text{form-sign relation}) \\
& \quad (word : struct_{mp}) \\
& \quad (proof_{ev} : word \equiv_{mp} word) \\
& \quad (proof_{m} : \pi_1\ word \leq_m cat_m) \\
& \quad (proof_{k} : klass(\pi_1\ word) \leq_k k) \\
& \quad (proof_{s} : meaning(existT\ Sns\ s_1\ \beta)(\pi_1\ word)) : Prop := \\
& \quad let\ \text{newsign} := \text{rule}\ word\ \beta\ proof_{ev}\ proof_m\ proof_k\ proof_s\ in \\
& \quad (klass\ \pi_1\text{stem} = klass\ \pi_1\text{word}) \wedge \\
& \quad (klass\ \pi_1\text{word} \leq_k \text{target}) \wedge \\
& \quad (isStem_{mp}\ \text{stem word}) \wedge \\
& \quad (SE_{sp}\ \text{sign}) \wedge \text{newsign} = \text{sign}
\end{aligned}
\tag{I}
$$

Just as in in_{mp}, the instance identifier is the target form class and a stem. Then there is the sign to be tested for membership. What I reduced in comparison to the coq definition are the arguments that define a form-sign relation rule, called *rule*, above. Below the *rule* are the arguments to the *rule*. The let expression defines a variable, a sign paradigm entry, newsign, through the application of *rule*. The initial portion of the predicate body should look familiar from in_{mp}. This establishes that the word supplied to the rules is of the expected class and has the expected stem. Finally, there is a check to ensure that *sign* is a valid sign paradigm entry and is equal to *newsign*, which was constructed using a word that was verified to be a member of the form paradigm instance.

To put it in a more concise way. For a sign paradigm entry to be a member of a paradigm instance, it must have been constructed according to a morphological pattern that is consistent with membership. The form paradigm entry it was derived from must be a member of the equivalent form paradigm instance for that to occur. Then there must be a rule that results in an equal sign paradigm entry using the word's form paradigm entry.

The in_{sp} predicate describes sign paradigm instance membership. That is its primary function, in addition to completing the definition of taxonomy.

An analyst can simply look at the definitions in the system, and know whether in_{sp} would be satisfied. That is because the structure of sign paradigm instances are defined by form-sign relations.

C.4 Sign Paradigm Structure

In §6.9 I provided many details of reasoning over form paradigms instances. That should make it clear that their structure is rigidly defined. All I said about sign paradigms is that there is a membership predicate. In the conclusion, I talked about various ways of improving the formalization of sign paradigms. This may have lead the reader to underestimate the extent to which their structure is well-defined.

Consider that the collection of constructed sign paradigm entries within a particular sign paradigm instance may be enormous. The notion that there are six and only six sign paradigm entries that correspond to intransitive English *walk* is a simplification. This is because *walk* may be polysemous. *Walk* doesn't seem like the best example of polysemy in intransitive uses, but there is an additional sense of 'identify with', such as 'I walk with the cool kids.' There is also a sense of 'leave' or 'quit', as in 'If you do it again, I'll walk.' These are sufficiently distinct and conventional that they merit individuation as meanings. For every one of the meanings, the number of sign paradigm entries for the *walk* instance is multiplied by six.

This is because there is a structural requirement. Any polysemous instance has six sign paradigm entry types, which have the same phenos and tectos as the others. The place where six is stipulated are the form-sign relations in SE_{sp}. Just as Paradigm Linking Theory (PLT) (Stump, 2016) provides relations such as $Corr$ and pm, which contain information specific to many paradigm instances under a single relation, I group the structural information for all sign paradigm instances under the SE_{sp} predicate.

There is an analogy that can be made with PLT. Remember that PLT has three paradigms, with the realized paradigm as an end-point. The middle paradigm, the form paradigm, corresponds to the role of the form-sign relation in SE_{sp}. This is due to the fact that the form paradigm represents a state where the structure of the realized paradigm has been established, prior to realization.

Given this analogy, one may wonder why, in Appendix B, there were not six form-sign relations corresponding to the six lexical entries. The T functions are used to reduce the number of rules. This obscures the paradigm structure description in some ways, even as it makes other patterns clearer.

Consider the rule that deals with a past tense sign paradigm entry. Repeated below.

$$\frac{\alpha \leq_m PAST_m \qquad \alpha \leq_k \mathsf{verb}_k \qquad meaning(\beta, \alpha)}{\langle(\lambda \mathrm{st.s} \bullet \mathrm{t})\alpha, (\lambda \mathsf{t.Nom} \multimap \mathsf{Fin})\alpha, (\lambda x.past(x))\beta\rangle} \text{pstSP} \tag{II}$$

The rule is a description of a number of qualities. First, it produces instances of verb_k, which would be better named $\mathsf{intrans}_k$ or something similar. One could use in_{sp} to prove class membership for a past tense sign, but the purpose of form-sign relations is to provide the description in a direct manner. Second, the conclusion serves as a template for a sign paradigm entry. To emphasize this, I simplify it and put in a blank where the template should be filled.

$$\lambda \mathrm{s.s} \bullet \underline{\qquad\quad}; \mathsf{Nom} \multimap \mathsf{Fin}; past \tag{III}$$

If one were to remove the T functions, there would be six cell-like objects defined for English intransitive verbs. The structure of instances of the intransitive verb paradigm are clearly stated.

For other paradigms, the number of intransitive entries may not equal 6. There may be overabundance, resulting in more than six entries, but the tectos and phenos, minus the string support, will be uniform. In general, paradigms bounded by inflection will be very uniform.

Paradigms such as derivational paradigms, or paradigms with lexical suffixes may be less uniform if lexical semantics filters out particular combinations, but lack of uniformity is only represented to the extent that it is true.

Note that the form-sign relations are primitives of the theory. They must be empirically discovered. They rely on other information, such as form classes and M-cats, but in their role of defining structures within syntactic paradigms, they are by no means secondary.

Appendix D

Peeke's Classifier Lists

There are two listings that I will reproduce here. The first was from Peeke (1968) and the second from Fiddler (2011). Both are from Catherine Peeke, the second listing was to be included in an unfinished dictionary. It may be particularly useful to list it here since it remains unpublished. I will include some additional annotations and notes to accompany the listings.

D.1 Peeke 1968

There are four columns. The first is headed by abbreviation used for classifiers in Peeke's dissertation. It lists the classifiers. I use Peeke's original orthography, except to replace the diaeresis with a tilde on nasal vowels. Peeke considered there to be a source noun for all items. These are in the second column. A verbal usage, if it exists, is listed in the third column. I believe (Nlz.) means that the verb is considered a nominalization. Additional combinations are in the last column.

{ncl}	Source Noun	With Verb	Other Combination
-ba	õba 'palm leaf'		adobake 'only one palm leaf'
-bæ	õbæ 'territory'		wabæka 'another territory'

-bẽ	õbẽ 'vine'	õkĩbẽ 'string he will make' (Nlz.)	kibẽdõ 'what vine, string'
-bo	ĩkeĩbo 'egg'		wabo 'another egg'
-bõ	õdõbõ 'eye, face'	ẽbõ 'to have a face'	adobõ 'same eye, face'
-dæ̃	kæ̃dæ̃ 'manioc'	kæ̃kidæ̃ 'manioc to be eaten'	wadæ̃ 'another manioc, plantain'
-de	õdõde 'mouth'	kæde 'to have pain in mouth'	adodeke 'one word'
-dẽ	õdõdẽ 'abdomen'	bẽdẽ 'to swell in abdomen'	
-dõ	oõdõ 'river'		kidõdõ 'what river'
-gæ̃	õgæ̃ 'genitals'	eõ togæ̃ 'to circumcise'	
-ga	baga 'tooth'	wæga 'tool he cries over' (Nlz.)	bĩga 'two teeth, tools'
-gõ	kagĩgõ 'corn (ear)'	kæ̃kigõ 'corn to be eaten' (Nlz.)	pikæ̃gõ 'old corn'
-ka	dika 'stone'	ẽka 'to have seed (stone)'	giyæ̃ka 'small stone'
-kã	õwẽkã 'flesh'	ĩkã 'to be an animal'	kĩkãbẽ 'what animal could it be'
-ko	weoko 'cloth'		kĩkodõ 'what cloth'
-kõ	õkõ 'dwelling'	godẽĩkõdẽ 'in the dwelling where he had gone' (Nlz.)	wakõdẽ 'in another dwelling'
-õdæ	õõdæ 'sky'	wæ̃kĩõdæ 'day he will die' (Nlz.)	ææõdædõ 'which day'
-pa	õõpa 'dart'	tæ̃dõpaka 'spear with which he spears' (Nlz.)	wapa 'another dart, spear'

-pæ̃	æpæ̃	bekĩpæ̃ ‘water to drink’ (Nlz.)	kĩpæ̃dõ ‘what water, liquid’
-po	õdõpo	kædẽpo ‘hand with which he had worked’ (Nlz.)	ææpodõ ‘what hand’; i.e., ‘how many’
-ta	õta ‘nail, claw’	yõdõta ‘table for laying things’ (Nlz.)	giyãta ‘small shell, clay bowl’
-tæ̃	batæ̃ ‘liver’		adotæ̃ke ‘only one liver’
-ti	õdõti ‘thigh’	giti ‘to become numb in thigh’ (Nlz.)	
-to	ooto ‘basket’		ææpotodẽdõ ‘in how many baskets’
-wa	õdõwa ‘foot’	kowa ‘to pierce foot’	adowake ‘only one foot’
-wæ̃	awæ̃ ‘tree’	wodõkiwæ̃ ‘stick to be hung up’ (Nlz.)	kĩwæ̃dõ ‘what manioc’
-wo	ẽbõwo ‘name’		wawo ‘another name’
-yõ	oyõ ‘leaf’		kĩyõdõ ‘what leaf’
-∅	unclassified noun	ko-∅ ‘to pierce’	kĩ-∅-dõ ‘what (unclassified thing)’
-bẽdẽ	õdõbẽdẽ ‘leg’	yãkæbẽdẽ ‘to tie around leg’	æ̃bæ̃bẽdẽ ‘the other leg’
-bõka	õdõbõka ‘ear’	gipobõka ‘to insert finger in ear’	ĩbæ̃bõka ‘this ear’
-gõpo	õdõgõpo ‘finger’	kægõpo ‘to have pain in finger’	bãdĩgõpo ‘that finger’
-tawẽ	õdõtawẽ	kætawẽ ‘to have pain in chest’	
-yabæ	õdõyabæ ‘back’	kæyabæ ‘to have backache’	tækæyabæ ‘in middle of back’

D.2 Fiddler 2011

There were a number of columns and annotations in Fiddler (2011) that were either unexplained or irrelevant to this work. I have condensed the two tables he provided, one for monosyllabic suffixes and the other for multisyllabic suffixes into a single listing. The first column is for the suffix. The second is a gloss. The third is an associated noun.

Affix	Gloss	Associated Noun
-a	‘covering, outer surface’ (e.g., skin)	
-ba	‘thing with parallel parts’	*õba* ‘palm leaf’
-bã	‘all’	
-bæ	‘land’	
-bæ	‘central part’	
-bẽ	‘long thin cylindrical object’	‘õbẽ’ ‘vine’
-bii	‘flat slab’	
-bo	‘round thing’	ĩkeĩbo ‘egg’
-bo	‘leaf’	õyabo ‘leaf’
-bõ	‘small round thing’	õdõbõ ‘face, eye’
-bõ	‘name’	ebõwo ‘name’
-dæ	‘hemisphere’ (e.g., bowl)	
-dæ	‘tail of bird’	
-dæ̃	‘staple food’	kæ̃dæ̃ ‘cassava’
-de	‘edge, opening’	õdõde ‘mouth’
-dẽ	‘inside part’	õdõdẽ ‘abdomen’
-do	‘forehead, dome-like covering’	
-dõ	‘way’	õdõ ‘river’
-ga	‘pointed thing’ (e.g., tooth, beak)	baga ‘tooth’
-gæ̃	‘spouse’	õgæ̃ ‘genitals’
-gæ̃	‘vessel, container’	

-gõ	'corn'	kagĩgõ 'corn'
-gõ	'long pointed thing'	
-gi	'ground, soil'	
-gĩ	'long thin thing' (e.g., string, hair)	yægĩkabẽ 'string'
-i	'semifluid' (pliable mass)	
-ka	'round extension' (e.g., head)	dika 'stone'
-kã	'physical part of person'	õwẽkã 'flesh'
-kõ	'building'	õkõ 'dwelling'
-koo	'collection' (e.g., bunch, cloth)	weokoo 'cloth'
-o	'area, part'	
-pa	'long flat thing' (e.g., board)	
-pa	'dart'	õõpa 'dart'
-pæ̃	'liquid'	æpæ̃ 'water'
-pæ̃	'branching thing' (e.g., stream)	õdõpæ̃ 'arm'
-po	'earth'	
-po	'hand' (e.g., paw, wingtip)	õdõpo 'hand'
-po	'canoe, boat'	wipo 'canoe'
-ta	'flat extended thing' (e.g., leaf)	õta 'fingernail, claw'
-tæ̃	'liver'	batæ̃ 'liver'
-ti	'thigh'	õdõti 'thigh'
-to	'stick, carrier' (e.g., pole, basket)	ooto 'basket'
-wa	'bottom part' (e.g., foot, tuber)	õdõwa 'foot'
-wæ̃	'plant'	awæ̃ 'tree'
-we	'red' (e.g., blood)	wepæ̃ 'blood'
-wẽ	'cassava'	kẽwẽ 'cassava plant'
-wo	'immaterial part of person' (e.g., soul)	ẽbõwo 'name'
-yæ̃	'appendage' (e.g., tail, appendix)	
-ye	'a certain type of waterfowl'	

-ye	'a certain type of insect'	
-ye	'bad thing' (e.g., taboo, need)	
-yõ	'material part of person'	
-yõ	'kind of palmate leaf'	oyõ 'leaf'
-bẽdẽ	'leg'	õdõbẽdẽ 'leg'
-bõka	'ear'	õdõbõka 'ear'
-depo	'year'	
-gadæ̃	'esophagus'	
-gade	'jaw'	
-gade	'stomach'	
-gẽbẽ	'nerve, ligament, tendon'	
-gĩdewa	'tongue'	õdõgĩdewa 'tongue'
-gõpo	'finger'	õdõgõpo 'finger'
-kade	'vessel'	
-kado	'nose'	õdõkado 'nose'
-kagĩ	'hair'	okagĩ 'hair'
-kapo	'knee'	
-mĩka	'throat'	õdõmĩka 'throat'
-miæ̃	'tail'	õdõkagimiæ̃ 'tail'
-nikæ̃	'stream'	
-õdæ	'day'	õõdæ 'sky'
-pade	'stream'	
-tagæ̃	'shell' (e.g., scale, husk)	
-tawẽ	'chest'	õdõtawẽ 'chest'
-tedæ̃	'time period' (e.g., season)	
-todepæ̃	'thigh'	õdõtodepæ̃ 'thigh'
-yabæ	'back'	õdõyabæ 'back'
-yebo	'small oblong thing' (e.g., thumb?)	

-yedẽ	‘time of occurrence’ (e.g., moment)
-yõbõ	‘place’

Appendix E

Linguistic Glosses

When possible, I follow the Leipzig glossing rules (Leipzig Glossing Rules, 2015). Otherwise, I utilize the conventions compiled by Wikipedia.[1]

One exception to the rules that are found in those sources is that I make a distinction between roots that are bound and those that are free. By root, I mean an unanalyzable morph, which may serve as a word-form alone, or from which a word-form may be derived by some process. The dash '-' follows suffixed free stems that are roots. The interpunct '·' follows suffixed bound stems that are roots. The dash is otherwise used as described in Leipzig.

Another exception is that I add an additional '.*name*' gloss to LS and CLF glosses to label the specific affix with an English mnemonic, where *name* may be something like *plant* or *cloth*. For instance CLF.leaf, *-yabo*, or LS.fruit, *-ka*. These names are used consistently, even though the meaning of the suffix is not necessarily associated with "leaf" or "fruit" in context. A classifier construction like *yẽdẽ-ka*, big-CLF.fruit, could refer to fruit, stones or other objects. The name "fruit" is used uniformly despite this. The choice to use LS or CLF is based on the role the lexical suffix plays in a particular context.

Two additional conventions used in this work are the us of '∅' and '?'. When a root has no obvious independent meaning, it is glossed as '∅'. When I do not know what meaning a root or affix carries, it is glossed '?'.

The glosses that are used in this work are listed below. An asterisk appears in the descriptive text when

[1] https://en.wikipedia.org/wiki/List_of_glossing_abbreviations

the glossing abbreviation does not appear in Leipzig.

E.1 Glossing Conventions

1 First Person: The affix *-bo* or the stem *bo* in the pronoun *botõ*. 2, 33, 44–45, 83–84, 86–87, 99, 101, 107, 109–111, 118–121, 134–135, 137–140, 143, 152, 161, 243, 323, 329, 333

1.INCL First Person Inclusive: The suffix *-bõ* or the stem *bõ-* for the short pronominal form *bõtõ*. 76, 323, 329, 333, 433

1.PL First Person Plural: The suffix *-bõdi* is plural and appropriate for exclusive usage. 64

2 Second Person: The suffix *-bi* for singular, or *-bĩ* with the dual or plural. 55, 88, 97, 109–110, 113–114, 117, 133, 312, 323, 329, 333

2.Mo * Second person with maternal kinship relation: The suffix *-bĩ*, which is also used for 2 dual and plural. 323, 329, 333

3.F Third Person Feminine: The suffix *-dã*, which is often, but not always, used to indicate that the person is feminine. 78, 82–83, 108–109, 138, 153, 290, 305, 323, 333

3.H * Third Person Sentient: The suffix *-kã*, which is used for human beings and sometimes animals. The suffix tends to be used similarly to *he, she* or singular *they* in English, in contrast to *it*. The suffix is also used as the lexical suffix for 'body'. 78, 82–83, 86–88, 100, 103–104, 106–110, 112, 116, 118, 121, 140, 152, 224, 290, 307, 323, 333

3.PL Third Person (Sentient) Plural: The suffix -*dãdĩ*. 80, 84–85, 100, 135, 137, 142

ACC Accusative: The accusative *ĩ*. 82–83, 110, 118, 141

.arm * The lexical suffix -*pẽ* (pre-merge /pæ̃/). 78, 117, 128

AUG * Augmentative: The suffix -*bo*, which is used to indicate that something is large. 100, 150

.board * The lexical suffix -*pa*. 3, 83, 116, 131, 146, 148, 322

.body * The lexical suffix -*kã*. The suffix is also used as the third person singular sentient. It may also indicate masculinity. 78, 85–86, 89, 112–114, 121, 127, 224, 307, 322

.butt * The lexical suffix -*bode*. 103, 128

.canoe * The lexical suffix -*po*. 68, 107, 110, 117–119, 125, 127–129, 137–138, 140, 143, 146, 148, 153, 161, 242–243, 322

CAUS Causitive: The affix -*dõ*. 83, 87, 107

CLF Classifier: Classifiers in Wao Terero are lexical suffixes that play a classifier role. 2–3, 6, 33, 43–46, 55, 79–80, 83–85, 89, 95, 99–107, 109–121, 125–126, 133, 137–138, 140, 142–143, 146, 148–149, 151–156, 160–162, 164–167, 243, 364

COL * Collective Number: The suffix -*idi*, which is used for collections of things, often with some typical member. 80, 85–86, 142

COMPL Completive Aspect: The affix -*ĩ* indicates that an action is complete. 86

COP Copula: The copular *ĩ*. 3, 43–44, 46, 78, 86–88, 97, 111, 114–117, 120, 134–135, 148–151, 154, 164–168

.cord * The lexical suffix -*bẽ* (pre-merge /bẽ/). 46, 98, 131, 149–150

DECL Declarative: The suffix -*pa* in Wao Terero is placed on the main verb of a declarative sentence. 2–3, 6, 33, 43–46, 66, 76, 78–80, 82–88, 99–101, 103–104, 106–107, 109–112, 115–116, 118–121, 134–135, 137–140, 142–143, 148–152, 160–161, 164–168, 243, 309, 315, 323–324, 333, 335

DEM Demonstrative: The longer form *bãdĩ* is generally provided as the translation for the Spanish proximal demonstrative, while the form *ĩ* is provided for the distal in many cases. Peeke (1968) also considered *bãdĩ* to be proximal. There are significant instances of speakers translating the demonstratives in the opposite direction. The *ĩ* also serves as the base for pronoun-like forms and is homophonous with the copula. 79, 88, 97, 106, 114–116, 142–143, 160

DESI * Desiderative: The suffix -*edẽ* expresses desire. The suffix -*kĩ* is also often used to express desire but may also be used to express future tense and is provided in translations of the Spanish infinitive for citation forms. 67

DU Dual: The suffix -*da* is used to form a dual for all persons. 121, 323, 329, 333

.ear * The lexical suffix -*bõka*. 101–103, 128, 132, 136

.egg * The lexical suffix -*bo*. 46, 99, 101, 127, 131, 146, 148–149, 151

.food * The lexical suffix *-dẽ*. 104, 127, 129, 322

.foot * The lexical suffix *-wa*. 120–121, 128

.forehead * The lexical suffix *-to*. 104, 128

.frond * The lexical suffix *-ba*. 94, 96, 146, 148

FUT Future Tense: The suffix *-ke*, the suffix *-kĩ*, or the periphrastic construction VERB.STEM-*ke* ke-PERSON.INFLECTION, such as *go-ke ke-bo*, ‘I will go.’. The *-kĩ* affix also expresses desire and is used as a translation of the Spanish infinitive for citation forms. 76, 87, 116, 323

GER * Gerundial: Called the gerundial since at least Peeke (1968), the suffix *-te* provides an adverb-like status to a verbal phrase. It is also used in constructions that indicate that a sentient being is a direct or indirect object. 82, 86–88, 100, 104, 110–111, 115, 117–118, 136, 323–324

.group * The lexical suffix *-koo*, which can be used for clothing and as a kind of collective plural. 6, 79–80, 85, 89, 115, 127, 146, 148, 150–152, 155, 160

.house * The lexical suffix *-kõ*. 139

IMP Imperative: The suffix *-ẽ* is used for positive commands issued to a single second person. 87, 97, 102, 106–107

INF Infinitive: The suffix *-kĩ* in Wao Terero may not be a grammatical infinitive but Wao speakers translate the Spanish infinitive citation form to verbal forms that end in *-kĩ*. 65–66, 76, 96

INS Instrumental: The suffix *-ka* is used to indicate that an object was used to perform an action, such as a knife is used for cutting. The affix is homophonous with the lexical suffix *-ka*. 83, 118–119, 137, 142–143, 161

.jaw * The lexical suffix *-gade*. 106–107, 128, 141

.knee * The lexical suffix *-kapo*. 115, 128

.leaf$_1$ * The lexical suffix *-yabo*. 9, 69, 83, 87, 94, 96, 129–131, 146, 148–149, 151, 322

.leaf$_2$ * The lexical suffix *-yo*. 94, 96, 131, 146, 148–149

.leg * The lexical suffix *-pẽdẽ*. 120, 128

LIM * Limitative: The suffix *-ke*, which is associated with a meaning of 'just' or 'only'. 78, 83, 152, 323, 335

.liquid * The lexical suffix *-pẽ* (pre-merge /pæ̃/). 9, 72, 83, 105, 117, 120, 127, 129, 145–146, 148, 322

LOC Locative: The suffix *-de* in Wao Terero functions as a locative when placed on nominals. 6, 76, 86, 107, 139

LS * Lexical Suffix: Lexical suffixes are suffixes with lexical meanings. 2, 6, 9, 33, 44, 46, 55, 68–69, 72, 78–79, 83–87, 89, 94, 96–101, 103–108, 110–111, 115–121, 127–141, 144–145, 147–155, 161, 224, 242–243, 305, 307, 322, 328, 335

.mouth * The lexical suffix *-de*. 103, 128, 137–138

NEG Negative: The suffix *-dãbaĩ*, which is used in the periphrastic negative or the negative word *wii*. 78, 86–88, 166–168

.nose * The lexical suffix -*kado*. 128

.penis * The lexical suffix -*gẽ*. 108

PL Plural (for persons): The affix -*di*. 312, 323, 329, 333, 433

.plant * The lexical suffix -*wẽ*. 2, 6, 33, 44, 79, 84–85, 107, 115–116, 121, 127, 129–131, 138, 140, 146, 148, 151, 155–156, 164, 166, 242, 322

PRO * Pronominal: The suffix -*tõ* on short pronominal forms. It will include the stem *tobẽ*- on long pronominal forms and *ĩ* with some short forms. 45, 76, 78, 81–83, 87–88, 97, 99, 103, 109, 111, 120–121, 134–135, 138, 140, 301, 325

PST Past Tense: The affix -*ta*. 2, 33, 44, 83–84, 87, 104, 107, 110, 118–119, 121, 137–140, 142–143, 161, 323–324, 333

PTCP Participle: The affix -*ĩ*. 139

PURP Purposive: The suffix -*ketãte* on a subordinate verb indicates an aim or purpose of the main verb action. 100

Q Question marker: The suffix -*dõ* occurs on question words. 85, 97, 113–114, 121, 154–155

.river * The lexical suffix -*pade*. 116

.road * The lexical suffix -*dõ*. 97, 105–106, 120, 127

.seed * The lexical suffix -*bõ*. 99–101, 111, 126, 129, 146, 153, 165–167, 305, 322, 328

.shell * The lexical suffix -*ta*. 68, 107, 117, 120, 127, 131, 134–135, 146, 148, 151, 153, 322

SIM Simultaneous: The suffix *-yõ* indicates that the event of the verb that it attaches to is concurrent with another verbal action. 86, 100

.spine * The lexical suffix *-yabe*. 128

.stone * The lexical suffix *-ka*. 2–3, 9, 43–46, 55, 68, 79, 98, 100–101, 105, 110–111, 127–131, 133, 142, 146, 148–155, 164, 322, 335, 364

.string * The lexical suffix *-gĩ*. 55, 98–100, 110, 128, 131, 133

.territory * The lexical suffix *-be* (pre-merge /bæ/). 2, 33, 44, 96–97, 121, 138, 145, 148

.thigh * The lexical suffix *-ti*. 120, 128

.thorn * The lexical suffix *-gõ*. 99, 118–119, 127–128, 131, 137–139, 143

.throat * The lexical suffix *-gadẽ*. 89, 103, 128

.tongue * The lexical suffix *-gẽdewa*. 109

.tooth * The lexical suffix *-ga*. 106, 127, 131

.vessel * The lexical suffix *-kade*. 104, 106–107, 153