LEFT-RIGHT ASYMMETRIES IN WORDS: A PROCESSING-BASED ACCOUNT

by

Yoshihiko Asao
February 1, 2015

A dissertation submitted to the
Faculty of the Graduate School of
The University at Buffalo, State University of New York
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

UMI Number: 3683004

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI 3683004

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Contents

Abstract

This thesis presents a novel account of the suffixing preference and related phenomena based on language processing. The suffixing preference refers to the typological fact that suffixes are more common than prefixes in the world's languages. This thesis argues that prefixes are avoided in languages because they are difficult to immediately recognize in language comprehension, due to their short length, unpredictability, and the lack of phonological cues. This idea is demonstrated by a simulation program based on a probabilistic model of morpheme segmentation, using corpus data of typologically diverse prefixing languages in the world, as well as a hypothetical language. We also argue that our processing-based account is compatible with the fact that there are considerable variations about the suffixing preference between languages, as well as between different grammatical morphemes. It is argued that our account can also explain the asymmetry in phonological integration of affixes, as well as parallel phenomena in clitics and compounds.

Chapter 1

Introduction

This thesis presents a novel account of the suffixing preference and related phenomena from a functional-typological perspective. The suffixing preference refers to the typological fact that there are more suffixes than prefixes in the world's languages.

A unique characteristic of our approach is that our claim is implemented as an explicit computational model. An external explanation of a linguistic fact often has a risk of being informal and ad hoc. A valid explanation should not arbitrarily pick an explanatory factor that happens to work, but produce an array of predictions that follow from its framework. A computational model is a way to achieve this goal. Our model will use probability. Recently there has been converging evidence that people are able to use probabilistic knowledge obtained from their previous experience in order to solve cognitive tasks (M. D. Lee & Wagenmakers, 2014). Our model is based on the assumption that people can use probabilistic knowledge in language comprehension. Stipulations about mechanisms specific to morphology or syntax are kept minimal.

The second goal of this thesis is to present a comprehensive account of the suffixing pref-

erence as a typological phenomenon, instead of just arguing for the general preference. For example, while it is true that there is a general preference for suffixing, prefixing is not categorically ruled out. In fact, we can find languages that rely heavily on prefixes in different corners of the globe. A unique feature of this thesis is that we conducted a computational study on those typologically diverse languages. The question why languages with apparent inefficient properties are still attested is not specific to the suffixing preference, but applies to any psycholinguistic account of typological facts.

1.1 Organization of the thesis

Chapter 2 presents the theoretical positions that we will take in this thesis, which lays the foundations for the discussions in the later chapters.

Chapter 3 is an extensive review of the data and existing approaches to explain the suffixing preference. Through this process, we will point out some aspects of the suffixing preference that a valid explanation needs to address. We will also point out some common pitfalls that functional approaches tend to fall into.

Chapter 4 presents our novel approach to the suffixing preference. Our approach is psycholinguistic in the sense that it claims that the suffixing preference is due to the temporal nature of language processing. Unlike previous approaches, our approach is implemented as a computationally explicit model.

The second half of this thesis will address additional issues to make our approach more comprehensive.

Chapter 5 will address the issue of **prefix-rich languages**. Although suffixes are more common than prefixes, there exist languages that rely heavily on prefixes. Thus, one might ask why such prefix-rich languages are thriving without difficulties, if the psychological disadvantage of prefixes is real. Through analyses of prefix-rich languages such as Swahili, it is argued in this dissertation that real-world prefixing languages have properties that compensate for the difficulties associated with prefixes.

Chapter 6 will discuss why the suffixing preference varies depending on **grammatical categories**. Psycholinguistic accounts predict only a general tendency towards suffixing. However, it is well known that the suffixing preference depends heavily on the grammatical categories of affixes. For example, while case markers show very strong suffixing preference, pronominal affixes on verbs show almost no suffixing preference. This chapter investigates whether the combination of historical and psycholinguistic factors can explain the typological pattern based on typological quantitative data.

Chapter 7 will examine whether our discussion can be extended to **clitics and compounds**, as well as **phonological asymmetries**. Because our model does not rely on the grammatical status of affixes, it predicts that similar asymmetries can be found in other types of morphological processes. This chapter discusses these predictions.

Chapter 2

Background

2.1 Introduction

This chapter presents basic theoretical assumptions of this thesis, which lay the foundation for later chapters.

This thesis takes a *functional-typological* approach. That is, we will regard grammatical facts as conventionalized functional preferences, rather than something that can be deduced from synchronic principles. The idea of understanding grammar as grammaticalized functional preferences is best formulated as the Performance-Grammar Correspondence Hypothesis (PGCH) (Hawkins, 2004, 2014), which is the topic of Section 2.2. In Section 2.3, we will shift our focus to the definition of *word* from a functional perspective. This is crucial because our discussions will hover around the boundary between morphology and syntax. Section 2.4 summarizes this chapter.

2.2 The Performance-Grammar Correspondence Hypothesis (PGCH)

One of the most elaborate theories that connect psycholinguistic consideration and typological facts is presented in Hawkins (1994, 2004, 2014). His basic tenet is formulated as the Performance-Grammar Correspondence Hypothesis (PGCH) in Hawkins (2004, 3).

> **Performance-Grammar Correspondence Hypothesis (PGCH)**
>
> Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments. (Hawkins, 2004, 3)

The most convincing argument for his position, in my opinion, is the parallelism we can often find between grammatical and usage facts. That is, a structure that is categorically ruled out in one language is often only a statistically less preferred option in another language. In such a case, a theory that regards the former as a hard-wired constraint governed by grammatical principles and the latter as the matter of usage will miss a significant generalization. A more parsimonious view is that a preference in usage is grammaticalized in some languages.

The measure of processing efficiency is quantitative in nature. On the other hand, it is sometimes claimed in the generative literature that a rule of grammar cannot *count* (Chomsky, 1965, 55).[1]

[1] Alternatively one may argue that grammaticality itself is graded. We are not particularly against that idea, but even in that case, however, we believe that the basic point of our discussion still holds. Grammatical judge-

Thus, what is grammaticalized is a grammatical property that correlates with a performance factor, not a performance factor itself. For example, in English usually an adjective must precede a noun, as in *a yellow book*. However, it must follow a noun when the adjective is followed by a modifying phrase:

(1) a. *a [yellow with age] book

b. a book [yellow with age]

(4.5) of Hawkins (2014, 81)

This can be thought of as an example of grammaticalization of a processing factor. On average we can expect that an adjective with a phrase is longer than an adjective without one, and a long preposed adjectival phrase creates a center-embedded structure. Because this is only a statistical expectation, it is not difficult to find a case where a preposed adjectival phrase is longer than a postposed adjectival phrase. The adjectival phrase is longer in [*very carefully articulated*] *question* than in *question* [*worth asking*], but grammar cannot rule out the former just by counting a number of words.

We will see similar examples from morphology. We will argue that stem-suffix order is more common than prefix-stem order, because an affix is shorter than a stem *on average*, and a short morpheme at the beginning is difficult to process. Of course, it is not difficult to come up with examples in which the affix has more syllables than the stem, such as *san-ity* or *inter-act*, but that does not change the overall grammaticalization pattern.

ments are often done with reference to a grammatical category, as opposed to a processing factor which usually refers to quantitative properties of a linguistic object, such as length and frequency.

An important feature of Hawkins' performance theory is that his principles are formulated clearly enough to calculate the relative efficiency of alternative constructions. His discussion is, however, focused on word order. Relatively little work has been done on the formalization of processing-based accounts of morphology. Hawkins (2004) discusses some issues on morphology based on his Minimize Form (MiF) principle, but its scope is limited to a few phenomena, such as the relation between frequency and syncretism. This thesis is intended to be an attempt to propose a formal machinery to connect psycholinguistic considerations with typological patterns in a domain other than word order.

A criticism against psycholinguistic approaches would be that they often lack discussions on how a psycholinguistic factor can result in a structural property of a language, or typological patterns. This is called the *problem of linkage* (Kirby, 1996). In Section 4.5, we will mention how the problem of linkage should be dealt with in our problem setting.

2.3 Emergent view of word

Word is an intuitive concept that has often been taken for granted in linguistic research. At the same time, it is one of the concepts which are notoriously difficult to define (Dixon & Aikhenvald, 2002; Haspelmath, 2011b). Haspelmath (2011b) presents the following list of the criteria for wordhood and points out that they do not coincide with each other.

(2) 1. speaker intuitions

2. semantic non-compositionality

3. orthography

4. phonology

5. morphosyntactic word criteria

   (a) potential pauses

   (b) free occurrence

   (c) external mobility and internal fixedness

   (d) uninterruptibility

   (e) non-selectivity

   (f) non-coordinatability

   (g) anaphoric islandhood

   (h) nonextractability

   (i) morphosyntactic idiosyncrasies

   (j) deviations from biuniqueness

While it is possible to arbitrarily pick one of the criteria (or a combination of criteria) and regard it as the definitional property of *word*, it is essentially a matter of terminology. A more interesting task from a functional-typological perspective would be to describe how these different wordhood criteria converge (or diverge), and explain why. This is compatible with the non-essentialist view of linguistic categories advocated in the functionalist literature. Dryer (1997), for example, argues that grammatical notions like *subject* and *object* are language-specific. In the Radical Construction Grammar (Croft, 2001), it is argued that parts of speech are language-specific. Haspelmath (2007, 2010) also argues for the position that categories

that are used to describe a language are language-specific. We can think of a similar argument for the concept of word.

This does not mean to deny that word is a basic concept for which even non-specialists often have intuitions (Sapir, 1921; Dryer, 2013). In order to explain the apparent universality of the word concept, however, it is not necessary to assume that word is an innate category which we are born with. Rather, we can think that words emerge in similar ways in different languages as a result of interacting factors.

An analogy with some other concept like *verb* may be useful here. While we can say that every language has verbs, this does not mean that there is a universally applicable formal criterion that delimits verbs from other parts of speech. Rather, prototypical verbs, that tend to take arguments and are used as predication in their unmarked forms, emerge as a result of how human cognition universally understands situations that involve actions. Languages can vary greatly with regard to whether words whose meanings deviate from a prototypical action exhibit the same formal properties, and languages also vary about what kind of formal properties can be used to distinguish verbs from other parts of speech. The concept *word* is similar: while prototypical examples of words can be identified in any language, what can be expressed within a word greatly varies according to languages, and there is no single formal criterion that can be used to determine whether a given expression is a word or not in any language.

In order to capture this kind of situation, the idea of the *Canonical Typology* (D. Brown & Chumakina, 2012) will be useful. The Canonical Typology describes each variable and then

observes their correlations. Using Haspelmath's list of wordhood criteria, we can characterize the canonical word as follows:

(3) **Canonical word**

1. speakers agree
2. semantically non-decomposable
3. written delimited by white spaces
4. forms a phonological unit
5. morphosyntactic criteria:
   (a) cannot be interrupted by pause
   (b) can form an utterance by itself
   (c) morpheme order is fixed inside, but free outside
   (d) cannot be interrupted by other words
   (e) selective restriction is strong inside, but free outside
   (f) a part cannot participate in coordination
   (g) a part cannot participate in anaphora
   (h) a part cannot be extracted
   (i) morphophonological idiosyncrasies only found internally
   (j) deviation from biuniqueness only found internally

With these criteria, *fish* and *kick* are canonical words, while *'ll*, *hot dog*, and *Alice in Wonderland* deviate from the canonical ideal in different ways. We expect that a canonical word satisfies many of the criteria above, and the less criteria it satisfies, the less word-like an expression is.

We argue that the criteria for wordhood are not an arbitrary collection of unrelated properties, but there are reasons why these properties correlate with each other and conspire to create prototypical *word*. In fact, there appear to be some logical relationships between the wordhood criteria. For example, anaphoric islands seem to presuppose compositionality. Anaphoric islands refer to the constraint that a part of a word cannot participate in anaphora (Postal, 1969). In order for an expression to participate in anaphora, its meaning must be accessed in comprehension. Thus, it is reasonable that if a word is not compositional and the meaning is retrieved from memory as a whole, then its subpart does not participate in anaphora.[2] On the other hand, phonological wordhood is arguably not directly relevant to anaphoric islands, and the correlation would be weaker. However, there are also motivations to align phonological and semantic units, which we turn to in the next section.

An affix can be understood as a grammatical morpheme which is smaller than a word. Using the definition of canonical word, we can define canonical affix: a canonical affix is smaller than word by every different criterion for wordhood. However, the boundary between an affix and what is not an affix is language-specific. Another consequence of the canonical view of *word* is that the historical development of an affix is not an instant reanalysis process

[2]See Ward, Sproat, and Mckoon (1991) for a pragmatic account of anaphoric islands that shares the basic idea advocated here.

(Haspelmath, 2011a). Rather, it is a process in which an expression gradually acquires more and more affix-like properties.

2.3.1 Phonological word

A phonological word is just one of the criteria in Haspelmath's (2011) list, but as Haspelmath himself mentions, the phonological word itself can be defined in many different ways. Hyman (2006, 229) lists the following different criteria for phonological wordhood:

(4) a. The demarcative word

A property marks the beginning or end of the word.

b. The culminative word

A property occurs only once per word.

c. The harmonic word

A property is realised throughout the word.

d. The metrical word

A word consists of hierarchically arrayed moras or syllables.

e. The minimum word

A word must consist of a minimum number of moras or syllables.

f. The maximum word

A word can consist of a maximum number of moras or syllables.

g. The phonotactic word

A word permits only certain output segments/sequences.

h. The morphophonotactic word

A word permits only certain input segments/sequences.

(Hyman, 2006, 229)

These different criteria do not always match, as quantitatively illustrated in Bickel, Hildebrandt, and Schiering (2009) based on the AUTOTYP database, which we will mention in Section 7.2.2. Just like wordhood in general, our position does not aim to claim that one of the criteria must be a universal definition of phonological word.

How, then, do these criteria for phonological wordhood correlate with grammatical wordhood? There is a huge amount of literature on the (mis)match between grammatical and phonological words (Nespor and Vogel (1986); Dixon and Aikhenvald (2002); Bickel et al. (2009), among others), but most of it is descriptive or formal in nature. Our question is whether there is a functional explanation for when and why grammatical and phonological words coincide.

We can think of a number of functional motivations as to why it is beneficial in language processing to align grammatical and phonological words. In language production, we can speculate that because a lexical item is retrieved from memory as a chunk, they are more easily pronounced as a chunk. In language comprehension and acquisition, benefits of aligning grammatical and phonological words are more evident, because it helps parse inputs. In psy-

cholinguistics, it has been revealed that people use a variety of cues to segment speech (See Jusczyk and Luce (2002) for an overview); each criterion of phonological wordhood corresponds to different segmentation strategies. The demarcative word (a), for example, is useful to identify a word boundary in a straightforward way. If, for example, word stress is fixed in the first syllable, it will be a reliable cue for word segmentation. In many of the world's languages, it is predictable where to put stress in a word (Hyman, 1977). In English, stress positions are not fixed at a specific syllable, but still it has been suggested that stress plays an important role in word segmentation (Cutler & Norris, 1988). The culminative word (4b) works in a similar way. When a word can only have a single main stress, two consecutive stresses mean that there is a word boundary between them. The harmonic word (4c), typical in vowel harmony, also helps segmentation. For example, when a language requires every vowel in a word to agree with regard to the feature [±back], the co-occurrence of vowels that do not match in [±back] will be a reliable cue for a word boundary between them (See Suomi, McQueen, and Cutler (1997) for experimental evidence of people's sensitivity to vowel harmony cues in Finnish). The phonotactic word (4g) helps parse too. If a language does not allow a word to have a certain sequence, say [nr], then the [nr] sound can be used as evidence that there is a word boundary between [n] and [r]. The role of phonotactics in speech segmentation, often probabilistic in nature, has been demonstrated in a number of experimental studies (Saffran, Newport, & Aslin, 1996; Saffran, Aslin, & Newport, 1996; Mattys & Jusczyk, 2001).

The minimal word (4e) works in a slightly different way. It may also help identify words, but it does not signal a word boundary. Instead, it helps us realize that there is *no* word boundary. For example, if we know that a word is at least bisyllabic, we can tell that the first syllable boundary from the beginning is *not* a word boundary.

In spite of these motivations, there can be discrepancies between a phonological and lexical unit, and thus languages have both productive morphology (i.e. phonological words larger than lexical units) and idioms (i.e. lexical units larger than phonological words). Our interest will be that there are certain left-right asymmetries in the patterns of mismatch.[3] In Chapter 4, we will discuss why left-right asymmetries arise, from the processing-based point of view.

2.3.2 Why canonical word?

The reason why the canonical view of word is useful in our discussion is twofold: (i) this justifies the use of the typological database, both in principle and in practice; and (ii) this enables us to understand the nature of our problem in a more coherent manner.

At a first glance, the canonical view of wordhood may appear to make a cross-linguistic comparison more difficult. However, we believe that this view justifies research based on typological databases. Each grammar author uses different formal criteria to distinguish affixes from words. If one takes the view that *word* must be defined with a single universal criterion, then we cannot rely on a database based on different descriptive grammars. In our approach, the lack of a universal definition is not fatal. As long as more prototypical words are treated

[3]The author is aware that the terms 'left' and 'right' are mere metaphors based on the writing convention of a particular culture. We will, however, occasionally use these terms as their use is well established in the literature and intuitive for the readers.

as words in each grammar, typological frequency comparisons make sense. It is a meaningful statement that suffixes are more common in the world's languages, even if the formal criteria to discern an affix from a word, for example, are not the same across different languages. A problem arises, however, when linguists are systematically biased about their way to apply wordhood criteria. For example, it would be a problem if linguists have a tendency to treat postposed grammatical morphemes as affixes more often than preposed grammatical morphemes, even when their grammatical behavior is the same. We will discuss this problem in Sections 3.2.1.

Recognizing that the boundary between affixes and words in subtle cases can be arbitrary in each grammar also enables us to better understand our predictions. Our processing-based account will predict that preposed grammatical morphemes tend to be more phonologically independent than postposed grammatical morphemes. A problem is that we do not know whether this difference in phonology is crucial for a grammar author in distinguishing affixes from words. In other words, we predict that preposed grammatical morphemes are more often *word-like*, but we do not know whether they are word-like enough to be described as words in each grammar. Assuming that sometimes they are described as words and sometimes they are not, we can make two empirical predictions. First, we expect more suffixes cross-linguistically, because sometimes the phonological difference between preposed and postposed grammatical words leads a grammar author to determine that the former is a word, while the latter is a suffix. Second, we will expect that prefixes are more phonologically independent than suffixes, because sometimes the phonological difference between preposed

and postposed grammatical morphemes do not affect a grammar author's judgment that they are both affixes (we will see such cases in Section 7.2). In this way, even without a universal criterion for wordhood, we can make an array of predictions, as long as each grammar is consistently written such that more canonical words are treated as words. We will discuss this issue again in Section 7.2.1.1.

2.4 Conclusion

This chapter laid the foundations of our position. In Section 2.2, we argued for the view in which grammar can sometimes be explained as conventionalized processing preferences. In Section 2.3, we advocated the emergent view of word.

Chapter 3

Suffixing preference: Data and previous approaches

3.1 Introduction

This chapter presents an extensive overview of the data on the suffixing preference and the existing approaches to explain them. When we talk about the suffixing preference, the following two hypotheses must be distinguished, as Bybee, Pagliuca, and Perkins (1990), as well as Himmelmann (2014), make clear: (i) the overall frequency asymmetry hypothesis: postposed grammatical morphemes are more common in general; and (ii) the boundedness asymmetry hypothesis: Postposed grammatical morphemes are more likely to be bound. We will argue that only the boundedness asymmetry hypothesis is supported by data. Many of the previous accounts are not aware of this distinction and sometimes give an explanation that works only for (i) the overall frequency asymmetry.

3.2 Suffixing preference: the data

Suffixes are cross-linguistically more common than prefixes. This observation is noted at least as early as Sapir (1921):

> On the three types of affixing—the use of prefixes, suffixes, and infixes—suffixing is much the commonest. Indeed, it is a fair guess that suffixes do more of the formative work of language than all other methods combined. (Sapir, 1921)

His impression was later confirmed by a number of databases, including Greenberg (1963); Hawkins and Gilligan (1988) and Dryer (2011e). In Greenberg's (1963) database of 30 languages, only one language is exclusively prefixing, while 12 languages are exclusively suffixing. Table 3.1 shows Greenberg's results. In the first table, his sample languages are classified by word order types; in the second table, they are classified by prepositional (Pr) and postpositional (Po) languages. We can see not only that suffixing languages are more common, but also that they correlate with the SOV word order or the use of postpositions.

Table 3.1: Greenberg (1963: 92)

	I (VSO)	II (SVO)	III (SOV)
Exclusively prefixing	0	1	0
Exclusively suffixing	0	2	10
Both	6	10	1

	Pr	Po
Exclusively prefixing	1	0
Exclusively suffixing	0	12
Both	15	2

Hawkins et al. (1988) present three different databases: Stassen's sample with 113 languages, Perkins and Bybee's data with 40 languages, and Gilligan's sample with 50 languages. In Stassen's sample, 4 languages are exclusively prefixing, and 47 languages are exclusively suffixing. In Perkins and Bybee's data, 2 languages are exclusively prefixing, while 7 languages are exclusively suffixing. In Gilligan's data, 2 languages are exclusively prefixing, while 15 languages are exclusively suffixing. Although different databases use different sampling methods and look at different sets of grammatical categories, suffixing preferences seem to be robust.

Another large, more recent database is the World Atlas of Linguistic Structures (WALS) (Dryer & Haspelmath, 2011). In addition to a number of chapters for individual grammatical categories, WALS has a chapter on the overall suffixing preference (Dryer, 2011e), which classifies the world's languages into 'strongly suffixing', 'weakly suffixing' and so on. The numbers taken from the WALS chapter are shown in Table 3.2, from which we can see that strongly suffixing languages are seven times as frequent as strongly prefixing languages.[1]

Table 3.2: Prefixing vs. Suffixing in Inflectional Morphology (Dryer 2011d)

Little affixation	141
Strongly suffixing	406
Weakly suffixing	123
Equal prefixing and suffixing	147
Weakly prefixing	94
Strongly prefixing	58
Total	969

[1] See Dryer (2011e) for the exact definition of 'strongly suffixing', 'weakly suffixing' and so on. The classification is based on the weighted sum of the feature values of the following ten categories: case, subject, tense-aspect, plural, possessive, definite or indefinite, object, interrogative, and adverbial subordinator.

This thesis mainly uses Matthew Dryer's database (p.c.), consisting of 1,836 languages. The database was constructed by consulting descriptive grammars. Some of the data points have been published as a part of WALS (Dryer, 2011e), but not all of them have been previously published. The complete table of data points can be found in Appendix. In this database, each language is tagged for the positions of bound and non-bound morphemes for the following 16 different grammatical categories:

(5) a. **Q (verbal):** Question particle on verbs.

b. **TA (verbal):** Tense and aspect markers on verbs.

c. **able (verbal):** Markers that denote ability on verbs.

d. **case (nominal):** Case markers on nouns.

e. **caus (verbal):** Causative markers on verbs.

f. **cop (nominal):** Copulas on nouns.

g. **def (nominal):** Definite markers on nouns.

h. **dem (nominal):** Demonstrative markers on nouns.

i. **indef (nominal):** Indefinite markers on nouns.

j. **neg (verbal):** Negation markers on verbs.

k. **obj (verbal):** Object pronominal affixes on verbs. For non-bound morphemes, the position of the full NP object was recorded, rather than pronouns. This must be

treated with caution because the positions of a full NP and a pronoun is not necessarily the same.

l. **pl (nominal):** Plural markers on nouns.

m. **poss (nominal):** Possessive markers on nouns. This is head-marking on the possessee, and should not be confused with genitive markers (which is dependent marking on the possessor).

n. **subj (verbal):** Subject pronominal markers on verbs. For non-bound morphemes, the position of the full NP subject was recorded, rather than pronouns. This must be treated with caution because the positions of a full NP and a pronoun is not necessarily the same.

o. **subord (clausal):** Adverbial subordinators on verbs.

p. **want (verbal):** Markers that denote desire on verbs.

We will call these 16 categories 'grammatical categories', and morphemes that instantiate those grammatical categories 'grammatical morphemes', although some of them may not be grammatical in the normal sense, as they include verbs for 'want', for example.

While our data include both bound and non-bound morphemes for those 16 grammatical categories, we focus on bound morphemes (i.e. affixes) in this section. We will discuss comparisons with non-bound equivalents in Section 3.2.2.

Table 3.3 summarizes the number of languages that use prefixes and suffixes, classified by grammatical categories.[2] Table 3.4 counts the number of *genera*, instead of *languages*, which is intended to prevent a group of languages that recently diverged and are highly similar to each other (such as Bantu and Oceanic) from being overrepresented. A *genus* is a set of languages which are historically related within about the same time depth, instead of a *language family*, whose time depth varies depending on data availability and research progress. See Dryer (1989) for more details of this concept. Note that a genus can be counted twice: a genus is regarded as having a property if there is at least one language in the genus that has the relevant property; thus the same genus can be counted multiple times if a genus is a mixture of different types of languages.

In Tables 3.3 and 3.4, we can see that more languages use prefixes for 15 out of 16 categories. The only category for which slightly more languages use prefixes is subject pronominal affixes in language-based counts (Table 3.3), while it is possessive affixes in genus-based counts (Table 3.4). We will discuss more in Chapter 6 about the differences among grammatical categories.

However, a serious problem of using these raw numbers is that there are not controlled for geography. Languages with a particular type may be concentrated in a single continent, which indicates that the figures may well be due to historical accidents, not a universal tendency. The databases cited in Hawkins and Gilligan (1988) try to avoid this pitfall by sampling unrelated languages as much as possible. In this thesis, we follow the method taken in Dryer (1989,

[2]Languages that use both prefixes and suffixes for the same grammatical category, or neither of them, are not included in the table.

Table 3.3: Position of grammatical morphemes classified by grammatical categories (language-based)

	prefixing languages	suffixing languages	% of prefixing languages
Q	8	**166**	4.6%
TA	180	**741**	19.5%
able	7	**19**	26.9%
case	43	**486**	8.1%
caus	133	**247**	35.0%
cop	1	**39**	2.5%
def	21	**75**	21.9%
dem	9	**28**	24.3%
indef	6	**26**	18.8%
neg	180	**231**	43.8%
obj	238	**316**	43.0%
pl	142	**583**	19.6%
poss	276	**389**	41.5%
subj	**438**	405	52.0%
subord	0	**68**	0.0%
want	5	**73**	6.4%

Table 3.4: Position of grammatical morphemes classified by boundedness and grammatical categories (genus-based)

	prefixing genera	suffixing genera	% of prefixing genera
Q	8	**97**	7.6%
TA	63	**310**	16.9%
able	6	**16**	27.3%
case	23	**207**	10.0%
caus	66	**132**	33.3%
cop	1	**33**	2.9%
def	13	**52**	20.0%
dem	8	**21**	27.6%
indef	6	**23**	20.7%
neg	63	**125**	33.5%
obj	112	**144**	43.8%
pl	46	**229**	16.7%
poss	**146**	126	53.7%
subj	170	**175**	49.3%
subord	0	**51**	0.0%
want	5	**55**	8.3%

1992a) based on large linguistic areas.

Dryer's method based on large linguistic areas is as follows. The world is divided into six large linguistic areas: Africa (Afr), Eurasia (Eur), Southeast Asia and Oceania (SEA&O), Australia-New Guinea (A-NG), North America (NA), and South America (SA). The numbers of languages of each linguistic area in our sample are summarized in Table 3.5:

Table 3.5: Total number of languages in our database

Afr	Eur	SEA&O	Aus-NG	NA	SA	Total
484	203	410	343	206	190	1,836

In order to see whether there is a universal tendency, we check whether the same asymmetry is observed in each of the six areas. For example, case suffixes are more common than case prefixes in all of the six language areas. This suggests that the suffixing preference of case affixes is in fact a universal tendency, not a historical accident. Following Dryer (1989, 1992a), we will use the simple non-parametric sign test to see whether the difference is significant. The probability of getting 6 heads out of 6 coin tosses is $1/2^6 \approx 1.56\%$, while the probability of getting at least 5 heads out of 6 coin tosses is 10.9%.[3] Thus, with the standard threshold of 5%, we can say that the difference is significant when case suffixes are more common than case prefixes in all of the six language areas. When one is more common than the other in five out of the six areas, then we cannot say that the result is statistically significant, but we will recognize that there is a tendency with some weaker standard in such a case.

[3]These figures will be doubled if we adopt two-sided tests, i.e. if we test the hypothesis that one is more common than the other, instead of the hypothesis that suffixes are more common.

We avoid the application of the simple T test or chi-square test to the numbers of languages or genera on purpose, as the data points are not independent from each other and thus they do not satisfy preconditions for those tests. There have been proposals of more sophisticated statistical tests of language universals, including Maslova (2000), (Bickel, 2008), (Cysouw, 2010) and Dunn, Greenhill, Levinson, and Gray (2011). The reasons why we do not adopt those advanced methods are the following. First, they are still experimental and there is no gold standard for establishing a statistical typological preference, and thus the premature use of fancy statistics could make our results more error-prone and difficult to reuse. Second, because our method is more conservative than most other approaches, the risk of obtaining false significant results would be small.

We will use the number of languages, instead of genera, for geography-controlled test of typological frequency biases hereafter. The reason of this decision is that results do not significantly change depending on whether we use languages or genera, as long as geography is controlled (Matthew Dryer, p.c.), and language is a more firmly established concept.

The results of applying this method to 16 grammatical categories are summarized in Table 3.6. The larger numbers are emphasized with bold for each pair. When there are six bold figures in a row for suffix, we can say that the suffixing preference is supported for that grammatical category, which is marked by an asterisk *. When there are five bold figures, we indicated them with (*).

In addition to subject pronominal affixes, we can see that some categories fail to show the suffixing preference in this more rigorous method: ‘able’, causative, demonstrative, negative,

object pronominal affixes, possessive, and 'want'.[4] However, 8 out of 16 categories still exhibit the suffixing preference, and none of them show the prefixing preference. This indicates that it is safe to say that there is an overall suffixing preference.

It must be noted that all the statistics above focus on inflectional morphemes. There appears to be no typological database that systematically counts the frequency of derivational affixes. Myler (2009a, 23) notes, however, that suffixes are more common in derivational morphology too, based on personal communications with Michael Cysouw and John Hawkins. We will briefly return to the issue of derivational morphology in Section 4.5.2.

3.2.1 Are the data reliable?

A potential problem in the use of the typological database is that whether a morpheme is *affix* or not is not determined consistently across across different grammar authors (Haspelmath, 2011b). As we argued in Section 2.3, *word* can be considered as a prototype category. While a prototypical word may be obvious in any language, there are moments when a grammar author needs to make subtler decisions about, for example, whether a morpheme is a clitic or affix. In such difficult cases, there is no single criterion that can be applied universally, and even if such a criterion were possible in principle, it is not practical to apply such a criterion to every language in our database. As we argued, we do not think this is a fatal problem for our approach. Even if there are cases where different authors might not agree on whether a given morpheme is an affix or clitic, it is still a remarkable fact that we systematically find

[4] As for causatives, prefixes outnumber suffixes only in Southeast Asia and Oceania. However, 57 out of the 70 languages that prepose causatives are Austronesian. Hence, we can say that causatives fail to show the significant result due to the single language family.

Table 3.6: Suffixing preference by large language areas

		Afr	Eur	SEA&O	A-NG	NA	SA	Total	
Q	Prefix	0	1	2	1	2	2	8	
	Suffix	**25**	**35**	**27**	**25**	**25**	**29**	166	*
TA	Prefix	80	1	48	13	36	2	180	
	Suffix	**92**	**143**	**83**	**206**	**102**	**115**	741	*
able	Prefix	**3**	0	1	1	1	1	7	
	Suffix	0	**6**	**8**	1	**2**	**2**	19	
case	Prefix	22	2	7	4	8	0	43	
	Suffix	**51**	**139**	**38**	**120**	**77**	**61**	486	*
caus	Prefix	14	9	**70**	10	22	8	133	
	Suffix	**45**	**63**	19	**26**	**56**	**38**	247	(*)
cop	Prefix	1	0	0	0	0	0	1	
	Suffix	**10**	**6**	**2**	**10**	**7**	**4**	39	*
def	Prefix	12	2	1	1	4	1	21	
	Suffix	**37**	**17**	**5**	**4**	**10**	**2**	75	*
dem	Prefix	0	**3**	1	**1**	4	0	9	
	Suffix	**20**	0	**3**	0	**5**	0	28	
indef	Prefix	4	1	0	0	1	0	6	
	Suffix	**10**	**4**	**3**	**3**	**4**	**2**	26	*
neg	Prefix	**78**	23	**46**	14	11	8	180	
	Suffix	39	**35**	26	**37**	**26**	**68**	231	
obj	Prefix	48	22	1	**65**	**68**	**34**	238	
	Suffix	**89**	22	**71**	61	40	33	316	
pl	Prefix	100	0	15	10	17	0	142	
	Suffix	**135**	**159**	**51**	**77**	**81**	**80**	583	*
poss	Prefix	11	8	22	**55**	**97**	**83**	276	
	Suffix	**137**	**55**	**109**	46	30	12	389	
subj	Prefix	**144**	12	**77**	74	**72**	**59**	438	
	Suffix	47	**124**	33	**112**	46	43	405	
subord	Prefix	0	0	0	0	0	0	0	
	Suffix	**5**	**12**	**12**	**9**	**13**	**17**	68	*
want	Prefix	0	0	0	0	4	1	5	
	Suffix	0	**6**	**7**	**9**	**25**	**26**	73	(*)

more suffixes than prefixes, which needs explanation.

However, the problem is that linguists' decision about wordhood could be systematically biased. In particular, linguists may have a tendency to regard postposed grammatical morphemes as suffixes, while preposed grammatical morphemes as independent words or clitics. We can think of a few reasons why this could happen. First, grammar descriptions may be under the influence of major European languages that most grammar authors are familiar with, which are (traditionally described as) suffixing languages. This bias could be present not only in distant languages, but in major European languages themselves. For example, the so-called pronominal clitics in Romance languages have been claimed to be better analyzed as affixes (see Spencer and Luís (2012a, 124) and the references therein), which leads us to suspect that grammar descriptions are often biased by traditions.

Second, linguists might want to avoid recognizing prefixes because they make the design of a dictionary more difficult. In most cases, people consult a dictionary in order to find a stem. If a language is prefixing, a dictionary user needs to split off prefixes first, which is not always easy when the user is unfamiliar with the language.[5]

While it is not practical here to examine data sources and determine whether they are not biased, we can point out indirect evidence that the suffixing preference captures the reality, rather than biases of linguists. There are considerable differences in the strength of the suffixing preference among grammatical categories, as shown in Tables 3.3, 3.4 and 3.6. For

[5]One could argue that this is indeed related to the reason why languages avoid prefixes. This idea is, in fact, similar to our approach we will uncover in the later chapters in that the identification of a prefix-stem boundary is a challenge for the parser. Our approach, however, claims that a prefix-stem boundary is difficult to identify due to the time course of language processing, rather than due to the lack of knowledge.

example, case marking has a clear suffixing preference, while subject pronominal affixes do not. Because both case and subject markers are suffixes in major European languages, this difference cannot be explained by the assumption that linguists are affected by the grammar of major European languages.

3.2.2 Overall frequency asymmetry or boundedness asymmetry?

There is an important distinction in the discussions of the suffixing preference. That is, the suffixing preference could be reduced to one of the following factors, or both:

(6) a. **Overall frequency asymmetry hypothesis:**

There are more postposed grammatical morphemes than preposed grammatical morphemes.

b. **boundedness asymmetry hypothesis:**

Postposed grammatical morphemes are more likely to be affixes than preposed grammatical morphemes are.

This point is made clear by Bybee et al. (1990) as well as Himmelmann (2014). Bybee et al. argue that both overall frequency asymmetry and boundedness asymmetry are present. Table 3.7 shows the frequency of grammatical morphemes in Bybee et al.'s database, classified by nonbound/bound and preposed/postposed. We can see that the overall frequency of preposed grammatical materials is lower than postposed grammatical materials (i.e. overall frequency asymmetry). At the same time, we can see that postposed grammatical morphemes

are more likely to be bound (i.e. boundedness asymmetry): while 426 (52.5%) out of 812 preposed grammatical morphemes (79.6%) are bound, 1,236 out of 1,552 postposed grammatical morphemes are bound.

Table 3.7: Relationship between the boundedness and position of grammatical morphemes from Bybee et al. (1990: 5) (simplified)

	Nonbound	Bound	All
Preposed	386	426	812
Postposed	316	1,236	1,552

However, our data does not support the overall frequency asymmetry. Table 3.8 shows the frequencies of affixes as well as non-bound grammatical morphemes with the equivalent meaning for each grammatical category. We can see that preposed grammatical morphemes are more frequent in 7 categories, while postposed grammatical morphemes are more frequent in 9 categories. While postposing is more common in slightly more grammatical categories, the margin is too small to be statistically meaningful.[6]

Table 3.9 tests the significance of the overall frequency asymmetry using language areas with the same method as in the last section. The symbol * indicates that postposing is more common, while # indicates that preposing is more common. When one is more frequent than the other in 5 out of 6 areas, we indicate it with (*) or (#). We can see that few categories have a clear tendency of preposing or postposing. Postposing is significantly more common in only two categories — question and plural. In one other category (subject), preposing is

[6]Due to data availability, subject and object pronominal affixes are compared to the positions of full NP subjects and objects respectively, rather than the positions of subject and object pronouns. There are languages in which full NP arguments and pronominal arguments occur in different places of a clause.

Another caveat is that sometimes the same language is counted twice for a single grammatical category. For example, when a language has both prepositions and case suffixes, it is counted in the rows of both 'preposed' and 'postposed' in the table. Such a language has no effect on the comparison of the totals.

Table 3.8: Overall grammatical morpheme frequencies (language-based)

	bound		non-bound		Total		
	prefix	suffix	pre	post	pre	post	% of pre
Q	8	166	157	357	165	**523**	24.0%
TA	180	741	322	242	502	**983**	33.8%
able	7	19	88	68	**95**	87	52.2%
case	43	486	564	634	607	**1,120**	35.1%
caus	133	247	49	36	182	**283**	39.1%
cop	1	39	195	221	196	**260**	43.0%
def	21	75	161	222	182	**297**	38.0%
dem	9	28	607	654	616	**682**	47.5%
indef	6	26	158	137	**164**	163	50.2%
neg	180	231	591	216	**771**	447	63.3%
obj	238	316	799	773	1,037	**1,089**	48.8%
pl	142	583	91	103	233	**686**	25.4%
poss	276	389	449	321	**725**	710	50.5%
subj	438	405	1357	192	**1,795**	597	75.0%
subord	0	68	432	107	**432**	175	71.2%
want	5	73	174	80	**179**	153	53.9%

significantly more common.

If there is no overall frequency asymmetry, then the suffixing preference must be due to the boundedness asymmetry. As expected, the boundedness asymmetry can be robustly observed in our data. Table 3.10 shows the proportions of languages that have bound grammatical morphemes. In the table, the percentage of languages with bound morphemes is calculated for each preposed/postposed grammatical categories; for example, 180 out of $180 + 322 = 502$ languages have bound preposed TA markers, which means that 35.9% of the languages with preposed TA markers have TA prefixes. We can see that postposed grammatical morphemes are more often bound in all of the 16 categories. Table 3.11 examines the boundedness asymmetry controlling for large linguistic areas. We can see that for 7 out of 16 categories, post-

Table 3.9: Overall frequency asymmetry by large language areas

		Afr	Eur	SEA&O	A-NG	NA	SA	Total	
Q	Preposed	30	25	33	22	37	18	165	
	Postposed	**156**	**63**	**132**	**81**	**48**	**43**	523	*
TA	Preposed	**209**	35	**134**	32	74	18	502	
	Postposed	120	**198**	133	**245**	**132**	**155**	983	
able	Preposed	**33**	11	33	4	**13**	1	95	
	Postposed	2	**20**	**49**	**6**	5	**5**	87	
case	Preposed	**234**	41	**205**	34	74	19	607	
	Postposed	136	**260**	141	**270**	**144**	**169**	1,120	
caus	Preposed	33	16	**85**	13	26	9	182	
	Postposed	**48**	**70**	35	**31**	**57**	**42**	283	(*)
cop	Preposed	**103**	23	39	7	17	7	196	
	Postposed	34	**56**	**53**	**43**	**41**	**33**	260	(*)
def	Preposed	27	**31**	46	22	**40**	**16**	182	
	Postposed	**149**	19	**64**	**32**	30	3	297	
dem	Preposed	59	**148**	110	95	**109**	**95**	616	
	Postposed	**299**	6	**213**	**110**	39	15	682	
indef	Preposed	24	**42**	**56**	8	**27**	**7**	164	
	Postposed	**58**	12	47	**29**	12	5	163	
neg	Preposed	**173**	**94**	**220**	**133**	**112**	39	771	(#)
	Postposed	133	46	73	73	35	**87**	447	
obj	Preposed	135	**164**	150	**295**	**146**	**147**	1,037	
	Postposed	**407**	62	**306**	105	131	78	1,089	
pl	Preposed	108	0	81	14	27	3	233	
	Postposed	**160**	**161**	**97**	**88**	**86**	**94**	686	*
poss	Preposed	80	**81**	140	**155**	**131**	**138**	725	
	Postposed	**323**	62	**176**	89	46	14	710	
subj	Preposed	**519**	**187**	**384**	**347**	**171**	**187**	1,795	#
	Postposed	75	129	95	122	110	66	597	
subord	Preposed	**140**	**60**	**130**	**30**	**58**	14	432	(#)
	Postposed	17	31	38	22	26	**41**	175	
want	Preposed	**53**	**22**	**64**	9	24	7	179	
	Postposed	7	17	28	**21**	**40**	**40**	153	

Table 3.10: boundedness asymmetries (language-based)

	bound		non-bound		% of bound	
	prefix	suffix	pre	post	pre	post
Q	8	166	157	357	4.8%	**31.7%**
TA	180	741	322	242	35.9%	**75.4%**
able	7	19	88	68	7.4%	**21.8%**
case	43	486	564	634	7.1%	**43.4%**
caus	133	247	49	36	73.1%	**87.3%**
cop	1	39	195	221	0.5%	**15.0%**
def	21	75	161	222	11.5%	**25.3%**
dem	9	28	607	654	1.5%	**4.1%**
indef	6	26	158	137	3.7%	**16.0%**
neg	180	231	591	216	23.3%	**51.7%**
obj	238	316	799	773	23.0%	**29.0%**
pl	142	583	91	103	60.9%	**85.0%**
poss	276	389	449	321	38.1%	**54.8%**
subj	438	405	1,357	192	24.4%	**67.8%**
subord	0	68	432	107	0.0%	**38.9%**
want	5	73	174	80	2.8%	**47.7%**

posed grammatical morphemes are more often bound, and there is no grammatical categories for which preposed grammatical morphemes are more often bound. The four exceptions that fail to show the boundedness asymmetry are ‘able’, demonstrative, object pronominal affixes and plural.

Some of the previous studies (for example, the acquisition-based accounts we will discuss in Section 3.4.5) are not aware of the distinction between the overall frequency asymmetry hypothesis and boundedness asymmetry hypothesis, and sometimes give an explanation only for the overall frequency asymmetry hypothesis. That is, some accounts of the suffixing preference try to explain why grammatical morphemes tend to be postposed. Our data suggests such approaches are misguided, because the suffixing preference arises not because there are

Table 3.11: Boundedness asymmetry by large language areas

		Afr	Eur	SEA&O	A-NG	NA	SA	Total	
Q	% bound (pre)	0.0%	4.0%	6.1%	4.5%	5.4%	11.1%	4.8%	
	% bound (post)	**16.0%**	**55.6%**	**20.5%**	**30.9%**	**52.1%**	**67.4%**	31.7%	*
TA	% bound (pre)	38.3%	2.9%	35.8%	40.6%	48.6%	11.1%	35.9%	
	% bound (post)	**76.7%**	**72.2%**	**62.4%**	**84.1%**	**77.3%**	**74.2%**	75.4%	*
able	% bound (pre)	**9.1%**	0.0%	3.0%	**25.0%**	7.7%	**100.0%**	7.4%	
	% bound (post)	0.0%	**30.0%**	**16.3%**	16.7%	**40.0%**	40.0%	21.8%	
case	% bound (pre)	9.4%	4.9%	3.4%	11.8%	10.8%	0.0%	7.1%	
	% bound (post)	**37.5%**	**53.5%**	**27.0%**	**44.4%**	**53.5%**	**36.1%**	43.4%	*
caus	% bound (pre)	42.4%	56.3%	**82.4%**	76.9%	84.6%	88.9%	73.1%	
	% bound (post)	**93.8%**	**90.0%**	54.3%	**83.9%**	**98.2%**	**90.5%**	87.3%	(*)
cop	% bound (pre)	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	
	% bound (post)	**29.4%**	**10.7%**	**3.8%**	**23.3%**	**17.1%**	**12.1%**	15.0%	*
def	% bound (pre)	**44.4%**	6.5%	2.2%	4.5%	10.0%	6.3%	11.5%	
	% bound (post)	24.8%	**89.5%**	**7.8%**	**12.5%**	**33.3%**	**66.7%**	25.3%	(*)
dem	% bound (pre)	**6.7%**	0.0%	**1.4%**	0.0%	**12.8%**	0.0%	1.5%	
	% bound (post)	0.0%	**2.0%**	0.9%	**1.1%**	3.7%	0.0%	4.1%	
indef	% bound (pre)	16.7%	2.4%	0.0%	0.0%	3.7%	0.0%	3.7%	
	% bound (post)	**17.2%**	**33.3%**	**6.4%**	**10.3%**	**33.3%**	**40.0%**	16.0%	*
neg	% bound (pre)	**45.1%**	24.5%	20.9%	10.5%	9.8%	20.5%	23.3%	
	% bound (post)	29.3%	**76.1%**	**35.6%**	**50.7%**	**74.3%**	**78.2%**	51.7%	(*)
obj	% bound (pre)	**35.6%**	13.4%	0.7%	22.0%	**46.6%**	23.1%	23.0%	
	% bound (post)	21.9%	**35.5%**	**23.2%**	**58.1%**	30.5%	**42.3%**	29.0%	
pl	% bound (pre)	**92.6%**	n/a	18.5%	71.4%	63.0%	0.0%	60.9%	
	% bound (post)	84.4%	98.8%	**52.6%**	**87.5%**	**94.2%**	**85.1%**	85.0%	
poss	% bound (pre)	13.8%	9.9%	15.7%	35.5%	**74.0%**	60.1%	38.1%	
	% bound (post)	**42.4%**	**88.7%**	**61.9%**	**51.7%**	65.2%	**85.7%**	54.8%	(*)
subj	% bound (pre)	27.7%	6.4%	20.1%	21.3%	**42.1%**	31.6%	24.4%	
	% bound (post)	**62.7%**	**96.1%**	**34.7%**	**91.8%**	41.8%	**65.2%**	67.8%	(*)
subord	% bound (pre)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
	% bound (post)	**29.4%**	**38.7%**	**31.6%**	**40.9%**	**50.0%**	**41.5%**	38.9%	*
want	% bound (pre)	0.0%	0.0%	0.0%	0.0%	16.7%	14.3%	2.8%	
	% bound (post)	0.0%	**35.3%**	**25.0%**	**42.9%**	**62.5%**	**65.0%**	47.7%	(*)

more postposed grammatical morphemes, but because postposed grammatical morphemes are more often bound.

3.3 Previous accounts

Many proposals have been made to explain why the suffixing preference exists, which can be classified into several threads of approaches:

(7) a. **Psycholinguistic accounts**

- **Comprehension-based accounts**

  Greenberg (1957); Cutler et al. (1985); Hawkins and Cutler (1988); Hall (1988, 1992)

- **Production-based accounts**

  Greenberg (1957); Cutler et al. (1985); Hawkins and Cutler (1988); Hall (1988, 1992); Himmelmann (2014)

- **Acquisition-based accounts**

  Greenberg (1957); Slobin (1973); Kuczaj (1979); St Clair, Monaghan, and Ramscar (2009); Hupp, Sloutsky, and Culicover (2009)

b. **Phonology-based accounts**

Greenberg (1957); Hall (1988, 1992)

c. **Historical accounts**

Givón (1979); Bybee et al. (1990)

d. **Formal accounts**

Myler (2009b); Biberauer, Holmberg, Roberts, and Sheehan (n.d.)

This classification is based on where the source of the asymmetry is claimed to be. Psycholinguistic accounts share the idea that the source of the asymmetry lies in language processing or acquisition. Namely, the proponents of psycholinguistic accounts believe that prefixes are more difficult or inefficient to produce, comprehend, or learn than suffixes are.

Historical approaches, on the other hand, share the idea that the suffixing preference can be reduced to the historical origins of affixes. These approaches rely on the hypothesis that most, if not all, affixes originate from an independent word, and thus the position of an affix reflects the word order in the past. We will call this idea the Fossilized Syntax Hypothesis, following Bybee et al. (1990). The asymmetries in word order could ultimately be reduced to psycholinguistic factors as well (Tomlin, 1986; Hawkins, 1994, 2004; Goldin-Meadow et al., 2008; Maurits et al., 2010; Gibson et al., 2013), but that is not our concern here.

By calling an approach psycholinguistic, we do not mean that it ignores the diachronic aspect. In fact, Hall (1988, 1992) explicitly argues for the need of incorporating the diachronic aspect, criticizing Cutler et al. (1985) who put aside the issue of how the psycholinguistic preference they propose can result in a typological pattern over time. Despite this, I classify his account as psycholinguistic because the asymmetry is still reduced to psycholinguistic factors that are at work in language change, not to their historical origins.

How can we evaluate different hypotheses? A successful account must reduce the suffixing preference into some known or independently motivated asymmetry. If one stipulates

another asymmetry on some invisible level only for explaining the suffixing preference, it could not be a valid scientific explanation. Previous accounts try to achieve this goal in a variety of ways. For example, it is reasonable to seek the source of asymmetry in psycholinguistics, because in on-line speech production and comprehension, there is an obvious left-right asymmetry: their temporal order. It is also reasonable to pursue a historical account, because we already know that there are asymmetries in word order. The formal approaches employ a UG principle (Kayne, 1994) which was proposed independently in syntax. Some acquisition-based accounts point out parallel asymmetries in non-linguistic domains.

In the later chapters of this thesis, we will discuss additional facts that a successful theory should be compatible with: (i) prefixes are not categorically ruled out, but they are in active use in many languages; (ii) each grammatical category behaves differently with regard to the suffixing preference; and (iii) there are other similar left-right asymmetries in phonology and morphology. We will argue that our theory is compatible with these major issues. However, no explanation, including the proposal in this thesis, can explain every fact that has been claimed to be related to the suffixing preference. Of course, some of the proposed parallelisms may be just a coincidence and have nothing to do with the suffixing preference. At the same time, however, it would not be surprising if more than one factor contributes to the suffixing preference.

In the next sections we will review each of the proposed accounts in greater detail.

3.4 Psycholinguistic accounts

3.4.1 Greenberg (1957)

An early attempt to explain the suffixing preference is found in Greenberg (1957). He suggests the following four possible explanations, which we tentatively call Hierarchy, Assimilation, Importance and Vividness, respectively.

(8) a. **Hierarchy**: People can more easily learn a pattern if more "diverse" stimuli are presented first (and stems are more diverse than affixes).

b. **Assimilation**: Because anticipatory assimilation is more common than progressive assimilation, prefixes are phonologically less stable than suffixes.

c. **Importance**: The listener prefers to hear stems first, which convey a larger amount of information.

d. **Vividness**: The speaker prefers to utter stems first, which convey more "vivid" information.

Among his proposals, (8b) is an argument from phonology. The rest of his proposals can be said to be psycholinguistic. More specifically, (8a) is based on acquisition, (8c) comprehension, and (8d) production. The idea of (8a) has recently gained attention again in St Clair et al. (2009); see Section 3.4.5 for discussions on acquisition-based accounts.

His second point (8b) has been criticized by Bybee et al. (1990). While it is true that there are more examples of anticipatory assimilation,[7] there is no evidence that prefixes are more often phonologically affected by the stem, as shown in Table 3.12.

Table 3.12: Bybee et al. (1990: 21)

	Affix affected by phonological processes conditioned by stem		Rate of effect
	yes	no	
prefixes	71	353	16.7%
suffixes	211	954	18.1%

Another issue of his phonological account (8b) is the logical relationship between the phonological unstability and the typological frequency, which we will return in Section 3.4.3.

His last two proposals, (8c) and (8d), must be said to be speculative. While it may be true that there are motivations to utter important and vivid information early (Givón, 1988), we can think of motivations to utter old information early as well. For speakers, it may be easier to utter given information first, because it is already activated in memory. To set the context first and let the listener prepare for less predictable information may be helpful for the understanding too. This also has an effect of distribute the amount of information evenly over the sentence (Jaeger, 2010). The topic-comment or old-new word order, i.e. the order in which vivid information is put off until a later part, has been considered to be common in the world's languages, although it is by no means universal (Mithun, 1987). In the English dative alternation, the constructions are chosen so that given information comes first (Bresnan,

[7]Empirical data that supports the claim that anticipatory assimilations are more common than progressive assimilation can be found in Javkin (1978, 129). He found that among 365 assimilation rules from 60 languages, 195 were anticipatory, while 89 were preseverative.

Cueni, Nikitina, & Baayen, 2007). When we can think of mutually contradictory motivations, it is risky to arbitrarily pick the one that happens to match the observation, as such an approach has no real predictive power (Hawkins, 2004, 13–14).

Furthermore, (8a,c,d) are all about the overall frequency asymmetry, not about the boundedness asymmetry. In other words, because these accounts only say that the postposing of grammatical morphemes is preferred in general, and do not mention the distinction between morphology and syntax, they do not explain why a grammatical morpheme is more often postposed in morphology than in syntax. While it would be possible that those factors have stronger effects on morphology than on syntax, such discussions are not provided.

3.4.2 Cutler et al. (1985), Hawkins and Cutler (1988), Hawkins and Gilligan (1988)

Cutler et al. (1985), Hawkins and Cutler (1988) and Hawkins and Gilligan (1988) provide what is probably the most well-known psycholinguistic accounts of the suffixing preference.[8] The essential part of their accounts can be summarized as follows:

(9) a. The human parser processes lexical information first, and syntactic information next

b. Thus the parser prefers to process a stem first, which conveys lexical information, and an affix next, which conveys syntactic information

c. Thus stem-suffix order is preferred over prefix-stem order

[8] These three papers are largely overlapping in their scope, and thus I will refer to Hawkins and Cutler (1988) as a representative of their approaches, when the difference among these papers do not matter.

Their crucial assumption is (9a). The idea behind this assumption is a bottom-up model of processing, where information flows from phonology to morphology to syntax.

Cutler et al. discuss two more factors that potentially work in favor of suffixes over prefixes, although they regard them as secondary compared to the main argument in (9). Both of the two additional factors are based on the observation that the word onset position is more salient than the rest of the word. They argue that the saliency of the beginning has the following two consequences. First, because the word onset position is more salient, there will be an additional motivation to put the stem at the word-initial position, which needs to be accessed first. Cutler et al. argue, however, that this additional factor cannot be a primary cause of the suffixing preference. They claim that prefix-stem order facilitates an earlier recognition of prefixes, just as stem-suffix order facilitates an earlier recognition of stems. Thus, in order to account for the advantage of the stem-suffix order, they argue, the computational priority of stems must be assumed anyway:

> Access to the affix information itself, on the other hand, is speeded in prefixed words as opposed to suffixed words. Given our assumption, based on the psycholinguistic evidence, that BOTH stems and affixes need to be processed, and processed SEPARATELY, early processing of either one should be equally useful UNLESS THERE IS A PREFERRED PROCESSING ORDER.
>
> Cutler et al. (1985, 751), emphasis in original

Second, because affixes are more predictable and convey less information than stems, the salient word onset position is better occupied by the stem. They, however, dismiss this idea

arguing that the predictability of affixes is not guaranteed.

A crucial problem of their approach is that assumption (9a) is speculative. While they review a wide variety of psycholinguistic papers, none of them directly supports this claim. It might turn out to be true that the lexical component of the language faculty has a preferred processing order, but without independent evidence it only puts off the question to another level. Even if we assume a purely bottom-up model of lexical processing, it does not necessarily mean that an early recognition of affixes should hamper processing.

They also discuss derivational affixes, but their argumentation appears to be overly simplistic. They argue in Hawkins and Cutler (1988, 306–307) that, for example, in order to understand the word *sadness*, the stem should have a computational priority over the affix, because the effect of the suffix *-ness* cannot be determined without knowing what the stem is. This argument, however, is too general, and it is again unclear why this factor is at work only in morphology. Unless it is more clearly formulated, it would predict that every abstract morpheme that does not make sense in isolation is preferably postposed, which does not seem to be the case.

The problem of their account is that they simply assume that the morphemes can be identified in the order of their presentation, no matter whether it is affix-stem or stem-affix order. However, as we will see in the following chapters, there are reasons to believe that affix-stem order and stem-affix order are not equally easy to understand, because stems and affixes are different in their frequency and length. We do not need to assume the computational priority of stems over affixes. Rather, the seeming priority of stems over affixes can be regarded *as a*

consequence of the processing difficulty associated with affix-stem order.

3.4.3 Hall (1988, 1992)

Hall basically accepts the psycholinguistic account of Hawkins and Cutler (1988). However, he criticizes the lack of the diachronic aspect in their account. His point is that a psycholinguistic account cannot be a complete story unless it specifies a diachronic path in which a psycholinguistic preference is grammaticalized (see Section 4.5 for discussions of this issue from our position). He proposes the following scenario:

(10) a. The word-final position more often undergoes phonological reduction than the word-initial position.

b. Affixes are more predictable than stems, and therefore convey less information.

Also, in his 1992 book, Hall attempts to relate his account of the suffixing preference to psycholinguistic models of word recognition. More specifically, he adopts the Cohort model (Marslen-Wilson, 1987), and argues that the parser has the monitors for lexical items that detect the match with an input. While suffixed forms can be handled by a single monitor, he argues, prefixed forms require multiple input monitors for the same stem, which adds the complexity for the parser.

A problem of Hall's account is that, along with Hawkins and Cutler (1988), it is simply hypothesized that preposed grammatical morphemes are more problematic for the parser in morphology than in syntax. In order to justify this assumption, Hall calls for the idea of *modularity* (Hall, 1992, 171–173), and speculates that the lexical processor, whose task is to

recognize a lexical item, prefers to process simpler words. It is not possible here to evaluate the idea of modularity in general, but without a solid method to empirically confirm what is going on in each module, it just leaves more room for ad hoc stipulations.

3.4.4 Himmelmann (2014)

Himmelmann (2014) provides a unique account of the suffixing preference based on language production and discourse (See also Section 6.3.1.5 for more discussions on Himmelmann's approach). His account goes as follows. Because functional words are more frequent and predictable, they are easier to retrieve from memory than content words. Thus, when dysfluency occurs in speech, it is likely to occur after a functional word, rather than before it. This makes a functional morpheme more tightly bound to the preceding word than to the following word, which results in a larger number of suffixes. He provides a survey of spoken data from English, German, Japanese and Tagalog, and demonstrates that dysfluencies occur more frequently after grammatical morphemes than before grammatical morphemes. The following is examples of dysfluency from English and Japanese. The English *the* is associated with the following noun (11a), while the Japanese *=no* is associated with the previous noun (11b). In both cases, dysfluency occurs *after* a grammatical word, irrespective of word order.

(11) a. And that's the end of the .. story.

b. ichioo zen- kokumin =no (0.7) joohoo =o nigitteru,
by and large all citizen GEN information OBJ have
'By and large, (they) have information about all the citizens,'

(5) and (6) of Himmelmann (2014, 935)

He further argues that there is a motivation for the speaker to produce a functional word even when the speaker has not fully constructed the subsequent phrase, from the interactional point of view: it can signal the speaker's intention to continue the utterance, and prevent the turn from being taken by the other speaker.

While we do not deny the possibility that his account plays a role in explaining the suffixing preference, we will argue in Chapter 5 that existing prefixes are designed for efficient parse in order to counterbalance their processing difficulty. This is not predicted by Himmelmann's approach.

3.4.5 Acquisition-based accounts

There have been two major threads of acquisition-based explanations for the suffixing preference: those based on the preferred stimulus order in associative learning (Greenberg, 1957; St Clair et al., 2009; Hupp et al., 2009; Ramscar, Dye, & Klein, 2013), and those based on the relative saliency of the word-final position in acquisition (Slobin, 1973; Kuczaj, 1979). See also E. V. Clark (2001) for a review of the literature on morphology acquisition and its relevance to typology.

3.4.5.1 Slobin (1973), Kuczaj (1979)

Based on acquisition data for a variety of languages, Slobin (1973) claims that preposed grammatical morphemes are more difficult to acquire than postposed grammatical morphemes. He points out, for example, that the locative suffix in Hungarian is acquired earlier than the locative prepositions in Serbo-Croatian, both by bilingual children and by monolingual children

of each language. Citing several studies that indicate children's sensitivity to word-final positions, he proposes the following principle:

> Operation Principle A: Pay attention to the ends of words.
>
> (Slobin, 1973, 191)

Kuczaj (1979) provides experimental support for Slobin's principle. He conducted four experiments where children learned words in an artificial language, and found that children more successfully acquired suffixes.

Slobin's claim apparently contradicts with Hawkins and Cutler (1988) simultaneously in two ways. First, while Hawkins and Cutler claim that the the word-initial position is the most salient part, Slobin claims that the word-final position is the most salient part. Second, Hawkins and Cutler claim that putting the stem in the most salient part facilitates processing, while Slobin claims that putting the affix in the most salient part facilitates acquisition. Because two contradictions cancel each other out, both authors predict that suffixes are preferred over prefixes.

Strictly speaking, because Slobin focuses on acquisition and Hawkins and Cutler do not, these two accounts may not be logically incompatible. However, unless a more explicit model of acquisition is presented, the relationship between positional saliency and ease of acquisition is unclear at best. Obviously, the acquisition of affixes is not the only problem children need to solve. Children need to acquire stems too, and the identification of the stem and affix should complement each other. Therefore it is not clear why the most salient part of a word is better occupied by an affix.

3.4.5.2 Greenberg (1957), St. Clair et al. (2009), Hupp et al. (2009)

One of the explanations proposed in Greenberg (1957) is based on acquisition ((8a) in Section 3.4.1, which is shown here again for convenience):

(12) **Hierarchy**: People can more easily learn a pattern if more "diverse" stimuli are presented first (and stems are more diverse than affixes)

His argument is based on the literature on associative learning (Osgood, 1949). The idea is that in stimulus-response (S-R) associative learning, so-called *convergent hierarchy* (a diverse stimulus followed by an identical response) is more quickly learned than *divergent hierarchy* (a unique stimulus followed by varying response). Because stems are more diverse than affixes, the acquisition of stem-affix combinations is facilitated when stems come first (convergent hierarchy).

Recently, this proposal has gained attention again (St Clair et al., 2009; Hupp et al., 2009). St Clair et al. (2009) demonstrated that people learned suffixes more successfully than prefixes in an artificial language experiment. Hupp and his colleagues conducted experiments on the learning of visual and music stimuli, in addition to linguistic stimuli. Based on the results, they claim that the suffixing preference is not language-specific, but domain-general.[9]

[9] See also Ramscar et al. (2013), who further develop the idea of St Clair et al. (2009), and claim that prefixes have their own benefits. They argue that prefixes help reduce uncertainties in language processing, while suffixes facilitate the abstraction of categories. At this point, the implication of their claim for typology is not clear.

3.4.5.3 Discussion

We discussed acquisition-based accounts as if they were incompatible with processing-based approaches. However, processing-based accounts such as Hawkins and Cutler (1988) can also be associated with difficulty in acquisition. This is because, if prefixes are difficult to produce/comprehend, it is probable that they are difficult to learn as well, although this is not a logical necessity because acquiring unknown morphemes is not the same task as processing known morphemes. Indeed, if one takes the view that the diachronic grammar change occurs as a result of 'incomplete' acquisition by a new generation, as often assumed in the generative literature, then one possibility is to assume that the processing factor is at work in language acquisition. Kirby's (1996) model of language change is such an approach. It makes references to Hawkins' processing-based theory of word order, but the locus of language change is assumed to be at acquisition.[10]

Some general notes on acquisition-based approaches are in order. First, all of the acquisition-based accounts seem to try to explain the overall frequency asymmetry, not the boundedness asymmetry. In other words, they try to explain why grammatical morphemes are preferably postposed, not why postposed grammatical morphemes are more likely to be an affix. This is serious drawback because, as we saw in Section 3.2, the true cause of the suffixing preference is not the former, but the latter. Second, the causal relationship between the relative acquisition difficulty and the suffixing preference has not been clarified. If every child eventually

[10]Not every scholar agrees with the idea that diachronic grammar change happens when a new generation 'fails' to acquire exactly the same grammar as the previous generations. In the usage-based perspective, even the adult grammar is flexible to some extent, and can lead language change. This position is defended in Chapter 6 of Bybee (2010), among others.

acquires prefixes, no matter how late, then there should be no consequence for the typological pattern. Third, as noted by the authors themselves, the participants of the artificial language experiments were English speakers, both in Kuczaj (1979) and St Clair et al. (2009). Thus, there is a risk that they found suffixes easier due to transfer from their native language. Interestingly, Kuczaj (1979) reports that children more easily acquired a preposed grammatical morpheme when it was of the future tense, which is enough to make us suspect that influence from English were at work in their experiments.

Despite these weaknesses, acquisition-based approaches should not be easily dismissed. The parallelism found in other cognitive domains is a fascinating finding, for which the other approaches have no explanation at this point. More future work will be needed to determine whether their experimental evidence of a domain-general asymmetry is what is behind the suffixing preference.

3.5 Phonology-based accounts

In the Optimality Theory framework, the fact that the initial position resists sound change is formalized as *positional faithfulness* or *positional markedness* (Beckman 1998, Smith 2002, among others). We can think of functional reasons why the final position is more vulnerable to sound change, both from the speaker's and listener's points of view. The speaker may be more likely to run out of energy for the later part of a phonological word; for the listener, the end of a phonological word may be more predictable and redundant, and thus there is less need for it to be accurately perceived. The latter is a part of Hall's (1988; 1992) account.

While Hall's account involves phonology, we can say that it is a part of his psycholinguistic account, rather than a purely phonological account.

It is not obvious, however, why the weakness of the end should result in more suffixes. If we believe that the end of a content word is likely to undergo sound change so that it is phonologically fused with the following grammatical word, then the question is why a phonological reduction occurs at the end of a content word, rather than at the end of a grammatical word. If the end of a grammatical word is reduced just like the end of a content word, then it should cause a phonological fusion with the next content word, and thus the suffixing preference cannot result. If we believe that a phonological reduction presupposes semantic redundancy as claimed by Hall, then we have to say that the last part of a content word is more predictable than the last part of a grammatical word, but this is not obvious.

Alternatively, if we believe that a phonological reduction occurs at the postposed grammatical morpheme which is at the end of a phonological unit, then the argument here presupposes that postposed grammatical morphemes form a phonological unit with the preceding stem from the beginning, which appears to be circular. This problem can probably be overcome by assuming that phonological word boundaries themselves are motivated by processing ease. There is little need to put a phonological word boundary between a stem and a grammatical morpheme that follows it. This is predicted by our claim in the following chapters, but it is not explicit in the previous literature.

Another problem is that the weakness of the end appears to cause not only the creation of a suffix, but also the *loss* of a suffix. In other words, we may expect that grammaticaliza-

tion is faster at the word-final position, but this does not logically entail that suffixes should outnumber prefixes typologically, unless one provides evidence that the creation of a suffixes outpaces the loss of suffixes, for example. In fact, Greenberg (1957, 93) mentions that suffixes are often reduced or lost, because they tend to become superfluous. French, for example, lost its inflectional endings in its history due to the phonological change at the end. Examples like this suggest that the phonological vulnerability of the end should sometimes predict fewer suffixes. Quite confusingly, Greenberg also uses phonological instability to explain the rarity of prefixes. As we saw in Section 3.4.1, he argues that prefixes are less common *because* they are more phonologically unstable due to the prevalence of anticipatory assimilation.

Despite these unsolved problems, once clearly formulated, the phonological account might be eventually compatible with our approach we will uncover later. Morphologization involves phonological change in any case, and thus a psycholinguistic account ultimately needs to incorporate a model of sound change.

3.6 Historical accounts

Historical accounts of the suffixing preference are based on the idea that many, if not all, affixes are historically derived from independent words, as a result of morphologization.[11] This is exemplified by the development of the Latin verb *habeo* 'have' into the future tense

[11] In this thesis, we will call this diachronic process *morphologization*, following Lehmann (1982) and Joseph (2004). The same concept is called *affixation* in Heine and Reh (1984), while Haspelmath (2011a) calls a similar concept *coalescence*. We take *grammaticalization* as a more general term, which subsumes formal as well as semantic changes, which morphologization can be a part of. Note that the term *morphologization* is not used here in the sense of Joseph and Janda (1988), where the term refers to a process in which a previously phonologically conditioned variation becomes morphological distinctions, a famous example being the development of a certain type of vowel alternation in Germanic languages, such as *foot/feet*.

suffix in French.

(13) *cantare habeo* (Latin) > *chanterai* (French)

The crucial point is that in the process of morphologization, a morpheme keeps its position. This idea is called the Fossilized Syntax Hypothesis (Bybee et al., 1990), and is also represented by Givón's (1979) slogan 'today's morphology is yesterday's syntax.' Because we know that there are asymmetries in word order universals, the suffixing preference can potentially explained without hypothesizing further asymmetries. Givón (1979) and Bybee et al. (1990) are two major pieces of work in this approach.

3.6.1 Givón (1979)

Givón's (1979) explanation can be summarized as follows.

(14) a. Most affixes are historically derived from independent words, and thus their positions reflect the word order in the past. (Fossilized Syntax Hypothesis)

b. Many languages had SOV word order in the past.

c. A grammatical morpheme tends to be postposed in an SOV language.

d. Thus, there are more suffixes today.

In his approach, the source of the asymmety is in the word order pattern in the past: the OV order used to be more common than it is today. While Hawkins and Cutler (1988, 310) as well as Hall (1992, 93-95) cast doubt on Givón's assumption about the dominance of OV

word order in the past, it has been suggested in numerous places that word order change from SOV to SVO is much more common than the other way around (Vennemann, 1973; Tai, 1976; Kiparsky, 1996). While the numbers of VO and OV languages are now similar, the number of VO languages may be inflated by the recent expansions of Niger-Congo and Austronesian languages (Dryer, 1989). Smaller language families more often have OV order, which can be more parsimoniously explained by hypothesizing that a few large language families switched to the VO order than hypothesizing the changes occurred in the other direction in many small language families (Dryer, 2012). The hypothesis that every language can be traced back to the SOV order has been claimed by some scholars as well (Newmeyer, 2000; Gell-mann & Ruhlen, 2011). It has also been suggested that there is something basic in the SOV order; for example, Goldin-Meadow et al. (2008) report that people use the SOV order when they are forced to use gestures to communicate.

We will not evaluate the hypothesis about word order frequency patterns in the past. It is still possible, however, to point out other evidence against the claim that the suffixing preference can be explained by the Fossilized Syntax Hypothesis. Bybee et al. (1990, 14–15) present an interesting counterargument against the proto-language approach. Their reasoning is that, if Givón's argument were correct, suffixes would be on average older in VO languages than OV languages, because suffixes in VO languages must be remnants of the older grammar, while in OV languages new suffixes are still emerging. They argue that this prediction is not borne out, based on their measure of the degree of grammaticalization.

Another problem with Givón's account is that not every grammatical category is postposed in OV languages. There are some grammatical categories that do show the boundedness asymmetry, but do not have a tendency to be postposed in the OV languages. We will return to this issue in Section 6.4.3. Givón's account also gives no explanation for phonological asymmetries we will discuss in Chapter 7.

3.6.2 Bybee et al. (1990)

Bybee et al. (1990) present critical evaluations of previous proposals, including Greenberg (1957); Hawkins and Cutler (1988) and Hall (1988), and propose an alternative approach. Their approach basically reduces the suffixing preference to a frequency asymmetry in the basic word order. Unlike Givón's account, however, Bybee et al. do not need the assumption that the frequency distribution of the basic word order was different in the past.

The crucial part of their explanation can be summarized as follows.

(15) a. A grammatical morpheme is most likely to become a part of verbal morphology when it is at the edge of a clause. When a grammatical morpheme is in the middle of a clause, whether it is fused into the verb depends on its semantic relevance to the verb (Bybee, 1985).[12]

b. In verb-final languages grammatical morphemes tend to be at the end of the clause, while in verb-initial languages they tend to be at the beginning of the clause. There are more verb-final languages than verb-initial languages. This results in the overall

[12]It must be noted that they only discuss verbal affixes, not nominal affixes.

higher frequency of suffixes.

In their data, preposed grammatical morphemes are more often bound than postposed grammatical morphemes are in verb-initial languages (Bybee et al., 1990, 6). If this is indeed the case, it is particularly convincing evidence for their claim: The 'clause-edge' factor can neatly explain why preposed morphemes tend to be bound in verb-initial languages, while all other accounts seem unable to provide an alternative explanation. Dryer (1992b), however, argues that Bybee et al.'s result is an artifact of their relatively small sample, which overrepresents Austronesian and Meso-American languages.

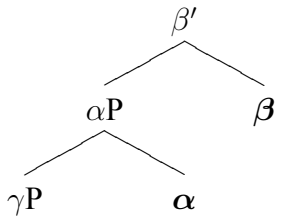
Furthermore, even if the 'caluse-edge' factor is in fact at work, it cannot be the whole story because of the following reasons. First, as they point out, suffixes are more frequent in verb-medial languages too. Second, they also mention that verb-final languages tend to be consistently suffixing, while not many verb-initial languages are consistently prefixing. Third, they do not discuss nominal affixes, such as case markers and plurals, which also show the suffixing preference. These facts suggest that there is still a general preference for suffixing, even if the factor they point out may further contribute to the higher frequency of suffixes in some cases.

3.7 Formal accounts

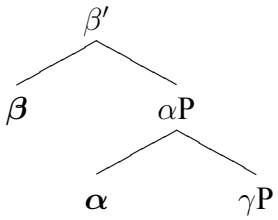
Recently there have been attempts to account for the suffixing preference by extending the Final-over-Final Constraint (FOFC) to the realm of morphology (Myler, 2009b, 2009a; Biberauer et al., n.d.). The FOFC is a formal constraint that prohibits a structure where the left

constituent of a right-headed branching is left-headed within a single extended projection. In other words, when α and β are within the same extended projection, (16a–c) are possible, but (16d) are forbidden. FOFC itself is claimed to be in turn derivable from more general principles of the Linear Correspondence Axiom (Kayne, 1994) and Relativized Minimality (Rizzi, 1990).

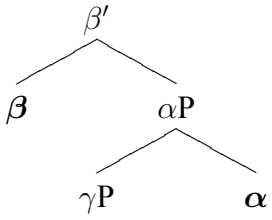
(16) a.



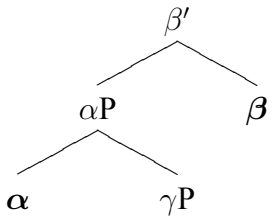
b.



c.



d. *



(Biberauer, Holmberg, & Roberts, 2014, 171)

Myler (2009b) attempts to extend this principle to morphology. FOFC predicts that prefixes are banned in certain conditions, due to the following reason. When a phrase level that

is immediately above the word level is head-final, then the outermost head of the word cannot be a prefix, because it corresponds to the situation (16d).[13] He examines Julien's (2002) database of word order and inflectional morphology consisting of 530 languages, and claims that the violation of FOFC is not attested.

It is difficult to statistically evaluate the performance of the formal approach here, because the predictions highly depend on the minimalist syntax analysis of each example. However, Myler (2009b, 68) admits that FOFC cannot be the whole story in explaining the suffixing preference. According to him, FOFC cannot predict the suffixing preference of some grammatical categories such as case and gender, as they are not considered as heads. FOFC does not work for derivational morphology either. In this thesis, we do not further investigate the formal approach, but we must note the possibility of formal constraints playing a role in explaining the suffixing preference.

3.8 Conclusion

We reviewed a wide variety of approaches to explain the suffixing preference. We saw that many of the accounts have their own strengths, and it would not be surprising if more than one factor is in fact playing a role in the overall distribution of affixes.

We also pointed out some common pitfalls in functional explanations. When we can think of more than one counteracting factors, we should not arbitrarily pick one of them that happens to match the observation we want to explain. Such an approach can 'explain' everything,

[13] Our account here involves simplification because there is an additional precondition that requires α and β to be within the same extended projection.

making it vacuous as a scientific theory.

Examples of counteracting motivations are as follows. First, Greenberg (1957) puts forward the idea that vivid information must be uttered first. However, it is also sometimes argued that given information is uttered first. Language probably has both motivations, and thus it is not explained why the former motivation should win over the latter motivation. Second, Greenberg claims that prefixes are fewer *because* prefixes are more vulnerable for sound change, but Hall (1988, 1992) argues that suffixes because the phonologically weaker (which should mean that suffixes are more vulnerable for sound change). Again, a clearly formulated theory will be needed to predict whether the phonological instability at a position leads to more or fewer affixes at that position. Third, Hawkins and Cutler (1988) claim that the saliency of the word-initial position, while Slobin claims the opposite; at the same time, Hawkins and Cutler claim that the salient position must be occupied by the stem, while Slobin claims that the salient position must be occupied by the affix. As we discussed, an explicit model of acquisition or processing is necessary to define what saliency means, and explain why it is better to locate a stem or an affix at the salient position.

The problem shared by these accounts is that the effect of a functional factor is only stated informally. In the next chapter, we will present a novel approach based on an explicit computational model of language processing.

Chapter 4

Suffixing Preference: A Probabilistic Account

4.1 Introduction

This chapter presents our new account of the suffixing preference based on language comprehension. Our central claim is that prefixes are difficult to immediately identify because of their shortness and relative position. This prediction is demonstrated by a simulation study of English and a hypothetical language, Reverse English.

This chapter is structured as follows. Section 4.2 reviews psycholinguistic studies of morphology and argues for the probabilistic approach to morphological processing. Section 4.3 argues that a probabilistic model of morphological processing predicts that the parser will have difficulty with prefixes. Section 4.4 proposes a computer programming experiment that can quantitatively evaluate the proposals in the previous section. Section 4.4.4 applies this simulation model to English data. The results that are solely based on English, however, may reflect not a language universal but the peculiarity of English. To overcome this issue, Section 4.4.5 presents the same simulation study in a hypothetical language, Reverse English. Section

4.5 discusses some additional issues including the so-called the 'problem of linkage', i.e. how a psycholinguistic effect can be related to typological frequencies. Section 4.6 concludes the chapter.

4.2 Background

This section summarizes psycholinguistic studies that motivate our approach to the suffixing preference.

4.2.1 Probabilistic nature of language processing

Recently, probabilistic models have seen successful applications in many different areas of cognitive science (M. D. Lee & Wagenmakers, 2014). Probabilistic models have also been applied to language in particular (Bod, Hay, & Jannedy, 2003; Chater & Manning, 2006). Levy (2008), for example, proposes a model in which processing difficulty in sentence comprehension is determined by the negative log probability of each word, or what he calls *surprisal*. Frank, Goldwater, Griffiths, and Tenenbaum (2010) model word segmentation in language acquisition using a Bayesian framework.

In morphological processing, MATCHECK (Baayen, Schreuder, & Sproat, 2000; Baayen & Schreuder, 2000) is an example of a model of visual morphological segmentation that uses probability. Balling and Baayen (2012) present a model of word recognition that takes morphologically complex words into account, by introducing the notion of surprisal to the phoneme level.

4.2.2 The incremental nature of word recognition

Our psycholinguistic model will assume that lexical access starts before the parser finishes listening to a word. This highly incremental nature of auditory word recognition is a relatively uncontroversial assumption. It has been demonstrated in a number of experimental paradigms, such as gating tasks and cross-modal priming, which was the very motivation for the Cohort model (Marslen-Wilson, 1987), one of the early successful models of word recognition. The Cohort model is used to account for the suffixing preference in Hall (1992), as we saw in 3.4.3.

Cohort has been criticized for the lack of consistent ways to deal with noisy input or word recognition in a larger speech stream, which could be overcome by connectionist models such as TRACE (Mcclelland & Elman, 1986) and Shortlist (Norris, 1994), as well as probabilistic models such as Shortlist B (Norris & McQueen, 2008). Our approach will be similar to Shortlist B in spirit, although our model segments speech into morphemes rather than words, unlike the original model that does not deal with morphology.[1]

4.2.3 Processing of a morphologically complex word

There is a huge amount of literature on morphological processing. Here we will discuss papers that are relevant to the differences between prefixes and suffixes. Unfortunately, the large majority of experimental work is conducted on predominantly suffixing languages, such as English. Due to this, most psycholinguistic studies that compare prefixes and suffixes are on

[1] See Balling and Baayen (2012) for a recent probabilistic approach to morphological processing, who point out the lack of morphological processing in Shortlist B.

derivational affixes.

A long-standing issue in morphological processing is whether a complex word is decomposed into parts, or is retrieved from memory as a whole. Two standard techniques to investigate this question are priming and the frequency effect. A priming effect that is only relevant to a part of a word can be used to investigate whether the word is decomposed or not. Another technique utilizes frequency effects. It is well known that frequency is a reliable predictor of reaction time in psychological experiments. Thus, we can investigate whether access time is better predicted by the frequency of the whole word or by the cumulative frequency of its parts.

Some claim that every derived word is stored in the lexicon, such as Butterworth's (1983) Full Listing Hypothesis. Others claim that prefixes are obligatorily decomposed before lexical access, such as in the Prefix Stripping model (Taft & Forster, 1975), one of the earliest explicit models of morphological processing. However, dual-route models, in which both a whole access and decomposition play roles (Caramazza, Laudanna, & Romani, 1988; Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 1995; Baayen, Dijkstra, & Schreuder, 1997; Hay, 2001, 2003), have recently gained more support (McQueen & Cutler, 2001).[2]

One of the earliest explicit models of morphological processing is the Prefix Stripping model (Taft & Forster, 1975). In their model, it is claimed that prefixes must be split off for every word prior to lexical access. The prefix stripping model has received a number of

[2]Those dual-route models vary in how the two routes divide the task of lexical recognition. In Caramazza et al. (1988), the whole-word route has the default and the decomposition route works as a back-up; in Frauenfelder and Schreuder (1992) and Hay (2001), the two routes compete and the faster route wins; in Schreuder and Baayen (1995) and Baayen et al. (1997) both routes simultaneously contribute to the interpretation.

criticisms (Tyler, Marslen-Wilson, Rentoul, and Hanney (1988) and Schriefers, Zwisterlood, and Roelofs (1991), among others). Cole, Beauvillian, and Segui's (1989) work is particularly interesting for our discussion, because they found an asymmetry between prefixes and suffixes and mention its potential relevance to the suffixing preference.

4.2.3.1 Cole et al. (1989)

Cole, Beauvillain, and Segui (1989) report experiments on morphological processing of English. Although their goal was not to explain typological preference, their discussion is worth reviewing here because they argue that their results provide evidence for the computational priority of stems over affixes.

In their lexical decision task experiment, they found the effect of root frequency on the reaction time for suffixed words, but not for prefixed words. Their results indicate that suffixed words are accessed via their root, while prefixed words tend to be accessed as a whole. According to them, this is because the lexicon is organized by stems. In a suffixed word, the recognition of the root activates its morphological family, and then the correct candidate will be selected from it. In a prefixed word, on the other hand, the information carried by the root is not available, because the processing of the root does not precede the processing of the whole word form.

But why should the lexicon be organized by stems, rather than affixes? Why should the parser not recognize the affix first and the stem next? One can attribute this asymmetry to a design feature of the human parser, but such an assumption would be unnecessary if we have an independent reason to believe that the processing of affix-stem order is difficult. We will

argue that this is indeed the case.

4.2.4 The issue of segmentation

As a necessary step of language comprehension, the parser must solve the non-trivial problem of speech segmentation. Namely, the parser needs to identify what morphemes are used in a message. A couple of studies suggest that parsing prefixes is a challenge for the parser, which provides a key insight for our approach.

4.2.4.1 Arguments from statistics on pseudo-affixes

Schreuder and Baayen (1994) argue against the idea of obligatory affix stripping (Taft & Forster, 1975), claiming that obligatory affix stripping is *inefficient*. Their argument is based on the abundance of pseudo-affixes in corpora. A pseudo-affix refers to a part of a word that happens to have the same form as an existing affix. Examples of words with pseudo-prefixes are *religion* and *relish*, and examples of pseudo-suffixes are *morning* and *herring*.[3] According to Schreuder and Baayen, pseudo-affixes are very common, especially in terms of token frequency. For the English prefixes they investigated, 81% of orthographical word forms and 83% of phonological word forms have pseudo-affixes. This means that obligatory affix stripping will end up with a garden path more often than successful parsing.

Similarly, Laudanna and Burani (1995) argue that their experimental results on the morphological processing of Italian can be explained in probabilistic terms. They found that the higher the proportion of truly prefixed words (as opposed to pseudo-prefixed words) is, the

[3] Some examples of pseudo-affixes are historically derived from affixes, but they are no longer analyzable in the synchronic point of view.

slower the reaction time for non-words with that prefix is. This finding indicates, according to them, that people are more likely to parse a word when the initial part of a word is more like a true affix; in the case of non-words, it leads the parser to an unsuccessful attempt to parse it and results in a slower reaction time in lexical decision.

What these studies suggest is that the parser needs to solve the non-trivial problem of *morpheme segmentation*, which has been overlooked in previous psycholinguistic approaches to the suffixing preference. The discussions in Greenberg (1957), Hawkins and Cutler (1988) and Hall (1992), as well as those of acquisition-based approaches, presuppose that morphemes are recognized in the order of their presentation. This assumption does not hold if a morpheme cannot be identified until it processes the subsequent part for disambiguation. The abundance of pseudo-affixation means that such a situation is very common.

4.2.4.2 Hay (2003)

Hay (2003) adopts a dual-route model of morphological processing and discusses how frequency, phonotactics, and other factors interact with it. In her dual-route model, the whole-word route and the decomposition route compete in lexical processing, and the faster route will win. For example, in order to process the word *insane*, the parser can either retrieve the whole word *insane* from memory, or decompose it into the prefix *in-* and the stem *sane*, access each item, and compositionally calculate the meaning. Whether the whole-word or decomposition route is faster involves a variety of factors, including semantic transparency, relative frequency, and phonotactics.

In Chapter 3 of her book, she investigates whether there is a correlation between prob-

abilistic phonotactics and 'prefixedness/suffixedness' (a measure of intuitive judgments of morphological complexity by native speakers proposed in Wurm (1997)), as well as semantic transparency. She found a significant correlation for prefixes, but not for suffixes. She discusses in Hay (2003, 63,67) why this asymmetry emerges, and suggests an account based on the amount of available information, which is roughly the same as our proposal. Due to the temporal order of language processing, the parser has more information when reaching a stem-suffix boundary than a prefix-stem boundary. Thus, she argues, phonotactics plays a smaller role for suffixed words in facilitating the decomposition. Here, the crucial fact is that a stem is longer than an affix on average, but this is not explicitly stated, and she does not further explore the consequence of this argument. We believe that this factor plays an important role in explaining the suffixing preference, which we turn to in the next section.

4.3 Why prefixes are difficult

Based on the discussions in the previous sections, this chapter presents our explanation for the suffixing preference. Our central claim is that prefixes are difficult to immediately identify, primarily because there is not enough information to disambiguate when the parser has only heard a small portion of a word.[4] This means that the parser needs to suspend the analysis and wait for more information, or commit to the wrong analysis and fix it later. Psycholinguists have firm evidence that these situations cause processing difficulties.

We argue that there are three factors that facilitate certain types of morphemes, but not

[4]In information theory, a coding system in which a meaningful unit can be identified without waiting for more information to disambiguate is called *prefix code* and is one of the desirable properties of a system.

prefixes: length, predictability, and phonological cues. These three factors and their consequences can be summarized as follows.

(17) a. **Length:** The end of a short morpheme is hard to detect, because a short morpheme is more likely to happen to match a part of other morphemes. This explains why the boundary after a prefix is harder to detect than the boundary after a *stem*.

b. **Predictability:** A subsequent morpheme is more predictable within a word than across a word boundary. This explains why the boundary after a prefix is harder to detect than the boundary after a *suffix*, because a suffix always has the preceding morpheme within a word.

c. **Phonological cues at word boundaries:** There are more phonotactic, suprasegmental, and phonetic cues for a word boundary than for a word-internal morpheme boundary. This makes the last morpheme of a word easier to identify, but a prefix is, by definition, never at the end of a word.

It must be noted that all of the three factors above are needed to explain the difficulty with prefixes. Each factor plays only a partial role. For example, length (17a) can explain that an affix is more difficult than a stem, but it does not predict that prefixes in particular are difficult compared to other grammatical morphemes. Predictability (17b) predicts that a morpheme at the beginning of a word is more difficult. This factor can explain why prefixes are harder than suffixes, but it does not explain why preposed grammatical words are not as problematic as prefixes. (17c) predicts that a preposed grammatical word is easier than a prefix, but it does

not distinguish prefixes from stems or suffixes that are not at the end of a word. Only the combination of those three factors predict that a prefix is the worst choice.

Another caveat is that what is relevant is whether the *end* of a morpheme is easy to identify, not the *beginning*. While whether the beginning of a morpheme is easy to identify or not is not completely irrelevant, it is not directly related to the ease of the processing of that morpheme. The parser's need is to identify the morpheme, and in most cases, the moment when the morpheme is identified is the moment when the end of the morpheme is detected, although, in principle, there is a small possibility that the end of the morpheme is identified earlier than the beginning of the morpheme.

Our claim is that the end of a prefix is difficult to identify. This is equivalent to saying that prefix-stem or prefix-prefix boundaries are difficult to identify than other types of boundaries. Note that the beginning of a prefix is not particularly difficult to identify, because phonological cues for the word boundary are available (at least when it is the first prefix of a word), but this fact does not mean that having a prefix is better than having a stem at the beginning of a word.

From the next section we will describe each factor in more detail.

4.3.1 Morpheme length

The first factor that works against the prefix-stem order, in favor of the stem-suffix order, is the fact that affixes are shorter than stems on average. The logic behind this claim is simple: shorter morphemes are harder to identify, because shorter morphemes are more likely to match a part of other morphemes by chance (Laudanna & Burani, 1995; see also Schreuder & Baayen

1994).

This argument can be illustrated by a simple hypothetical example. Consider the English word form *pipes* /paʲps/, which has the plural suffix *-s*. After hearing the first three segments /paʲp/, one can be very sure that the morpheme *pipe* is used, and can be ready to hear the next morpheme.[5] On the other hand, imagine a hypothetical language where everything is the same as English except that the plural morpheme is a prefix *s-*, and suppose that one hears the plural of *pipe*, which is *s-pipe* /spaʲp/. When hearing *s-*, one cannot be sure whether this is the plural prefix or a part of another morpheme. After hearing /spaʲ/, one is still not sure whether one has already passed a morpheme boundary, or a morpheme such as *spy* or *spike* is being uttered. When hearing the entire sequence /spaʲp/, one is finally able to notice that there was in fact a morpheme boundary in an earlier stage, assuming that there is no morpheme that begins with /spaʲp/. This means that the fast recognition of a morpheme boundary near the beginning of a phoneme sequence tends to be difficult.

For an isolated word with a single affix, the length difference between stems and affixes would be enough to explain that a prefixed word is harder to process than a suffixed word. Because stems consist of a large number of items with low token frequency, it is expect from information theory (Zipf, 1935; Shannon, 1948) that they are longer on average than closed class items such as affixes. This means that a prefix–stem boundary typically appears near the beginning of a word, while a stem–suffix boundary typically appears near the end. Thus, we predict that the detection of a prefix–stem boundary involves more ambiguity.

[5]Other possible morpheme sequences that match this input, such as *pie plate*, can be thought of, but here we assume that their probabilities are low enough.

The length-based account, however, cannot be the whole story if we consider the morpheme segmentation of a sequence of words, rather than an isolated word. When processing an isolated word with a single affix, like *pipe-s*, the issue of detecting the boundary *after the suffix* does not come to the surface because that is the end of the utterance. However, in sentence comprehension in general, the parser has to identify not only the end of prefixes but also the end of suffixes. This raises a problem for our discussion above, because the length-based account only predicts that the end of a short morpheme is difficult to identify, no matter whether it is a prefix or suffix. Consider, for example, the following pair of two-word sequences. We assume that the prefix and suffix have the equal length, but the stems are longer.

(18) a. [stem] [[prefix]-[stem]]

b. [[stem]-[suffix]] [stem]

In both (18a) and (18b), a stem is followed by an affix, which is followed by another stem. Thus, if length is the only relevant factor, the parsing of (18a) and (18b) must be equally difficult, in spite of the fact that (18a) is a prefixing language, while (18b) is a suffixing language. We will argue that there are two additional factors that lead to a preference of (18b) over (18a).

4.3.2 Morpheme predictability

The first additional factor that further disfavors prefixes is predictability.[6] A morpheme is easy to recognize when the use of that particular morpheme is predictable from its preceding context. A prefix is less predictable from the preceding context than a suffix, because the majority of prefixes are at the beginning of a word.

To illustrate this point, consider the examples of *pipe-s* and *s-pipe* again. In English, when the parser hears /s/ that follows *pipe*, there is a fair chance of /s/ being a plural marker, and this expectation leads to higher confidence at the morpheme boundary after /s/. On the other hand, in the hypothetical language with a plural prefix, there is no or limited supporting evidence in the preceding context that /s/ is a plural marker, and therefore it can only have a lower confidence at the morpheme boundary after /s/. This means that the boundary after a prefix is harder to detect than the boundary after a suffix, even if their lengths are the same.

One might argue that, given that we are now concerned with sequences of multiple words, a plural prefix should also have a preceding context that helps predict it. In principle this is true; semantic contexts or syntactic factors (i.e. agreement) could also be used as cues to expect a plural marker to follow. However, this argument does not invalidate the claim that *a plural suffix is more predictable than a plural prefix*, because in the former, the host noun is already given, in addition to other possible cues. Furthermore, there are reasons to believe that the host noun of a plural marker has more predictive power than syntactic and semantic

[6]The argument here should not be confused with the predictability of affixes as opposed to stems, which has already been discussed in Greenberg (1957) and Cutler et al. (1985). Our claim in this section is that prefixes are less predictable than suffixes in a sentence.

contexts. For example, suppose that the plural affix /s/ has allomorphs, as in fact it does in English. In a suffixing language, the parser can tell that /s/ is the right allomorph for the noun *pipe*, and thus it further increases confidence. In a prefixing language, the parser cannot tell that /s/ is the right allomorph for the upcoming noun, thus the allomorphy cannot facilitate the parsing of the plural prefix. The factor that plays a crucial role here is that allomorphy is typically conditioned within a word, rather than across a word boundary, although exceptions could be pointed out (as in the English indefinite article *a/an*, in which allomorphs are selected based on the phonology of the following word).

In general, relationships between morphemes within a word are characterized by fixed order, allomorphy, limited productivity, and lexical idiosyncrasy (Haspelmath, 2011b). Again, these are only tendencies, and none of them can be a definitive property of *word*, as discussed by Haspelmath himself. However, for our purposes, we only need to assume that these are statistically more often characteristics of the relationship between morphemes within a word, rather than morphemes across word boundaries. All of these factors make a following morpheme more predictable within a word than across a word boundary. For example, the fixed morpheme order narrows down possible next morphemes so that the identification of the next morpheme will be facilitated; a plural marker can be more easily predicted in a language where it always immediately follows a noun, than in a language where an arbitrary phrase can be inserted between a noun and a plural marker. By definition, a suffix always has a preceding morpheme within a word, while a prefix is at the beginning of a word, unless more than one prefix is stacked. Thus, we can expect that predicting a prefix is more difficult than predicting

a suffix.

4.3.3 Additional cues for word boundaries

In this section, I argue that there is yet another factor that further facilitates the processing of suffixes or grammatical words, but not prefixes.

The psycholinguistic literature has revealed that a variety of cues are employed in the process of word segmentation, including phonetic, prosodic, and phonotactic cues (Jusczyk & Luce, 2002). Only some of them are available for word-internal morpheme boundaries. This further helps to identify a morpheme at a word-final position, but not a prefix, whose end is by definition not a word boundary.

Phonotactics is another factor that can be used as a cue for morpheme segmentation. Some phoneme sequences are unusual within a single morpheme, which signals that there is a morpheme boundary (Hay, 2003). For example, in English a phoneme sequence like /nr/ is unusual within a single morpheme, and therefore it can be used as a cue to detect a morpheme boundary in sequences like *in reality*. However, languages often have assimilation processes that effectively eliminate such low-probability phoneme sequences and blur the morpheme boundary. For example, the negative prefix *in-* undergoes assimilation and ends up with /ɪ/ when it is followed by /r/, as in *irregular*. As a result, the word loses the phonotactic cue for its morpheme boundary. While assimilation processes across word boundaries are possible, they are only occasional and/or incomplete (Dilley & Pitt, 2007).

It must be noted that, as discussed in Section 2.3, different criteria for phonological wordhood do not always match. This means that some word-boundary cues may be available even

for boundaries that are word-internal according to another criterion. However, this is not particularly problematic for our discussion here, because it is unlikely that more phonological cues are available at a word-internal boundary than at a word boundary, as long as word is properly defined so that a unit with more evidence for wordhood is described as a word. Furthermore, when there is a mismatch in wordhood criteria, there is an asymmetry in the direction of the mismatch: prefixes tend to be more phonologically independent than suffixes do. We will discuss in Chapter 7 the fact that this asymmetry is predicted from our approach.

4.3.4 Interim summary

To summarize, a variety of factors work against the immediate recognition of prefixes: (i) length, (ii) predictability and (iii) phonological cues. Their effect on different types of morphemes can be summarized as in Table 4.1. In the table, morphemes are classified into five categories: preposed grammatical words, prefixes, stems, suffixes, and postposed grammatical words. The checkmark (✓) indicates that the detection of the morpheme boundary after a particular morpheme type is facilitated by the processing factor in that row.

Table 4.1: facilitating factors by morpheme types

facilitates the recognition of ..	(preposed grammatical word)	(prefix-	stem	-suffix)	(postposed grammatical word)
length	–	–	✓	–	–
predictability	–	–	–	✓	(✓)**
phonological cues	✓	–	(✓)*	(✓)*	✓

* Facilitated only when they are at the end of a word

** Less predictable than suffix

Table 4.2: facilitating factors by morpheme types: prefix vs. suffix

facilitates	(prefix-	stem)	facilitates	(stem	-suffix)
length	–	✓	length	✓	–
predictability	–	–	predictability	–	✓
phonological cues	–	✓	phonological cues	–	✓

First, the length factor predicts that the morpheme boundary after a stem is easier to identify compared to other types of morphemes (i.e. grammatical morphemes). Because prefixes and suffixes are both short, this factor appears to have no direct relevance to the difference between prefixes and suffixes. However, it does predict that a stem–suffix boundary is easier to detect than a prefix–stem boundary, which means that the use of a suffix instead of a prefix does improve the detection of word-internal morpheme boundaries; if length is not taken account, we do not have an answer for why we will be better off having a stem at the beginning of a word than having a prefix. This point is illustrated in Table 4.2, which compares a word with only one prefix and a word with only one suffix. When the affix comes first, none of the three factors facilitate its processing; when the stem comes first, it is less problematic thanks to the length factor. The postposed item is not as problematic as the preposed item no matter whether it is the stem or suffix, because of multiple factors that help its detection. Second, the predictability factor predicts that suffixes are more predictable from the previous context than the other types of morphemes.[7] Postposed grammatical words are probably also predictable

[7]We expect that a prefix or stem would not be as predictable as a suffix even if it is in the middle of a word. This is because while a suffix follows a stem, a prefix or stem only follows prefixes. On average, a prefix is unlikely to make the next morpheme highly predictable, while a stem can make the following suffix highly predictable. This is because the type frequency of affixes is much smaller than that of stems, and therefore the information carried by an affix is small (in the information-theoretic sense). In this sense, the length and predictability factors are related. Because there are many stems with low token frequency, they are longer on average, and also each of them is more informative.

to some degree, but not as predictable as suffixes, because suffixes are always adjacent by definition while postposed grammatical words may be intervened by other words. Third, we argued that more phonological cues are available for word boundaries. This helps the parsing of grammatical words, as well as the last morpheme of a word, which is either a stem or suffix, but is never a prefix.

These three factors conspire to make prefixes the worst choice. Each type of morpheme has some factors that help its identification, except prefixes. Grammatical words are likely to have phonological cues that signal their boundaries. Stems are longer and more likely to be uniquely identified just by the match with the phonological representation of a lexical item. Suffixes are likely to be predicted from the preceding stem. Only prefixes are not helped by none of these factors.

To take another view of this table, consider the historical morphologization of a grammatical word into an affix. When a postposed grammatical word is morphologized into a suffix, it might lose some phonological cues, but it becomes more predictable from the preceding stem, and thus it does not become a burden for the parser.[8] On the other hand, when a preposed grammatical word is morphologized into a prefix, it only loses its phonological cues. It does not become more predictable, and thus its identification only becomes more difficult. This is why the development of prefixes is hindered and the suffixing preference results.

Furthermore, our account predicts that prefixes are *not* categorically ruled out. This is

[8]The loss of phonological cues may occur either before or after it becomes obligatorily adjacent. A morphologization process is the loss of word-like properties we saw in 2.3, and we assume that there is no fixed order of the word-like properties in which they are lost. When the loss of phonological cues occur first, it roughly means that it goes through the stage of clitic.

because the difficulty of a prefix depends on its statistical properties; they are difficult only *on average*. See Chapter 5 for more discussions as to why existing prefixes are not incompatible with our account.

Unlike previous processing-based approaches like Hawkins and Cutler (1988) and Hall (1988, 1992), this account does not need to assume that the lexical processing module is designed to deal with stems and affixes in different ways. Our predictions simply follow from their length, frequency and positions, and the parser that tries to identify morphemes as early as possible using its probabilistic knowledge.

The discussion here also suggests that experimental results, such as Cole et al. (1989), can be explained by the difference in length between stems and affixes, and there is no need to hypothesize the priority of stem as a design feature of the mental lexicon or parser. This conclusion can be obtained if we assume a dual-route model of morphological processing, where both whole-word and decomposition routes are available for the parser. Cole et al.'s experimental results suggest that a suffixed word is often processed via the decomposition route, while a prefixed word tends to be processed via the whole-word route. This can be understood if we take into account the fact that the detection of a prefix is often delayed. In other words, when the parser can uniquely identify the prefix being used, it is likely that it has already received a part of the stem. This means that the parser may have almost reached the point where the whole word is uniquely identified. Thus, it can be simpler for the parser to use the whole-word route, rather than identifying the overlooked prefix and then solving the separate problem of identifying the stem. On the other hand, in suffixed words, stems are

more likely to have been uniquely identified when the parser reaches their end, and thus it can activate possible following suffixes. This makes processing of suffixed words via the decomposition route more efficient. In this account, we need not assume any a priori difference between stems and affixes for the parser; the asymmetry arises simply because of the length difference between stems and affixes. In order to make these claims more explicit, however, we will need a quantitative investigation of the lexicon of a language (See Balling and Baayen (2012) for a recent quantitative analysis of unique identification points of derived words).

In the next section, we will present a simulation model that can demonstrate our points quantitatively.

4.4 A simulation study

In this section, we will present the design of our simulation study that is aimed at demonstrating the processing difficulty with prefixes discussed in the previous sections.

4.4.1 Language model

This thesis adopts the bigram model of *morphemes* as a language model. In a bigram model, it is assumed that the probability of an item only depends on the immediately preceding item. In other words, the probability of getting a sequence $\langle w_0, w_1, ..., w_n \rangle$ is approximated as follows, w_0 and w_{n+1} being special symbols that stand for the beginning and end of the sequence respectively:

$$P(w_0, w_1, ..., w_n, w_{n+1}) \approx \prod_{i=1}^{n+1} P(w_i|w_{i-1})$$

While an N-gram model appears to be an overly simplistic model of human language, it is immensely useful in practice (Jurafsky & Martin, 2008). When adopting a bigram model, the set of possible segmentations can be represented as a lattice. Figure 4.1 shows an example of a lattice structure taken from Norris and McQueen (2008), who propose a probabilistic model of word segmentation. In the lattice, each connection has a value that corresponds to the transitional probability between the two lexical items it connects. Then the parser's task is to find the route(s) with the largest overall probability. In this kind of model, the search for the best route can be achieved by efficient dynamic programming techniques such as the Viterbi algorithm.

Because we are interested in morpheme segmentation within a word, it is not enough to segment speech into words. A unique characteristic of the current approach is that bigrams are constructed based on *morphemes*, instead of words, without distinguishing word boundaries from word-internal morpheme boundaries. In this thesis a *morpheme boundary* refers to both a word boundary and a word-internal morpheme boundary. This move plays a crucial role in explaining suffixing preferences without assuming a priori differences between syntactic and morphological processing. An example lattice is shown in Figure 4.2.

Our approach has a couple of limitations. First, it can only deal with morpheme-based concatenative morphology. In the following simulations, examples of non-concatenative morphology, such as vowel alternations, are not analyzed; for example, irregular past-tense forms

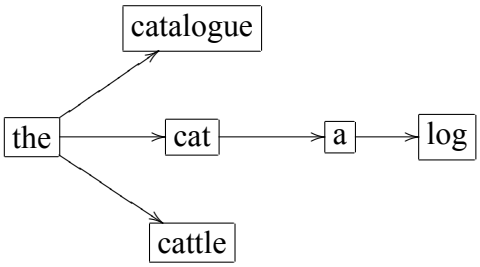


Figure 4.1: A part of the lattice structure for the input "The catalogue in a library,", taken from Norris & McQueen (2006)

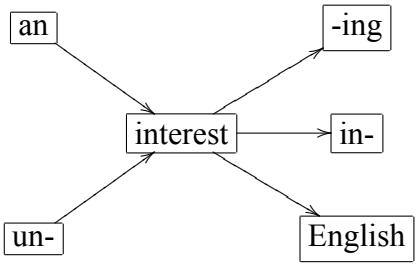


Figure 4.2: An example of lattice structure for the input *uninteresting* in a model with no distinction between word and word-internal boundaries

like *took* are simply treated as monomorphemic. In order to deal with a variety of morphological processes in a consistent manner, including non-concatenative morphology, allomorphy, irregularity and suppletion, it is necessary to incorporate a more full-fledged model of computational morphology, such as Finite State Morphology (Koskenniemi, 1983; Beesley & Karttunen, 2003), which is beyond the scope of this thesis.

Second, our model ignores many different types of cues that are potentially useful for speech segmentation. In language comprehension, both lexical cues (match with known lexical items) and sublexical cues (evidence from phonotactic, suprasegmental, and any other perceptual cues) are employed to segment speech into morphemes, and a complete processing model should integrate both types of cues (Mattys, White, & Melhorn, 2005). A serious consequence of discarding sublexical cues is the zero-frequency problem: in realistic settings

there is always a possibility that a parser encounters an unknown word, in which case it is necessary to have some mechanism to assign a probability for it. A system that only relies on the match with existing words will fail at such an unknown word, because it can only assign a zero probability for it. In natural language processing, this is typically solved by preparing a word model which is used to calculate how word-like a form is.[9] This move, however, would add considerable complexity to our model. In order to avoid this complication, our study is based on artificial data which was generated by the same bigram model. This also enables us to avoid the complication caused by smoothing, i.e. techniques to assign probabilities for missing n-grams. This simplification would not be appropriate if our goal were to measure the performance of a learning algorithm, but it should not be a huge problem for our purpose, because our goal is only to observe what happens in the process of identifying known morphemes, and the small chance of encountering a novel morpheme would not heavily affect the picture.

4.4.2 Parsing

Based on the language model discussed above, we constructed a parser that received speech input and identified morphemes being used. The following paragraphs explain how the parser works.

The input given in the simulation model is phoneme sequences. For example, when the

[9]For example, P. E. Brown, Pietra, Mercer, Pietra, and Lai (1992) assume a word model in which the distribution of word length follows a Poisson distribution. In Nagata (1999), a word model is constructed based on a character bigram for Japanese; In G. G. Lee, Lee, and Cha (2002), a word model is based on a syllable trigram for Korean.

original message is *ice cream*, the phonological input will be /aʲskrim/.[10] The task of the parser is to recover the original message from this phoneme sequence. Although a challenge in a realistic speech recognition system lies in how to convert sound waveforms into phoneme sequences, we simply assumed that phoneme sequences were given.[11]

In our study, bigram frequencies are calculated based on existing language resources. For English, frequencies are based on the Google Web Ngram. The data sources will be explained in the section for each language. The simulation program attempts to recover the original message given a phonological input by identifying morphemes being used. The program is *incremental* in the sense that it performs inferences as soon as a new piece of evidence comes in, without waiting for the end. This is in accordance with the psycholinguistic literature we reviewed in Section 4.2.2.

The simulation program holds all the possible morpheme sequences that match the partial phonological input, each of which has its probability. For example, according to our bigram model for English, when the initial part of the phonological input is /aʲh/, the most likely phoneme sequence that matches this input is *I have* with the probability of 61.3%. *I had* is the second most likely sequences with the probability of 20.0%, which is followed by *I hope*, with 7.5%, which is in turn followed by *I hate*, with 2.7%, which is further followed by numerous

[10]The use of phonemic representations means that our model ignores allophonic variations. In *ice cream*, for example, /k/ is pronounced as aspirated [kʰ], which may be used to disambiguate this message from *I scream* with unaspirated [k]. In this example, the aspiration can be considered as one of the phonological cues for word boundaries which we cannot directly implement into our model (although the real picture is more complicated than this, because an aspirated stop can also occur at the beginning of a stressed syllable, regardless of morpheme boundary).

[11]An industry-level speech recognition system typically has another probabilistic model that connects each phoneme or sequence of phonemes with acoustic patterns.

low-probability sequences. When the parser perceives /æ/ as the third morpheme, all of these probabilities are updated. The probability of *I have* increases to 74.8% and remains at the first place, and *I had* remains at the second place with the probability of 24.3%. *I happen* becomes the third most likely candidate with 7.5%. Sequences like *I hope* and *I hate* disappear from the list as they do not match the input anymore.

How can this simulation be used to make predictions about the difficulty of morpheme segmentations? In order to measure how confident the simulation program is about the presence of each morpheme boundary, the probability the model assigns to each morpheme boundary, henceforth *confidence*, is calculated. Here, the confidence at a particular morpheme boundary is equal to the sum of the probabilities of all paths in the lattice that have a morpheme boundary at that position. Low confidence means that a probabilistically optimal parser would not readily recognize the existence of the morpheme boundary, at least at the moment when the speech input reaches that morpheme boundary. Thus, confidence, as formulated above, can be conceived of as a measure of how easy it is to identify the morpheme that ends at that morpheme boundary.[12]

Table 4.3: An example of confidence calculation

partial input	/hi/	/hidɪs/	/hidɪslaʲk/	/hidɪslaʲks/
type	word boundary	prefix	suffix	word boundary
confidence	93%	10%	100%	5.8%

[12]This is, of course, not the only possible way to measure the difficulty of morpheme segmentation. We might also have to measure how long the detection of a morpheme boundary delays after it passed the morpheme boundary. We might also be interested in how often the parser mistakenly detects a morpheme boundary when there is no boundary. It might be possible to calculate the overall performance of the probability distribution the parser holds using an information-theoretic measure, but we leave it as a future challenge.

To take a closer look, consider the sentence *He dislikes me* /hidıslaʲksmi/. The first morpheme boundary is located after the initial two phonemes /hi/. Because of the high frequency of *he* in the sentence-initial position, the simulation program is confident about the presence of the morpheme boundary, with a probability of 93%. Although /hi/ could be a part of other words like *heal* or *heath*, they have much lower probabilities. The program is much less confident about the second morpheme boundary at /hidıs/. The most likely morpheme sequence that matches this phonological input was *He'd especial-* in our model with a probability of 26.8%, and *He decide-* is second, with a probability of 12.5%. The correct analysis for the current input, *He dis-*, is only the fourth place with the probability of 10.0%. Because *He dis-* is the only candidate that has a morpheme boundary after /s/, the confidence of the morpheme boundary is equal to 10.0%. Note that it is possible that more than one analysis identifies a morpheme boundary at the same position; in that case the probability is summed for all the analyses that have the boundary. Suppose that, for example, English has a prefix *is-* /is/. Then, the input /hidıs/ could have two analyses: *He dis-* and *He'd is-*. In this case, the probabilities of *He dis-* and *He'd is-* are summed and the total is considered to be the confidence value for the boundary after /s/.

Table 4.3 summarizes the confidence values for all the morpheme boundaries in the input /hidıslaʲksmi/. In this study, morpheme boundaries are categorized into the following three categories: (i) word boundaries, (ii) boundary after a prefix, and (iii) boundary before a suffix. The input /hidıslaʲksmi/ contains two word boundaries: one boundary after a prefix, and one boundary before a suffix.

Test sentences were randomly generated based on the bigram model, in a way similar to examples in Section 4.3 of Jurafsky and Martin (2008). Because of the limitations of a bigram model, it does not always produce grammatical sentences. The reason why artificial data were used instead of a real corpus is the issue of zero frequency, as explained in the previous section. A few examples from English include:

(19) a. Workshops followed me out.

b. If they received and yelling.

4.4.3 Procedure

Our simulation study was conducted following the procedure in (20) for each language we will discuss below. Our procedure is also schematically shown in Figure 4.3.

(20) a. Construct a morpheme-based bigram using existing language resources.

b. Generate 1,000 random sentences based on the bigram created in (a).

c. Let the simulation program parse the sentences generated in (b), using the language model in (a). The simulation program is only given a phonological representation of the sentences and identifies morpheme boundaries. The confidence value for each morpheme boundary is computed and summarized.

The nature of the language resources used for the simulation vary depending on languages, due to restrictions on the availability of data; we will explain them in the sections on each language. The programs for constructing morpheme-based bigrams and test sentences, as

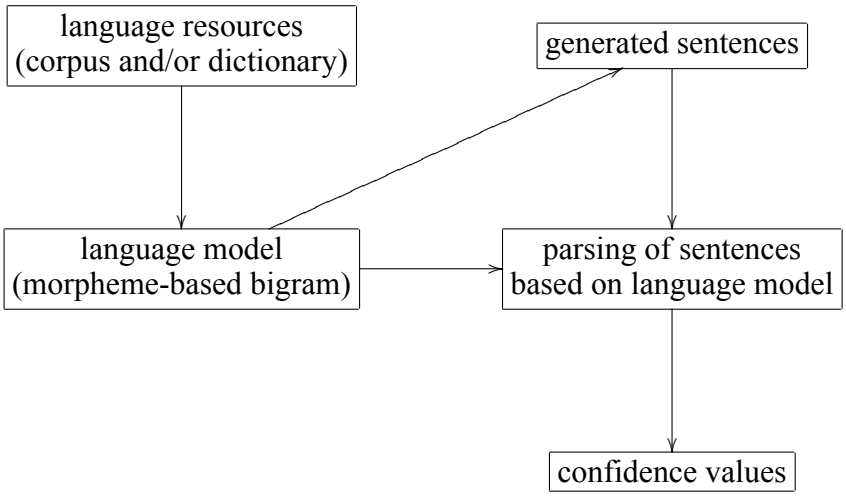


Figure 4.3: Flow of the simulation experiment

well as the parser for simulation studies, were all written by the author using the Python programming language.

The simulation model was applied to English, Reverse English, a hypothetical prefixing language that is made up for the purpose of this study, as well as to more languages in the next chapter.

4.4.4 Experiment 1: English

A morpheme-based bigram model was constructed by combining the following two resources: Google Web NGram (Brants & Franz, 2006) and the CELEX2 Lexical Database (Baayen, Piepenbrock, & Gulikers, 1996). The frequency count of each NGram was reduced to 1/1000 of the original, and a word pair whose frequency was rounded to 0 were discarded, in order to speed up the simulation. The former database provided information about bigram frequency, while the latter database provided information about phonology and morphology. In CELEX2, morphological boundaries are only given to the spelling data, not to the pronunci-

ation data. Due to this constraint, we identified morphological boundaries in pronunciation, combining the data from phonology and morphology. For example, for the word form *making*, CELEX2 has the information that this is pronounced as /meʲkɪŋ/, as well as the information that its morphological structure is *mak-ing*. It was necessary for the purpose of this study to infer /meʲk/+/ɪŋ/ from those pieces of information. In our experiments we followed CELEX2's definition of affixes; it has been pointed out that there are inconsistencies in the morpheme segmentation of CELEX2, which could be a potential problem for our method. Compounds are not segmented and are simply treated as monomorphemic in this study.

Figure 4.4 shows our results. The figure classifies morphological boundaries into three types: word boundary, boundary *after prefix*, and boundary *before suffix*. The graph shows the confidence of each type of boundary, averaged over tokens.

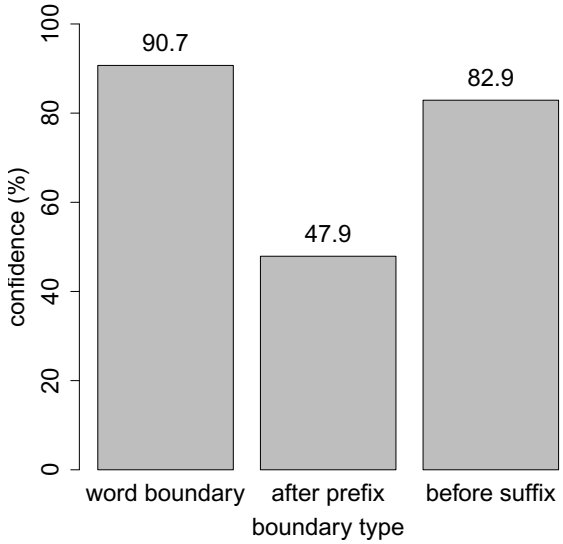


Figure 4.4: Average boundary confidence for English

We can see a striking asymmetry between different boundary types. While morpheme boundaries between words and before suffixes were recognized with about 80% to 90% confidence, those after prefixes show a much lower average confidence around 50%. This indicates that the immediate recognition of prefixes is much more difficult than that of the other types of morphemes.

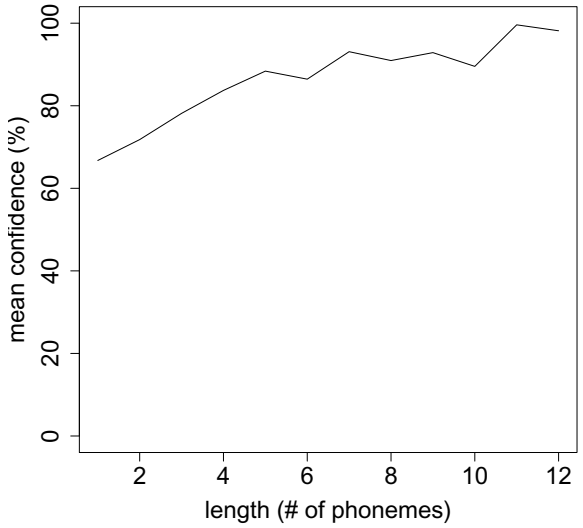


Figure 4.5: Mean confidence by length (English)

The results were further analyzed to see if the factors we mentioned earlier in this chapter were in fact at work. Figure 4.5 shows the relationship between the length of a morpheme and the confidence about the boundary after that morpheme based on our simulation results for English. Here the length of a morpheme is defined as the number of phonemes it contains. For each length, the confidence values are averaged over *morpheme types*.[13] We can see that the

[13]For example, if there are only two morphemes with only one phoneme, and their average confidence values are 50% and 70% respectively, then the average confidence value for one-phoneme morphemes is 60%, no

shorter a morpheme is, the harder it is to detect. The confidence for two-phoneme morphemes is 66.7% and that for three-phoneme morphemes is 71.8%, while the mean confidence exceeds 80% when a morpheme has more than three phonemes.

However, as discussed in Section 4.3, length is not the only factor that works against prefixes. Figure 4.6 shows the same data as Figure 4.5, but this time the data is classified by types of morpheme boundaries. Again, morpheme boundaries are classified into three types: boundaries after prefixes, boundaries before suffixes, and word boundaries. We can see that boundaries after prefixes have lower confidence values even when we compare morphemes with the same length. The other types of morpheme boundaries are not significantly worse, even when they are short.

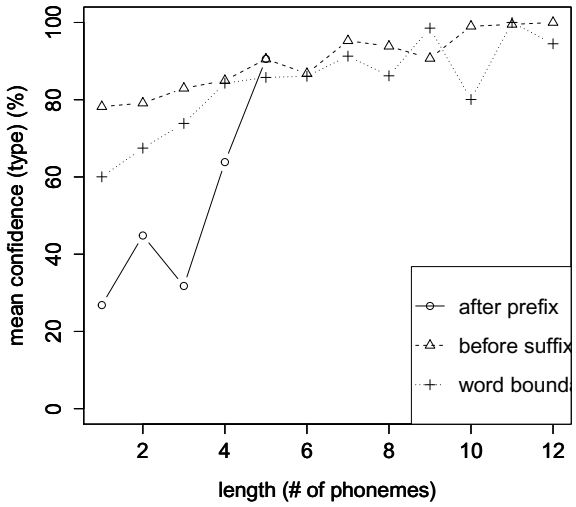


Figure 4.6: Mean confidence by length and boundary type (English)

The difference between prefixes and other types of morphemes, even when controlled

matter the token frequency of each morpheme.

for length, can arguably be reduced to the two additional factors we argued in Section 4.3: predictability and phonology. In order to confirm that predictability is playing a role, the transition probability between each pair of morphemes was calculated for our English test data, which we will explain later. We found that the mean transition probability between morphemes within a word is as high as 47.8%, while the mean transition probability across a word boundary is 3.9%. This means that a following morpheme within a word is highly predictable, while a following morpheme across a word boundary is not.

The third factor we argued for in Section 4.3 is phonology. Because our simulation model does not use sublexical cues (such as phonotactics and suprasegmental evidence) for segmentation, the effect of phonological cues is not directly observable in our data. Nonetheless, we can see that phonological wordhood does have an influence on our results. The prefix *in-* becomes just /ɪ/ when it is followed by /n/, /m/, /l/ or /r/, as in *immature*, *illegal* and *irregular*. This makes immediate recognition of the prefix harder because, as a logical necessity, more morphemes start with /ɪ/ than with /ɪn/, and therefore there is more ambuiguity after /ɪ/ than after /ɪn/. In our English simulation results, the mean confidence for the boundary after the prefix /ɪn/ is 61.4%, while the mean confidence for the morpheme boundary after /ɪ/ is very low at 0.7%. This suggests that assimilation does have an effect of lowering the confidence for prefixes in our model.

4.4.5 Experiment 2: Reverse English

It was demonstrated in the last section that prefixes are more difficult to immediately recognize in English, as we predicted. However, data from English alone is not enough to support a

typological claim. The results that we saw in the last section may be merely a property of English, not of human language in general. More seriously, it might be the case that the parser was less confident about prefixes only because prefixes are less frequent than suffixes in English. If so, the causal relationship would be the opposite: prefixes are harder because they are rare, rather than that prefixes are rare because they are harder.

Another problem is that the data includes derivational affixes. In English, all the prefixes are derivational, while a large share of suffixes are inflectional (in terms of token frequency). Including derivational affixes may be problematic, because derivational affixes might not need to be always parsed out for successful comprehension, as we will discuss in 4.5.2. If we omit derivational affixes from our data, it is not possible to compare prefixes and suffixes.

More generally, there is a problem of using a real language for the purpose of this study. Our claim is that prefixes are more difficult to process than suffixes, *all other things being equal*. It is difficult to keep other things equal when analyzing a real language. A real language may be "designed" so that the difficulty with prefixes is counterbalanced by other factors (this is what we will investigate in the next chapter). If these concerns are indeed correct, it is difficult to take asymmetries in the simulation on a real language as evidence for the suffixing preference.

To overcome this issue, this chapter makes use of Reverse English, a hypothetical language made up for this study. In Reverse English, everything is the same as English, except that phonemes are ordered backwards in each sentence. Because all regular inflections are performed via suffixes in English, all inflections are performed via prefixes in Reverse En-

glish. The phonotactics of Reverse English would be unnatural as a natural language, but that is presumably irrelevant for our purpose. Below is an example sentence of Reverse English.

(21) im s-kaʲl ih
me 3SG-like he
'He likes me.'

If prefixes are still more difficult than suffixes in Reverse English, then it will suggest that the difficulty with prefixes in English was not due to their relative infrequency or some other factor that is peculiar to English, but is really due to their position.

The resources and procedure used for Reverse English were the same as with English, except that phoneme order was reversed when constructing a language model. Figure 4.7 shows the results. Remarkably, the results were very similar to those of English: the average confidence for the boundaries after prefixes was around 50%, while the confidence for the word boundaries and the boundaries before suffixes was around 90%, despite the fact that the language contains much more prefixes than suffixes. Figure 4.8 indicates that short prefixes are difficult to recognize immediately in Reverse English, too.

Because derivational affixes may not need to be parsed in comprehension (see Section 4.5.2), we ran the simulation that focuses on inflectional affixes, without analyzing derivational affixes. The mean confidence value for prefixes was .399 (type-based) and .464 (token-based), while the mean confidence value for the other types of morphemes was .850 (type-based) and .849 (token-based). Again, it was demonstrated that prefixes are more difficult to identify than the other types of morphemes. Table 4.4 shows in more detail the confidence values for inflectional prefixes of verbs and nouns in Reverse English. Figure 4.9 is a mosaic

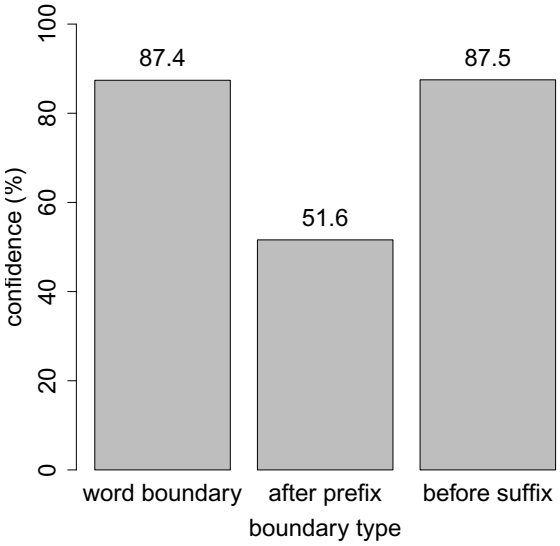


Figure 4.7: Average boundary confidence for Reverse English

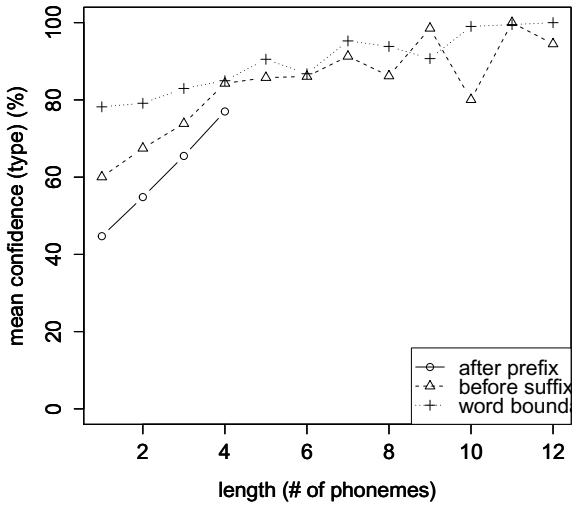


Figure 4.8: Mean confidence by length and boundary type (Reverse English)

Table 4.4: Confidence for prefixes in Reverse English

prefix	corresponding English suffix	token freq.	mean confidence
d-	*-ed*	83	.233
dɪ-	*-ed*	84	.461
ŋɪ-	*-ing*	120	.698
s-	*-s*	127	.398
t-	*-ed*	24	.114
z-	*-s*	313	.504
zɪ-	*-es*	53	.386
prefix		804	
type mean			**.399**
token mean			**.464**
others		6,864	
type mean			**.850**
token mean			**.849**

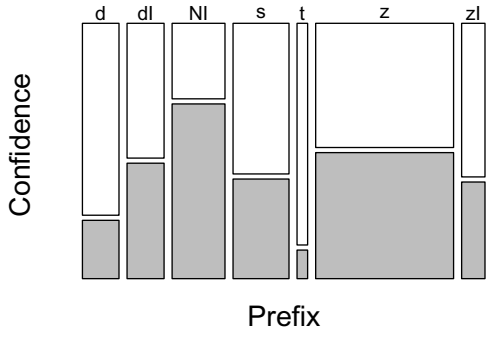


Figure 4.9: Confidence for prefixes in Reverse English

plot of the same data, in which the height of a gray area represents the confidence and the width the token frequency, so that we can see how much each prefix contributes the mean confidence. Each allomorph is counted separately. As we can see, NI (that corresponds to English -*ing*) shows the highest confidence, while /t/ and /d/ show the lowest mean confidence.

We can think of a number of reasons why prefixes are still hard to process in Reverse English, even though they are more frequent. Aside from general factors against prefixes, we can point out the following points:

(22) a. All inflectional prefixes in Reverse English alternate with zero. This means that even if one knows that one is listening to a verb or noun, one cannot automatically determine whether it has a prefix or not.

b. English has a lot of morphemes that happen to end with /t,d,s,z/. This means that Reverse English has a lot of morphemes that happen to start with /t, d, s, z/, which makes these sounds ambiguous between prefixes and the initial part of other morphemes.

c. Due to the allomorphy d-/dɪ- as well as z-/zɪ-, the parser cannot immediately determine which allomorph is being used right after it heard the consonant /d/ or /z/.

Additionally, a prefix is not easily predicted from context in Reverse English. A prefix is relatively easy to predict when, for example, it is a subject person prefix, and the subject precedes the verb. This situation does not happen in the neutral word order in Reverse English,

which is OVS (as a result of the flipping of the order).[14] This factor, however, is not directly reflected in our simulation due to the limitation of the bigram model.

4.5 The problem of linkage

Every psycholinguistic account of linguistic structure must eventually answer the following question of how the psycholinguistic preference can be conventionalized. This is what H. H. Clark and Malt (1984) and Kirby (1996) call the *problem of linkage*. Bybee (1988) and Hall (1988, 1992) point out, more specifically, that a psycholinguistic account of the suffixing preference would be insufficient, unless one provides an explanation of how a psycholinguistic factor can result in the typological frequency difference over time.

A potential problem with our approach is that our explanation is solely based on *comprehension*, rather than *production*. It is difficult to think that our segmentation-based account is directly related to the difficulties in production, because there is no need for the speaker to segment the speech signal into morphemes. If one structure is preferred over another in production, it is easier to see how that preference drives language change. However, if the account is based on comprehension difficulty, it is less straightforward. Speaker-oriented accounts, including Himmelmann (2014), have an advantage in this point.

A possible way to overcome this issue is to hypothesize *speaker altruism*, namely the idea that the speaker can choose the form that is most understandable for the listener.[15] This view,

[14] There may be some cases where a prefix can become more predictable due to agreement in Reverse English. For example, if the verb is in a non-third-singular present form, then it slightly increases the probability that the plural prefix follows.

[15] The term *altruism* may be misleading because being successfully understood by the listener has an immediate benefit for the speaker too. We simply follow the term used in Kirby (1996) among others here.

however, has been criticized by a number of authors (Kirby (1996, 76-79); Wasow (2002, 42-47); Song (2012, 260-264)). It has been argued that natural utterances are full of fillers and other signs of dysfluency, which indicates that an utterance is not planned ahead of time. It is difficult to believe that the speaker has enough resources to compute which form will be most understandable for the listener.

However, even without speakers' altruism, the same results could be achieved by subtler processes. It might be the case, for example, that a construction without prefixes is better understood and memorized, and therefore more likely to be reproduced later. In his simulation study of language change, Kirby (1996) assumes that primary linguistic data is filtered by their parsing complexity when an individual acquires a language. We take a similar view, although unlike Kirby, we do not commit that language change must take place at language acquisition. Rather, our position is that even if it is not always possible for a speaker to calculate on the fly which form is most listener-friendly, it is still possible that more understandable expressions survive over time.

4.5.1 Phonological cues and the grammatical status of affixes

We argued in Section 4.3 that the crucial factor that distinguishes prefixes from preposed grammatical words is the presence or absence of phonological cues. This appears to mean that the relevant factor is phonology, not the grammatical distinction between affixes and grammatical words. For example, there may be no obvious perceptual difference between the preposition *in* and the prefix *in-* in English, and thus it appears that whether a morpheme is an affix or word has no direct impact on the availability of phonological cues.

Our answer is that although there is no necessary relationship between the grammatical and phonological statuses of a morpheme, they motivate each other. We discussed motivations behind the correlation between lexical and phonological units in Section 2.3.1. We can say that prefixes are less likely to develop because preposed grammatical morphemes tend to keep the following phonological boundary in order to help their immediate recognition, which in turn prevents their morphologization in other respects as well.[16] Actually, a preposed grammatical morpheme tends to be phonologically independent even when they have become morphologized in other respects. We will discuss this point in Chapter 7. In case of derivational morphology, we could think of an alternative story. Instead of saying that prefixes are less likely to develop, we can say that prefixes are more likely to lose their productivity, as a result of repeated failures to access their lexical representations. This is the topic of the next section.

4.5.2 Inflection vs. derivation

The loss of a derivational affix can follow a path that is quite different from that of an inflectional affix. Because inflection is basically obligatory, it must be available for every word in a category, or else it must be absent altogether. For example, when English lost its case distinctions, they were eliminated from all the nouns. A derivational affix, on the other hand, can leave a large number of lexicalized words even after it loses productivity. To take an example, the productivity of the nominalization suffix *-ment* in English lost its productivity

[16] To avoid teleological wording, we can more accurately say that preposed grammatical morphemes that keep the following phonological boundary is better understood, and is more likely to survive.

by the late 19th century (Bauer, 2001, 181), but many words with *-ment* are still in active use.

This is because for a derived word, there is the option to memorize a whole word without analyzing it into parts. In order to successfully use the word *government*, for example, it is not necessary to know that this is a nominalized form of *govern*. In inflection, on the other hand, it is necessary to know the correspondence between present and past forms, singular and plural forms and so on, although their phonological forms can be irregular or even suppletive.

A number of researchers have argued for dual route models of morphological processing, in which both decomposition and whole-word routes are available for the processing of the same word (Caramazza et al., 1988; Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 1995; Hay, 2001).[17] These kinds of models hypothesize, for example, that the lexicon has not only entries for morphemes *un-* and *common*, but also a redundant whole-word entry, *uncommon*. In these models, the failure to detect the prefix *un-* does not necessarily mean processing difficulty. The parser can still successfully interpret the word via the whole-word entry without analyzing the morphological structure of the word.

What kind of historical change is expected in these kinds of models? Assuming that the memory of a morpheme is activated and reinforced when it is accessed, we can expect that the activation of a derivational prefix will diminish over time and eventually lose its status as a productive prefix.[18] For example, we can hypothesize that, in order to understand the

[17]Models vary in what roles the two routes play. In Caramazza et al. (1988), the whole-word route has default status and the decomposition route works as a back-up; In Schreuder and Baayen (1995), both routes simultaneously contribute to the interpretation; In Frauenfelder and Schreuder (1992) and Hay (2001), the two routes compete, and the faster route wins.

[18]Frauenfelder and Schreuder (1992) discuss a diachronic process of the lexicalization of a whole word based on the dual-route model, which is similar to ours.

word *illegal*, it is faster to retrieve the whole word *illegal* from memory, rather than recognizing the morpheme boundary between *i-* and *legal* retrospectively and calculate the meaning compositionally. In other words, the entry for the prefix *i-* can be ignored even if it is in the mental lexicon, and this results in the diminishing productivity of the prefix. This story indicates that the dual route model can potentially explain the loss of derivational prefixes, without concerning processing difficulty.

4.6 Conclusion

In this chapter, we presented reasons to believe that prefixes are harder to identify than other types of morphemes. This prediction simply follows from the distributional facts of affixes and the assumption that people use their probabilistic knowledge to identify morphemes. Then our point was demonstrated by simulation studies on English, as well as a hypothetical language, Reverse English.

Chapter 5

A closer look at typological minorities

5.1 Introduction

A psycholinguistic account of a typological universal attempts to show that a typologically common pattern is more efficient or easier to process than a rare pattern. One might ask, however, why languages with typologically rare characteristics, which are claimed to be inefficient, have existed with no problem. Are there efficient and inefficient languages?

To take an example other than the suffixing preference, it is unusual for a VO language to have a prenominal relative clause (Relative clause-Noun order, henceforth RelN). In the WALS data (Dryer, 2011f), only five VO languages (Mandarin, Cantonese, Hakka, Bai and Ami) are RelN, as opposed to 416 VO languages that has Noun-Relative clause order (henceforth NRel). Hawkins' (2002) performance-based principles of Maximize Online Processing (MaOP) and Minimize Domians (MiD) provide a straightforward account for this typological skewing. In essence, MiD predicts the correlation of VO and NRel, and MaOP predicts that the dominance of NRel over RelN. Because VO languages with prenominal relative clauses violate both principles, we expect that such languages are typologically rare.

But if VO&RelN is that inefficient, why are the languages with this property, though small in number, thriving without falling into malfunction? Comrie (2008) points out VO&RelN has been a stable feature of Chinese for a long time, as well as in Formosan languages possibly independently, and questions the validity of the claim that their word order is inefficient, although he does not provide an alternative answer for the question why the combination of VO&RelN is overwhelmingly rare.[1]

In a similar vein, Harris (2008) discusses the historical origins of the unusual split case marking system in Georgian as well as endoclitics in Udi, and claims that a typologically rare feature of a language is often due to a coincidental co-occurrence of normal historical changes. According to her, typologically rare features are rare because there are limited historical paths on which they can come into being, not because they are difficult to process or acquire, or are discouraged by innate principles. Considering that they have been stable features of Georgian and Udi for more than one and half millenia, she argues, a theory that reduces their typological rarity to psycholinguistic difficulties is untenable.

The same argument would apply to psycholinguistic accounts for the suffixing preference. Jacques (2013) is an example of discussions of this line. He describes the development of the so-called associated motion prefixes in Japhug rGyalrong, an unusual prefixing OV language in Southwest China. In this language, preposed motion verbs are grammaticalized into prefixes, despite the fact that there is an alternative construction where the motion verbs follow

[1] While it has been suggested that contacts with Altaic languages have played a significant role in the development of word orders of Chinese (Hashimoto, 1986; Dryer, 2003), this fact alone does not explain why Chinese must choose the most unlikely combination of word order.

the main verb, which could have led to them being developed into suffixes. He argues that this single example is enough to falsify the validity of the Fossilized Syntax Hypothesis (ibid, 212).

This kind of argument seems to lead to the conclusion that a psycholinguistic account is always invalid if the universal is only statistical. Such a view seriously limits the explanatory role of psycholinguistic factors. Is there no way to reconcile a psycholinguistic account with the fact that there are exceptional cases?

One possible way to save psycholinguistic accounts is to argue that some languages are in fact slightly inefficient, but not to the extent that it cannot be sustainable. Even if the effect of a psycholinguistic preference is small enough for some languages to survive with an inefficient property for thousands of years, it is still possible that the psycholinguistic preference creates a typological frequency asymmetry in a long run. This argument is similar to the position taken in Dryer (2011a, 375).

Another way to save psycholinguistic accounts is to argue that there is a trade-off with another factor. For example, among the six basic stop consonants /p,t,k,b,d,g/, /p/ and /g/ are found in a smaller number of languages (Maddieson, 2013). Ohala (1983) proposes an articulation-based account for why /p/ and /g/ are less preferred. But if it is the case that /p/ and /g/ are less ideal consonants, why do a lot of languages have them? The intuitive answer would be that there is a trade-off between articulation effort and perceptual distinctiveness. Making fewer distinctions would reduce the speaker's effort, but make speech difficult to understand. In order to seek a balance between these two counteracting factors, each language needs to

have some 'less ideal' consonants, in addition to the consonants that are easiest to produce. However, it would not be clear how this type of argument can be applied to the problems we discussed earlier. It is not clear, for example, what kind of benefit Chinese enjoys at the expense of adopting an apparently inefficient word order.[2] Similarly, we do not know what is the merit of having prefixes, instead of suffixes, for Japhug rGyalrong.

A similar, but slightly different argument is that a language with an 'inefficient' property has yet another property that mitigates the inefficiency. Such an argument has been suggested for the Chinese word order. In their reading time experiments of Cantonese, Matthews and Yeung (2001) and Ching (2008) demonstrated that when a sentence has a long relative clause that modifies the object noun, a topicalization construction is processed faster. The topicalization construction has the word order of OSV, which can avoid the center embedding caused by the prenominal relative clause in O. They suggest that this processing factor is one of the motivations behind the use of the topicalization construction in Cantonese.[3]

This is the kind of argument we would like to pursue in this chapter.[4] Prefixes are harder to process, *other things being equal*. Here 'other things being equal' means that if we just flip the position of the affix, without changing its phonological or morphosyntactic status, it becomes more difficult to process. Thus, simply getting rid of prefixes can be one of the strategies that a language can utilize, which results in the typological infrequency of prefixes. However,

[2]Preposed relative clauses in Chinese may have an effect of maintaining the fact that a noun phrase always ends with the head noun, which could help parsing.

[3]A further alternative hypothesis would be that Chinese has another way to convey the equivalent meaning, which does not use a relative clause.

[4]See also Hall (1992, 173–174) for the defense of psycholinguistic accounts for the statistical universal (and the suffixing preference in particular).

difficulty in prefix detection can also be overcome by other means, such as by maintaining a suprasegmental cue between a prefix and stem, or by making the occurrence of a prefix more predictable, and so on. In this section, we will empirically confirm these predictions by statistically investigating 'typological minorities', i.e. prefix-rich languages. Our view can be summarized as follows.

(23) a. Prefixes are harder to process, other things being equal.

b. Thus, in a hypothetical language, the parser faces difficulty processing prefixes. This is what we saw in Reverse English in Section 4.4.5.

c. On the other hand, real prefix-rich languages have properties that facilitate the processing of prefixes. This will be illustrated by the investigation of typologically diverse languages with prefixes: Walman, Swahili, Tariana, and Japhug rGyalrong.

The languages are chosen based on the availability of tagged corpora with the sufficient frequency of inflectional prefixes, considering their historical, geographical and typological diversity. Languages with rich non-concatenative morphology or complex morphophonological processes were avoided, as it is difficult to handle them with our basic parsing model.[5]

In general, we may predict the following typological tendencies that facilitate the identification of prefixes:

[5]It is difficult to estimate the effect of excluding those languages. Intuitively, complex morphophonology makes processing harder, but there is no particular reason to believe that they work particularly against prefixes. Thus, our prediction that prefixes are not particularly difficult to process in real prefix-rich languages is probably not affected by including languages with more complex morphophonology. Furthermore, as we will discuss in Chapter 7, prefixes tend to be phonologically more independent from the stem than suffixes. This means that there are fewer morphophonological processes at work at the prefix-stem boundary, which would make prefixes easier to process, but not more difficult.

(24) a. Prefixes are more often obligatory and uniform in length than suffixes, so that prefixes can automatically be split off

b. Prefixes are more likely to have an infrequent segment or sequence of segments than suffixes, so that they can be easily identified

c. Prefixes are more predictable from the preceding *words* than suffixes, so that (again) they can be easily identified

d. Prefixes are more often excluded from the phonological word formation than suffixes. In other words, prefixes leave more perceptual cues at their boundaries with the stem

Among these, the prediction about phonological word formation (24d) will be discussed in Chapter 7.

These empirical predictions play a crucial role in distinguishing our approach from other psycholinguistic approaches such as those ofHawkins and Cutler (1988) and Himmelmann (2014). First, Hawkins and Cutler (1988) do not predict that prefixing languages are designed such that prefixes are easily parsed out. Their argument relies on the assumption that in lexical processing, a stem is preferably processed earlier than an affix. If this is the case, making prefixes more immediately identifiable should have no benefit for the parser. Of course, the fact that there is a need to help parse prefixes out does not directly contradict with their assumptions, but then the fact that prefixes are otherwise difficult to process is enough to explain the suffixing preference, and the assumption about preferred processing order becomes redun-

dant. Himmelmann's (2014) account can predict (24d). His claim is that dysfluency more often happens after a grammatical morpheme, and thus there is a reason why a phonological boundary after a prefix is often maintained. However, there is no obvious reason to expect (24a-c) from his approach.

The following sections describe the results obtained for each language examined.

5.2 Walman

Walman is a Torricelli language in Papua New Guinea, which is classified as 'weakly prefixing' in WALS (Dryer, 2011e). The description of the Walman grammar is based on L. Brown and Dryer (2008).

In Walman, the most frequent affixes are pronominal affixes on verbs. Tables 5.1 and 5.2 summarize the person affixes in Walman. While all the subject markers are prefixes, some object markers are prefixes and some are suffixes. The third person object markers appear as infixes in a limited number of verbs, which were regarded as irregular verbs and not morphologically analyzed in our simulation.

Table 5.1: subject affixes in Walman

	singular	plural
1st person	*m-*	*k-*
2nd person	*n-*	*ch-*
3rd person	*n-* (masc.), *w-* (femin.), *l-* (dimin.)	*y-*

Other verbal affixes include the applicative suffix *-ro*. The template for Walman verbs can be represented as in (25).

Table 5.2: object affixes in Walman

	singular	plural
1st person	*p-*	*p-*
2nd person	*ch-*	*ch-*
3rd person	*r-* (refl.), *-n* (masc.), ∅ (femin.), *-l* (dimin.)	*-y*

(25) subj-(obj)-stem-(appl)-(obj)

There is also an infrequent imperative prefix, whose position and phonological realization vary, depending on the environment. Besides the verbal morphology, a genitive prefix *w-* can be analyzed, which is used only with pronouns. This was, however, not analyzed as a prefix in our text data and thus was not included in our analysis.

Walman data is taken from narrative data courtesy of Matthew Dryer (p.c.), which consists of 25,521 morphemes. An example sentence taken from the data is shown in (26):[6]

(26) Nyanam y-ete-y, y-aru-l y-ekiel
children 3PL-see-3PL 3PL-break-3SG.DIMIN 3PL-go.south

y-nare-n Chamul: Pakol nta y-ara elieu.
3PL-tell-3SG.MASC Chamul Pakol this 3PL-come war

'When the children saw them, they ran away back up and told the Chamul: The Pakol people have come here to fight.'

Table 5.3 shows the results for Walman. When we calculate the averages per type, the confidence value for prefixes was relatively low at .585, as compared to the confidence value for the other morphemes, which was .740. When we use token-based measures instead, the

[6]Sometimes glosses are not entirely accurate. For example, in (26), *nare* 'tell' must be analyzed as *na* 'speak' plus *-re* (applicative). However, we present the output of the parser on which our simulation study is based.

difference is smaller: it is .740 for prefixes and .793 for the other morphemes. The low type-based mean confidence is heavily affected by the especially low confidence of *p-*. However, this prefix occurs only twice in our test data, and thus the type-based mean confidence may not be very reliable. The data is also represented as a mosaic plot in Figure 5.1.

Table 5.3: Confidence for prefixes in Walman

prefix	token freq.	mean confidence
ch-	41	.655
k-	44	.408
l-	10	.885
m-	56	.498
n-	440	.721
p-	2	.119
r-	20	.436
w-	257	.661
y-	452	.886
prefix	1,322	
type mean		**.585**
token mean		**.740**
others	2,989	
type mean		**.862**
token mean		**.793**

Why are prefixes in Walman not as difficult as those in Reverse English? We can think of a couple of reasons. First, subject person prefix is obligatory and does not alternate with zero. Second, it always corresponds to a single phoneme. Thus, the parser can automatically parse the first consonant out, as long as it can tell that it is a verb. This is in contrast with the situation in Reverse English, where the parser cannot tell beforehand whether a verb has an affix or not. Third, while an object prefix is not always present and thus cannot be automatically parsed out, it often yields an unusual consonant cluster, as noted in L. Brown and Dryer (2008) with

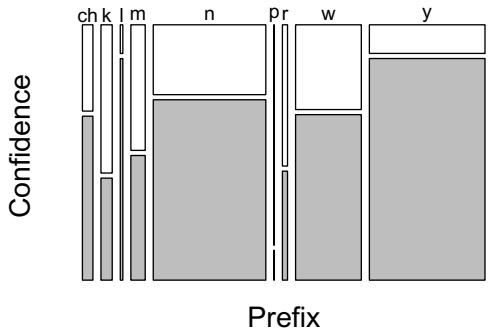


Figure 5.1: Confidence for prefixes in Walman

the following example:

(27) Mnon n-m-p-klwaro.
GEN.3SG 3SG.M-IMPER-1.OBJ-deceive
'let him lie to me'

(L. Brown & Dryer, 2008, 533)

This can be used as a reliable cue for the morpheme boundary. In Walman, there are three object prefixes, *p-* (1st person), *ch-* (2nd person) and *r-* (3rd person reflexive). For example, the word-initial sequence of *n-p* is always the sequence of two prefixes in our data. Because verb stems that start with *ch* or *p* are very few, when the parser can successfully identify the subject person prefix and it is followed by *ch* or *p*, the parser can confidently identify it as the object person prefix. For the reflexive prefix *r-*, on the other hand, there are some verbs that happen to start with *r*. In our corpus, there are two verbs (*rcho* 'pick up' and *ralkue* 'say goodbye') that can potentially be confused with the reflexive prefix. However, the total

frequency of these verbs is only four in our corpus, compared to the reflexive prefix which occurs 60 times in our data. Thus, these verbs would not seriously hinder the processing of the reflexive prefix.[7]

Furthermore, because the language is SVO (although the alternative order of SOV is also common), subject pronominal prefixes are relatively predictable from the context, especially when an overt subject noun phrase is present. This effect, however, was only incompletely reflected in our results because of the limitations of the bigram model.

5.3 Swahili

Swahili is a Bantu language spoken in a number of countries in Eastern Africa. While both Walman and Swahili are classified as 'weakly prefixing' in WALS (Dryer, 2011e), in Swahili it is more common to see a word consisting of many morphemes. It is well known that Bantu nouns are classified into more than ten noun classes, and a verb and adjective show agreement phenomena called *concord*. These agreement phenomena are all realized by prefixes. The verb template for Swahili can be represented as follows:

(28) subject-tense-(relative)-(object)-STEM-(extension)-(final vowel)

Below is an example of a Swahili phrase.

(29) Kitabu ni-li-cho-m-som-esh-a mtoto
book 1/2.SG1.SBJ-PAST-7/8.SG.REL-1/2.SG3.OBJ-read-CAUS-IND child
'the book which I made a child read'

[7]Despite this, the confidence values for those object prefixes were relatively low in our result. The reason appears to be due to the design decision of the parser: it cannot distinguish homonymous affixes from each other, such as the pair of *y-* (3PL.SUBJ) and *-y* (3PL). The performance should improve if the parser can hold different representations for homonymous morphemes.

(Sato, 2012)

Swahili is arguably the prefixing language with the largest corpus available. In this study, the book part of Helsinki Corpus of Swahili,[8] which contains 1,055,425 words, was employed. While morpheme boundaries are not directly given in the corpus, each word comes with its morphosyntactic information extracted by the technology of SALAMA (Hurskainen, 1999). Using this information, the author wrote a script to identify morphological boundaries with pattern match techniques similar to those employed in De Pauw and Schryver (2008).

The results are summarized in Table 5.4. In Swahili, we can see that the confidence for prefixes and for the other types of morphemes is almost the same, measured by both type and token. It must be noted that polysemous prefixes are not distinguished from each other in our data, and thus the same figures are repeated for morphemes with the same sound (such as *u*-) in the table. While the table only lists verbal third-person prefixes, figures may include other prefixes with the same sound.

There are a number of reasons why prefixes are not particularly a problem for a parser in Swahili. First, almost all prefixes are monosyllabic, and because the phonotactics of Swahili is simple, there is little ambiguity over syllable boundaries.[9] This means that the parser's task is to simply parse each syllable off. Interestingly, a suffix does not correspond to a syllable, as illustrated by the causative suffix *-esh* and the final vowel *-a*.

[8](Institute for Asian and African Studies (University of Helsinki) & CSC – IT Center for Science, 2004)

[9]The object prefix for class 1, *m*-, is also syllabic. In our result, the confidence value of *m*- is low, but this is probably due to the fact that our data does not distinguish syllabic and non-syllabic /m/. If the perceptual difference of these two sounds is obvious, it should not be difficult to identify the object prefix *m*- immediately.

Table 5.4: Confidence values for Swahili prefixes (detail only for verbal person prefixes)

class	prefix	token freq.	mean confidence
1 (subj)	a-	89	.687
1 (obj)	m-	146	.345
2	wa-	151	.674
3	u-	146	.727
4	i-	62	.551
5	li-	165	.805
6	ya-	42	.724
7	ki-	81	.603
8	vi-	35	.643
9	i-	62	.551
10	zi-	28	.541
11	u-	146	.727
15	ku-	230	.765
16	pa-	3	.406
17	ku-	230	.765
18	mu-	1	.212
prefixes (not limited to the list above)		1,632	
	type mean		**.641**
	token mean		**.657**
others		5,240	
	type mean		**.660**
	token mean		**.663**

5.4 Japhug rGyalrong

Japhug rGyalrong is a Tibeto-Burman language with unusually rich prefixes. Japhug rGyalrong is an SOV language, but it has more prefixes than suffixes. This is remarkable because it appears to violate both the correlation with basic word order and the suffixing preference.[10] It has a complex verbal template shown in Table 5.5.

Table 5.5: The Japhug verbal template (Table 2 of Jacques 2013, 197)

1	2	3	4	5	6	7	8	9
a-	*mɯ-* *mɤ-*	*ɕɯ-* *ɣɯ-*	*tɤ-* *pɯ-* etc.	*t-*	*wɣ-*	*ɣ-*	*s-*	*rɤ-*
10	11	12	13	14	15	16	17	18
nɤ-	*a-* *sɤ-*	*nɯ-*	*ɣɤ-* *rɯ-* etc.	noun	Σ	*-t*	*-a*	*-nɯ* *-ndʑi*

As mentioned at the beginning of this chapter, Jacques (2013) argues against the psycholinguistic account of the suffixing preference, based on the observation of the grammaticalization of associated motion prefixes, which correspond to Slot 3 of the template in Table 5.5. In the following example, the associated motion prefixes are glossed as TRANSLOC, and translated as *go*.

(30) mphrɯmɯ ɕ-pɯ-sɯ-re tɕe,
divination TRANSLOC-IMP-CAUS-look[III] CNJ

ɕ-tɤ-the ra
TRANSLOC-IMP-ask[III] NPST.have.to

[10] Another notable example of a prefixing SOV language is Athabaskan languages.

'You have to go to make him do a divination, go to ask him.'

(9) of Jacques (2013, 200)

For the purpose of simulation, we used all the Japhug rGyalrong texts with glosses available from Collection Pangloss.[11] The texts consisted of 586 sentences, 5,331 words, and 9,340 morphemes. The number of different morpheme types was 1,006.

The simulation results are summarized in Table 5.6.

Table 5.6: Confidence values for Japhug rGyalrong

prefix	
type mean	0.939483
token mean	0.948169
suffix	
token mean	0.934185
type mean	0.936206
stem	
token mean	0.962328
type mean	0.988685

We can see that the confidence values for prefixes in Japhug rGyalrong are no different from those of the suffixes. Like Swahili, the fact that most prefixes are monosyllabic may be the primary reason why prefixes are not difficult to parse out in Japhug rGyalrong. In fact, Japhug rGyalrong shares an interesting property with Swahili: while most prefixes have the simple CV structure, suffixes do not.

[11] http://lacito.vjf.cnrs.fr/archivage/tools/list_rsc.php?lg=Japhug&aff=japhug

5.5 Tariana

Tariana is an Arawak language spoken in Northwestern Brazil. A Tariana verb has one prefix slot, which is occupied by either a pronominal, negative, or relative affix. The verb morphology template is as follows:

(31) a. Cross-referencing prefixes (A/S_a), Negative *ma-* or Relative *ka-*.

b. ROOT

c. Thematic syllable

d. Causative *-i*

e. Negative *-(ka)de*

f. Reciprocal (rarely: reflexive) *kaka*

g. *-ina* 'almost, a little bit'

h. Topic-advancing *-ni*, or Passive *-kana*, or Purposive non-visual *hyu* or Purposive visual *-kaɾu*

i. Verbal classifiers

j. Benefactive *-pena*

k. Relativisers or nominalisers.

(Aikhenvald, 2006, 253)

The paradigm for person prefixes is shown in Table 5.7. These prefixes are used for marking not only the person of a verb, but also the possessor of a noun.

Table 5.7: Person prefixes in Tariana (Table 6.1 of Aikhenvald (2006, 122), simplified)

	singular	plural
1	*nu-*	*wa-*
2	*pi-*	*i-*
3nf	*di-*	*na-*
3f	*du-*	
impersonal	*pa-*	-
indefinite	*i-*	-

We used the narrative text published in Aikhenvald (1999), which was available in an electronic format, as a corpus. The text consisted of 615 sentences, 6,219 words, and 13,241 morphemes. The number of different morpheme types was 1,442.

The results are shown in Table 5.8. Again, the confidence values for prefixes were not particularly low.

Table 5.8: Confidence values for Tariana

prefix	
type mean	0.809519
token mean	0.918156
suffix	
token mean	0.936814
type mean	0.957409
stem	
token mean	0.918936
type mean	0.945191

A potential problem with the Tariana data is that there are a considerable amount of cases where a prefix-stem boundary is blurred by phonological processes, including H-metathesis and vowel fusion:

(32) prefix *di-* + root *-híma* (hear) → *dhi-ima* → *dhíma* 'he hears'

(Aikhenvald, 2006, 46)

At this point, we do not have a consistent way to handle cases where phonological or morphophonological processes obscure word-internal boundaries.

5.6 Conclusion

In this chapter, we saw that prefixes are not as difficult as suffixes in real languages with inflectional prefixes. We indicated a number of reasons why a prefix is not particularly difficult, as opposed to the constructed language Reverse English that we observed in the last chapter.

One might wonder why a language should avoid prefixes in the first place, if there are rich examples of prefixes that are easy for the parser to process. If an affix is very frequent, there will be functional pressure to shorten it (Zipf, 1935; Bybee, 2001). English *-ed* ([t] and [d]) and *-s* ([s] and [z]) achieve this in the sense that these affixes have the shortest forms that can be distinguished from zero. This was, arguably, easier because they are suffixes. If they were prefixes, the parser would struggle to parse them out, as we demonstrated in Reverse English. We can speculate that such a single-phoneme prefix can be stable only when certain conditions are met. In Walman, a subject person prefix is obligatory and is always monophonemic, and thus it can be parsed out without ambiguity, as long as the parser can guess that it is a verb. The matter is less straightforward when an object person prefix is stacked, but in that case, it will end up with a consonant cluster which is unusual within a single morpheme, thus creating a clear phonotactic cue. In the other languages we investigated, phonotactics is simpler and

many prefixes have the form of CV, making them relatively straightforward to identify. In either case, we can speculate that prefixes are more stable in a language where they have certain morpheme structure constraints.

Among the languages we saw in this chapter, Walman, Swahili, and Tariana all contain subject person prefixes. It is known that subject person marking shows little suffixing preference; as we will argue in Section 6.3.1, this fact itself does not need a special explanation other than the Fossilized Syntax Hypothesis. However, subject person markers may be especially predictable from the context, because S precedes V in the majority of languages, and when S is omitted it is typically given. This might be a reason why subject person marking proves no problem for the parser. Unfortunately, our simple language model is not powerful enough to capture a structural relation like agreement, and it is difficult to find quantitative support for this hypothesis. Related to this, Cysouw (2009) points out that person is more often marked by a prefix when a paradigm is small (i.e. the number of types of different person affixes is small). This is consistent with our view: when a paradigm is small, it is easier to predict a particular morpheme is to follow, and thus the difficulty seen with prefixes will be mitigated.

The number of languages we have investigated is still small, and because the languages were chosen so that they could be handled by our simulation model, there is a risk of bias in our data. In order to deal with a wider range of languages in a consistent manner, a principled way to handle phonological and morphophonological processes is needed.

Chapter 6

Differential suffixing preferences among grammatical categories

6.1 Introduction

It has been argued that a problem shared by psycholinguistic approaches to the suffixing preference is that they only predict the general tendency towards suffixing, and do not explain why the strength of the suffixing preference widely varies across grammatical categories (Myler, 2009b). For example, it has been pointed out that case markers show a very strong suffixing preference (Hawkins & Gilligan, 1988; Dryer, 2011d), while pronominal affixes on verbs lack one (Hawkins & Gilligan, 1988; Siewierska & Bakker, 1996; Enrique-Arias, 2002).

The following list summarizes the types of affixes with and without suffixing preferences according to Hawkins and Gilligan (1988). No category shows a prefixing preference.

(33) a. **Nominal Affixes**

- Categories with Suffixing Preferences:
  - Case

  - Gender
  - Plural
  - Nominalization
  - Indefiniteness
  - Definiteness
- Categories without Suffixing Preferences:
  - Possessive

b. **Verbal Affixes**

- Categories with Suffixing Preferences:
  - Mood
  - Tense
  - Aspect
  - Valence
  - Causative
- Categories without Suffixing Preferences:
  - Person-Marking (Subject)
  - Person-Marking (Object)
  - Negation
  - Voice

Existing psycholinguistic accounts of the suffixing preference do not provide direct accounts for these differences, with the possible exception of Himmelmann's (2014) account; we will return to his account in Section 6.3.1.5. This chapter explores the idea that the combination of historical and psycholinguistic factors can explain the differences between grammatical categories. While the psycholinguistic force against prefixes is at work for every category, some categories tend to be preposed from the beginning, while some other categories do not. Our claim is that the interaction of psycholinguistic and historical factors can explain the current distribution of affixes.[1]

This idea is already suggested in Dahl (1979) and Dryer (2011c) for negation affixes. Negation affixes are one of the categories that do not exhibit suffixing preferences in Hawkins and Cutler (1988) (nor in our data). This fact could be explained by the combination of psycholinguistic and historical factors. Suppose that a negation affix used to be an independent negation word. Non-bound negation words are more often preposed (Table 6.1). If preposed and postposed words are equally likely to be morphologized, we expect that there should be more prefixes than suffixes for negation. This is not the case. Why does the preference for preposing disappear in morphology? This can be readily explained if there is an independent psychological factor which hinders the morphologization of preposed negation markers. The psychological factor need not to be the one we argue for in this thesis. It can be any account

[1]The importance of integrating psycholinguistic and historical aspects of the suffixing preference has been emphasized in Hall (1988, 1992), but the historical aspect does not play a role in explaining the typological patterns in his account. Hawkins and Cutler (1988) is also similar in spirit to our approach in that they try to capture the distribution of affixes as a combination of the suffixing preference and its correlation with word order. However, they reduce the word order factor to the overarching Head Order Principle (HOP), which they claim is *not* a historical factor. See Section 6.4.2 for our criticism of Hawkins and Cutler's (1988) approach.

as long as it predicts that the development of prefixes is hindered.

Table 6.1: Position of negation markers, bound and non-bound

(language-based)			(genus-based)		
	preposed	postposed		preposed	postposed
non-bound	591	216	non-bound	232	90
bound	180	231	bound	63	125
total	771	447	total	295	215

Why, then, are preposed negation words more common than postposed negation words in the first place? One possibility is that a negation marker that appears after its scope can cause the semantic garden path effect — if a negation appears at the end, then a partially constructed semantic representation will turn out to be false, which may cause a processing difficulty. While this is an interesting hypothesis, the ultimate cause of the word order asymmetry is not our primary concern here.

This chapter is an attempt to investigate whether the same logic works for other grammatical categories. Our prediction is that, while psycholinguistic factors cause the boundedness asymmetry for every grammatical category, some grammatical categories tend to be preposed in the first place. As a result, some categories show no or only weak suffixing preference, while some categories show a strong suffixing preference.

Our approach should not be confused with the more common approach in which the synchronic correlation with the affix position and word order is investigated (Hawkins & Gilligan, 1988; Siewierska & Bakker, 1996). We do *not* claim, for example, that if a language has SV order, then the subject affix should precede the verb. Rather, our claim is that if SV order is cross-linguistically more common, it should result in relatively common preposed subject

affixes. Because we do not have statistical data as to how often word order changes, we do not make predictions as to whether the subject is still on the same side of the verb as the subject affixes now, which reflect the past word order.

6.2 Data

The typological database that we employ in this chapter is the the same as that of Chapter 3: Matthew Dryer's data for 16 grammatical categories, consisting of 1,836 languages. Table 6.2 shows whether each grammatical category shows the suffixing preference, overall postposing tendency, and boundedness asymmetry. Detailed figures are available in Tables 3.6, 3.9, and 3.11 from Chapter 3. As discussed in Chapter 3, the suffixing preference can be due to either (i) overall frequency asymmetry, or (ii) boundedness asymmetry, or the combination of both.

In Table 6.2, the meanings of the symbols *, #, (*), and (#) are the same as in the tables in Chapter 3. In each column, the symbol * means that in all of the six language areas, suffixes are more common than prefixes, postposed grammatical morphemes are more common than preposed grammatical morphemes, or postposed grammatical morphemes are more often bound, respectively. (*) means that the same asymmetry is observed only in five out of the six areas. # and (#) mean the opposite trend, i.e. the factor that makes prefixes more common. No category shows the prefixing preference, or the boundedness asymmetry where preposed morphemes are more often bound. However, some categories do show the overall *preposing* tendency: subject markers, negative morphemes, and subordinators.

To take a closer look, consider tense/aspect markers as an example, summarized in Table

Table 6.2: Summary of the differences among grammatical categories

	suffixing preference?	overall postposing?	boundedness asymmetry?
Q	*	*	*
TA	*		*
able			
case	*		*
caus	(*)	(*)	(*)
cop	*	(*)	*
def	*		(*)
dem			
indef	*		*
neg		(#)	(*)
obj			
pl	*	*	
poss			(*)
subj		#	(*)
subord	*	(#)	*
want	(*)		(*)

Table 6.3: tense aspect markers

		Afr	Eur	SEA&O	A-NG	NA	SA	Total	
	Prefix	80	1	48	13	36	2	180	
	Suffix	**92**	**143**	**83**	**206**	**102**	**115**	741	*
overall	Preposed	**209**	35	**134**	32	74	18	502	
	Postposed	120	**198**	133	**245**	**132**	**155**	983	
% bound	Preposed	38.3%	2.9%	35.8%	40.6%	48.6%	11.1%	35.9%	
	Postposed	**76.7%**	**72.2%**	**62.4%**	**84.1%**	**77.3%**	**74.2%**	75.4%	*

6.3. In the table, the top two rows show whether there is a suffixing preference, the middle two rows show whether there is an overall postposing tendency, and the bottom two rows show whether there is a boundedness asymmetry. We can see that tense aspect markers are more often preposed in Africa and Southeast Asia and Oceania (although the margin is very small in the latter area), and thus there is no overall postposing tendency. However, postposed tense/aspect markers are consistently more often bound, which results in the suffixing preference in all six areas.

Based on Table 6.2, we can classify the 16 grammatical categories in the following way.

(34) a. Categories that show the suffixing preference thanks to the boundedness asymmetry: **Q**, **TA**, **case**, **caus**, **cop**, **def**, **indef**, **subord**, **want**

b. Categories that lack the suffixing preference because they tend to be preposed but the boundedness asymmetry cancels it out: **subj**, **neg**, **poss**

c. Categories that lack the suffixing preference because they lack the boundedness asymmetry: **obj**

d. Categories that we probably do not have enough data for: **able**, **dem**, **pl**

The categories in (34a) have the boundedness asymmetry (either * or (*)), and also have the suffixing preference (either * or (*)). The categories in (34b) have the boundedness asymmetry (either * or (*)), but fail to show the suffixing preference. The reason why the categories (34b) fail to show the suffixing preference despite the fact that they show the boundedness

asymmetry is that they tend to be *preposed* in syntax. This can be seen from the table as for negative morphemes and subject markers. While possessives do not show the overall preposing tendency, we include it in (34b) because the failure to find the suffixing preference for possessives is due to the fact that in Australia-New Guinea and South America, there are strong tendencies for a possessive to be preposed, and while postposed possessives are more often bound in these areas, it is not powerful enough to reverse the numbers in morphology.

The categories in (34a) and (35b) confirm our idea that while the boundedness asymmetry is at work for every category, the suffixing preference varies because some categories are more often preposed in the first place.

There is, however, one caveat here. Although subject markers appear to follow a pattern similar to negative markers as long as we see Table 6.2, the underlying story may be quite different. We will see in Section 6.3.1 that there have been a number of other stories where postposed subject suffixes develop in SV languages.

We did not find the boundedness asymmetry for the following four categories: 'able', demonstratives, object markers, and plurals (see Table 3.11). Among these, 'able', demonstratives and plurals (34d) could be attributed to the lack of data. For 'able', suffixing languages do not outnumber prefixing languages in Africa and Australia-New Guinea. For demonstrative, suffixing languages do not outnumber prefixing languages in Eurasia, Australia-New Guinea, and South America. However, the total numbers of languages that have a relevant affix are at most three in these cases. In addition, plural markers also fail to show the boundedness asymmetry, but this may be because that no preposed plural marker is reported in Eurasia.

Without a single example of preposed plural marker, we cannot calculate the proportion of bound morphemes.

On the other hand, object pronominal affixes show no boundedness asymmetry (34c), which is a puzzle for our approach. This indicates that there are additional complicating factors for object markers too, in addition to subject markers (see 6.3.1 again for more discussions).

Besides, among the categories in (34a), some of them appear to show a particularly strong boundedness asymmetry. The most notable example is adverbial subordinators, which are more often preposed in total, but all of the bound adverbial subordinators are suffixes. We will discuss adverbial subordinators in Section 6.3.2. Case markers and tense/aspect markers are also remarkable in that postposed morphemes are more often bound with a wide margin in every large language area. It has already been pointed out that the suffixing preference is particularly strong for case marking (Greenberg, 1963; Dryer, 2011d). We will discuss this in Section 6.3.3.

6.3 Discussion of some grammatical categories

This section discusses grammatical categories that may need special treatment. Section 6.3.1 is dedicated to pronominal affixes. Pronominal affixes have been claimed to lack the suffixing preference. For the subject pronominal affixes, this might not need a special treatment; the lack of the suffixing preference can simply be explained by the fact that the subject is more often preposed in syntax. However, it has also been argued that pronominal affixes do not

necessarily follow the Fossilized Syntax Hypothesis. We will review previous studies on the peculiarity of pronominal affixes and its consequences for our approach. Sections 6.3.2 and 6.3.3 discuss two grammatical categories that appear to have a particularly strong boundedness asymmetry: adverbial subordinators and case markers.

6.3.1 On pronominal affixes on verbs

A number of authors have pointed out that the suffixing preference for pronominal affixes is absent or weak (Hawkins & Gilligan, 1988; Siewierska & Bakker, 1996; Enrique-Arias, 2002).[2] Our data agrees with those observations: both subject and object pronominal affixes fail to show the suffixing preference. For the subject pronominal affixes, the lack of the suffixing preference does not need a special explanation. The SV order is much more common than the VS order, but because the boundedness asymmetry cancels it out, no clear preposing or postposing can be observed in morphology. Object prefixes are more puzzling for our approach, because they do not show the boundedness asymmetry.

It has been pointed out that pronouns do not always follow the 'Fossilized Syntax Hypothesis' (Comrie, 1980; Enrique-Arias, 2002): they can flip their positions before they are morphologized. Givón (1977) points out that sometimes pronominal affixes are historically attached to auxiliaries rather than verbs, which also complicates the picture. This section reviews the relevant literature and discusses consequences for our approach. Both position flips from left to right (e.g. Buryat) and from right to left (e.g. Spanish) have been reported, and

[2]Siewierska and Bakker (1996) presents a detailed typological quantitative analysis of agreement marker positions. They argue that the combination of the Fossilized Syntax Hypothesis and suffixing preference perform best, while the Head Ordering Principle (HOP; see Section 6.4.2) is less successful (when following Hawkins and his colleagues' assumption that all affixes are heads).

thus their overall effect on the suffixing preference is unclear at this point.

In the following sections, we will review previous discussions on cases where pronouns do not appear to preserve older positions in more detail. Sections 6.3.1.4 and 6.3.1.5 are critical reviews of Enrique-Arias (2002) and Himmelmann (2014), who argue the lack of the suffixing preference of pronominal affixes can be explained by the psycholinguistic factors they propose.

6.3.1.1 Mobility of pronominal marking

It has been pointed out that pronominal clitics, presumably an earlier stage of pronominal affixes, are often mobile and do not necessarily reflect the basic word order in the past (Comrie, 1980; Siewierska & Bakker, 1996; Enrique-Arias, 2000, 2002).

For example, some object pronominal proclitics (or prefixes) are preposed in Romance languages. While it might be tempting to speculate that this is a remnant of the older SOV order of Latin, it is a fairly recent development that happened much later than when the SVO basic order was established (Enrique-Arias, 2002). Similarly, Buryat, a Mongolic language, developed subject suffixes despite its SOV word order, based on the marked construction with a right-dislocated pronoun (Comrie, 1980), which we will discuss more in the next section. Other cases where pronouns do not preserve the older word order is examples of so-called Wackernagel's Law, or second-position clitics (See Chapters 3 and 8 of Spencer and Luís (2012b) for a recent overview). Steele (1977), for example, illustrates how second-position clitics in Uto-Aztecan languages developed out of independent pronouns without preserving their original positions.

6.3.1.2 Buryat subject person suffixes

Comrie's (1980) discussion on Buryat, a Mongolic language mainly spoken in Russia, is worth discussing here because it is a clear case where the Fossilized Syntax Hypothesis does not work. Buryat has the following set of person suffixes, which can relatively easily be seen as etymologically related to pronouns:

Table 6.4: person suffixes in Buryat (Comrie, 1980, 88)

	Pronoun (nom.)	Verb ending	Pronoun (gen.)	Noun ending
Sg1	bi	-b	mi´nī	-m(ni), -ni
Sg2	ši	-š	ši´nī	-š(ni)
Pl1	bide	-bdi	manai	-(m)nai
Pl2	ta	-t	tanai	-tnai

Buryat is an SOV language where the subject precedes the verb. He argues that, given historical evidence, Buryat had already been SOV when pronominal suffixes developed, and thus pronominal suffixes are not remnants of the older word order. He argues that pronominal suffixes developed out of the marked right-dislocated construction, rather than the unmarked SOV sentence. Comrie points out the following three factors as to why the right-dislocated construction was chosen as the source of grammaticalization:

(35) a. Because the language had already been overwhelmingly suffixing, introducing more suffixes was preferred over introducing prefixes.

b. The language lacks a construction where an adjunct without stress precedes its head constituent.[3]

[3]While Comrie uses the term *head* here, we believe that it refers to *stem* in this context.

c. A preposed subject was not always adjacent to the verb, while a right-dislocated subject was.

As Bybee (1988) argues, the validity of (35a) can be questioned. As she puts it, 'it is certainly not possible for the speakers who were tending toward this change to test out the results, compare them with existing typologies, and decide against the change' (Bybee, 1988, 356). As we will see later in Section 6.3.1.4, Enrique-Arias (2002) makes an apparently contradictory claim: prefixes are preferred when a language already has many suffixes. The same criticism may apply to (35b) too; while his generalization may be true as a description of existing patterns, it is not clear whether this can be a motivating force that flips the position of pronouns.

6.3.1.3 Person marking on auxiliaries

Another factor that has been suggested to account for the seemingly unexpected position of agreement markers is that they may be originally attached to the auxiliary, not to the verb.

Givón (1977) argues how subject person suffixes were developed out of the SOV word order in Indo-European as well as in Semitic languages. He sketches the development of person suffixes in an SOV language as follows:

(36) a. The most common universal source for tense-aspect morphemes in languages is from a certain group of main verbs which then become 'auxiliary', often take infinitival complements, and eventually fuse to the verb stem as prefixes or suffixes.

b. Of these, the most common precursor for perfective aspects — and thus for eventual past tenses — are ‘be’ and ‘have’.

c. In an OV language, such an auxiliary would appear *after* the main verb, the latter being its complement.

d. In a subject-first language, the subject pronominal agreement — which appears on the main-auxiliary verb rather than on its non-finite complement — will naturally appear as a prefix to that auxiliary verb.

(Givón, 1977, 484–485)

While Givón focuses on Indo-European and Semitic languages, we can find equivalent examples in other language families as well. For example, Chepang, a rGyalrong (Tibeto-Burman) language in China, has an enclitic *=teʔ* that marks second-person subject (Jacques, 2012, 103–106).

(37) ʔamh bəyh=teʔ=newʔ
food give=2=N.PST:3O
‘You give food.’

(Caughley 1982, 89 cited in Jacques 2012, 103)

Jacques hypothesizes that in Proto-rGyalrong, the second person marker was a prefix **t-*. He argues that *=teʔ* in Chepang was derived from **t-leʔ*, a copula morpheme with a prefixed second person marker (Jacques, 2012, 105).

Another example is from the first-person singular actor marker of Choctaw (Mithun, 2003). In Choctaw, all person markers are prefixes, except the first-person singular marker,

-li.

(38) písa-li-tok

see-1sg.agent-past

'I saw it'

Ulrich (1986) cited in Mithun (2003)

She argues that the past tense marker *-tok* can be traced back to **tah* 'finish' in Proto-Muskogean, and *li-* was a prefix on it. This eventually developed into a sequence of suffixes. These examples are strikingly similar to what was proposed by Givón for Indo-European and Semitic languages.

6.3.1.4 Do affixes want to be at both sides of the stem?

Enrique-Arias (2000, 2002) focuses on person marking and provides an account of why it does not show the suffixing preference. He argues for the integration of historical and psycholinguistic factors, as Hall (1988, 1992) does, but from a quite different perspective. His idea can be summarized as follows:

(39) a. Person clitics are often mobile and do not necessarily reflect word order in the past.

b. Because person affixes are semantically less relevant to the stem, they usually appear outside TAM markers (Bybee, 1985), and develop later.

c. When a language already has a lot of affixes on one side of the stem, adding more affixes on the same side is discouraged because a sequence of multiple affixes causes

processing difficulty. Thus, a new affix tends to develop on the other side of the stem.

d. Because TAM markers show the suffixing preference, person markers, which develop later, are preferred to be prefixes.

His argument is mainly based on Spanish, where object personal pronouns came to be preposed in some stage of its history, despite the fact that the VO basic word order had been established earlier than that. Thus Spanish instantiates an example of mobile pronouns (39a). His claim (39a) is well supported by other languages too, including the Buryat example we saw earlier (Comrie, 1980), as well as the examples of Wackernagel's Law observed in Uto-Aztecan (Steele, 1977) and other languages. However, the basis of his central psycholinguistic claim (39c) — affixes stacked on one side of the stem will cause processing difficulty, and therefore it is preferred to add a new affix on the other side of the stem — is unclear at best.[4]

Let us look at his claim (39c) in more detail. First, he argues that a morphological relation is harder to process than a syntactic relation, citing Hall (1992, 176). He also points out in Enrique-Arias (2002, 13) a couple of cases where a periphrastic option is preferred for a longer stem, when a language has both morphological and periphrastic options. For example, the English comparative affix *-er* is used when the stem is short, while the periphrastic construction *more* + adj is used when the stem is longer; the Korean negative proclitic *an=* is less acceptable when the stem is longer. These examples, however, only suggest that a longer inflected word is avoided, and say nothing about affix positions.

[4]See also Myler (2009b, 71–72) for a criticism of Enrique-Arias' approach.

He then claims that it is preferred to put affixes on both sides of the stem when there are multiple affixes, citing the following pieces of evidence:

(40) a. **Data from Indo-European languages.** In Indo-European languages, a high morpheme-to-word ratio correlates with a high prefix-to-suffix ratio (Cowgill, 1966, 131).

b. **Bybee et al.'s comment on verb-final languages.** "Bybee, Pagliuca, and Perkins (1990: 8) observe that the only V-final languages with a high ratio of prefixing are those that have a high degree of affixation" (Enrique-Arias, 2002, 14)

c. **Nouns vs. verbs.** Prefixing is more common on verbs than it is on nouns (Hall, 1988, 332). According to Hall, while 60.7% of verbal morphology is suffixes only, 81.6% of nominal morphology is suffixes only. Enrique-Arias claims that this is because typically, verbs have more affixes than nouns.

d. **Georgian.** In Georgian, most agreement markers are prefixes, but only the third person subject is a suffix. Two agreement markers (for subject and object) co-occur only when a third person subject is involved, i.e. they occur on the different sides of the stem.

e. **Correlation with the number of affixes.** Languages with more affixes are more likely to have both prefixes and suffixes.

A serious problem with the evidence he cites is that they are just additional distributional facts, but not evidence for the proposed psycholinguistic factor behind them. This means

that his argument has the risk of being circular. For example, the fact that prefixes are more common on verbs (40c) may be just another way to capture the high frequency of person prefixes, which is exactly what we want to explain.

Even worse, a closer examination of the claim (40c) reveals that the higher ratio of suffixes in nominal morphology is an artifact of the analysis. From Table 6.5 we can see the three most common classes of nominal morphology are case, possessive, and plural. Among them, case and plural show a strong suffixing preference, while possessive has more prefixes, which is expected from the Fossilized Syntax Hypothesis alone. In addition, case markers probably have another independent reason to be more often postposed (see Section 6.3.3). The problem is that his largest data source (Stassen's data given in Hawkins and Gilligan (1988)) does not contain the data of possessive markers. Due to this, the two most common nominal affixes in his data are case and plural, which happen to have the strong suffixing preference. A further problem is that the third largest class in his data is nominalizers, which also show a strong suffixing preference. However, including nominalizers here is not appropriate for the purpose of Enrique-Arias' argument, because a nominalizer is not an affix attached to a noun.

His evidence (40e) is also problematic. He calculates the ratio of exclusively prefixing or suffixing languages classified by the existence of the following five grammatical categories: tense, mood, aspect, subject agreement and object agreement. As can be seen from Table 6.6, when a language has more types of affixes, it is less likely that it is exclusively prefixing or suffixing.

Table 6.5: Counts of nominal affixes (Hall, 1988, 332)

	prefix	both	suffix
Stassen			
case	0	0	65
definite	7	0	18
indef	1	0	6
Gilligan			
definite	5	0	5
possessive	10	4	14
genitive	5	0	13
plural	5	3	29
nominalizer	3	2	30
case	0	0	20

Table 6.6: Number of exclusively suffixing or prefixing languages (Enrique-Arias, 2002, 16)

# of affixes	1		2		3		4		5	
only suf or pref	20	91%	27	82%	26	49%	11	28%	4	20%
both suf and pref	2	9%	6	18%	27	51%	28	72%	16	80%

However, it is expected that, just by chance, the more affixes a language has, the less likely they are to be all prefixes or suffixes. In fact, if we assume that the position of each affix is independently determined by a toss of a fair coin, the probability that five affixes are all prefixes or all suffixes is $2/2^5 = 6.3\%$. Thus, the fact that 20% of the languages in Enrique-Arias' data are exclusively suffixing or prefixing indicates that affixes tend to occur on the same side of the stem more than expected by chance, contrary to his claim.[5] Furthermore, the fossilized syntax hypothesis alone predicts that those five grammatical categories do not line up on the same side. O/V correlates with V/Aux (Dryer, 1992a), and thus it is predicted that

[5]This calculation involves a number of simplifications, and a closer approximation may increase or decrease the chance of getting an exclusively prefixing or suffixing language. For example, if suffixes are three times more common than prefixes, then the probability that five affixes are all prefixes or all suffixes is $(1/4)^5 + (3/4)^5 = 23.8\%$. On the other hand, a single grammatical category can include both prefixes and suffixes, which lowers the probability of an exclusively prefixing or suffixing language. In any case, it is too premature to claim that the number that Enrique-Arias obtained is significant.

the tense/aspect marker and object marker are typically on different sides.

Romance languages are special in that the basic word order changed from OV to VO while maintaining TAM suffixes. This created the situation where the flip of the position of the object markers can prevent exclusive suffixing. Thus, in order to say that the position flip plays a crucial role in explaining the weaker suffixing preference of person markers, one needs to show that the Spanish-like scenario has repeated many times around the globe. Even though the word order change from OV to VO has been claimed to be relatively common (see Section 6.4.3), it is difficult to believe that this scenario had an overwhelming effect on the current frequency distribution, compared to other possible scenarios.

A further problem with Enrique-Arias' account is that the exact opposite case has been claimed, i.e. a morpheme changes its position to become a suffix *because* the language was already suffixing. Buryat (Comrie, 1980) we saw earlier is such an example. We saw the three factors that he points out in (35), which are shown here again:

(41) a. Because the language had already been overwhelmingly suffixing, introducing more suffixes was preferred over introducing prefixes.

b. The language lacks a construction where an adjunct without stress precedes its head constituent.

c. A preposed subject was not always adjacent to the verb, while a right-dislocated subject was.

Among his three points, (41a,b) are contradictory with Enrique-Arias' discussions. Comrie argues that a suffixing language prefers to have more suffixes, while Enrique-Arias argues that a suffixing language prefers to have some prefixes. Again, we have to say that their account is evoked only for interpreting a single case in a post-hoc way.

Because Enrique-Arias' claim is based solely on Spanish while Comrie's claim is based solely on Buryat, and because neither of the papers discusses how affixes can cause a problem for prosodic structure in detail, we do not have evidence to support one account over the other; it is possible that depending on the prosodic profile, some languages prefer to stack affixes on one side, while others do not. In any case, we are still far away from being sure how their points affect the overall typological frequency of prefixes and suffixes.

6.3.1.5 Himmelmann's account on person marking

Himmelmann (2014) provides another account for why person marking lacks a suffixing preference. He argues that some kinds of preposed grammatical morphemes can signal a particular category to follow. He calls such constructions *target-specific constructions*, and lists examples as in Table 6.7.

Table 6.7: Target-specific constructions in English (Himmelmann 2012: 28)

Marker	Target
Article	Noun (e.g. the bicycle)
Auxiliary	Non-Finitite verb (e.g. might be going, was done)
Preposition	Noun (e.g. by chance, of the city, to John)

In contrast, personal pronouns do *not* initiate such a target-specific construction, because it is difficult to predict what follows. He points out examples where a clause which is initiated

by a personal pronoun is canceled after hesitation, like the following:

(42) Pear Story 9

[.7] one of them .. whistles back to the guy on the bicycle,

"Here's you hat," [sic!]

[.35] or **he** [.4] I don't know,

and he goes and takes it,

His claim is that, when a morpheme initiates a target-specific construction, it is less likely to be grammaticalized into an affix. This explains, he argues, why the morphologization of preposed articles, prepositions, auxiliaries and so on are hindered, while we find as many person prefixes as person suffixes.

His definition of the target-specific construction is not entirely clear. However, his argument seems to fail for adverbial subordinators. Because an adverbial subordinator takes a clause, it is difficult to predict which lexical category follows. Himmelmann himself argues that a word that projects a clause is not an initiator of a target-specific construction. Hence, it should predict that adverbial subordinator prefixes are common, which contradicts with what we found (Section 6.3.2). Indeed, it has been argued that adjacency with a particular lexical category is one of the preconditions for the development of an affix (Comrie, 1980). This appears to imply that a target-specific construction is more likely to develop a prefix, contrary to Himmelmann's claim.

6.3.2 Adverbial subordinator

Adverbial subordinators show an unusually strong suffixing preference. Overall, preposed subordinators are more common than postposed subordinators. Nonetheless, when it is an affix, it is always a suffix; no clear example of an adverbial subordinator prefix is attested. This striking pattern may be explained as follows.[6]

First, the distribution of adverbial subordinators can be captured by (i) the dominance of preposing *and* (ii) the correlation with the order of the verb and object. As a result of the interaction of these two factors, postposed subordinators in VO languages are extremely rare, as can be seen from Table 6.8. The table includes both bound and non-bound subordinators.

Table 6.8: Correlation between V/O and the position of adverbial subordinator

basic word order	InitSub	FinalSub
OV	66	155
VO	303	5

This situation is similar to the order of the relative clause and noun (N/Rel): the distribution of N/Rel can be captured by the dominance of NRel and the correlation between VO and NRel. As mentioned in Section 5.1, Hawkins' two principles can neatly explain this: MaOP explains the dominance of NRel, and MiD explains the correlation between V/O and N/Rel. The same explanation can apply to adverbial subordinators: MaOP explains the dominance of preposed subordinators, while MiD explains its correlation with V/O.

Second, in order for a word to become an affix, it must be adjacent to a particular lexical

[6]Our data does not include languages in which the adverbial subordinator occurs in the middle of a clause. We suppose that this decision does not seriously affect the overall picture, as such languages are rare. In WALS (Dryer, 2011b), among 502 languages with non-bound adverbial subordinators, only 8 languages place them not at the edge of the subordinate clause.

category (Comrie (1980); *exclusive adjacency* in Hall (1988, 334); Hall (1992, 162)). In a strict verb-final language with a postposed subordinator, it is guaranteed that the verb and subordinator are adjacent to each other, thus providing a precondition for the development of an affix. On the other hand, in a verb-final language with a preposed subordinator, the lexical category of the word that immediately follows the subordinator is not consistent. The majority of VO languages are SVO, so the lexical category of the clause-initial word is not consistent either.

Here our explanation is similar to that of Bybee et al. (1990), which we reviewed in Section 3.6.2: the fact that verb-final languages outnumber verb-initial languages play a crucial role. While Bybee and her colleagues use this for any kind of (verbal) affix, the frequency difference between verb-initial and verb-final languages has a particularly strong effect on the suffixing preference of the subordinator, because the subordinator tends to be at the edge of a clause more often than other categories.

This still does not explain, however, why a subordinator prefix is never attested, even among verb-initial languages. Table 6.9 shows correlations between the positions of subordinators and S/V/O basic word order, taken from our database. We can see that our database has 93 verb-initial languages with a preposed subordinator. It is rather surprising that none of these languages has developed a subordinator prefix, even when we accept the idea that there is a universal psycholinguistic force against prefixing. Closer examinations of these verb-initial languages will be needed.

Table 6.9: Correlation between basic word order and the position of adverbial subordinator

	bound		non-bound	
basic word order	initial subord.	final subord.	initial subord.	final subord.
SOV/OSV	0	57	63	96
SVO	0	1	210	3
VSO/VOS	0	0	93	1
OVS	0	0	3	2

6.3.3 Stronger suffixing preference in case marking?

It has been noted that case marking shows a particularly strong suffixing preference (Greenberg, 1963; Dryer, 2011d). Our data basically confirm this. Languages with case suffixes outnumber languages with case prefixes with a wide margin in all of the six large linguistic areas. Kahr (1976) argues that new case forms only develop through the suffixation of postpositions, based on languages in Eurasia. Reh (1986) points out interesting cases in Niger-Congo languages where a preposition develops into a verbal suffix, instead of a nominal prefix. It appears that there is some force that prevents a preposition from being morphologized into prefixes. These studies, however, do not provide explanations as to why there is an asymmetry.

There is one possible reason to believe that case prefixes are especially unlikely to develop. It has been pointed out that in order to signal argument relations, verb-initial languages tend to employ head-marking (i.e. person marking on verbs), while verb-final languages tend to employ dependent-marking (i.e. case marking on nouns) (Siewierska & Bakker, 1996; Hawkins, 2002; Dryer, 2002). From Table 6.10, taken from Dryer (2002), we can see that verb-final languages employ case marking (including both case affixes and adpositions) more often than

verb-medial or verb-initial languages.

Table 6.10: Proportions of languages with case distinction (Table 4 of Dryer 2002: 154)

	SOV	SVO	V-initial
% of genera	62% (85/138)	20% (16/82)	41% (15/37)
% of languages	72% (181/253)	14% (26/190)	47% (28/59)

We can confirm the same tendency from our data, as shown in Table 6.11. It is puzzling that even among VO languages case suffixes are more common than case prefixes, but we can still say that the overwhelmingly high frequency of case suffixes can largely be reduced to the fact that case affixes themselves are much more common in OV languages.

Table 6.11: Correlation of case and word order

	case prefix	case suffix	preposition	postposition
SOV&OSV	4	300	14	495
SVO	23	50	333	36
VSO&VOS	12	26	132	8
OVS	0	2	3	8

Hawkins argues that this divide can be predicted by the Maximize Online Processing (MaOP) principle that he proposes. MaOP states that the parser prefers structures where more syntactic and semantic properties of a sentence can be established early in a sentence. For argument relations, this means that case marking is encouraged in verb-final languages, while person marking is encouraged in verb-initial languages. However, MaOP does not explain why case marking is even less common in SVO languages than in verb-initial languages. As discussed by Dryer (2011d), this might be explained by another functional factor: in SVO languages, S and O can be distinguished by word order alone, even when either S or O is zero. In verb-initial or verb-final languages, some overt marking is necessary to disambiguate them.

This explains why the suffixing preference of case marking is stronger than expected. While case prefixes are more likely to develop in VO languages, VO languages have less motivation to have case marking in the first place. This works as an additional factor why case prefixes are rare.

An assumption behind this argument is that while case affixes are less needed in VO languages, adpositions are equally well employed in OV and VO languages. This assumption is justified because while case affixes typically mark core arguments, adpositions typically mark a wider range of oblique arguments or adjuncts. (We can speculate that this fact can, in turn, be reduced to the fact that core arguments are more frequent and thus there is more motivation to reduce their forms.) While SVO languages can distinguish two core arguments by their positions relative to the verb alone, they cannot distinguish more arguments or adjuncts; thus, adpositions are more equally needed in VO and OV languages, while case marking is less needed for VO languages.

A possible complicating factor that we might have to consider is that there are languages with case affixes that do not distinguish subject and object; our discussion would be weakened if such languages are the majority. Table 6.12 shows the number of languages with case systems that do not distinguish subject and object (NoSuOb) and those with case systems that distinguish subject and object (SuOb), classified by basic word order (data based on Matthew Dryer, p.c.). The data is compatible with our expectation: among languages with case affixes, the distinction of subject and object cases is more common in OV languages than in VO languages. In other words, if we focus on case systems that have subject/object distinction,

the tendency that OV languages employ them more often is stronger.

Table 6.12: Case systems with and without subject/object distinctions

		Afr	Eur	SEA&O	A-NG	NA	SA	Total
OV	NoSuOb	2	7	3	8	7	6	33
	SuOb	17	60	19	35	19	12	162
	% SuOb	89.5%	89.6%	86.4%	81.4%	73.1%	66.7%	83.1%
VO	NoSuOb	17	3	4	3	7	3	37
	SuOb	13	18	2	9	7	0	49
	% SuOb	43.3%	85.7%	33.3%	75.0%	50.0%	0.0%	57.0%

Note that we do not expect a similar effect for other grammatical categories, such as tense, aspect and plural, because their primary function is to mark properties of the object or situation being denoted, rather than establishing grammatical relations.

6.3.3.1 Person marking and word order

Hawkins (2002) also claims that head-marking of argument relations is more common in verb-initial languages. At a glance, this appears to make the wrong prediction: in verb-initial languages, both the subject and object follow the verb, so we should expect stronger suffixing preferences for pronominal markers on verbs. This is not supported by our data; suffixing preferences are weak for pronominal markers.

However, it is unlikely that the advantage of head marking in verb-initial languages predicts that stronger suffixing preferences for pronominal markers on verbs. First of all, unlike case marking, the frequency difference is small, if any, as can be seen from Table 6.13 (Dryer, 2002).

Second, because verb-initial languages are much less common than verb-final languages, its

Table 6.13: Proportions of langugaes with rich verb agreement (Table 4 of Dryer 2002: 154)

	V-initial	SVO	SOV
% of genera	62% (29/47)	47% (49/104)	54% (80/147)
% of languages	56% (48/86)	47% (94/213)	49% (140/283)

effect on the overall typological frequency is expected to be small. In fact, because verb-final languages more often have person markers than verb-medial languages do, it is even possible that the correlation with the basic word order has an effect of the other direction, because verb-final languages are more common than verb-initial languages. Third, as we already discussed, person markers often do not preserve the past word order, and thus the consequence of the correlation between word order and head marking for the suffixing preference will be even less clear.

6.4 Further discussion

Section 6.4.1 discusses a possible problem with the assumption that an affix develops out of a word with the equivalent meaning. Section 6.4.2 examines Hawkins' claim that their Head Ordering Principle (HOP) cannot be replaced by a historical factor. Section 6.4.3 goes back to Givón's claim that SOV was once more frequent, and discusses its relevance to our approach.

6.4.1 Unusual sources of grammaticalization

We have assumed so far that an affix develops from a non-bound morpheme with the equivalent meaning. This, however, is not always the case. An affix may emerge as a result of a reanalysis of existing words, or may undergo a semantic change after it is established as an affix.

To take an example, we have assumed that case markers develop from adpositions, but there are other possible sources of case markers, including adjectives, adverbials and indexicals (Kulikov, 2008). The Georgian nominative (absolutive) case *-i*, for example, was established as a result of a reanalysis of the postposed demonstrative pronoun *-igi*.

To take another example, we have assumed that bound tense and aspect markers develop out of auxiliaries. While this has been claimed to be the most common pattern (Givón, 1977), there are other grammaticalization paths as well. Mithun (2003), for example, points out that in Navajo, the iterative aspect prefix *ná-* is historically derived from an adverb meaning 'again', which occurs before the verb. This type of development appears to be behind the unusually rich prefix system of Athabaskan languages, although we do not have an explanation as to why this particularly happened in Athabaskan languages (See also Givón (2000)). An example of the mirror image of Navajo can also be found. Bybee et al. (1990, 12–13) discuss Car (Nicobarese) (India; Austroasian), a verb-initial suffixing language. In this language, they argue, aspect suffixes are derived from locative and directional adverbs or prepositions that follow the verb, rather than from auxiliaries.

It is difficult to quantitatively estimate how these exceptional historical sources would affect our overall account. We can at least say that exceptional historical sources appear to play important roles in unusual prefixing OV languages, such as Athabaskan and rGyalrong, as well as some suffixing VO languages, such as Car.

6.4.2 On the Head Order Principle

The Head Order Principle (HOP) is one of the two components of Hawkins and Cutler (1988)'s account of the distribution of affixes, the other being the psycholinguistic suffixing preference. They welcome the integration of the diachronic aspect, but claim that the Fossilized Syntax Hypothesis cannot replace HOP (ibid; 311). In this section we critically evaluate the role HOP plays in their account and argue that the Fossilized Syntax Hypothesis can replace HOP, and the reference to the notion of 'head' is superfluous and misleading.

HOP is concerned with the correlation between affix position and word order, rather than the dominance of suffixing over prefixing. In their account, these two statistical generalizations, correlation and dominance, can complementarily explain the overall distribution of affixes. This results in a 'three quarters' situation shown in Table 6.14. Suffixes in OV languages are the most common type, as they satisfy both their correlation with the basic word order and the suffixing preference, followed by prefixes in VO languages or suffixes in VO languages.

Table 6.14: The 'three quarters' relation of affix position and basic word order

	prefix	suffix
VO	✓	✓
OV	-	✓

This is similar to the explanation for the order of the head noun and relative clause (N/Rel); the distribution is described by the dominance of NRel and the correlation with V/O, each of which has been given a psycholinguistic account (Hawkins, 2002).

They formulate HOP as follows:

(43) **Head Ordering Principle (HOP)**

> The affixal head of a word is ordered on the same side of its subcategorized modifer(s) as P is ordered relative to NP within PP, and as V is ordered relative to a direct object NP. (Hawkins & Cutler, 1988, 290)

A problem with HOP is that *head* is a theoretically charged notion, and because of that, its empirical consequence highly depends on a particular theoretical framework. Hawkins and Cutler (1988) assume that all affixes are heads. This means that, in their account, HOP simply predicts that OV languages tend to be suffixing, while VO languages tend to be prefixing. The prediction would be completely different if we adopt the Righthand Head Rule (Williams, 1981), which states that the the head is always on the right side in morphology. Myler (2009b) makes different predictions for different grammatical categories by closely following the minimalist literature.

What Hawkins and Cutler (1988) try to capture with HOP is the simple fact that OV languages tend to be suffixing, while VO languages tend to be prefixing. The notion of *head* is defined such that this correlation can be captured. However, there could be a psycholinguistic or historical explanation for this correlation. Unless the explanation crucially relies on a syntactic theory that uses the notion of head, it could be superfluous and misleading to concern the definitional issue of *head*.[7]

Just as there are psycholinguistic, historical, and formal approaches for the suffixing preference, we can think of psycholinguistic, historical, and formal approaches to the correlation

[7] See also Ch.3 of Hall (1992) for an extensive critical evaluation of the notion of *head* in the context of the suffixing preference.

between affix position and word order.[8] First, there could be psycholinguistic advantages in aligning affix position and word order in a certain way. Vennemann (1973, 40ff.) advocates the Natural Serialization Principle, which states that a language tends to use consistent word order for the operand–operator relation. Hawkins' Cross-Category Harmony (CCH), or more recently Minimal Distance (MiD) (Hawkins, 2002, 2004, 2014), also predicts consistent word order across categories in a language. Thus, one possibility is that these psycholinguistic principles are also at work in morphology. However, a simple head/non-head relationship is not the best way to capture word order universals (Dryer, 1992a); instead, Dryer proposes the Branching Direction Theory (BDT):

(44) Branching Direction Theory (revised version):

> Verb patterners are nonphrasal categories or phrasal categories that are not fully recursive, and object patterners are fully recursive phrasal categories in the major constituent tree. That is, a pair of elements X and Y will employ the order XY significantly more often among VO languages than among OV languages if and only if X is not a fully recursive phrasal category in the major constituent tree and Y is a fully recursive phrasal category in the major constituent tree.

(16) of Dryer (1992a, 114)

While BDT is basically compatible with Hawkins' processing-based account of word order (Dryer, 1992a, 131–132), it would make the idea of applying it to morphology less promis-

[8] We do not discuss formal approaches in any detail here, but the FOFC-based accounts of the suffixing preference (Myler, 2009b; Biberauer et al., n.d.) clearly make predictions about the correlation between affix position and word order.

ing. Because there is no 'recursive phrasal category' in morphology, BDT does not make predictions about morphology. If we assume that morphology can be 'fully recursive' and extend BDT so that not only 'phrasal' but any fully recursive categories are relevant, then it loses the ability to make predictions in syntax, as many of 'verb patterners' can include recursive morphology.

Hawkins' own account of word order based on his processing-based principles such as MiD is also difficult to extend to morphology. As we saw in the example of English adjectival phrases in Section 2.2, the fact that a simple adjective without a phrase is preposed in English is explained by arguing that a short preposed element does not cause serious processing difficulty associated with center-embedding. Hawkins (2014, 101) argues that because an adjectival phrase is on average shorter than a relative clause or possessor noun, its center-embedding is less often avoided in the environment [Prep [[ ] N]], as can be seen from Table 6.15.

Table 6.15: Noun phrase word order in prepositional languages (Hawkins 2014, 102)

AdjN	32%	NAdj	68%
PosspN	12%	NPossp	88%
RelN	1%	NRel	99%

A corollary of this argument is that the effect of MiD on morphology must be even weaker than N/Adj, as the smaller the average size of the relevant category is, the smaller the effect of MiD on morpheme order is.

Another problem of psycholinguistic accounts is that there is no historical evidence that a configuration is chosen over other possibilities so that it does not violate HOP. As Bybee et al. put it, '..grammatical material develops in whatever position it happens to be in when

grammaticization occurs' (Bybee et al., 1990, 19).

Second, the correlation could be reduced to a historical factor. More specifically, it could be reduced to the Fossilized Syntax Hypothesis, namely the assumption that morpheme position reflects the word order in the past. Typical historical sources of affixes, such as auxiliaries and postpositions, are clear *verb patterners* in the sense of Dryer (1992a, 2009); their position relative to the verb or noun they take correlates with the position of the verb object. This appears to explain why OV languages have more suffixes than VO languages do.

Hawkins and Cutler (1988, 311) reject the idea that the historical account can replace HOP, based on the claim that some grammatical categories, including definiteness, indefiniteness, gender and plural, are more often historically derived from modifiers, rather than heads, citing Greenberg (1978), but they are still more suffixing in OV languages. This argument seems to be based on the simplistic assumption that VO languages are head-initial, while OV languages are head-final, for every category. However, articles and plural markers have been claimed to be verb patterners in Dryer (1992a, 2009), i.e. they tend to be preposed in VO languages and postposed in OV languages, no matter whether they should be analyzed as modifiers or not. Thus, the fact that they are more often suffixes in OV languages than in VO languages follows from the Fossilized Syntax Hypothesis, and the introduction of HOP only complicates the picture.

6.4.3 Givón's approach revisited

An implicit assumption of the discussion so far is that word order frequencies do not radically change over time. We simply take the contemporary statistic as a hypothetical past state from

which affixes develop. This assumption can be challenged, however. In fact, Givón's (1979) account on the suffixing preference crucially relies on the assumption that the frequencies of the word order types were different in the past, as we saw in Section 3.6.1. This section investigates whether Givón's historical account can provide an alternative explanation for the differential suffixing preference.

While our approach and Givón's approach starts from quite different assumptions, it is not straightforward to empirically distinguish the two. Taking a tense/aspect marker as an example, the logic of these two approaches can be summarized as follows:

(45) a. **psycholinguistic account:**

Tense/aspect affixes tend to be postposed, because postposed grammatical morphemes are more likely to be morphologized.

b. **Givón's historical account:**

Tense/aspect affixes tend to be postposed, because postposed grammatical morphemes were more frequent in the past, and today's morphology reflects the older word order.

Although these two approaches are based on quite different assumptions, they often make the same prediction, because the OV word order correlates with suffixing (see the discussions on the Head Order Principle in the previous section). The lack of overall frequency asymmetry and the presence of the boundedness asymmetry (Chapter 3) are also basically compatible with Givón's account; If we follow Givon's approach, there were the overall fre-

quency asymmetry in the past, but the asymmetry disappeared due to the increasing number of VO languages. However, because affix positions mirror older word order, affixes are skewed towards postposing, and thus the boundedness asymmetry can be observed.

However, the predictions of the two approaches above diverge when we look at grammatical categories that do not have a tendency to be postposed in OV languages. Not every grammatical category is more often postposed in OV languages. Dryer (1992a, 2009) presents lists of pairs that correlate and do not correlate with V/O. According to him, among the grammatical categories we are investigating, postposing of the following four categories does not correlate with, or inversely correlate with, the OV order:

(46) a. No correlation with V/O:

- Demonstrative

b. More often preposed in OV languages:

- Subject
- Object (by definition)
- Possessive

Because the grammatical categories discussed in Dryer (1992a, 2009) do not exactly match our dataset, we replicated the same calculations based on our data, which are summarized in Table 6.16. Our methodology is the same as what we did in Chapter 3. In each of the six large language areas, we checked whether each grammatical morpheme is more

often preposed in VO languages than in OV languages. We marked the row with symbols * (postposed) and # (preposed) for significant results.

As we argued above, the boundedness asymmetry follows from both our approach and Givón's approach. However, while our approach considers that the boundedness asymmetry is due to a universal psycholinguistic preference, Givón's approach argues that the boundedness asymmetry arises because grammatical morphemes tend to be postposed in OV languages, which were more common in the past. Givón's approach, then, should predict that when a grammatical category has no tendency to be postposed in OV languages, then it should not show the boundedness asymmetry.

Table 6.17 summarizes the relation between the suffixing preference, overall postposing tendency, boundedness asymmetry, and postposing tendency in OV languages. The first three columns are repetitions of 6.2; here the fourth column is added to facilitate the comparison with Givón's approach.

Our approach predicts the universal boundedness asymmetry, but the four categories we did not find any boundedness asymmetries are 'able', demonstrative, object and plural. We have argued that the failure of detecting the boundedness asymmetry for 'able', demonstrative and plural is probably due to the scarceness of data (Section 6.2). When excluding these three categories, the only truly problematic category is object pronominal affixes, which may be due to additional complicating factors that flip the position of pronominal affixes, as we discussed in Section 6.3.1.

Givón's approach appears to face a problem in more grammatical categories. For example,

Table 6.16: Correlation between V/O and proportions of postposed grammatical morphemes (non-bound)

		Afr	Eur	SEA&O	A-NG	NA	SA	
Q	% postposed in VO	81.1%	5.9%	68.8%	**81.3%**	26.7%	0.0%	
	% postposed in OV	**89.5%**	**77.4%**	**94.3%**	74.1%	**60.0%**	**58.3%**	(*)
TA	% postposed in VO	14.3%	6.7%	10.1%	50.0%	16.7%	38.5%	
	% postposed in OV	**36.4%**	**94.2%**	**90.9%**	**72.1%**	**88.5%**	**82.1%**	*
able	% postposed in VO	3.7%	0.0%	20.0%	33.3%	0.0%	n/a	
	% postposed in OV	**33.3%**	**93.3%**	**94.3%**	**80.0%**	**75.0%**	100.0%	(*)
case	% postposed in VO	9.1%	16.7%	0.5%	11.5%	6.2%	46.7%	
	% postposed in OV	**91.5%**	**96.4%**	**98.0%**	**97.0%**	**100.0%**	**97.6%**	*
caus	% postposed in VO	11.8%	0.0%	0.0%	**100.0%**	0.0%	**100.0%**	
	% postposed in OV	**33.3%**	**100.0%**	**87.5%**	57.1%	**100.0%**	75.0%	
cop	% postposed in VO	8.3%	4.8%	5.3%	66.7%	48.1%	37.5%	
	% postposed in OV	**66.7%**	**100.0%**	**94.0%**	**83.3%**	**95.0%**	**92.3%**	*
def	% postposed in VO	**88.2%**	5.9%	49.4%	50.0%	16.1%	0.0%	
	% postposed in OV	88.0%	**12.5%**	**86.4%**	**60.5%**	**78.9%**	**12.5%**	(*)
dem	% postposed in VO	**91.2%**	**15.2%**	**84.3%**	51.4%	**33.8%**	**22.2%**	(#)
	% postposed in OV	51.7%	0.9%	32.7%	**59.7%**	22.0%	12.1%	
indef	% postposed in VO	**71.7%**	0.0%	33.3%	60.0%	0.0%	0.0%	
	% postposed in OV	57.1%	**33.3%**	**76.0%**	**87.0%**	**61.5%**	**37.5%**	(*)
neg	% postposed in VO	**55.0%**	10.0%	19.8%	**38.7%**	0.0%	10.0%	
	% postposed in OV	29.0%	**14.6%**	**29.5%**	25.3%	**30.0%**	**62.5%**	
obj	% postposed in VO	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	#
	% postposed in OV	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
pl	% postposed in VO	68.0%	n/a	27.1%	20.0%	9.1%	25.0%	
	% postposed in OV	**100.0%**	100.0%	**92.0%**	**100.0%**	**100.0%**	**100.0%**	(*)
poss	% postposed in VO	**83.1%**	**15.4%**	**59.4%**	**47.6%**	**56.5%**	**28.6%**	#
	% postposed in OV	36.6%	6.1%	2.7%	28.3%	15.8%	0.0%	
subj	% postposed in VO	**9.1%**	**13.5%**	**26.7%**	**15.4%**	**79.5%**	**40.5%**	#
	% postposed in OV	1.2%	0.0%	0.7%	0.9%	0.0%	6.7%	
subord	% postposed in VO	1.7%	0.0%	0.8%	0.0%	0.0%	9.1%	
	% postposed in OV	**30.8%**	**43.2%**	**77.4%**	**50.0%**	**86.7%**	**92.0%**	*
want	% postposed in VO	4.3%	0.0%	0.0%	20.0%	5.9%	50.0%	
	% postposed in OV	**50.0%**	**68.8%**	**87.0%**	**66.7%**	**92.9%**	**76.9%**	*

Table 6.17: Summary of the differences among grammatical categories

	suffixing preference?	overall postposing?	boundedness asymmetry?	postposing in OV?
Q	*	*	*	(*)
TA	*		*	*
able				(*)
case	*		*	*
caus	(*)	(*)	(*)	
cop	*	(*)	*	*
def	*		(*)	(*)
dem				(#)
indef	*		*	(*)
neg		(#)	(*)	
obj				#
pl	*	*		(*)
poss			(*)	#
subj		#	(*)	#
subord	*	(#)	*	*
want	(*)		(*)	*

we did not find a tendency that negation markers are more often postposed in OV languages, but as discussed at the beginning of this chapter, negation markers show the boundedness asymmetry. Givón's account cannot explain why postposed negation markers are more often bound. However, because the overall preposing tendency or boundedness asymmetry of negation markers is not statistically significant in our language-area-based method, this counterargument may not be very powerful.

More problematic categories for Givón's approach are subject and object markers, as well as possessives. These categories are preposed significantly more often in OV languages, which means that they should show the *prefixing preference* if we follow Givón's account, contrary to the data. These three categories are all pronominal, and thus the lack of the prefix-

ing preference may be due to the additional factors we discussed in Section 6.3.1. In order for this account to work, however, we have to say that the development of pronominal suffixes in OV languages is significantly common across the world. While we have seen such examples in Section 6.3.1.2 as well as Section 6.3.1.3, it is difficult to determine whether this is in fact the dominant pattern at this point.

6.5 Conclusion

In this chapter, we argued that if we take into account the frequency differences of non-bound counterparts of affixes, we can basically account for the difference in the degree of the suffixing preference between grammatical categories.

The argument put forward in this chapter is basically compatible with any psycholinguistic account, as long as it predicts that suffixes are preferred in general, regardless of their meaning or grammatical function. While a psycholinguistic account cannot make predictions about the differences between grammatical categories, the combination of psycholinguistic and historical factors can make better predictions than a historical account alone.

Chapter 7

More asymmetries in morphology

7.1 Introduction

So far we have focused on a particular kind of morphological asymmetry, namely the suffixing preference. In this chapter, we extend our scope and discuss the following two additional asymmetries in words.

(47) a. Prefixes are phonologically more independent from stems than suffixes.

b. Similar asymmetries can be found in clitics and compounds. That is, (i) a short morpheme at the beginning is less common than a short morpheme at the end, and (ii) a short morpheme at the beginning is phonologically more independent from the rest than a short morpheme at the end.

These two predictions are direct consequences of our account of the suffixing preference. First, as we have already discussed, when a language has prefixes, they are expected to have properties that counterbalance their inherent processing difficulty. One of the strategies a language can utilize is to have a phonological boundary between the prefix and stem, so that

prefixes are easily separated out from the rest of the word. In other words, a prefix is expected to behave as an independent *phonological domain*. This chapter discusses the phonological independence of prefixes, based on the literature on phonological wordhood and an examination of the AUTOTYP database.

Our processing-based account also predicts (47b). Unlike some other approaches, our processing-based explanation for the suffixing preference does not rely on the semantic or grammatical status of stems or affixes. Thus, our approach naturally extends to cases where the short morpheme is not an affix, but a clitic or a part of a compound. A short morpheme at the beginning of a unit is harder to detect, no matter their semantic or syntactic properties. The predictions are two-fold. First, preposed short morphemes are less common than postposed short morphemes. Second, preposed short morphemes are more phonologically independent than postposed short morphemes.

Unfortunately, available typological databases of clitics and compounds are limited, so it is difficult to quantitatively investigate their typological asymmetries. Inevitably, our discussion will focus on a few well-studied languages. We can see that existing evidence, at least, is in accordance with our prediction. For clitics, we predict that preposed clitics (i.e. proclitics) are less common than postposed clitics (i.e. enclitics). For compounds, we predict that compounds are more often left-branching.

Our predictions about the position and phonology of morphemes can also be summarized in the following way. Consider the following two structures in (48). Within a phonological unit, a semantic boundary near the end is more easily identifiable than a boundary near the

beginning, because more information is available when the parser reaches the boundary. This means that the structure (48a) is preferred over the structure (48b).[1]

(48) a. (long] [short)

b. (short] [long)

When the long morpheme is the stem and the short morpheme is the affix, this is equivalent to the suffixing preference; the stem-affix order is preferred over the affix-stem order. Our claim in this chapter is that there is no reason to limit our prediction to the combination of a stem and affix.

Alternatively, we can formulate our prediction in the following way. If there is a morpheme boundary soon after the beginning, it is likely to be signaled by phonological cues. This is because it is difficult to detect the boundary if no phonological cues are available. A morpheme boundary near the end is less often signaled by phonological cues, because the parser is more likely to have collected enough information to identify the boundary even if there is no phonological cue. Therefore we expect that the structure (48a) is more often realized as a single phonological unit, as in (49a), while the structure (48b) is likely to be marked by a phonological boundary.

(49) a. [long] [short] → (long short)

b. [short] [long] → (short) (long)

[1] In this chapter, we will use [ ] to indicate morphological/syntactic bracketing, while ( ) indicate phonological bracketing.

We will argue in the following sections that these predictions hold not only for affixes, but also for clitics and compounds.

This chapter is structured as follows. Section 7.2 discusses cross-linguistic phonological asymmetries between prefixes and suffixes. We will point out that prefixes tend to be phonologically more independent from the stem. Sections 7.3 and 7.4 discuss clitics and compounds respectively. Section 7.5 discusses bracketing paradox, which we believe an extreme case of the mismatch between phonology and morphology/syntax.

7.2 Phonological asymmetries in affixation

In Section 2.3.1, we argued that they are functional reasons why syntactic and phonological words tend to coincide. However, not every mismatch between syntactic and phonological words is equally troublesome for the parser. Sometimes the lack of a phonological cue of a morpheme boundary is more problematic for the parser. We have argued that a morpheme boundary soon after the beginning is difficult to detect, because a short morpheme is likely to have the same sound as the initial part of other morphemes just by coincidence (we assume that the location of the 'beginning' is obvious from some larger context). Phonological cues in such cases are expected to greatly help parsing. Phonological cues can be suprasegmental or phonotactic cues, each of which corresponds to a different way to define *phonological word* we reviewed in Section 2.3.1. On the other hand, phonological demarcation is less needed after a long morpheme, because it is more likely that the parser can uniquely identify the morpheme even if there is no phonological cues.

In this section, we will examine whether phonological asymmetries between prefixes and suffixes are real based on the following two approaches. Section 7.2.1 discusses phonological asymmetries revealed by the Prosodic Hierarchy Theory, a formal theory that attempts to capture the hierarchical structure of phonology. Section 7.2.2 uses the AUTOTYP database, a typological database of phonological processes.

7.2.1 Insights from the Prosodic Hierarchy theory

The Prosodic Hierarchy Theory is an attempt to capture the hierarchical structure of phonology by assuming a fixed number of universal phonological categories, such as the Prosodic Word and Phonological Phrase, although scholars do not necessarily agree on the details. These categories are hierarchically structured following the universal principle of the Strict Layer Hypothesis.

Some researchers who work on the Prosodic Hierarchy Theory take the so-called end-based approach, in which the range of a phonological unit is determined by aligning it with an edge of a specific grammatical unit (Selkirk, 1986; Selkirk & Shen, 1990). This approach has been applied to morphology as well (Kang, 1992). For example, when a language has a left-end parameter $_{\text{lex0}}[$ for a prosodic word, it means that a prefix forms an independent prosodic word, while a suffix does not. As Wennerstrom (1993) discusses, we can find more examples of left-end parameters than right-end parameters in morphology, which means prefixes are more often phonologically independent from the stem. Whether it is written in the end-based approach or not, it has been reported that prefixes have more independent status than suffixes in a variety of languages (Kang, 1992; Wennerstrom, 1993).

A note on terminology is in order here. In the following examples, different authors use different terms for word-like phonological units, such as 'phonological word' (Booij and Rubach (1984); Nespor and Vogel (1986), among others), 'prosodic word' (Wennerstrom (1993); Hyman (2006), among others) or even 'minor phrase' (Poser, 1990). We will simply call them 'phonological unit', as our aim here is not to illuminate the status and relationship of these concepts within the phonological theory. We are just concerned with whether there are asymmetries between prefixes and suffixes by *some* phonological criteria. As we discussed in Section 2.3.1, different criteria for phonological wordhood do not necessarily match, and we do not have to commit the idea of universal phonological units that can be identified cross-linguistically. Section 7.2.1.1 has more discussions of the indeterminacy of word and its relevance to the current chapter.

English Booij and Rubach (1984) propose a solution of bracketing paradoxes in English by assuming that prefixes form a separate phonological unit (See Section 7.5 for more discussions on bracketing paradoxes). Wennerstrom (1993) argues that while prefixes in English can carry contrastive stress, suffixes cannot. He also argues that this is because prefixes form separate phonological units.

(50) a. This function is DEcreasing here, but INcreasing there.

b. Andy is smart, but Mike is even SMARTer. (*smarTER)

(1a) and (6a) of Wennerstrom (1993, 311-312)

Italian A couple of pieces of evidence suggest that stem-suffix sequences form a phonological unit in Italian, while prefixes are excluded from it (Nespor & Vogel, 1986, 124–134). Intervocalic s-voicing of Italian occurs within a stem, or between a stem and a suffix, but not between a prefix and a stem:[2]

(51) a. a[z]ola (‘button hole’)

b. ca[z]-ina (house-DIM, ‘little house’)

c. a-[s]ociale (NEG-social, ‘asocial’)

Hungarian In Hungarian, two rules illustrate the asymmetry between prefixes and suffixes with regard to phonology: vowel harmony and palatalization. In Hungarian vowel harmony, the stem and suffix assimilate with the feature [±back], as in (52a). In contrast, a prefix does not participate this process, as in (52b) (Nespor & Vogel, 1986, 122ff.).

(52) a. (ölelés-nek)

embracement-DAT.SG

b. (fel)-(ugrani)

up-jump ‘to jump up’

[2]They argue that intervocalic s-voicing does occur between a prefix and stem when a prefixed form is lexicalized. Their example is *pre-[s]entire* ‘to hear in advance’ and *pre-[z]entire* ‘to have a presentiment’ (Nespor & Vogel, 1986, 128). The difference between vowel-ending and consonant-ending prefixes will be another complicating factor.

French In French, differences between prefixes and suffixes are evident in two phenomena: glide formation and nasalization. Both processes indicate that prefixes are phonologically more independent than suffixes (Hannahs, 1995). Glide formation refers to the phenomenon where /i/ becomes /j/ when followed by a vowel-initial morpheme. This process does not occur between a prefix and stem.

(53) a. colonie + -al → (colonial) [kɔlɔnjal]

'colonial'

b. anti- + alcoolique → (anti)-(alcoolique) [ãtialkɔlik]

'antialchohol'

Japanese In most cases, Japanese does not show an asymmetry between prefixes and suffixes with regard to phonology. In Japanese, a word-like phonological unit can be defined by the domain of pitch accent. The following example shows that both prefixes and suffixes become a part of the pitch accent domain.

(54) a. (*fu-shizen*)

un-natural 'unnatural'

b. (*kindai-ka*)

modern-ization 'modernization'

However, Japanese has a special class of prefixes, dubbed *Aoyagi prefixes* in Poser (1990), which resist accent domain formation with the following stem (Poser, 1990; Kageyama, 2001).[3] There is no equivalent class for suffixes.[4]

(55) Examples of Aoyagi prefixes (Kageyama, 2001)

hon- 'this'

moto- 'former'

zen- 'immediately preceding'

gen- 'current'

kaku- 'each'

boo- 'a certain'

doo- 'above-mentioned'

zen- 'all'

ryoo- 'both'

ko- 'deceased'

han- 'anti-'

hi- 'non-'

(56) hi-gooriteki

(hi) (gooriteki)

[3]See Poser (1990) and Kageyama (2001) for discussions on why these preposed morphemes are prefixes, not words, despite their phonological independence. However, as we will argue in Section 7.2.1.1, it is not crucial for our position whether they should be treated as prefixes.

[4]Some postposed items, such as *-nado* 'etc.', can optionally carry the pitch accent (Tomoyuki Kubo, p.c.).

non-rational ‘irrational’

Bantu languages Unfortunately, all languages we discussed above are suffixing languages, and thus one might suspect that the markedness of the prefix phonology is only the reflection of their markedness in terms of frequency. However, asymmetries have been reported in prefixing languages as well. In a number of Bantu languages, a phonological analysis where suffixes are more tightly bound to the stem than prefixes has been proposed, including reconstructed proto-Bantu (Hyman, 2008, 324ff.). An even more striking case is Kukuya (Paulian, 1975), cited in Hyman (2008). Hyman cites the following three remarkable facts about Kukuya:

(57) a. There is a “pause’’, however slight, before every C1 [stem-initial] consonant

b. A C1 nasal or /l/ is automatically geminated

c. Prefixes join the preceding stem to form a clitic-group-like P-domain

(Hyman, 2008, 333)

Paulian’s point (57a) is remarkable in that a pause directly helps identify a morpheme boundary. Her point (57c) indicates that they are almost like ditropic clitics we will see in Section 7.3.

Right-end languages To my knowledge, no language has been given an analysis where the stem forms a phonological unit with a prefix, but not with a suffix. There are a few examples of

languages in which suffixes form an independent phonological unit, but they are exclusively suffixing languages, and thus are not exceptions (Kang, 1992; Wennerstrom, 1993).

For example, a Nimboran verb can have two prosodic domains defined by the stress assignment, as in (58):

(58) ngedóu-k-be-k-u → (ngedóu) (kebekú)
draw-DuSubj-6Loc-Past-1
'We two drew from here to above.'

(Inkelas, 1993, 563)

If Nimboran had prefixes that form a single prosodic domain with the stem, we would obtain a structure where suffixes are phonologically more tightly bound to the stem than prefixes, contrary to our expectations. However, Nimboran is a strongly suffixing language, as we can see from the verb morphology template:

(59) 0 (root) - 1 (PlSubj) - 2 (DuSubj/PlObj) - 3 (MObj/Dur/Part) - 4 (IncDuSubj) - 5 (Loc) - 6 (Iter) - 7 (tense) - 8 (SubjPers)

(Inkelas, 1993, 561)

Yidiɲ is another case where suffixes are claimed to form an independent phonological unit, which is defined by the phenomenon of penultimate lengthening:

(60) gumari-daga-ɲu → (guma:ri) (daga:ɲu)
red-Inch-Past
'to have become red' (Nespor & Vogel, 1986, 135)

However, Yidiɲ is also an exclusively suffixing language (Dixon, 1977, 204). Therefore these languages do not constitute exceptions to our generalization.

7.2.1.1 The indeterminacy of word revisited

A note on the definition of *word* is in order here. As we discussed in Sections 2.3 and 3.2.1, there is no single criterion that can define word cross-linguistically, and no matter whether this position is correct, existing descriptive grammars are not based on a unique cross-linguistic criterion anyway, from which our typological data are taken. Our position is that differences of wordhood criteria between languages is not a fatal problem, as long as each grammar treats preposed and postposed morphemes in a consistent manner.

How do different ways to define words affect our discussion of the relative phonological independence of prefixes in particular? Suppose that a language has two grammatical morphemes, one preposed and the other postposed. Our prediction is that preposed grammatical morphemes tend to be phonologically more independent. Suppose that this is indeed the case in this language. In such a situation, there are two possibilities: one possibility is that the phonological difference between preposed and postposed morphemes lead the grammar author to determine that the preposed morpheme is *not* a prefix but a word, while the postposed morpheme is a suffix. In this case, what we will find is the suffixing preference itself. The other possibility is that the grammar author determines that the given phonological difference is not crucial for distinguishing an affix from a word, and thus both preposed and postposed morphemes are regarded as affixes (or both are regarded as words, in which case they are irrelevant to the discussion of the current section). This is the situation where we expect that

prefixes are phonologically more independent than suffix. As we saw, such examples are in fact common.

7.2.2 AUTOTYP database

Schiering, Bickel, and Hildebrandt (2010) and Bickel et al. (2009) argue for the emergent view of phonological words based on the Word Domains subpart of AUTOTYP, the typological database they constructed. While the proponents of the Prosodic Hierarchy Theory claim that domains of phonological processes can be reduced to a universally fixed set of categories, Bickel and his colleagues take a bottom-up approach in which the phonological processes in a language are exhaustively described, aside from the question of whether they fit into a theoretically motivated phonological domain. They argue that the typological pattern does not show a sign of converging into a fixed number of categories. Furthermore, they point out that sometimes the domains of two phonological process are not in an inclusion relation (i.e. when two types of domains overlap, one must always include the other as a subpart), contrary to the prediction of the Strict Layer Hypothesis proposed in the Prosodic Hierarchy literature.

AUTOTYP lists 679 phonological processes from 70 languages.[5] It must be noted that AUTOTYP is not meant to be a typologically balanced database at this point, and certain language families are overrepresented, such as Indo-European, Sino-Tibetan and Austroasiatic.

We excluded 533 out of 679 phonological processes that only apply to a particular lexical stratum, as it is not practical to fully recover subset relations among different lexical strata

[5] http://www.spw.uzh.ch/autotyp/projects/wd_dom/wd_dom.html

in each language. In addition, we exclude languages for which we cannot determine whether prefixes are more closely attached to suffixes or the other way around.

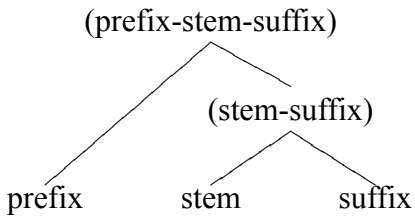
For some languages, we can determine whether suffixes are closer to the stem or not. For example, Manange, a Sino-Tibetan language spoken in Nepal, has the following two phonological processes:

(61) a. **prefix-stem-suffix**: Stress

b. prefix-**stem-suffix**: Obstruent voicing assimilation

There is no phonological process that takes the prefix-stem sequence as its domain. Hence, we can unambiguously determine the following hierarchy:

(62)



Not every language allows us to unambiguously determine a hierarchical structure in this way. For example, Limbu (Sino-Tibetan) has the following two phonological processes, among other patterns (Hildebrandt, 2007; Bickel et al., 2009; Schiering et al., 2010):

(63) a. prefix-**stem-suffix**

/m/ → [ŋ] / [velar]___

(haŋ-**ŋ**ʔna) 'sent'

send-PASS.PTPCL

b. **prefix-stem**-suffix

/n/ → [ŋ] / ___[velar]

(kɛ-**ŋ**-g^{ɦ}o:s)-u-n ‘You didn’t find it’

2-NEG-find-3P-NEG

Because the domains of those two processes are not in an inclusion relation, we cannot draw a simple tree that represents the domains of both processes. We exclude languages like this. In addition, in some languages no phonological process that involves prefixes or suffixes is reported; we also exclude those languages.

As a result, we found that in the following 11 languages, prefixes are phonologically outside of suffixes: Armenian (Eastern), Chukchi, German, Nama (Khoekhoe), Yimas, Persian, Turkish, Kinnauri, Kham, Swedish, Manange and Lithuanian.

Table 7.1: 11 languages with the structure (prefix)-(stem-suffix)

Armenian (Eastern)
Chukchi
German
Nama (Khoekhoe)
Yimas
Persian
Turkish
Kinnauri
Kham
Swedish
Manange
Lithuanian

In only one language, Semelai (Austronesian), prefixes are closer to the stem than suffixes are. This language has a rule stating that a glottal stop is deleted at the stem initial position

(Kruspe, 2004, 52, 67). This rule does not apply to the stem-suffix boundary.

(64) pa- + ʔyəŋ → par-yəŋ

However, Semelai belongs to Mon-Khmer languages, where suffixes are uncommon. All of the suffixes in Semelai are borrowed from Malay (Kruspe, 2004, 68). This may suggest that the exceptional behavior of Semelai can be explained by assuming that all suffixes belong to different lexical strata.[6]

We should note that the data given in Bybee et al. (1990) appears to contradict our claim. We mentioned in Section 3.4.1 that prefixes are not more phonologically affected by the stem than suffixes, which is compatible with the hypothesis that prefixes are phonologically more independent. However, they also found that stems are more often affected by prefixes than suffixes (Table 7.2). Moreover, based on their measure of the degree of grammaticalization, they argue that prefixes are more grammaticalized than suffixes, and also that non-bound preposed grammatical materials are more grammaticalized than non-bound postposed grammatical materials.

Table 7.2: Bybee et al. (1990: 21)

	Affix conditions change(s) in Phonological expression of stem		Rate of effect
	yes	no	
prefixes	120	305	28.2%
suffixes	231	966	19.2%

[6]This kind of scrutiny, however, should also be applied to the languages that conform our prediction. For example, Turkish is a strongly suffixing language, and most of the prefixes are due to borrowing, which may suggest they belong to a different lexical stratum.

Their findings are bewildering, because they contradicts not only the dominance of anticipatory assimilation, but also the common idea that a stem-initial position is more resistant to sound change (See Section 3.5). We do not have an explanation as to why their data goes against other evidence. A potential cause of the discrepancy could be due to the fact that their sample is relatively small with 71 languages, or the fact that they count the number of affixes, not the number of languages.

7.3 Asymmetries in clitics

As discussed earlier, our explanation of the suffixing preference does not rely on the grammatical status of affixes, but only on the *shortness* of affixes. Then the same discussion should apply to clitics too. We expect that enclitics are typologically more common than proclitics.[7] We do not have comprehensive typological frequency data of clitics, but the data from adpositions and plural (Matthew Dryer, p.c.) in Tables 7.3 and 7.4 clearly show that enclitics outnumber prolicitcs.

Table 7.3: Positions of adposition clitics (genus-based)

	Afr	Eur	SEA&O	A-NG	NA	SA	Total	
proclitic	4	2	3	0	3	2	14	
enclitic	**11**	**6**	**9**	**40**	**9**	**18**	93	*

Another fact that may suggest that enclitics are more prolific is that enclitics have a wider distribution than proclitics. Klavans (1985) is an attempt to capture the possible positions of clitics with the following three parameters:

[7] See also Himmelmann (2014) on the left-right asymmetries of clitics and their relevance for the suffixing preference.

Table 7.4: Positions of plural clitics (genus-based)

	Afr	Eur	SEA&O	A-NG	NA	SA	Total	
proclitic	0	0	0	0	2	1	3	
enclitic	**15**	**3**	**9**	**9**	**3**	**10**	49	*

(65) a. P1: Initial/Final

b. P2: After/Before

c. P3: Proclitic/Enclitic.

P1 controls whether the clitic is adjacent to the first or last element of its syntactic domain. P2 controls whether the clitic precedes or follows that element, which is necessary to allow the so-called second-position (Wackernagel) clitics. P3 is a phonological parameter that indicates whether a clitic precedes or follows its phonological host. Using these three parameters, we can recognize eight different types of clitics, as in Table 7.5.

Table 7.5: clitic classification based on Klavans (1985)

type	P1	P2	P3	schematic rep.	attested?
1	initial	before	enclitic	(W[=c) X Y Z]	✓
2	initial	before	proclitic	[(c=X) Y Z]	✓
3	initial	after	enclitic	[(X=c) Y Z]	✓
4	initial	after	proclitic	[X (c=Y) Z]	
5	final	before	enclitic	[X (Y=c) Z]	(✓) *see body text
6	final	before	proclitic	[X Y (c=Z)]	
7	final	after	enclitic	[X Y (Z=c)]	✓
8	final	after	proclitic	[X Y Z (c=]W)	

For example, *'ll* in English is put before its syntactic host (a verb phrase), but it is an enclitic. Thus, its parameter settings are Intial&Before&Enclitic (type 1). While Klavans shows examples for all of the eight possible configurations, subsequent work has questioned some

of his analyses, and pointed out that not every combination of parameter values is equally well attested (Marantz, 1988; Sadock, 1991; Halpern, 1992, 1998; Himmelmann, 2014). According to Halpern (1998), types 4, 5, 6, and 8 are only questionably attested. In Table 7.5, the attested cases are marked by ✓following Halpern. More recently, Peterson (2001) has reported that the Ingush clause-chaining clitic *ʔa* is an example of type 5 clitics, which is indicated by (✓) in the table.

Among the attested configurations, only type 2 is proclitics, and all of the rest are enclitics. Examples of types 1, 2, 3 and 7 from Klavans' paper are the following. Type 3 corresponds to the so-called second-position (Wackernagel) clitics.

(66) a. Type 1 (Kwakwala NP markers)

nəp'idi=da gənanəm x̣a guk^{w} sa t'isəm
throw=DEIC child OBJ house OBL rock
'The child hit the house with a rock by throwing.'

(Klavans, 1985, 106)

b. Type 2 (Greek article)

hoi=agathoi Spartioi

'The strong Spartans'

(Klavans, 1985, 117–118)

c. Type 3 (Ngiyambaa enclitics)

girbadja=ndu mamiyi gambira
kangaroo=2.NOM catch.PAST yesterday
'You caught a kangaroo yesterday.'

(Klavans, 1985, 99)

d. Type 7 (Spanish pronominal clitics)

dá=me=lo
give=me=it!

(Klavans, 1985, 98)

The only cases where proclitics are well attested are when a clitic is at the beginning of its domain (type 2). For type 2, there is a functional motivation for a clitic to be more strongly associated with the following word than the preceding word: if they were enclitics, there would become type 1, in which there is incompatible bracketing between phonological and syntactic units. Thus, we can summarize possible configurations in the following way: enclitics are preferred in general, but proclitics are also motivated when it is semantically associated with the following phrase.

Types 1 and 8 above are sometimes called *ditropic clitic*, although type 8 is not clearly attested.[8] These are particularly interesting because the phonological and syntactic bracketing looks incompatible. The following is another example from K^{w}ak'^{w}ala (Anderson, 1984):

(67) kw'iχʔid=ida bəgwanəma=χ=a q'asa=s=is t'əlwagwayu
clubbed=the man=OBJ=the otter=INSTR=his club
'The man clubbed the otter with his club.'

(Anderson, 1984)

[8]Cysouw (2005) defines ditropic clitic as a subtype of type 1 and 8. While English auxiliaries like *'ll* also belong to type 1, he excludes them from the category of ditropic clitic, based on the fact that its phonological host is invariably the last word of the subject noun phrase.

Syntactically [*=ida bəgwanəma*] and [*=ẋ=a q'asa*] form units, while phonologically (*bəg-wanəma=ẋ=a*) and (*q'asa=s=is*) form units. Thus, when we take a look at a phonological unit, a major syntactic boundary is near the end (68a). On the other hand, if we focus on a syntactic phrase like [*=ida bəgwanəma*], we can see preposed grammatical morphemes are demarcated by a phonological boundary (68b). This is in accordance with the direction of asymmetry we propose.

(68) a. (bəgwanəma][=ẋ=a)

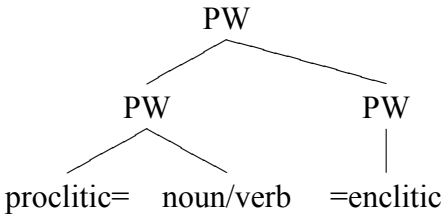
b. [=ida) (bəgwanəma]

Ditropic clitics are attested in a variety of languages, including Udi, Ingush and Greek, among others (Cysouw, 2005). In every case, the clitic is grammatically associated with the following word but phonologically associated with the preceding word (type 1 in Table 7.5). The mirror image (type 8 in Table 7.5) is not attested (ibid., 18).

We should note that there is some evidence that appears to go against our generalization. First, Bybee et al. (1990) argue that preposed grammatical morphemes are more 'grammaticalized' than postposed grammatical morphemes, not just for affixes, but also for non-bound words. Again, we do not have an answer for why Bybee et al.'s finding seemingly contradicts with our claim.

A similar problem for our approach is the analysis of Bantu languages, such as Shona and Luganda presented in Hyman (2008, 340), where proclitics are claimed to be phonologically closer to the stem than enclitics.

(69)



(Hyman, 2008, 340 (56))

However, there is a reason to believe that Hyman's analysis does not contradict our claim. Interestingly, the sole evidence for the phonological wordhood of the proclitic=stem sequence given by Hyman is a minimality condition. In Shona, proclitic=stem should be at least bisyllabic. In the following example, *mu=* is a proclitic, while *=i* is an enclitic. (70a) is allowed because the proclitic=stem sequence has two syllables, while (70b,c) are ruled out. When the word is monomorphemic and there is no proclitic, as in (69b), an ephenthetic vowel *i-* is added to satisfy the minimality condition (*i-pá* 'give!').

(70) a. mu=mbá 'in the house'

b. *pá 'give!'

c. *pá=i 'give (pl.)!'

(Hyman, 2008, 340)

As we discussed in Section 2.3.1, the minimality condition is the only criterion of wordhood which does *not* help identify its end in language comprehension, at least in a direct manner. Rather, it helps recognize that it has *not* ended. This means that while a Shona word can be analyzed as (proclitic=stem)(=enclitic), its phonological boundary does not help parse the enclitic out.

7.4 Asymmetries in compounds

Unlike cases of affixation or cliticization, in which grammatical morphemes are expected to be shorter than stems on average, a compound consists of more than one stem, and we have no obvious reason to believe that one is consistently shorter than the other. However, when a compounds have three stems, we expect that the branching constituent will be phonologically longer than the non-branching constituent on average. Namely, we expect that, on average, [A B] is longer than [C] in (71a), while [B C] is longer than [A].

(71) a. [[A B] C] (left-branching)

b. [A [B C]] (right-branching)

What does the parser have to do in order to understand these three-stem compounds? When the whole compound is stored in the lexicon, the parser does not have to parse it, in which case it is irrelevant to our discussion. When [A B] or [B C] is stored in the lexicon, the parser needs to detect the boundary between B and C in (71a), and between A and B in (72b). In this case, the situation is quite similar to the case of affixes. Because our account of the suffixing preference relies not on the grammatical status of affixes but on its relative shortness, and because the sequence of two stems is on average longer than a single stem, we expect that (71a) — analogous to a suffixed word — is more common than (71b) — analogous to a prefixed word.

What if none of the combinations is stored in the lexicon, and the parser needs to identify all A, B and C? Because we have modeled only morpheme segmentation, not the task

of recovering a constituent structure, we do not have clear predictions at this point. We can speculate, however, that (71a) is still preferred over (71b) from a different reason. Parsing would be more complicated if, in (71b) for example, the parser first recognizes A+B and C as units, and later realizes that B+C forms a constituent, compared to the case where the parser recognizes A and B+C as units from the beginning. Given that A and B become available earlier than C, the parser would commit the analysis where A+B is a unit, without any evidence that tells otherwise. The reason why the parser detemines that A+B is a unit is probably that two-stem compounds are probabilistically most likely, or it is more costly to suspend the analysis of constituency when there is a way to solve it (i.e. recognizing a two-stem compound). In either case, this means that (71b) is likely to trigger misparse.[9] This argument is similar to what Hay (2003) proposes for the bracketing paradox, which we will return to in Section 7.5.2.

We can also expect phonological asymmetries for compounds. Suppose that [A B] in (71a) or [B C] in (71b) is stored in the lexicon. Then, the logic will be the same as in case of affixes. The boundary near the beginning is more likely to be overlooked, and thus a phonological demarcation between A and B in (71b) assures that the parser can identify A. On the other hand, it is easier for the parser to identify the boundary between B and C in (71a) due to the length of [A B], and thus a phonological demarcation is less needed. Thus, we expect that (71a) is more likely to be realized without a phonological word boundary, while

[9]This does not mean that right-branching always induces misparse in language processing. It induces misparse in compounding in particular because there is no immediately available evidence that tells that A+B is not a constituent. Unlike compounds, for example, the parser does not mistakenly recognize *this is* in *this is a pen* as a constituent, because the parser knows that *is* is a head of a verb phrase [*is* NP].

(71b) is marked by a phonological boundary:

(72) a. [[A B] C]

(A B C)

b. [A [B C]]

(A) (B C)

We can expect the same asymmetry when none of the combinations of stems is lexically stored. In (72b), as we discussed above, there is a risk that the parser would commit the wrong analysis where A+B forms a constituent. This misanalysis can be circumvented if there is a phonological boundary cue between A+B by which the parser can tell that A+B is not a simple compound. Again, we expect that right-branching (72b) is more likely to be divided into two phonological words.

There have been a couple of typological studies of recursive compounds (Mukai, 2008; Bauer, 2009; Štekauer, Valera, & Kortvélyessy, 2012), but it is difficult to evaluate the hypotheses above in a quantitative manner at this point. The discussion in this section inevitably focuses on a few well-studied languages. It is thus dangerous to draw conclusions about typological tendencies at this point, but we will see that available data are at least not contradictory with our expectation.

7.4.1 Asymmetries in frequency

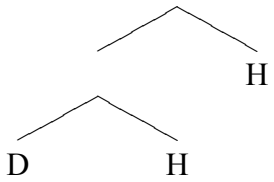
Left-branching compounds are more common in English (Chomsky & Halle, 1968, 92), as well as in Japanese (Kubozono, 1993, 35,46,65). A similar observation has been made for

Northern Germanic languages in Mukai (2008).

A problem is that compounds are right-headed in all of these languages. Thus one might suspect that the abundance of left-branching compounds is just a consequence of right-headedness. If we suppose that compounds are consistently left-headed or right-headed in each language, there are the following four possible configurations for a three-stem compound, as shown in (73a–d).

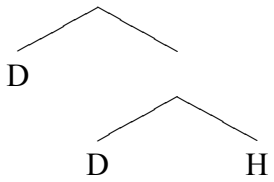
(73) a. right-headed / left-branching / multiple-head

e.g. [[*computer class*] *instructor*]



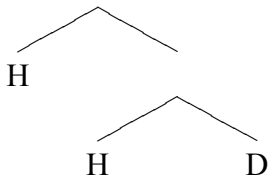
b. right-headed / right-branching / single-head

e.g. [*evening* [*computer class*]]



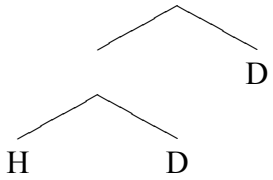
c. left-headed / right-branching / multiple-head

e.g. [*instructor* [*class computer*]]



d. left-headed / left-branching / single-head

e.g. [[*class computer*] *evening*]



Our prediction is that left-branching is more common, no matter where the head is. Thus, we expect that (74a) and (74d) are common types. However, it has been argued that right-branching, left-headed compounds (73c) may be more common than left-branching, left-headed compounds (73d) in Romance languages (Haruo Kubozono, p.c.). In fact, left-branching, left-headed compounds (73d) are claimed to be impossible in Haider (2001). If this is the case, then the reality would be that (73a) and (73c) are common types, then the generalization would be that a multiple-head compound is better than a single-head compound. However, it is difficult to draw a conclusion at this point, as compounding is less productive in Romance languages than Germanic languages no matter whether it is left-branching or right-branching, and we do not know what the picture would be if we consider more typologically diverse languages.

It is worth mentioning here that Mukai (2008) proposes an alternative processing-based account for why (73a) is more common than (73b). In a right-branching right-headed compound like (73b), she argues, the parser should wait for the final constituent to solve all unresolved dependencies, which is captured by Hawkins' (2004) MiD principle.[10] In order to further investigate this issue, we will need a more systematic approach to collect typological

[10]We are, however, skeptical with the idea MiD plays a role in restricting recursive compounds. If a head-final structure with multiple dependents is not preferred by the parser, then a verb-final clause with multiple arguments and adjuncts must be more seriously penalized, as arguments and adjuncts of a verb are on average longer than constituents of compounds. However, verb-final word order is quite common in the world's languages.

data on compounds.

7.4.2 Phonological asymmetries in compounds

We expect that the leftmost constituent in the right-branching compounds tends to be more phonologically independent than the rightmost constituent in the left-branching compounds. In other words, we expect a phenomenon parallel to the phonological independence of prefixes we saw in Section 7.2, although a constituent of a compound is not an affix.[11]

In this context, a curious phonological asymmetry has been reported for Japanese recursive compounds (Kubozono, 1993, 51). When a compound is left-branching, as in (74a), the whole compound is a single phonological word. On the other hand, in a right-branching structure like (74b), the first element of the compound forms a separate phonological word, which can be determined by pitch accent domains and also by sequential voicing in some cases (Otsu, 1980). In the examples below, the symbol ˈ indicates the locus of an accent nucleus. Each phonological word can have at most one accent nucleus.

(74) a. [[nihon buyoo] kyookai]
[[Japan dance] association]
(nihon buyoo kyoˈokai)
'association for Japanese dance'

b. [nihon [buyoo kyookai]]
[Japan [dance association]]
(nihoˈn) (buyoo kyoˈokai)
'dance association of Japan'

(57) of Kubozono (1993, 51)

[11] When a compound is left-branching but, nonetheless, the first constituent is phonologically independent, then we have an example of bracketing paradoxes. In Japanese, this happens when the first two constituents are in the coordination relation (Kubozono, 1993). The direction of bracketing paradoxes is compatible with our discussion in Section 7.5.

Similar asymmetries have been pointed out in English. In English compounds, left constituents carry main stress, but a right-branching compound is an exception (Chomsky & Halle, 1968). For example, in [*evening* [*computer class*]], *computer* receives the main stress. It has been proposed that stress is used to detect word boundaries in English (Cutler & Norris, 1988), which suggest that the stress pattern of right-branching compounds is motivated by the need to clearly demarcate the morpheme boundary between the first two stems in right-branching compounds. These patterns are in accordance with the prediction we made above.

A problem is that in both Japanese and English, right-branching compounds are known to be less common. There is a possibility that right-branching compounds are phonologically marked in these languages only because they are less frequent (Kubozono, 1993). Evidence from more languages will be necessary to make a typological claim.

7.5 Bracketing paradoxes

The bracketing paradox refers to morphologically (or syntactically) complex expressions for which morphosyntactic and phonological criteria lead to contradictory constituent analyses (Bolinger (1967); Siegel (1974); Allen (1978), among others). The paradox has spurred discussions on how the morphological and phonological aspects of word formation must be simultaneously represented. This section discusses how the bracketing paradox can be understood in our processing-based approach.

More specifically, we will claim that the bracketing paradox is an extreme case of the boundary position preferences we discussed above. We argued that a morphological boundary

is preferably near the end of a phonological unit, as in (75a), while a phonological boundary is preferably near the beginning of a morphological unit, as in (75a). Again, [] stands for morphological bracketing, and () phonological bracketing.

(75) a. (A B] [C)

b. [A) (B C]

A bracketing paradox arises when (75a) and (75b) happen in the same sequence at the same time. In such a situation, the middle item is morphologically/syntactically associated with the first item, while it is phonologically associated with the last item:

(76) a. [A B] [C]

b. (A) (B C)

We will see that the direction of mismatch in known examples of bracketing paradox follows this pattern.

Many of the previous approaches on the bracketing paradox are descriptive or formal in nature and do not aim to provide a functional explanation; an exception is Hay (2003), whose approach we basically agree with. We will review her approach in Section 7.5.2.

7.5.1 Classification of bracketing paradoxes

Following Beard (1991), we argue that at least three types of the bracketing paradox must be recognized in English:

(77) a. **Type I: generative grammarian**

- [*generative grammar*]-*ian*: semantics
- *generative* (*grammarian*): phonology

b. **Type II: unhappier**

- [*unhappi*]-*er*: semantics
- *un*-(*happier*): phonology

c. **Type III: ungrammaticality**

- [*ungrammatical*]-*ity*: semantics and morphology (subcategorization)
- *un*-(*grammaticality*): phonology (level ordering)

Examples of Type I include *generative grammarian* and *nuclear physicist*. The expression *generative grammarian* appears to consist of two phonological words, *generative* and *grammarian*. When we consider the semantics, however, [*generative grammar*]-*ian* appears to be the correct constituent structure, because it means a person who studies generative grammar, rather than a grammarian who is generative. Type II is another celebrated example of the bracketing paradox (Pesetsky, 1985). Because the morphological comparative *-er* requires that the stem be disyllabic at most, phonologically it appears to have the structure *un*-(*happier*). On the other hand, because it is interpreted as the comparative of *unhappy*, the morphological structure appears to be [*unhappi*]-*er*. In Type III, the bracketing [*ungrammatical*]-*ity* is motivated by the intuition that *ungrammaticality* is the nominalized form of

ungrammatical. Because the prefix *un-* takes an adjective, this analysis is also supported by the subcategorization.[12] On the other hand, the theory of level ordering (Siegel, 1974; Allen, 1978) predicts that *-ity* must be attached to the stem first. In the theory of level ordering, English affixes are classified into two classes, Level I and Level II. Level I affixes are closer to the stem and can cause a shift of the stressed syllable, palatalization, and other phonological processes, while Level II affixes cannot. Because *-ity* can cause these phonological processes, we can see that it belongs to the Level I affix, while *un-* belongs to the Level II affix.

A striking commonality that we can find in (77) is that phonological criteria always suggest that the right two constituents form a unit, while morphological/syntactic considerations suggest that the left two constituents form a unit. This is in accordance with our predictions; in morphology, a short morpheme at the end is easier to understand than a short morpheme at the beginning; a short morpheme at the beginning is easier to understand when it is phonologically independent. However, it is not obvious whether Types I and II must be understood as mismatches between morphology/syntax and phonology, as we discuss below. Type III is the only clear case for which the processing-based asymmetry can give an explanation.

The reason why Type I should not be analyzed as a mismatch of bracketing is that the bracketing [*generative grammar*]-*ian* only motivated by its interpretation. There is a possibility that its meaning is obtained by mechanisms of lexical semantics, which have nothing to do with their morphological structure. A convincing argument for the lexical semantics approach is that a similar interpretation can be obtained in phrases with morphologically non-

[12]But see Allen (1978) and Fabb (1984), who take the approach in which *un-* can subcategorize for nouns. Note that the prefix *un-* can also take verbs but this is irrelevant here.

derived head nouns, such as *old friend* and *chemical engineer* (Beard, 1991). An old friend can not only mean a friend who is old, but also a person whose friendship is old; a chemical engineer is not an engineer who is chemical, but a person who is working on chemical engineering, in a way similar to *generative grammarian*; however, unlike in the case of *generative grammarian*, these interpretations of *old friend* and *chemical engineer* cannot be expressed by bracketing, at least in a straightforward manner. Instead, Beard proposes the Principle of Decompositional Composition:

> The semantic features of an attribute subjoin with one and only one semantic feature of its head.
>
> (Beard, 1991, 208)

Harris (2006) also gives a semantic solution for a bracketing paradox in Georgian.

Another issue of the bracketing approach for Type I is that a phrase like *nuclear physicist* has a French equivalent, *physicien nucléaire*. Because adjectives are postposed in French, [physic- nucléaire] must be analyzed as a discontinuous constituent in order to obtain its interpretation directly from bracketing, just like the English [nuclear physic]-. This problem does not arise if it is reduced to an issue of lexical semantics.

Another case that should not be treated as an example of the mismatch between phonological and morphological structures is discussed in Kageyama (1993, 331–335). Consider the following example from Japanese:

(78) senzo-no haka-mairi
ancestor-GEN grave-visiting
'visiting of ancestor's grave'

Although *haka-mairi* 'grave visiting' forms a phonological word, *senzo* 'ancestor' appears to only modify a part of it, *haka* 'grave.' It is therefore tempting to assign the syntactic structure [*senzo-no haka*]-*mairi*. However, Kageyama argues that this kind of structure is only allowed when the non-head of the compound, *haka* in this particular case, is a relational noun and the external modifier is its argument. In the example above, the noun *haka* 'grave' requires its possessor. To support this claim, he points out that the following example is not acceptable.

(79) *mikageishi-no haka-mairi
granite-GEN grave-visiting
'visiting of granite grave' (intended)

This suggests that the overall interpretation is better analyzed by lexical semantics, rather than morphological bracketing, in this example too, although it is somewhat different from the English examples we discussed above, because *grammar* is not a relational noun that takes *generative* as an argument.

The Type II bracketing paradox, such as *unhappier*, also conforms our prediction in that the first part is more phonologically independent while the last part is more morphologically independent. However, there are few words that belong to this category, and thus we do not know if it is only a coincidence. It has been pointed out by Spencer (1988) and Beard (1991) that aside from *unhappier* and *unluckier*, other examples of the equivalent structure are awkward (?*unwarier*, ?*uncannier*). This may indicate that *unhappier* and *unluckier* are lexically listed exceptions.

On the other hand, we think that the Type III bracketing paradox such as *ungrammaticality* is a genuine example of the mismatch between phonological and morphological structures, for which our processing-based approach can give an explanation. Because *un-* is a Level II affix and more phonologically independent from the stem than *-ity*, the stem-suffix boundary is more difficult to detect, other things being equal. However, the detection of the stem-suffix boundary is helped by the fact that the stem-suffix boundary is near the end of the word. When the parser reaches the stem-suffix boundary, it is not difficult for the parser to see that it has something to do with *ungrammatical*, which means that the parser successfully identifies the boundary.

The mirror image of examples like *ungrammatical* is not attested (Spencer, 1991). For example, if there were a word like **insuccessful*, the morphological structure would be like (80a), as the prefix *in-* takes an adjective. The phonological structure would be like (80b), because the prefix *in-* is a Level I affix, while *-ful* is a Level II affix.

(80) a. in-[successful]

b. (in-success)-ful

In order to make the contrast between *un-* and *in-* clearer, consider an (illegitimate) example that exhibits assimilation, like **immeaningful*:

(81) a. im-[meaningful]

b. (im-meaning)-ful

As discussed in Chapter 4, it is quite difficult to immediately tell that a morpheme boundary is immediately after the initial phoneme [ɪ]. Compared to this, it is relatively easy to detect the suffix *-ful* when it reaches the end, because it is phonologically more independent (and also, more productive). However, parsing *-ful* first ends up with the wrong analysis [im-meaning]-ful, which cannot be interpreted.

This account is quite similar to Hay's (2003) account of bracketing paradoxes, that we will explain in the next section.

7.5.2 Hay's (2003) account on bracketing paradoxes

A processing-based account of the bracketing paradox is proposed in Hay (2003, 182–184). We have already discussed the relevance of her theory to the suffixing preference in Section 4.2.4.2. Although she does not discuss its consequence for typology, her account of the English bracketing paradoxes is quite similar to ours.

Her account is built upon her theory of affix ordering: A more *parsable* affix occurs outside a less parsable affix (Hay, 2003; Hay & Plag 2004; see also Plag & Baayen, 2009). A variety of factors affect parsability, including frequency and phonological cues. She argues that her parsability-based account enables us to make finer-grained predictions about affix ordering of English than the dichotomy between Level I and Level II affixes proposed by the theory of level ordering.

In addition, she argues that a prefix-stem sequence is more likely to be perceived as a unit, due to the temporal nature of speech comprehension. Even when the prefix is more separable from the stem than the suffix in other respects such as phonology and frequency, a prefix-stem

sequence is more likely to be perceived as a unit, because it becomes available earlier. She classifies the following four logically possible configurations of a word with both a prefix and suffix. The two configurations of (82) are non-paradoxical cases where less parsable affixes are closer to the base. Assuming that more parsable (i.e. more phonologically independent or frequent) affixes can be easily parsed out, the word can be interpreted in a straightforward manner, as a parsed affix plus the rest. On the other hand, the two configurations of (83) are trickier cases where more parsable affixes are closer to the base.

(82) non-paradoxical configurations:

a. [(marginally parsable affix) (base)] (highly parsable affix)

e.g. *im-polite-ness*

b. (highly parsable affix) [(base) (marginally parsable affix)]

e.g. *un-famili-ar*

(83) paradoxical configurations:

a. [(highly parsable affix) (base)] (marginally parsable affix)

e.g. *un-grammatical-ity*

b. *(marginally parsable affix) [(base) (highly parsable affix)]

e.g. **in-success-ful*

(11) and (12) of Hay (2003, 183–184); the examples are inserted by the current author

Hay argues that the temporal nature of processing plays a role in (83). She writes:

> Because we process speech in a temporal manner, the prefix-base sequence is *available* as input to suffixation in (12a) [(83a) above], even if its [sic] not the preferred input. The configuration in (12b) [(83b) above], however is essentially non-parsable. Both the left-to-right nature of speech, and the marginal parsability of the prefix, will lead to an overwhelming tendency to parse the first two elements as a constituent, contrary to the intended analysis, in which the last two elements are a constituent.
>
> (Hay, 2003, 184)

We made a similar argument based on the 'early availability' of a constituent for compounds in Section 7.4. Again, while the argument based on the 'early availability' of the prefix-stem sequence may work, an alternative approach based on our original explanation of the suffixing preference would work too: A morpheme boundary near the beginning is difficult to parse. This effect is not included in her definition of parsability, and thus if we take them together, we can see why (83b) is the worst configuration: the prefix is less parsable not only because of the factors Hay points out (phonology, frequency and so on), but also because *it is a prefix*, in the situation where the prefix must be parsed out first in order to interpret the structure of (83b).

7.5.3 Bracketing paradox beyond a word

While Hay only discusses derivational affixes, her argument must be able to be extended to a wider range of phenomena, if it is just a consequence of the ease of segmentation. Con-

sider the following examples in Japanese from Kageyama (1993), which consist of multiple grammatical words:

(84) a. [maboroshi-no chosha] -sagashi
[phantom-GEN author] -search
(maboroshi-no) (chosha-sagashi)
'search for the unknown author'

b. [yutakana umi] -zukuri
[rich sea] -making
(yutakana) (umi-zukuri)
'making a rich sea'

In these examples, the last morpheme (*-sagashi* 'search' or *-zukuri* 'making') appears to take a phrase, although the last two morphemes, *chosha-sagashi* 'author-search' or *umi-zukuri* 'sea-making', form a phonological unit. We can see that the direction of bracketing mismatches is the same as other examples of bracketing paradoxes we have seen so far.

At a first glance, these examples look similar to *senzo-no haka-mairi* 'visiting ancestors' grave' (78), but while *senzo-no* must be an argument of *haka*, there is no restriction in the modifying part in these examples. Rather, the acceptability depends on the lexical choice of the head part; *-sagashi* 'search' and *-zukuri* 'making' accept phrasal complements.

Again, the mirror image of these patterns is not attested. It is possible to find a structure where a prefix appears to be outside of a phrase, but unlike their suffixal equivalents, prefixes form an independent phonological domain, as can be seen from the following examples.[13]

[13]Despite their phonological independence and apparent phrasal scope, they are still prefix-like because (i) they are exceptional usage of the items that are normally used as more typical prefixes and (ii) the following phrase is lexicalized, as can be seen from the fact that they are the titles of a book or a TV program in (85), and they cannot be interrupted by other random words. However, if one takes a view that these preposed items are words, not prefixes, then that would not contradict our position, as discussed in Section 7.2.1.1.

(85) a. shin- [hana-no hyakumeizan]
new [flower-GEN hundred.famous.mountain]
(shin) (hana-no) (hyakumeizan)
'New 100 Famous Mountains of Flower' (booktitle)

b. zoku- [hoshi-no kinka]
sequel- [star-GEN coin]
(zoku) (hoshi-no) (kinka)
'Star's Coins: Sequel' (title of a TV drama show)

Here, prefixes *shin-* and *zoku-* form independent phonological units; both the pitch accent pattern and the lack of sequential voicing suggest their phonological wordhood. They can be considered as examples of Aoyagi prefixes we saw in Section 7.2.1, although *shin-* is not listed as an Aoyagi prefix in the previous literature.

In English, examples like the following have a structure similar to the Japanese examples above.

(86) a. pre-[industrial revolution]

b. [anti-[nuclear weapons]] movement

In these examples, prefixes carry an independent stress (Bates, 1988; Kageyama, 2001), which indicates, again, that they form a separate phonological unit. These examples could be considered English equivalents of Aoyagi prefixes, as Kageyama suggests. Typically, even in spelling, these wide-scope prefixes are written with a space or hyphen after them. The Japanese examples are similar: the prefixes *shin-* and *zoku-* are clearly demarcated by an interpunct (·) in their official names. This never happens in the case of wide-scope postposed items like (84). This suggests that the same processing effect may be at work in reading, not just in listening.

To summarize, we have some evidence that the bracketing paradox has the same asymmetry beyond a word boundary. Again, an obvious weakness of the current analysis is that we do not have data from typologically diverse languages. More future work will be needed to expand the data sources and see whether our generalizations are still tenable.

7.6 Discussion

We have seen that preposed short morphemes are less common than postposed short morphemes, and that preposed short morphemes tend to be phonologically detached from the following item. These tendencies can be seen no matter whether the relevant morpheme is an affix, clitic, or a non-branching constituent of a compound. In some cases it is premature to argue typological frequencies due to the limitation of available data, but known examples are at least compatible with our expectations. This section gives brief discussions of whether other approaches to the suffixing preference can explain these facts.

It is difficult to see what kind of predictions historical, formal, and acquisition-based approaches make about the phenomena we saw in this chapter. Givón's (1979) account may be compatible with the fact that prefixes are phonologically more independent. His account predicts that suffixes are older on average, because they are more often remnants of older word order. Thus, if one assumes that older affixes tend to be more phonologically fused with the stem, the fact that suffixes are phonologically less independent may be explained.[14] We are skeptical about this approach, however. Examples of ditropic clitics (Section 7.3) clearly

[14]This is similar to the hypothesis considered in Bybee et al. (1990, 14ff.), although their conclusion is negative.

show that grammatical morphemes tend to be phonologically attracted by the preceding item, even when they are grammatically associated with the following item. This suggests that the phonological asymmetry cannot be reduced to the fact that more grammatical morphemes were postposed in the past.

We will discuss the consequence of our discussion in this chapter for the two processing-based approaches, Hawkins and Cutler (1988) and Himmelmann (2014), in more detail below.

7.6.1 Hawkins and Cutler (1988)

We have already argued in Section 3.4.2 the fact that Hawkins and Cutler's (1988) account is dangerously close to stipulating an asymmetry in an invisible level in order only to account for a visible asymmetry (i.e. the suffixing preference).

The discussion of this chapter reveals two further problems with their approaches. First, their account does not explain why prefixes tend to be phonologically more independent from stems than suffixes. A phonological demarcation between a prefix and stem may facilitate the identification of the stem, but it also guarantees that the prefix is identified earlier than the stem. This contradicts their claim that the affix-stem processing order must be avoided.

Second, their argument cannot be extended to clitics and compounds. Unlike affixes, which can be said to convey the syntactic information of the lexical item, a clitic is not a locus of the syntactic information of its phonological host. Consider an example of the ditropic clitic we saw in Section 7.3, which we show here again:

(87) kw'iχ̣ʔid=ida bəgwanəma=χ̣=a q'asa=s=is t'əlwagwayu
clubbed=the man=OBJ=the otter=INSTR=his club
'The man clubbed the otter with his club.'

(Anderson, 1984)

Here, [=*χ̣=a q'asa*] is just a sequence of different words, and thus Hawkins and Cutler's model makes no prediction about its preferred order. Alternatively, we might take a view in which =*χ̣=a* is a prefix that conveys the syntactic information of *q'asa*; then, again, Hawkins and Cutler's model has no explanation as to why they are phonologically attached to the previous word, because that does not change the prefix-stem order which needs to be avoided according to their account. They do not explain why the mirror image is not attested, i.e. why suffixes or postposed grammatical morphemes are never phonologically attached to the next word.

7.6.2 Himmelmann (2014)

Himmelmann's (2014) production-based account can basically make the same predictions about the claims in this chapter.[15] First, just like the account presented here, his account does not directly rely on the meaning or grammatical function of affixes, but on the *frequency* (and relative ease of the production) of affixes. Thus, we can expect that the same argument can be extended to clitics. His account explains the phonological asymmetry too. Because his account predicts that more pauses or other signs of dysfluency occur after grammatical morphemes than before them, it is predicted that grammatical morphemes are more likely to be phonologically independent from the following item.

[15]His paper itself argues that clitics can be handled in the same way as affixes.

7.7 Conclusion

We argued in the previous chapters that preposed grammatical morphemes are on average harder to identify than postposed grammatical morphemes. There are multiple strategies for a language to overcome this difficulty. The first strategy is to avoid such a construction altogether, which results in the typological rarity of preposed grammatical morphemes. The second strategy is to clearly demarcate the phonological boundary between the preposed grammatical morpheme and the following item, so that the morpheme boundary is easily identified. We saw that this second strategy is actually employed in a variety of languages.

Chapter 8

Conclusion

In this thesis, we presented a novel account of a long-standing problem in the typology of morphology, i.e. the suffixing preference. Our account is based on the ease of segmentation in language comprehension. The core idea is that a short morpheme at the beginning of a word (or a similar unit) is difficult to recognize immediately, as such a morpheme is highly ambiguous.

We believe that our account is parsimonious in that it only relies on visible properties of morphemes such as length and frequency, as well as a parser equipped with probabilistic knowledge. The assumption that people can use their probabilistic knowledge to solve their cognitive tasks has been proved to be useful in many areas of cognitive science (M. D. Lee & Wagenmakers, 2014). Our account assumes a very simple language model, and does not stipulate a processing mechanism specific to morphology or to particular types of morphemes.

We have argued that our account is compatible with the following facts:

(88) a. Boundedness asymmetry: The suffixing preference is observed not because grammatical morphemes are more often postposed, but because postposed grammatical

morphemes are more likely to be bound (Chapter 4).

b. Psycholinguistic evidence: derivational prefixes are less likely to be parsed out than suffixes are in language comprehension (Cole et al., 1989), and whether a prefix is parsed out is more heavily affected by phonotactics (Hay, 2003) etc (Chapter 4).

c. Prefixes do not cause processing difficulty in every language (Chapter 5).

d. The strength of the suffixing preference varies depending on grammatical categories. This fact is explained in our approach by taking into account the word order asymmetries of the historical sources of affixes (Chapter 6).

e. Prefixes tend to be phonologically more independent from the stem than suffixes (Chapter 7).

f. Asymmetries similar to the suffixing preference can also be found in clitics and compounds. Enclitics are more common than proclitics. In compounds with three stems, the first element tends to be phonologically more independent than the last item (Chapter 7).

Most of the previous work has not been as extensive as our account. However, our account has not explained every fact that has been argued to be relevant, including:

(89) a. Asymmetry in learning, especially in domains other than language (Hupp et al., 2009)

b. Asymmetry in the pattern of dysfluency in spontaneous speech (Himmelmann, 2014)

This might suggest that more than one factor conspires to the suffixing preference. Of course, some of the points above might be just coincidences and might turn out to have nothing to do with the suffixing preference.

8.1 Future directions

In this last section, we will discuss some future directions.

8.1.1 More sophisticated measures of processing difficulty

Our study only calculated how confident the parser is about a morpheme boundary when a morpheme boundary is in fact there. This measure does not fully capture the degree of mismatch between the parser's belief and the correct parse from the following reasons. First, it ignores *when* the parser reaches the correct identification of a morpheme, which can be either before or after the morpheme boundary. Second, it ignores the effect of the possibility that the parser identifies a morpheme boundary in the wrong place. Third, in some cases, the parser could commit to a particular wrong analysis, while in some other cases, the parser could just have no idea, confronted with multiple possibilities. Our study did not have a principle way to handle consequences of those different situations.

The processing difficulty caused by the gap between the correct parse and the expectations of the parser has been formalized as *surprisal*, the negative log probability of input, in the recent psycholinguistic literature (Hale, 2001; Levy, 2008) and it has been applied to morphology too (Balling & Baayen, 2012). The implementation of the surprisal model will allows us to track the time course of the gap between the parser's expectation and the correct

parse.

There are other threads of thoughts as well about what causes difficulty in language comprehension. One is theories that resort to memory limitation: the parser faces difficulty when there are too many unresolved dependencies, for example. This type of account has been invoked to explain the low acceptability of center-embedding sentences in Miller and Chomsky (1963), and more recently it has been formalized as the Dependency Locality Theory (Gibson, 1998), or applied to typology by Hawkins (1994, 2004). Yet another potential cause of processing difficulty is interference (Gordon, Hendrick, and Johnson (2001), among others): difficulty arises when more than one equally reasonable interpretation is available. Each theory will yield different predictions, and their comparisons would only become possible when these ideas are fully formalized and implemented.

This move will also enable us to include sublexical cues in the model in a principled way. A weakness of our model was that it could not directly handle sublexical cues for morpheme segmentation, such as phonotactic, suprasegmental, or phonetic cues. The moment that those cues become available often does not coincide with the moment that the parser reaches the morpheme boundary. For example, when a language has a fixed stress on the first syllable, it must be a powerful cue for a word boundary, but this cue becomes available only after the syllable following that boundary is heard. This suggests that in order to quantify the effect of these suprasegmental cues, it is necessary to track the time course of the mismatch between the parser's belief and the correct parse in its entirety, not just to calculate confidence values at morpheme boundaries.

Another goal related to the issue above is to directly model the effect of a loss of phonological cues, which will open the possibility that language change itself is implemented as a part of the model.

8.1.2 Finite State Morphology

The model presented in this thesis has no consistent way to deal with allomorphy, let alone infixes, circumfixes, ablaut or templatic morphology, which seriously limited the target languages the simulation program could handle. In order to deal with the full diversity of morphology in the world's languages, we need a more sophisticated model of computational morphology, such as Finite State Morphology (Koskenniemi, 1983; Beesley & Karttunen, 2003). In Finite State Morphology, words are parsed into a stem plus a set of features, no matter whether it is concatenative morphology or not. For example, *took* can be translated to *take*+PAST. However, a simple application of Finite State Morphology cannot supersede our approach, because that means that we will introduce a distinct model for morphology, which eliminates the merit of our model in which word boundaries and word-internal morpheme boundaries can be compared within the same model. More work will be needed to figure out how a more advanced model of computational morphology can be integrated to our account.

8.1.3 Left-right asymmetries in phonology

A number of left-right asymmetries are known in phonology, including the dominance of anticipatory assimilation over progressive assimilation. While some of the phonological asymmetries may be due to articulatory (e.g. MacNeilage (1970)) or perceptual (e.g. Ohala (1990))

factors, there have also been proposals that sound change is driven by the redundancy of information carried by a speech segment (Aylett & Turk, 2004; Jurafsky, Bell, Gregory, & Raymond, 2000; Jaeger, 2010; Cohen Priva, 2012).

For example, it has been argued that some parts of a word are more resistant to sound change, as mentioned in Section 3.5. A possible explanation for this is that sound change at some positions is statistically more damaging for the identification of morphemes in language comprehension. Our simulation model can be used to evaluate how severe the effect of a sound change is for the identification of a morpheme, thereby explaining phonological universals with the same framework as the one for the suffixing preference.

Another issue that we were not able to investigate in this thesis is the possible link between the development of affixes and the prosodic profile of a language. The prosody of a language has been argued to have an immense effect on what morphologization looks like. For example, Donegan and Stampe (2004) argue that the difference in prosody is behind many of the typological contrasts between Munda and Mon-Khmer languages in Austroasian, including their affixation patterns. While Mon-Khmer languages are VO languages with limited morphology with a few prefixes, Munda languages are OV languages that have richer morphology with both prefixes and suffixes. They argue that those typological contrasts result from a change from a rising to a falling rhythm that happened in the history of Munda. We had almost no discussion about the effect of prosody on the development of affixes, but if Donegan and Stampe's discussion is correct, prosody should not be ignored in explaining the frequencies of affixes. Schiering (2006) also discusses how the rhythm-based typological

classification of a language can capture the pattern of phonological reduction of clitics, which may have implications about in what context an affix is likely to develop.

References

Aikhenvald, A. Y. (1999). *Tariana Texts and Cultural Context*. Munich: Lincom Europa.

Aikhenvald, A. Y. (2006). *Grammar of Tariana, from Northwest Amazonia*. Cambridge: Cambridge University Press.

Allen, M. (1978). *Morphological Investigations*. Unpublished doctoral dissertation, University of Connecticut.

Anderson, S. R. (1984). Kwak'wala syntax and the Government-Binding theory. In E. D. Cook & D. B. Gerdts (Eds.), *The Syntax of Native American Languages* (pp. 21–75). New York: Academic Press.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(Pt 1), 31–56.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and Plurals in Dutch: Evidence for a Parallel Dual-Route Model. *Journal of Memory and Language*, *37*(1), 94–117.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *CELEX2*. Philadelphia: Linguistic Data Consortium.

Baayen, R. H., & Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, *358*(1769), 1281–1293.

Baayen, R. H., Schreuder, R., & Sproat, R. (2000). Morphology in the mental lexicon: a computational model for visual word recognition. In F. van Eynde & D. Gibbon (Eds.), *Lexicon Development for Speech and Language Processing* (pp. 267–293). Kluwer Academic Publishers.

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension

of morphologically complex words. *Cognition*, *125*(1), 80–106.

Bates, D. E. (1988). *Prominence Relations and Structure in English Compound Morphology*. Unpublished doctoral dissertation.

Bauer, L. (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, L. (2009). Typology of compounds. In R. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Compounding* (pp. 343–356). Oxford: Oxford University Press.

Beard, R. (1991). Decompositional composition: the semantics of scope ambiguities and 'bracketing paradoxes'. *Natural Language & Linguistic Theory*, *9*(2), 195–229.

Beckman, J. N. (1998). *Positional Faithfulness*. Unpublished doctoral dissertation, Univeristy of Massachusetts, Amherst.

Beesley, K. R., & Karttunen, L. (2003). *Finite-State Morphology: Xerox Tools and Techniques*. Center for the Study of Language and Information.

Biberauer, T., Holmberg, A., & Roberts, I. (2014). A syntactic universal and its consequences. *Linguistic Inquiry*, *45*(2), 169–225.

Biberauer, T., Holmberg, A., Roberts, I., & Sheehan, M. (n.d.). *The Final-over-Final Constraint*. Cambridge, MA: MIT Press.

Bickel, B. (2008). A refined sampling procedure for genealogical control. *Language Typology and Universals*, *61*, 221–233.

Bickel, B., Hildebrandt, K. A., & Schiering, R. (2009). The distribution of phonological word domains: a probabilistic typology. In J. Grijzenhout & B. Kabak (Eds.), *Phonological Domains: Universals and Deviations* (pp. 47–75). Berlin: Mouton de Gruyter.

Bod, R., Hay, J., & Jannedy, S. (2003). Introduction. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (Vol. 104, pp. 1–8). Cambridge, MA: MIT Press.

Bolinger, D. (1967). Adjectives in English: attribution and predication. *Lingua*, *18*, 1–34.

Booij, G., & Rubach, J. (1984). Morphological and prosodic domains in lexical phonology. *Phonology Yearbook*, *1*, 1–27.

Brants, T., & Franz, A. (2006). *Web 1T 5-gram Version 1*. Philadelphia: Linguistic Data Consortium.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Boume, J. Kraemer, & G. Zwarts (Eds.), *Cognitive Foundations of Interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Brown, D., & Chumakina, M. (2012). What there might be and what there is: an introduction to Canonical Typology. In D. Brown, M. Chumakina, & G. G. Corbett (Eds.), *Canonical Morphology and Syntax* (pp. 1–19). Oxford: Oxford University Press.

Brown, L., & Dryer, M. S. (2008). The verbs for 'and' in Walman, a Torricelli language of Papua New Guinea. *Language*, *84*(3), 528–565.

Brown, P. E., Pietra, V. J. D., Mercer, R. L., Pietra, S. A. D., & Lai, J. C. (1992). An estimate of an upper bound for the entropy of English. *Computational Linguistics*, *18*(1), 31–40.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language Production Vol 2* (pp. 257–294). London: Academic Press.

Bybee, J. (1985). *Morphology*. John Benjamins.

Bybee, J. (1988). Morphology as lexical organization. In *Theoretical Morphology* (pp. 119–141). Academic Press.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J., Pagliuca, W., & Perkins, R. D. (1990). On the asymmetries in the affixation of grammatical material. In W. Croft, K. Denning, & S. Kemmer (Eds.), *Studies in Diachronic Typology for Joseph H. Greenberg* (pp. 1–42). John Benjamins.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297–332.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–44.

Ching, H.-l. H. (2008). *The advantages of topicalization: an experimental study of sentence processing in Cantonese.* Unpublished MA thesis. University of Cambridge.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Clark, E. V. (2001). Morphology in Language Acquisition. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology.* Blackwell.

Clark, H. H., & Malt, B. C. (1984). Psychological constraints on language: A commentary on Bresnan and Kaplan and on Givón. In W. Kintsch, J. R. Miller, & P. G. Polson (Eds.), *Methods and Tactics in Cognitive Science* (pp. 191–214). Hillsdale: Laurence Erlbaum.

Cohen Priva, U. (2012). *Sign and Signal: Deriving Linguistic Generalizations from Information Utility*. Unpublished doctoral dissertation, Stanford University.

Cole, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language*, *28*(1), 1–13.

Comrie, B. (1980). Morphology and word order reconstruction: problems and prospects. In J. Fisiak (Ed.), *Historical Morphology* (pp. 359–372). The Hague: Mouton.

Comrie, B. (2008). Prenominal relative clauses in verb-object languages. *Language and Linguistics*, *9*(4), 723–733.

Cowgill, W. (1966). A search for universals in Indo-European diachronic morphology. In J. H. Greenberg (Ed.), *Universals of Language* (pp. 114–141). Cambridge, MA: MIT Press.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Cutler, A., Hawkins, J. A., & Gilligan, G. (1985). The suffixing preference: a processing explanation. *Linguistics*, *23*, 723–758.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology. Human Perception and Performance*, *14*(1), 113–121.

Cysouw, M. (2005). Quantitative methods in typology. In *Quantitative Linguistics: An International Handbook* (pp. 554–578). Berlin: Mouton de Gruyter.

Cysouw, M. (2009). The asymmetry of affixation. *Snippets*, *20*, 10–14.

Cysouw, M. (2010). Dealing with diversity: Towards an explanation of NP-internal word order frequencies. *Linguistic Typology*, *14*, 253–286.

Dahl, O. (1979). Typology of sentence negation. *Linguistics*, *17*, 79–106.

De Pauw, G., & Schryver, G.-M. de. (2008). Improving the computational morphological analysis of a Swahili corpus for lexicographic purposes. *Lexikos*, *18*, 303–318.

Dilley, L. C., & Pitt, M. a. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America*, *122*(4), 2340–53.

Dixon, R. M. W. (1977). *A Grammar of Yidiny*. Cambridge: Cambridge University Press.

Dixon, R. M. W., & Aikhenvald, A. Y. (2002). Word: a typological framework. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Word: A Cross-linguistic Typology* (pp. 1–41). Cambridge: Cambridge University Press.

Donegan, P., & Stampe, D. (2004). Rhythm and the synthetic drift of Munda. In R. Singh (Ed.), *The Yearbook of South Asian Languages and Linguistics* (pp. 3–36). Berlin: Mouton de Gruyter.

Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language*, *13*(2), 257–292.

Dryer, M. S. (1992a). The Greenbergian word order correlations. *Language*, *68*(1), 81–138.

Dryer, M. S. (1992b). *Verb-initial languages and the suffixing preference.* Paper presented at 1992 Linguistic Society of America Annual Meeting.

Dryer, M. S. (1997). Are grammatical relations universal? In J. Bybee, J. Haiman, & S. Thompson (Eds.), *Essays on Language Function and Language Type: Dedicated to T. Givón* (pp. 115–143). Amsterdam: John Benjamins.

Dryer, M. S. (2002). Case distinctions, rich verb agreement, and word order type (Comments on Hawkins' paper). *Theoretical Linguistics*, *28*, 151–157.

Dryer, M. S. (2003). Word order in Sino-Tibetan languages from a typological and geographical perspective. In G. Thurgood & R. LaPolla (Eds.), *Sino-Tibetan Languages* (pp. 43–55). Richmond: Curzon Press.

Dryer, M. S. (2009). The Branching Direction Theory of Word Order Correlations Revisited. In S. Scalise, E. Magni, & A. Bisetto (Eds.), *Universals of Language Today* (pp. 185–207). Berlin: Springer.

Dryer, M. S. (2011a). The evidence for word order correlations. *Linguistic Typology*, *15*(2), 335–380.

Dryer, M. S. (2011b). Order of adverbial subordinator and clause. In M. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Munich: Max Planck Digital Library.

Dryer, M. S. (2011c). Order of negative morpheme and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Munich: Max Planck Digital Library.

Dryer, M. S. (2011d). Position of case affixes. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Munich: Max Planck Digital Library.

Dryer, M. S. (2011e). Prefixing vs. suffixing in inflectional morphology. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Munich: Max Planck Digital Library.

Dryer, M. S. (2011f). Relationship between the order of object and verb and the order of relative clause and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Munich: Max Planck Digital Library.

Dryer, M. S. (2012). *Word order and language family size.* working paper presented at University at Buffalo.

Dryer, M. S. (2013). *In defense of words.* working paper presented at Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S., & Haspelmath, M. (2011). *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79–82.

Enrique-Arias, A. (2000). Spanish object agreement markers and the typology of object agreement morphology. In S. Dworkin & D. Wanner (Eds.), *New Approaches to Old Problems: Issues in Romance historical linguistics* (pp. 149–164). Amsterdam: John Benjamins.

Enrique-Arias, A. (2002). Accounting for the position of verbal agreement morphology with psycholinguistic and diachronic explanatory factors. *Studies in Language*, *26*(1), 1–31.

Fabb, N. A. J. (1984). *Syntactic Affixation*. Unpublished doctoral dissertation, MIT.

Frank, M. C., Goldwater, S. J., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–25.

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 165–183). Kluwer Academic Publishers.

Gell-mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *PNAS*, *108*(42), 17290–17295.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(20), 8051–6.

Givón, T. (1977). On the SOV origin of the suffixal agreement conjugation in Indo-European and Semitic. In A. Juilland (Ed.), *Linguistic Studies Offered to Joseph Greenberg on the Occasion of His Sixtieth Birthday* (pp. 481–503). Saratoga: Anma Libri.

Givón, T. (1979). *On Understanding Grammar*. Academic Press.

Givón, T. (1988). The pragmatics of word-order: predictability, importance and attention. In M. Hammond, E. A. Moravcsik, & J. Wirth (Eds.), *Studies in Syntactic Typology* (pp. 243–284). John Benjamins.

Givón, T. (2000). Reconstructing grammar, comparative linguistics and grammaticalization. In S. Gildea (Ed.), *Reconstructing Grammar: Comparative Linguistics and Grammaticalization* (pp. 107–160). John Benjamins.

Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9163–8.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*(6), 1411–1423.

Greenberg, J. (1957). *Essays in Linguistics*. Wenner-Gren Foundation.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In *Universal of Language: Volume 2* (pp. 73–113).

Greenberg, J. (1978). How does a language acquire gender markers? In J. H. Greenberg (Ed.), *Universals of Human Language: Word Structure (Vol. 3)* (pp. 47–82). Stanford: Stanford University Press.

Haider, H. (2001). Why are there no complex head-initial compounds? In C. Schaner-Wolles, J. Rennison, & F. Neubart (Eds.), *Naturally!* (pp. 165–174). Torino: Rosenberg & Sellier.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL 2* (pp. 159–166).

Hall, C. J. (1988). Integrating diachronic and processing principles in explaining the suffixing preference. In J. A. Hawkins (Ed.), *Explaining Language Universals.* Blackwell.

Hall, C. J. (1992). *Morphology and Mind*. Routledge.

Halpern, A. L. (1992). *Topics in the Placement and Morphology of Clitics*. Unpublished doctoral dissertation, Stanford University.

Halpern, A. L. (1998). Clitics. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 101–122). Blackwell.

Hannahs, S. J. (1995). The phonological word in French. *Linguistics*, *33*(6), 1125–1144.

Harris, A. C. (2006). In other words: external modifiers in Georgian. *Morphology*, *16*(2), 205–229.

Harris, A. C. (2008). On the explanation of typologically unusual structures. In J. Good (Ed.), *Linguistic Universals and Language Change* (pp. 54–78). New York: Oxford University Press.

Hashimoto, M. (1986). The altaicization of Northern Chinese. In J. McCoy & T. Light (Eds.), *Contributions to Sino-Tibetan Studies* (pp. 76–97). Leiden: Brill.

Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, *11*(1), 119–132.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, *86*(3), 663–687.

Haspelmath, M. (2011a). The gradual coalescence into "words" in grammaticalization. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Grammaticalization* (pp. 342–355). Oxford: Oxford University Press.

Haspelmath, M. (2011b). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, *45*(1), 31–80.

Hawkins, J. A. (1994). *Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (2002). Symmetries and asymmetries: their grammar, typology and parsing. *Theoretical Linguistics*, *28*(2), 95–149.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hawkins, J. A. (2014). *Cross-Linguistic Variation and Efficiency*. Oxford: Oxford University Press.

Hawkins, J. A., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins (Ed.), *Explaining Language Universals* (pp. 280–317). Blackwell.

Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua*, *74*, 219–259.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, *39*(6), 1041–1070.

Hay, J. (2003). *Causes and Consequences of Word Structure*. Routledge.

Hay, J., & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and procesing restrictions in derivational morphology. *Natural Language & Linguistic Theory*, *22*(3), 565–596.

Heine, B., & Reh, M. (1984). *Grammaticalization and Reanalysis in African Languages*. Hamburg: Buske Helmet Verlag Gmbh.

Hildebrandt, K. A. (2007). Prosodic and grammatical domains in Limbu. *Himalayan Linguistics*, *8*, 1–34.

Himmelmann, N. P. (2014). Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language*, *90*(4), 927–960.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, *24*(6), 876–909.

Hurskainen, A. (1999). SALAMA Swahili Language Manager. *Nordic Journal of African Studies*, *8*(2), 139–157.

Hyman, L. M. (1977). On the nature of linguistic stress. In L. M. Hyman (Ed.), *Studies in Stress and Accent* (pp. 37–82). Los Angeles: Department of Linguistics, University of Southern California.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*(02), 225–257.

Hyman, L. M. (2008). Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics*, *46*(2), 309–350.

Inkelas, S. (1993). Nimboran position class morphology. *Natural Language & Linguistic Theory*, *11*(4), 559–624.

Institute for Asian and African Studies (University of Helsinki), & CSC – IT Center for Science. (2004). *Helsinki Corpus of Swahili.*

Jacques, G. (2012). Agreement morphology: The case of Rgyalrongic and Kiranti. *Language and Linguistics*, *13*(1), 83–116.

Jacques, G. (2013). Harmonization and disharmonization of affix ordering and basic word order. *Linguistic Typology*, *17*, 187–215.

Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Javkin, H. R. (1978). *Phonetic Universals and Phonological Change*. Unpublished doctoral dissertation, University of California, Berkeley.

Joseph, B. D. (2004). Morphologization from syntax. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 1–16). Blackwell.

Joseph, B. D., & Janda, R. D. (1988). The how and why of diachronic morphologization and demorphologization. In M. Hammond & M. Noonan (Eds.), *Theoretical Morphology*

(pp. 193–210). San Diego: Academic Press.

Julien, M. (2002). *Syntactic Heads and Word Formation*. Oxford: Oxford University Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2000). Probabilistic relations between words: evidence from reduction in lexical production. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229–254). Amsterdam: John Benjamins.

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing*. Pearson Prentice Hall.

Jusczyk, P. W., & Luce, P. A. (2002). Speech perception and spoken word recognition: past and present. *Ear and Hearing*, *23*(1), 2–40.

Kageyama, T. (1993). *Bunpo to Gokeisei [Grammar and Word Formation]*. Tokyo: Hituzi.

Kageyama, T. (2001). Word plus: The intersection of words and phrases. In *Issues in Japanese phonology and morphology* (pp. 245–276).

Kahr, J. C. (1976). The renewal of case morphology: sources and constraints. *Working Papers on Language Universals*, *20*, 107–151.

Kang, O. (1992). *Korean prosodic phonology*. Unpublished doctoral dissertation, University of Washington.

Kayne, R. S. (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Kiparsky, P. (1996). The shift to head-initial VP in Germanic. In H. Thrainsson, J. Peter, & S. Epstein (Eds.), *Comparative Germanic Syntax.* Kluwer Academic Publishers.

Kirby, S. (1996). *Function, Selection and Innateness: the Emergence of Language Universals*. Unpublished doctoral dissertation, University of Edinburgh.

Klavans, J. L. (1985). The independence of syntax and phonology in cliticization. *Language*, *61*(1), 95–120.

Koskenniemi, K. (1983). *Two-level morphology: A general computational model for word-form recognition and production.* University of Helsinki, Department of General Linguistics.

Kruspe, N. (2004). *A Grammar of Semelai*. Cambridge: Cambridge University Press.

Kubozono, H. (1993). *The Organization of Japanese Prosody*. Tokyo: Kurosio.

Kuczaj, S. A. (1979). Evidence for a language learning strategy: on the relative ease of acquisition of prefixes and suffixes. *Child Development*, *50*, 1–13.

Kulikov, L. (2008). Evolution of case systems. In A. Malchukov & A. Spencer (Eds.), *The Oxford Handbook of Case* (pp. 439–457). Oxford: Oxford University Press.

Laudanna, A., & Burani, C. (1995). Distributional properties of derivational affixes: implications for processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 345–364). Laurence Erlbaum Associates.

Lee, G. G., Lee, J.-h., & Cha, J. (2002). Morpheme segmentation and estimation for hybrid part-of-speech tagging of Korean. *Computational Linguistics*, *28*(1), 53–70.

Lee, M. D., & Wagenmakers, E.-j. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.

Lehmann, C. (1982). *Thoughts on Grammaticalization* (No. 9). University of Cologne, Institut für Sprachwissenschaft.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–77.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological review*, *77*(3), 182–96.

Maddieson, I. (2013). Voicing and Gaps in Plosive Systems. In M. S. D. Haspelmath & Martin (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Marantz, A. (1988). Clitics, morphological merger, and the mapping to phonological structure. In M. Hammond & M. Noonan (Eds.), *Theoretical Morphology* (pp. 253–270). San Diego: Academic Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102.

Maslova, E. (2000). A dynamic approach to the verification of distributional universals. *Linguistic Typology*, *4*(3), 307–333.

Matthews, S., & Yeung, L. Y. Y. (2001). Processing motivatons for topicalization in Cantonese. In *Cognitive-Functional Linguistics in an East Asian Context* (pp. 81–102).

Tokyo: Kurosio.

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 644–655.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology. General*, *134*(4), 477–500.

Maurits, L., Perfors, A., & Navarro, D. (2010). Why are some word orders more common than others? A uniform information density account. *Advances in Neural Information Processing Systems*, *23*, 1585–1593.

Mcclelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McQueen, J. M., & Cutler, A. (2001). Morphology in Word Recognition. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology.* Blackwell.

Miller, G. a., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 419–491). New York: Wiley.

Mithun, M. (1987). Is basic word order universal? In R. S. Tomlin (Ed.), *Grounding and Coherence in Discourse* (pp. 281–328). Amsterdam: John Benjamins.

Mithun, M. (2003). Why prefixes? *Acta Linguistica Hungarica*, *50*(1), 155–185.

Mukai, M. (2008). Recursive compounds. *Word Structure*, *1*, 178–198.

Myler, N. (2009a). *FOFC and morphology.* University of Cambridge.

Myler, N. (2009b). *Form, function and explanation at the syntax-morphology interface: agreement, agglutination and post-syntactic operations.* Unpublished MA thesis. University of Cambridge.

Nagata, M. (1999). A part of speech estimation method for Japanese unknown words using a statistical model of morphology and context. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* - (pp. 277–284). Morristown, NJ, USA: Association for Computational Linguistics.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Newmeyer, F. J. (2000). On the reconstruction of 'proto-world' word order. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language* (pp. 372–388). Cambridge University Press.

Norris, D. (1994). Shortlist: a connectionist model continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological review*, *115*(2), 357–95.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech.* New York: Springer-Verlag.

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge University Press.

Osgood, C. E. (1949). The similarity paradox in human learning: a resolution. *Psychological review*, *56*(3), 132–43.

Otsu, Y. (1980). Some aspects of rendaku in Japanese and related problems. *MIT Working Papers in Linguistics*, *2*, 207–227.

Paulian, C. (1975). *Le kukuya: langue teke du Congo*. Paris: SELAF.

Pesetsky, D. (1985). Morphology and logical form. *Linguistic Inquiry*, *16*(2), 193–246.

Peterson, D. A. (2001). Ingush ʔa: the elusive type 5 clitic? *Language*, *77*(1), 144–155.

Plag, I., & Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language*, *85*(1), 109–152.

Poser, W. J. (1990). Word-internal phrase boundary in Japanese. In S. Inkelas & D. Zec (Eds.), *The Phonology-Syntax Connection* (pp. 279–288). Chicago: University of Chicago Press.

Postal, P. (1969). Anaphoric islands. In *CLS 5* (pp. 205–239).

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*(6), 1017–23.

Reh, M. (1986). Where have all the case prefixes gone? *Afrikanistische Arbeitspapiere*, *5*, 121–134.

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.

Sadock, J. M. (1991). *Autolexical Syntax: A theory of parallel grammatical representations*. Chicago: University of Chicago Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Sapir, E. (1921). *Language*. New York: Harcourt, Brace and Company.

Sato, H. (2012). Swahili-go ni okeru doshi fukugotai no keitaiso kaiseki no kokoromi. *Corpus-ni motozuku gengogaku kyoiku kenkyu hokoku*, *9*, 129–142.

Schiering, R. (2006). *Cliticization and the Evolution of Morphology: A Cross-linguistic Study on Phonology in Grammaticalization*. Unpublished doctoral dissertation, University of Konstanz.

Schiering, R., Bickel, B., & Hildebrandt, K. A. (2010). The prosodic word is not universal, but emergent. *Journal of Linguistics*, *46*(03), 657–709.

Schreuder, R., & Baayen, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, *33*, 357–375.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 131–154). Lawrence Erlbaum Associates.

Schriefers, H., Zwisterlood, P., & Roelofs, A. (1991). The indentification of morphologically complex spoken words: Continous processing or decomposition? *Journal of Memory and Language*, *30*, 26–47.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405.

Selkirk, E., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. In S. Inkelas & D. Zec (Eds.), *The Phonology-Syntax Connection* (The Univer ed., pp. 313–337). Chicago.

Shannon, C. E. (1948). The mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423, 623–656.

Siegel, D. C. (1974). *Topics in English Morphology*. Unpublished doctoral dissertation, MIT.

Siewierska, A., & Bakker, D. (1996). The distribution of subject and object agreement and word order type. *Studies in Language*, *20*(1), 115–161.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of Child Language Development* (pp. 175–208). New York: Holt, Rinehart & Winston.

Smith, J. L. (2002). *Phonological Augmentation in Prominent Positions*. Unpublished doctoral dissertation, Univeristy of Massachusetts, Amherst.

Song, J. J. (2012). *Word Order*. Cambridge: Cambridge University Press.

Spencer, A. (1988). Bracketing paradoxes and the English lexicon. *Language*, *64*(4), 663–682.

Spencer, A. (1991). *Morphological Theory*. Blackwell.

Spencer, A., & Luís, A. R. (2012a). The canonical clitic. In D. Brown, M. Chumakina, & G. G. Corbett (Eds.), *Canonical Morphology and Syntax* (pp. 123–150). Oxford: Oxford University Press.

Spencer, A., & Luís, A. R. (2012b). *Clitics: An Introduction*. Cambridge: Cambridge University Press.

St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: the suffixing preference and grammatical categorization. *Cognitive Science*, *33*(7), 1317–29.

Steele, S. (1977). Clisis and diachrony. In C. N. Li (Ed.), *Mechanisms of Syntactic Change* (pp. 539–579). Austin: University of Texas Press.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, *36*(3), 422–444.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*, 638–647.

Tai, J. H.-Y. (1976). On the change from SVO to SOV in Chinese. In S. B. Steever, C. A. Walker, & S. S. Mufwen (Eds.), *Papers from the Parasession on Diachronic Syntax.* Chicago Lingusitic Society.

Tomlin, R. S. (1986). *Basic Word Order: Functional Principles*. Croom Helm.

Tyler, L. K., Marslen-Wilson, W. D., Rentoul, J., & Hanney, P. (1988). Continuous and Discontinuous Access in Spoken Word-Recognition: The Role of Derivational Prefixes. *Journal of Memory and Language*, *27*(4), 368–381.

Ulrich, C. (1986). *Choctaw Morphophonology*. Unpublished doctoral dissertation, UCLA.

Vennemann, T. (1973). Explanation in syntax. In J. Kimball (Ed.), *Syntax and Semantics 2* (pp. 1–50). New York: Seminar Press.

Štekauer, P., Valera, S., & Kortvélyessy, L. (2012). *Word-Formation in the World's Languages*. Cambridge: Cambridge University Press.

Ward, G., Sproat, R., & Mckoon, G. (1991). A pragmatic analysis of so-called anaphoric islands. *Language*, *67*(3), 439–474.

Wasow, T. (2002). *Postverbal Behavior*. Chicago: University of Chicago Press.

Wennerstrom, A. (1993). Focus on the prefix: evidence for word-internal prosodic words. *Phonology*, *10*(2), 309–324.

Williams, E. (1981). On the notions "Lexically related" and "Head of a word". *Linguistic Inquiry*, *12*(2), 245–274.

Wurm, L. H. (1997). Auditory processing of prefixed English words Is both continuous and decompositional. *Journal of Memory and Language*, *461*(37), 438–461.

Zipf, G. K. (1935). *The Psychobiology of Language*. Houghton-Mifflin.

Appendix

In this thesis, we used a typological database consisting of 1,836 languages, mainly in Chapters 3 and 6. The database is of courtesy of Matthew Dryer (for more details, see Section 3.2). Some of the data points have been previously published as a part of WALS (Dryer, 2011e); the references to most of the original descriptive grammars are also available in WALS. The complete table of all data points is presented from the next page. In the table, 'p' stands for the presence of prefixes, 's' suffixes, 'P' preposed non-bound morphemes, and 'S' postposed non-bound morphemes; tags for bound and non-bound morphemes are separated by /. The languages are classified by large linguistic areas, language families and genera. Each language is tagged for 16 grammatical categories, as explained in Section 3.2. Sometimes a language has both bound and non-bound morphemes for the same grammatical category. The absence of a symbol may mean (i) the lack of information, (ii) the absence of the relevant grammatical morphemes, or (iii) the presence of both preposed and postposed morphemes.

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Africa																
Afro-Asiatic																
BEJA																
Beja	s/	s/S	/	s/S	p/	s/	p/	/	/	p/	s/P	s/	s/S	/P	s/	/
BERBER																
Berber (Ayt Seghrouchen Middle Atlas)																
	/	/	/	/P	/	/	/	/S	/	/	/S	/	/	/	/	/
Berber (Chaouia)	/	/	/	/	/	/	/	/S	/	/P	/S	/	s/	/	/	/
Berber (Figuig)	/P	p/P	/	p/P	/	/P	/	/P	/	/P	/S	/	s/S	/S	/P	/
Berber (Middle Atlas)																
	/P	/P	/	p/P	p/	/	/	/	/	/	s/S	/	s/	/S	/P	/
Berber (Mzab)	/	/	/	/P	/	/	/	/	/	/P	/	s/	/	/	/	/
Berber (Rif)	/P	p/P	/	p/P	/	/P	/S	/S	/P	/P	s/S	/	s/S	/P	/P	/
Berber (Siwa)	/	/	/	/P	/	/	/	/S	/	/	s/	s/	s/	/	/	/
Kabyle	/	/	/	/P	/	/	/	/	/	/	/S	/	/	/S	/	/
Tamashek	/P	/P	/P	p/P	/	/P	/	/S	/	/P	/S	s/	s/	/S	/P	/P
Tashlhiyt	/	p/	/	p/	p/	/P	/	s/	p/	/	/S	/	s/	/	/	/
Tuareg (Air)	/	/	/	/P	/	/	/	/S	/	/	/S	/	s/S	/S	/	/
Tuareg (Ghat)	/	p/	/	/P	p/	/	/	/S	/	/P	/S	/	/S	/	/P	/
Zenaga	/	/	/	/P	/	/	/	/	/	/	s/	/	s/	/	/	/
BIU-MANDARA																
Buduma	/S	p/P	/	/P	/	/	/	/S	/	/S	s/S	s/	s/S	p/P	/S	/
Ga'anda	/	/	/	/P	/	/	s/	/S	s/	/	/S	s/	/S	/	/	/
Giziga	/	/	/	/P	/	/	/	/	/	/S	s/S	/S	s/	/	/	/
Gude	/S	s/P	/	/P	s/	/	s/	s/	/	/P	/S	s/	s/P	/S	/P	/
Hdi	/S	/P	/	/P	/	/	/	/	/	/	/S	s/	s/	/S	/	/P
Lagwan	/S	p/P	/	/P	/	/	s/	s/	/	/S	s/S	s/	s/	p/P	/	/
Lamang	/S	/	/	/P	/	/	/S	/	/	/S	s/S	s/	s/	s/S	/P	/
Mada (in Cameroon)	/	/	/	/	/	/	/	/S	/	/S	/S	/P	/S	/P	/	/
Malgwa	/	/	/	/P	/	/	/	/S	/	s/	/S	s/	s/	s/P	/P	/
Margi	/S	/P	/P	/P	/	/P	/S	/S	/	/S	s/S	/	s/S	/P	/P	/P
Mbara	/S	/P	/	/	/	/P	s/	/S	/	/S	s/S	s/	s/S	/P	/P	/
Mina	/S	/S	/	/P	/	/	/	/S	/	/S	/S	/	/S	/P	/P	/P
Mofu-Gudur	/	/P	/	/P	/	/P	/S	/S	/	/	s/S	/S	/S	/P	/	/
Musgu	/S	/	/	/P	/	/P	/S	/S	/S	/S	s/S	s/	s/	/P	/P	/P
Podoko	/S	/	/	/P	/	/	/	/P	/P	/S	/S	/	s/S	/S	/	/
Tera	/S	/P	/	/P	/	/	/S	/S	/S	/S	s/S	/	s/S	s/P	/P	/
Uldeme	/S	p/	/	/P	/	/	/S	/S	/	/S	/S	/S	/S	p/P	/P	/
Zina	/S	/S	/	/P	/	/	/S	/S	/	/S	/S	s/	/S	/P	/	/
CENTRAL CUSHITIC																
Bilin	s/	s/	/	/S	/	/	/	/	/	s/	s/P	s/	/	s/	/	/
Kemant	s/	s/	/	s/S	s/	/S	/	/P	/	s/	/P	s/	/	s/P	/	/
EAST CHADIC																
Bidiya	/	/	/	/P	/	/	/	/S	/	/	/S	s/	s/	s/P	/	/
Dangaléat	/S	/	/	/	/	/	/	/	/	/S	/S	s/	s/	p/P	/	/
Kera	/S	s/	/	s/P	/	/	/S	/S	/	/S	s/S	p/	s/S	/P	/P	/
Lele	/S	s/	/	/P	/	/	/	/S	/S	/S	/S	s/	s/	/P	/P	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Migama	/	/	/	/P	/	/	/	/	/	/S	/	/	s/	/	/	/
Mokilko	/	/	/	/P	/	/	s/	/	/	/	/	/	s/	p/	/	/
EASTERN CUSHITIC																
Arbore	/	/P	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Dahalo	/	s/	/	/S	/	s/	/	/S	/	/P	/P	s/	s/S	s/P	/	/
Dhaasanac	/	/	/	s/	/	/	/	/S	/	/	/P	/	/S	/P	/	/
Oromo (Harar)	/	s/S	/	s/S	s/	/	/	/S	/	/	/P	/	/S	s/P	/P	/
Oromo (Waata)	/	s/P	/	s/S	s/	/S	/	/S	/	/	/P	/	/S	s/P	/	/
Qafar	/	/	/	/S	/	/	/	/P	/	p/	/P	/	/	/P	/	/
Rendille	/	/	/	s/	/	/	/	/S	/	/	/P	/	/S	/P	/	/
Saho	/	/	/	s/S	p/	/	/	/P	/	/	/	/	/	/	/	/
Sidaama	/	s/S	/	s/S	/	/S	/	/P	/	/P	/P	s/	/P	s/P	/	/
Somali	/	s/S	/	/	s/	/S	s/	s/	/	/P	/P	s/	s/	/P	/P	/
EGYPTIAN-COPTIC																
Coptic	/P	/P	p/	/P	/	/S	/P	/	/P	p/	s/S	/	/P	p/P	/P	/P
MASA																
Masa	/S	/	/	/P	/	/	/S	/	/	/S	s/S	s/	/	/P	/	/
NORTH OMOTIC																
Gamo	/	/	/	s/S	/	/	s/	/P	/	/	/P	s/	/P	/P	/	/
Gimira	s/	s/S	/	/S	s/	s/	/S	/S	/	s/	/P	s/	/P	s/P	/S	/
Koorete	/	s/	/	/S	s/	/	/S	/P	/	s/	/P	s/	/P	s/P	/	/
Maale	s/	s/	/	s/S	/	s/	/	/P	s/	s/	/P	s/	/	/P	s/	/
Wolaytta	s/	s/	/	s/	/	/	/	/P	/	s/	/P	s/	/	s/P	/	/
Yemsa	s/	s/	/	s/S	/	s/	s/	/P	/P	s/	/P	/S	s/P	s/P	/	/
Zayse	/	s/	/	s/	/	/S	/P	/P	/	s/	/P	s/	/P	s/P	/	/
ONGOTA																
Ongota	/S	/	/	/	/	/	/	/	/	/	/P	/	/S	/P	/	/
SEMITIC																
Amharic	/	/	/	s/	/P	/S	/S	/P	/P	/	s/P	s/	s/	/P	/S	/
Arabic (Egyptian)	/P	/P	/P	/P	/	/P	p/	/S	/	/P	s/S	s/	s/S	/P	/P	/
Arabic (Gulf)	/	/P	/	/P	/	/P	p/	/	/	/P	s/S	s/	s/S	/P	/P	/
Arabic (Iraqi)	/	/P	/P	/P	/	/	p/	/	/P	p/	s/S	s/	s/	/P	/P	/P
Arabic (Kuwaiti)	/	/P	/	/P	/P	/	p/	/	/	/P	/S	s/	/	/P	/P	/P
Arabic (Modern Standard)	/P	/P	/	s/P	/P	/P	p/	/P	/	/P	s/S	/	s/	s/S	/P	/P
Arabic (Moroccan)	/P	/P	/	/P	/	/	p/	/P	p/	/	s/S	s/	s/S	/S	/P	/P
Arabic (Syrian)	/S	/P	/	/P	/	/P	p/	/	/	/P	s/S	s/	s/S	/	/P	/P
Chaha	/S	/S	/	/	/	/S	/S	/S	/P	p/	s/P	/	s/	/P	/	/
Chaldean (Modern)	/	s/	/	/P	/	/	/	/P	/	/	s/	s/	s/	s/	/	/P
Hebrew (Modern)	/P	/P	/	/P	/	/P	p/	/S	/S	/P	s/S	s/	s/S	s/P	/P	/
Mehri	/	/	/	/P	/	/	/P	/	/	/S	/S	s/	/S	/	/P	/P
Neo-Aramaic (Arbel Jewish)	/	s/	/	/P	/	/	s/	/P	/P	/P	s/P	s/	s/S	s/P	/P	/
Soddo	/	/	/	/	/	/	/	/	/	p/	/P	/	s/	s/P	/	/
Tigrinya	/	/	/	/P	/	/	/P	/P	/P	/	s/P	s/	s/	/P	/	/
Tigré	/	/S	/	/P	p/	/S	/P	/	/P	p/	/P	s/	s/	s/P	/	/
SOUTH OMOTIC																
Aari	s/	s/	/	s/S	s/	/	s/	/S	/	s/	/P	/	/S	s/P	/	/
Dime	s/	s/	/	/S	s/	/S	/S	/S	/	s/	/P	/	/P	s/P	/	/S

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
SOUTHERN CUSHITIC																
Burunge	s/	s/P	/	/	/	/P	s/	s/	/	s/	s/S	s/	s/S	s/P	/	/
Iraqw	s/	/P	/	/P	s/	/P	/	/S	/	s/	/P	s/	s/S	s/P	/	/S
WEST CHADIC																
Angas	/S	/P	/	/P	/	/	/	/S	/	/S	/S	/S	s/S	/P	/	/
Goemai	/S	/	/	p/P	/	/P	/S	/S	p/	/S	/S	/P	/S	/P	/P	/
Hausa	/S	/P	/P	/P	/P	/S	/S	/	/P	/P	/S	s/	s/S	/P	/P	/P
Kanakuru	/S	s/P	/	/P	/	/P	s/	/S	/	/	s/S	s/	/S	s/P	/P	/
Kwami	/	/	/	/P	/	/	/	/	/	/	/	/	/	/	/	/
Miya	/S	/P	/	/P	/	/	/S	/P	/P	/	s/S	s/	s/S	/	/P	/
Mupun	/S	/P	/	/P	/	/P	/S	/S	/P	/S	/S	/S	/S	/P	/P	/
Ngizim	/	s/P	/	/P	/	/	/S	/S	/S	/S	s/S	s/	s/	/P	/	/
Pa'a	/	s/P	/	/P	s/	/S	/P	/	/P	/S	s/S	s/	s/S	/P	/P	/
Pero	/S	/	/	/P	/	/P	s/	/	/	/	s/S	/	s/	p/P	/	/S
Ron	/S	/	/	/	/	/	/S	/S	/	/S	/S	/	/	/P	/	/
Tangale	/	s/	/	/	/	/	s/	/	/	s/	/S	/	s/	/P	/	/
Yiwom	/	/	/	/P	/	/	/S	/S	/	/	/S	/	/S	/P	/	/
Bangime																
BANGIME																
Bangime	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Kadugli																
KADUGLI																
Kadugli	/	p/	/	/	/	/	/	/S	/	/P	/S	p/	s/	p/P	/	/
Katcha	/	p/P	/	/	/	/P	/	/S	/	/P	/S	p/	/S	/P	/	/
Krongo	s/	p/P	/	p/	/	/P	/	/S	/	/	/S	p/	/S	/S	/P	/
Khoisan																
!UI																
/Xam	/	/	/	/	/	/	/	/P	/	/P	/S	/	/	/P	/	/
N|Uuki	/	s/P	/	/	p/	/P	/	/S	/	/P	/S	/	/P	/P	/	/
=|HOAN																
=|Hoan	/	/	/	/	/	/	/	/S	/	/P	/S	/	/	/P	/	/
JU (NORTHERN KHOISAN)																
Ju|'hoan	/	/P	/	/S	/P	/P	/P	/S	/S	/P	/S	/	/P	/P	/P	/P
KHOE (CENTRAL KHOISAN)																
//Ani	/S	s/	/	/S	/	/	/	/P	/	/S	s/	s/	/P	/P	/P	/
Khoekhoe	/	/P	/S	/S	/S	/	/	/P	s/	/S	s/P	s/	/P	/P	/S	/S
Korana	/S	/	/	s/S	/	/S	/	/P	/	/S	/P	s/	/	/	/S	/
Kxoe	/S	/	/	/S	/	/	/	/P	/	/S	/P	s/	s/	/P	/	/
SANDAWE																
Sandawe	/	s/	/	/S	s/	/	s/	/P	/S	s/	s/P	s/	/P	/P	/P	/
TAA																
!Xóõ	/P	/	/	/	/S	/S	/	/S	/	/	/S	s/	/P	/P	/	/
Laal																
LAAL																
Laal	/P	s/P	/	/P	/	/	/S	/S	/	/S	s/S	/	s/S	/P	/P	/
Niger-Congo																
ADAMAWA																
Day	/	/	/	/	/	/P	/S	/S	/	/	/S	/S	/S	/P	/	/
Doyayo	/S	/P	/	/	s/	/	/S	/	/	/	s/S	s/	s/	/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Gula Iro	/	/	/	/	/	/	/	/	/	/P	/S	/	/	/	/	/
Mambai	/S	s/	/	/P	/	/	/	/	/	/	/	s/	s/	/	/	/
Mbum	/S	/	/	/P	/	/S	/	/S	/	/S	/S	/S	/S	/P	/P	/
Mumuye	/S	/	/	/P	/	/	/S	/S	/S	/S	/S	/S	/P	/P	/	/
Mundang	/	/	/	/P	/	/	/S	/S	/S	/S	/S	/	s/S	/P	/	/
Samba Leko	/S	/	/	/S	/	/	/S	/S	/	/	/S	/S	/S	/P	/	/
Tupuri	/	/	/	/P	/	/	/	/S	/S	/S	/S	/	/	/P	/	/
BANTOID																
Aghem	/	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/	/
Akoose	/S	/	/P	/P	/	/	/P	/P	/	/	/S	p/	/P	p/P	/P	/
Babole	/	/	/	/P	/	/	/	/	/	p/	/S	/	/S	p/P	/	/
Babungo	/S	p/	/	/P	/	/P	/S	/S	/	/	/S	/	/S	/P	/	/
Bafia	/	/	/	/P	/	/	/S	/S	/	/	/	p/	/	/	/	/
Bafut	/	/	/	/P	/	/	/	/	/	/	/S	/	/	/P	/	/
Bakueri	/	p/	/	/	/	/	/	/P	/	p/	p/S	p/	/S	p/P	/	/
Bakundu	/S	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Bangala	/	p/	/	/P	/	/	/	/S	/	/S	/S	p/	/S	p/P	/	/
Bankon	/	/	/	/P	s/	/	/	/	/	/	/S	p/	/S	/P	/	/
Basaá	/	/	/	/P	/	/	/	/S	/	/S	/S	p/	/	s/P	/	/
Bemba	/P	p/	/	/	/	/	/	/	/	/	p/	/	/	p/	/	/
Bembe	/	/	/	/P	/	/	/	/P	/	/	/	/	/P	/	/	/
Bena-Lulua	/	/	/	/P	/	/	/	/	/	/	/S	p/	/	/	/	/
Benga	/	/	/	/P	/	/	/	/	/	/	/	p/	/	/	/	/
Bira	/	/	/	/P	/	/	/	/	/	p/	/S	/	s/	p/P	/	/
Bobangi	/	/	/	/P	s/	/	/	/	/	/	p/S	p/	/S	p/P	/	/
Bolia	/	/	/	/P	/	/	/	/S	/	p/	/S	/	/	p/P	/	/
Bubi	/	/	/	/	/	/	/	/P	/	/	/S	/	/	/P	/	/
Budu	/	/	/	/P	/	/	/S	/S	/	/	p/S	p/	/S	p/P	/P	/
Bujeba	/	/	/	/	/	/	/	/S	/	/	/	/	/S	/	/	/
Bulu	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Bushoong	/	p/	/	/P	/	/	/	/S	/	p/	p/S	p/	/	p/P	/P	/
Chaga	/S	/	/	/	/	/	/	/S	/	p/	/S	/	/S	p/P	/	/
Chichewa	/P	p/	p/	/P	s/	/	/	/S	/	p/	p/S	p/	/S	p/P	/P	/
Digo	/	/	/	/	/	/	/	/P	/	/	/	/	/S	/	/	/
Ding	/	/	/	/P	/	/	/	/S	/	/	/	/	/	/	/	/
Doko	/	/	/	/P	/	/	/	/S	/	/	/S	/	/	/P	/	/
Duala	/	/	/	/	/	/	/	/P	/	/P	/S	/	/	/P	/	/
Duma	/	/	/	/	/	/	/	/S	/	/S	/S	/	/S	/P	/	/
Ekoti	/	/	/	/	/	/	/	/	/	/	/S	/	/	/P	/	/
Enya	/	/	/	/	/	/	/	/	/	p/	/	p/	/	/	/	/
Eton	/P	p/P	/	/P	/	/P	/	/S	/S	p/	/S	p/	/S	p/P	/P	/
Ewondo	/P	p/P	/	/P	/P	/P	/P	/S	/	/P	/S	p/	/S	p/P	/P	/
Fang	/	/	/	/	/	/	/S	/S	/	/	/	/	/S	/	/	/
Haya	/	p/	/	p/P	/	/	/	/S	/	p/	p/S	p/	/S	p/P	/	/
Hehe	/	/	/	/P	/	/	/	/	/	p/	/S	/	/	p/P	/	/
Hemba	/	/	/	/P	/	/	/	/S	/	/	/S	p/	/	/P	/	/
Holoholo	/	/	/	/	/	/	/	/	/	p/	/S	/	/	/P	/	/
Hunde	/P	p/P	/	p/	/	/P	p/	/S	/	p/	/S	p/	/S	p/P	/P	/
Ifumu	/S	/	/	/P	/	/	/	/S	/	/	/S	p/	/S	/P	/P	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Ila	/	p/	/	/P	/	/	/	/S	/	p/	/S	p/	s/S	p/P	/	/
Iwoyo	/	/	/	/P	s/	/	/	/S	/	/	p/S	p/	/S	p/P	/P	/
Jarawa (in Nigeria)	/S	/	/	/P	/	/	/	/S	/	/S	/S	p/	s/	/P	/	/
Kagulu	/	/	/	/P	/	/	/	/S	/	p/	p/S	p/	/S	p/	/P	/
Kako	/	/	/	/P	/	/	/S	/S	/P	/	/	/	/S	/	/	/
Kamba	/	p/	/	/	/	/	/S	/S	/	p/	p/S	p/	/	p/P	/	/
Kami	/	/	/	/P	/	/	/	/	/	p/	/	/	/	/	/	/
Kanyok	/	/	/	/P	/	/	/S	/S	/	/	/	p/	/	/	/	/
Kenyang	/	/	/	/	/	/	/	/S	/	/	/	/	/S	/	/	/
Kete	/	/	/	/P	/	/	/	/S	/	/	/S	/	/	/P	/	/
Kikuyu	/	/P	/	/P	/	/P	/	/S	/	p/	p/S	p/	/S	p/P	/P	/
Kiluba	/S	/	/	/P	/	/	/	/	/	p/	/S	/	/	/P	/P	/
Kinga	/	/	/	/P	/	/	/	/	/	p/	/S	p/	s/	p/P	/	/
Kinyarwanda	/	p/P	/	/P	s/	/	/	/P	/	p/	/S	/	/	p/P	/	/
Kol	/	/	/	/	/	/	/	/S	/	/	/S	/	/S	/	/	/
Kongo	/	/	/	/P	/	/	/	/S	/	/	p/S	p/	/S	p/P	/	/
Kwangali	/S	p/	/	/P	s/	/	/	/S	/	/P	p/S	p/	s/S	p/P	/P	/
Lamba	/	/	/	/P	s/	/P	/	/	/	/	p/	p/	/S	p/	/P	/P
Langi	/	/	/	/P	/	/	/	/S	/	p/	/S	/	/S	p/P	/	/
Lebeo	/	/	/	/P	/	/	/S	/S	/	p/	/S	/	/S	/P	/	/
Lega	/	/	/	/	/	/	/	/	/	p/	p/S	p/	/	p/P	/	/
Lingala	/	/	/	/	/	/	/	/	/	/S	/S	/	/	p/P	/	/
Londo	/S	/	/	/P	/	/	/	/S	/	p/	/S	p/	/S	p/P	/	/
Lozi	/P	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Lucazi	/	/	/	p/	/	/	/	/	/	p/	p/S	/	/	p/P	/	/
Luganda	/	p/P	/	/P	/	/	/	/S	/	p/	p/	p/	s/	p/P	/P	/
Lunda	/	/	/	p/P	/	/	/	/	/	/	p/S	p/	s/	p/P	/P	/
Luvale	/	p/P	/	/P	s/	/P	/	/S	/	/	p/S	p/	s/S	p/P	/P	/S
Luyia	/	/	/	/	/	/	/	/S	/	/	/	/	/S	/	/	/
Makaa	/	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/	/
Makonde	/	/	/	/	/	/	/	/	/	p/	/S	/	/	p/P	/	/
Makua	/	p/	/	/P	/	/	/	/	/	p/	p/S	p/	/S	p/P	/	/
Mambila	/	/	/	/	/	/	/	/	/	/	/S	/	/	/P	/	/
Mambwe	/	/	/	/	/	/	/	/	/	p/	/	/	s/S	p/	/	/
Mankon	/	/	/	/	/	/	/	/S	/	/	/S	/	/S	/P	/	/
Matuumbi	/	p/	/	/	/	/	/S	/S	/	/S	p/S	p/	/S	p/P	/	/
Mbalanhu	/	/P	/	/	/	/	/	/	/	/P	/S	/	/	/P	/	/
Mbere	/	/	/	/P	/	/	/	/S	/	/S	/S	/	/S	/P	/	/
Mbili	/	/	/	/	/	/	/S	/S	/	/P	/S	p/	/	/P	/	/
Mbole	/	/	/	/	/	/	/	/S	/	/	/S	p/	/	p/	/	/
Mbosi	/	/	/	/	/	/	/	/S	/	/S	/S	/	/S	p/P	/	/
Mmen	/	/	/	/	/	/	/	/P	/	/	/	/	/S	/	/	/
Mongo	/	/	/	/P	/	/	/	/S	/	/	/S	/	/S	/P	/	/
Mpongwe	/	/	/	/	/	/	/	/S	/	/P	/S	/	/	/	/	/
Mundani	/	/	/	/	/	/	/S	/S	/S	/	/	/	s/	/	/	/
Mungaka	/S	p/	/	/	/	/	/	/	/	/S	/S	p/	/	/P	/	/
Mwera	/	p/	/	/	s/	/	/	/	/	p/	p/S	p/	/	p/P	/	/P
Nande	/	/	/	/P	/	/	/S	/S	/	p/	/	p/	/	/	/	/
Nchane	/S	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Ndebele (in South Africa)	/S	/	/	p/	/S	/	/	/S	/	/	p/S	p/	/S	p/P	/	/
Ndonga	/	s/P	/	p/P	s/	/P	/	/S	/	/P	/S	p/	s/S	/P	/P	/P
Ndumu	/	p/P	/	/P	/	/	/	/S	/	/	/S	p/	/S	/P	/	/
Ngombe	/	/	/	/	/	/	/	/	/	/	/S	/	/	/P	/	/
Ngoni	/	/	/	/P	s/	/	/	/S	/	/	p/S	p/	/	p/P	/P	/
Nkem	/S	/	/	/	/	/	/	/S	/	/	/S	p/	/S	/P	/P	/
Nkore-Kiga	/S	p/P	/	/P	s/	/P	/	/S	/	p/	p/S	p/	/S	p/P	/	/
Noni	/S	s/P	/	p/	/	/P	/	/S	/	/	/S	p/	/S	/P	/P	/P
Nsenga	/	/	/	/P	/	/	/	/	/	/	/S	/	/	p/P	/	/
Ntomba	/	/	/	/	/	/	/	/	/	p/	/S	/	/S	/P	/	/
Nyakyusa	/	/	/	/	/	/	/	/S	/	/	/	/	/S	/	/	/
Nyamwezi	/	/	/	/	/	/	/S	/S	/S	p/	p/S	p/	/S	p/	/	/
Nyanga	/	/	/	/	/	/P	/	/S	/	/	/S	/	/	/P	/	/
Nyiha	/	/	/	/	/	/	/	/	/	p/	/	/	/S	/	/	/
Nzadi	/P	/P	/	/P	/	/P	/	/S	/S	/P	/S	p/	/S	/P	/P	/
Pangwa	/	/	/	/P	/	/	/	/	/	p/	/S	/	/	p/P	/	/
Pare	/	p/	/	/P	/	/	/	/P	/	p/	p/S	p/	s/S	p/P	/	/
Pogoro	/	/	/	/P	/	/	/	/	/	/S	/S	/	/	p/P	/	/
Rimi	/	/	/	/	/	/	/	/	/	p/	/S	/	/	p/P	/	/
Ronga	/	/	/	/P	/	/	/	/S	/	/	/S	/	/	/P	/	/
Rundi	/	/	/	/	/	/	/	/	/	p/	/	/	/	/P	/	/
Runyankore	/	p/P	/P	/P	s/	/P	/	/S	/	p/	p/S	s/	/S	p/P	/P	/P
Runyoro-Rutooro	/	p/	/	/	/	/	/	/	/	p/	/S	/	/	/P	/	/
Sangu	/	/	/	/	/	/	/	/S	/	p/	/S	/	/S	p/P	/	/
Sena	/	/	/	/P	/	/	/	/	/	/	/	/	s/S	/	/	/
Sengele	/	/	/	/P	/	/	/	/	/	/	/S	/	/	p/P	/	/
Sesotho	/	/	/	/P	/	/	/	/S	/	/P	/S	/	/S	/P	/	/
Shambala	/	p/	/	/	/	/	/S	/P	/	p/	/S	p/	s/S	p/P	/	/
Shona	/	p/P	/	/P	/	/	/	/S	/	p/	p/S	p/	/S	p/P	/P	/
Songe	/	/	/	/P	/	/	/	/	/	/	/	/	/	/	/	/
Sotho (Northern)	/	/	/	/	/	/	/	/	/	/	p/S	p/	/	p/P	/	/
Subiya	/	p/	/	/	/	/	/	/	/	/	/	p/	/	p/	/	/
Suku	/	/	/	/	/	/	/	/S	/	/	/S	/	/	p/P	/	/
Sukuma	/	/	/	/	/	/	/	/	/	/	/S	/	/	/P	/	/
Swahili	/P	p/	/	/P	s/	/P	/P	/S	/	p/	p/S	p/	s/	p/P	/P	/
Swati	/S	/	/	/	/	/	/	/P	/	/	/S	/	/	/	/P	/
Tabwa	/	/	/	/P	/	/	/	/	/	p/	/S	p/	/	p/P	/	/
Talinga	/	/	/	p/P	/	/	/	/	/	p/	/S	/	/	/P	/	/
Tetela	/S	/	/	/P	/	/	/	/S	/	p/	p/S	/	/S	p/P	/	/
Tikar	/	/	/	/	/	/	/	/S	/S	/	/	/	/S	/P	/	/
Tiv	/	/	/	/P	/	/	/	/S	/	/S	/S	/	/S	/P	/	/
Tonga (in Zambia)	/	p/P	/	p/	/	/	/	/S	/	p/	p/S	p/	s/	p/P	/	/
Tsogo	/	/	/	/P	/	/	/S	/S	/	/P	/	p/	/S	/	/	/
Tsonga	/P	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Tuki	/	p/	/	/	/	/	/	/	/	p/	p/S	/	/	p/P	/	/
Tunen	/S	s/P	/	/	/	/P	/	/P	/	/P	/P	p/	/P	/P	/P	/
Venda	/	p/	/	/	/	/	/	/	/	p/	/S	/	/	p/P	/	/
Vili	/	/	/	/P	/	/	/	/S	/	/	/	/	/S	/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Xhosa	/	/	/	p/P	/	p/	/S	/	/	p/	p/S	p/	/S	p/P	/	/
Yansi	/	p/P	/	/	/	/	/	/S	/	/	/	/	/S	p/	/P	/
Yao (in Malawi)	/	p/	/	/	/	/	/	/S	/	p/	p/S	/	/S	p/P	/	/
Yeyi	/	p/	/	/P	s/	/P	/	/S	/	/	p/S	p/	/S	p/P	/	/
Zulu	/S	p/P	p/	p/P	s/	/P	/	/	/	p/	p/S	p/	/S	p/P	/P	/P
CROSS RIVER																
Efik	/P	/P	/	/	/	/	/	/S	/	s/	/	/	/S	p/	/P	/P
Gokana	/P	/	/	/P	/	/	/S	/S	/	/P	s/S	/P	s/P	/P	/P	/
Ibibio	/	/	/	/	/	/	/S	/S	/	/	/S	/	/	/P	/	/
Kana	s/	/P	/P	p/P	/P	/P	/P	/S	/	/P	/S	/P	/P	p/P	/P	/
Leggbó	/	/	/	/	/	/	/S	/S	/	p/	/S	/	/S	/P	/	/
Obolo	/P	p/	/P	/P	/	/P	/	/S	/S	p/	/S	/	/S	p/P	/P	/P
Ogbronuagum	/S	p/	/	/P	/	/	/S	/S	/	p/	/S	p/	/S	p/P	/	/
DEFOID																
Yoruba	/	/P	/	/P	/	/	/	/S	/	/P	/S	/P	/S	/P	/P	/P
DOGON																
Donno So	/S	s/S	/	/S	s/	/S	/S	/S	/	s/	/P	/	/	s/P	/S	/S
Jamsay	/S	s/	/	/S	/	/	/	/S	/S	s/	/P	/S	/	s/P	/S	/
EASTERN MANDE																
Boko	/	/	/	/S	/	/	/	/S	/	/S	/P	/	/	/P	/	/
Busa	/S	/P	/	/S	/	/	/S	/	/	/S	/P	/	/P	/P	/	/
Dan	/S	s/P	/	/S	/	/	/	/S	/	/P	/P	/S	/P	/P	/	/
Mano	/	/S	/	/S	/	/	/	/	/	/P	/P	s/	/P	/P	/P	/
Mona	/	/	/	/	/	/	/S	/P	/	/P	/P	/S	/	/P	/	/
Wan	/	/	/P	/S	/	/	/S	/	/	/S	/P	/	/	/P	/	/
EDOID																
Bini	/	/	/	/P	/P	/P	/	/S	/	/P	/S	/	/S	/P	/	/
Degema	/	/	/	/	/	/	/S	/S	/	/	/S	/	/S	/P	/	/
Emai	/	/	/	/P	/	/	/P	/S	/	/	/S	/	/	/P	/	/
Engenni	/S	/	/	/P	/	/	/S	/S	/	/	/S	/	/S	/P	/	/
GBAYA-MANZA-NGBAKA																
Gbaya Kara	/	/	/	/	/	/	/	/S	/	/S	s/S	p/	s/S	p/P	/	/
Gbeya Bossangoa	/S	s/P	/P	/P	/	/P	s/	/S	/P	/S	s/S	/P	s/S	s/P	/	/P
Mbodomo	/S	s/P	/	/P	s/	/	/	/S	/	/S	/S	p/	/S	/P	/	/
GUR																
Bimoba	s/	s/P	/P	/	/P	/P	/S	/S	/	/P	s/S	/	/P	/P	/	/P
Buli (Ghana)	/	/	/	/	/	/	/	/	/	/	/S	/	/	/P	/	/
Dagaare	/	/	/	/S	/	/	/P	/S	/	/P	/S	/	/	/P	/	/
Dagbani	/S	/	/	/S	s/	/	/S	/S	/	/P	/S	s/	/P	/P	/P	/
Gangam	/	/P	/P	/S	/	/	/	/S	/S	/P	/S	/	/P	/P	/S	/
Kabiyé	/S	/	/	/P	/	/	/	/	/	p/	s/S	/	/	p/P	/	/
Kirma	/	/P	/	/S	/P	/	/	/S	/	/P	/S	s/	/	/P	/P	/
Konni	/	s/	/	/	/	/	s/	/S	s/	/P	/S	s/	/	/P	/	/
Koromfe	/S	s/	/	/	/	/P	/S	/S	/	/P	s/S	s/	/P	p/P	/P	/
Kàsim	/S	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/	/
Mbèlimè	/	/	/	/	/	/	/	/S	/	/	/	/	/P	/	/	/
Mooré	/S	s/P	/	s/	/	/P	s/	s/	/	/	/S	/	/P	/P	/	/P
Nanerge	/	/	/	/	/	/	s/	/P	s/	/S	/	s/	/P	/	/	/
Nateni	/S	/	/	/	/	/	/	/S	/	/	/	/	/P	/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Ncàm	/S	/	/	/	/	/P	/	/S	/	/P	/S	/	p/	/P	/	/
Sisaala	/	/	/	/	/	/	/S	/S	/	/P	/S	s/	/P	/P	/	/
Supyire	/S	/P	/	/S	/P	/P	s/	/P	/S	/S	/P	s/	/P	/P	/P	/P
Tenyer	/S	s/P	/	/S	/	/	/	/	/	/P	/P	s/	/	/P	/	/
Toussian (Northern)	/	p/P	/	/S	/	/P	/	/	/S	/P	/P	s/	/	/P	/P	/
Waama	/S	/	/	/	/	/	/	/S	/	/	/S	/	/P	/P	/	/
HEIBAN																
Moro	/	p/P	/	/	/	/P	/	/S	/	/P	s/S	p/	s/	p/P	/P	/P
Otoro	/S	p/	/	s/	s/	/	/	/S	/	/	/S	p/	/S	/P	/	/
Tira	/S	s/	/	s/	/	/	/	/	/	/S	/S	p/	/S	p/P	/P	/
IDOMOID																
Idoma	/	/	/	/	/	/	/	/S	/	/	/	/	/	/	/	/
Igede	/S	/S	/	/	/	/	/	/S	/	/S	/S	/	/S	p/P	/	/
IGBOID																
Igbo	/	s/P	/	/P	/	/P	/	/S	/	/	/S	/	/	/P	/P	/P
Iká	/	/	/	/	/	/	/	/S	/	/	/	/	/	/	/	/
Izi	/S	/P	/P	/P	/	/P	/	/S	/	/	s/S	/	/S	p/P	/P	/
IJOID																
Defaka	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Ijo (Kolokuma)	s/	s/S	/P	s/	/	/	s/	/P	s/	/	p/P	s/	p/P	/P	/	/
JUKUNOID																
Jukun	/S	/	/S	/P	/	/P	/S	/S	/S	/S	/S	p/	/S	/P	/	/
Kuteb	/S	/	/	/P	/	/P	/S	/S	/S	/	/S	p/	s/S	/P	/	/
KAINJI																
Amo	/S	p/	/P	/	/	/P	/	/S	/	/	s/S	p/	s/S	p/	/P	/P
Duka	/S	s/P	/	/P	/P	/P	/	/P	/P	/S	/S	/	/S	/P	/P	/
KATLA-TIMA																
Katla	/	s/P	/	/	s/	/	/	/S	/	/	/S	p/	s/	p/P	/	/
Tima	/	/	/	/	/	/	/	/S	/	/	/S	p/	/S	/P	/	/
KRU																
Dadjriwalé	/	/	/	/	/	/	s/	/S	/S	/	/	/	/	/P	/	/
Grebo	/P	s/P	/	/S	s/	/S	/	/S	/	/P	/S	s/	/P	/P	/	/
Klao	/	/	/	/	/	/	/S	/	/	/P	/S	s/	/P	/P	/	/
Neyo	/S	/	/	/	/	/	/S	/S	/	/	/	/	/P	/P	/	/
Seme	/	s/P	/	/S	/	/	/	/P	/	/S	/P	/	/	/P	/	/
KWA																
Adioukrou	/S	p/S	/P	/S	/	/P	/S	/S	/	s/	s/S	p/	/P	p/P	/	/
Ajagbe	/S	/S	/P	/	/	/P	/S	/S	/S	/S	/S	/S	s/P	/P	/P	/P
Akan	/P	p/S	/P	/	/	/P	/S	/S	/	p/	s/S	/	p/	p/P	/	/P
Anufo	/S	/	/	/	/	/	/S	/P	/S	/S	/S	/	/P	/P	/	/
Anyi	/	p/	/	/	/	/	/	/	/	/	/S	p/	/	/P	/	/
Avikam	/S	/	/	/	/	/	s/	/S	/	/	/S	/S	/	/P	/	/
Baule	/	/	/	/S	/	/	/	/S	/	/	/S	/	/	/P	/	/
Chumburung	/	/	/	/S	/	/	/	/S	/	/	/S	/	/	/P	/	/
Ega	/S	/	/	/	/	/	/	/S	/	/S	/S	/	/	/P	/	/
Ewe	/S	/	/	/S	/	/P	/S	/S	/S	/	s/S	/	s/P	p/P	/P	/
Fongbe	/	/	/	/	/	/	/S	/S	/S	/P	/S	/S	/	/P	/	/
Gungbe	/S	/	/	/	/	/	/S	/	/	/	/S	/	/	/P	/	/
Gã	/	/	/	/	/	/	/S	/P	/S	s/	/S	s/	/	p/P	/P	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Igo	/	/	/	/	/	/	/S	/S	/S	/	/	/	/	/	/	/
Kposo	/	/	/	/S	/	/	s/	/	s/	/	/S	/	/	/P	/	/
Lelemi	/	p/	/	/S	/	/P	/S	/S	/S	p/	s/S	p/	/P	p/P	/P	/
Logba	/	/	/	/	/	/P	/S	/S	/S	/	/S	p/	/P	p/P	/P	/
Nkonya	/S	p/	/	/S	/	/	/S	/S	/S	p/	/S	p/	/	p/P	/P	/
NORTHERN ATLANTIC																
Balanta	/	/	/	/	/	/	/S	/S	/	p/	/S	/	/S	p/P	/	/
Banjal	/	/	/	/	/	/	/	/S	/	/	/P	p/	s/S	/P	/	/
Basari	/	/	/	/	/	/	/	/	/	/	/P	/	/	p/P	/	/
Diola-Fogny	/	/P	/P	/P	s/	/	s/	/S	/S	s/	s/S	p/	s/	p/P	/P	/P
Fula (Cameroonian)	/S	s/P	/	/P	s/	/P	/S	/S	/	s/	s/S	s/	/S	p/P	/P	/P
Ndut	/	s/	/	/P	s/	/	/	/S	/	s/	/S	s/	/S	/P	/P	/
Noon	/S	s/P	/P	/P	/	/P	s/	/S	/	s/	s/S	s/	s/	/P	/P	/
Wolof	/P	/P	/	/P	/	/S	/S	/S	/P	/	/S	/S	s/P	/P	/P	/P
NUPOID																
Ebira	/S	/P	/	/	/	/	/	/	/	/P	/S	/	/	/P	/	/
Gwari	/S	/P	/	/	/	/	/S	/S	/S	/	/	/	/	/P	/P	/
Nupe	/P	/P	/	/P	/	/	/	/S	/S	/S	/S	/	/S	/P	/	/
PLATEAU																
Birom	/S	p/	/P	/P	/	/P	/S	/S	/S	/S	/S	p/	/S	p/P	/P	/
Migili	/	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/	/
PLATOID																
Fyem	/S	s/P	/	/P	/	/	/S	/S	/S	/S	s/S	p/	s/	/P	/P	/
Gworok	/	/P	/	/P	/	/	/	/S	/S	/S	/S	/	/	/P	/	/
RASHAD																
Orig	/	/	/	/	/	/	/	/S	/	p/	/P	/	/S	p/P	/	/
Rashad	/	/	/	/	s/	/	/	/S	/	p/	p/P	s/	s/	p/P	/	/
SOUTHERN ATLANTIC																
Baga Sitemu	/	/	/	p/S	/	/	/	/S	/	/	/S	/	/	/P	/	/
Kisi	/	/P	/	/P	/P	/P	/	/S	/	/	/	/	/S	/P	/P	/
Temne	/S	/P	/P	/P	s/	/	p/	/S	p/	s/	/S	p/	/S	/P	/P	/P
TALODI PROPER																
Masakin	/	/	/	/	/	/S	/	s/	/	/	/S	p/	/S	/	/	/
TEGEM																
Lafofa	/	/	/	/	/	/	/	/	/	/	/P	p/	/P	/	/	/
UBANGI																
Baka (in Cameroon)	/S	s/	/	/P	/	/	/S	/S	/P	/S	/S	/	s/	/P	/P	/
Barambu	/	p/P	/	/P	/	/	/S	/S	/	p/	s/S	p/	s/	/	/	/
Dongo	/	s/P	/	/	/	s/	/	/S	/	/S	s/	s/	s/S	p/P	/	/
Linda	/S	/	/	/P	/	/P	/S	/S	/	/	/S	/P	/S	/P	/P	/
Ma	/	p/P	/	/	/	/	/	/	/	/	/S	s/	s/P	p/P	/	/
Mayogo	/	/P	/	/P	/	/	/	/S	/	/S	/S	p/	/S	/P	/P	/
Mba	/	p/P	/	s/	/	s/	/	s/	/	/S	/S	s/	p/P	p/P	/	/
Mondunga	/	/	/	/P	s/	/	/	/	/	/	/S	s/	/S	p/P	/P	/
Munzumbo	/	/	/	/P	/	/	/S	/S	/	/	/	/	/S	/	/	/
Ngbaka (Ma'Bo)	/S	/S	/	/	/	/	/	/	/	/S	/S	/	/	/P	/	/
Ngbaka (Minagende)	/S	/	/	/P	/	/	/	/S	/P	/	/S	/	/S	/P	/	/
Ngbandi	/	/	/	/P	/	/	/	/S	/	/	/S	/	/	/P	/	/
Nzakara	/	p/	/	/P	/	/	/	/S	/	/S	s/S	p/	s/S	p/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Sango	/S	/P	/P	/P	/	/P	/S	/S	/	/S	/S	p/	/S	p/P	/P	/P
Zande	/	p/	/	s/P	/	/	/	/	/	/	s/S	p/	s/P	p/P	/	/
UKAAN																
Ikaan	/	/	/	/	/	/	/	/S	/	/	/S	/	/S	p/P	/	/
WESTERN MANDE																
Bambara	/S	s/P	/	/S	p/	/P	/S	/	/	/P	/P	s/	/	/P	/	/P
Bobo Madaré (Northern)																
	/S	/	/	/S	/	/	/P	/	/	/	/	/	/	/	/	/
Bobo Madaré (Southern)																
	/	/	/	/	/	/	/S	/P	/	/	/	/	/P	/	/	/
Bolon	/	/	/	/	/	/	/	/S	/	/	/P	/	/P	/P	/	/
Dzùùngoo	/S	/	/	/S	/	/P	/S	/	/	/	/P	/	/P	/P	/P	/
Jalonke	/	/	/	/S	/	/	/S	/P	/	/	/P	/	/	/P	/	/
Jeli	/	/	/	/S	/	/	/	/	/	/	/	/	/P	/	/	/
Kpelle	/	/	/	/	/	/	/	/S	/	/	/	s/	/P	/	/	/
Mandinka (Gambian)	/	s/P	/	/S	/	/	/	/P	/	/P	/P	s/	/	/P	/P	/
Maninka (Western)	/S	/P	/	s/S	/	/	/	/S	/	/P	/P	/	/	/P	/S	/
Mauka	/S	/P	/	/S	p/	/P	/S	/S	/	/P	/P	/S	/P	/P	/P	/
Mende	/	s/	/	/S	/	/	s/	/S	/	/P	/P	s/	/	/P	/P	/P
Soninke	/S	/	/	/	/	/	/	/	/	/P	/P	/	/	/P	/	/
Susu	/	s/	/	/S	/	/	/	/P	/	/	/P	/	/	/P	/P	/
Vai	/	s/P	/	/S	/	/	/S	/S	/	/P	/P	/S	/P	/P	/P	/P
Xasonga	/P	/	/	/S	/	/	/	/P	/	/	/P	s/	/	/P	/P	/
Nilo-Saharan																
BERTA																
Berta	/S	s/	/	/	/	/	/	s/	/	p/	s/S	s/	s/	/P	/	/
BONGO-BAGIRMI																
Bagirmi	/S	/S	/P	/P	/P	/P	/S	/S	/	/S	s/S	/	s/S	/P	/P	/
Bagiro	/	p/	/	/	/	/	/S	/S	/	/S	/S	/S	s/	p/P	/	/
Baka (in Sudan)	/	/	/	/P	/	/P	/S	/S	/	/S	/S	/	/S	/P	/P	/
Binga	/	p/	/	/	/	/	/	s/	/	/S	/S	s/	s/	/	/	/
Bongo	/S	p/	/	/P	/	/P	/	/	/	/S	/S	/	s/S	p/P	/	/
Gula (in Central African Republic)																
	/S	/	/	/	/	/	/S	/S	/S	/S	/S	/S	/S	p/P	/	/
Jur Mödö	/	/S	/	/P	/	/P	/S	/S	/	/S	/S	/S	/S	/P	/P	/
Kabba	/	/	/	/	/	/	/	s/	/	/	/S	/	/S	/P	/	/
Kara (in Central African Republic)																
	/S	s/P	/	/P	/	/P	/	/S	/	/S	s/S	/	s/S	/P	/	/P
Kenga	/S	/P	/	s/	/	/	/S	/S	/	/S	s/S	/	s/	/P	/	/
Mango	/	/	/	/	/	/	/	/S	/	/	/	s/	/	p/P	/	/
Mbay	/S	/	/	s/	/	/P	/	/S	/	/S	s/S	/	s/S	/P	/	/
Ngambay	/S	/	/P	s/P	/P	/P	/	/S	/	/S	s/S	s/	s/S	p/P	/P	/P
Yulu	/S	/S	/	/P	/	/P	/	/S	/	/S	/S	/	s/	/P	/	/
DAJU																
Laggori	/	/	/	/	/	/	/	/S	/	/	s/	s/	s/	p/	/	/
Shatt	/	/P	/	/	/	/	/	/S	/	/S	s/S	s/	s/S	/P	/	/
Sila	/	/	/	/	/	/	/	/	/	/P	/P	s/	/	/P	/	/
EASTERN JEBEL																
Gaahmg	/S	s/	/	/	/	/	/S	/S	/S	/S	s/S	s/	p/	/P	/P	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
EASTERN SAHARAN																
Beria	s/	/	/	/S	/	/	/S	/S	/S	s/	p/P	/	s/S	s/P	/S	/
FUR																
Fur	/	s/P	/	s/S	/	/S	/	/P	/	/	/P	/	/	p/P	/	/P
GUMUZ																
Gumuz	s/	/S	/	/P	s/	/	s/	/S	/	/S	/S	/	s/	/P	/	/
KOMAN																
Gwama	/	/	/	/P	/	/	/	/S	/	/	/S	p/	/	/P	/	/
Komo	/	/P	/	/	/	/	/	/S	/	/P	s/S	p/	/	s/P	/	/
KRESH																
Aja	/	p/	/	/P	/	s/	/	/	/	/S	s/S	/	/S	p/P	/	/
Kresh	/S	/	/	/P	/	/	/S	/	/	/S	s/S	/P	s/	p/P	/	/
KULIAK																
Ik	/S	s/S	/	s/P	/	/	/	s/	/	/P	/S	s/	/	s/S	/	/P
So	/S	p/P	/	s/P	/	/	/	/S	/	/P	/S	s/	s/	s/S	/	/
KUNAMA																
Kunama	s/	s/S	/	/S	/	/	s/	/S	/	s/	p/P	/	s/	/P	/	/
LENDU																
Lendu	/	s/	/P	/S	/	/P	/	/P	/	/	/	s/	p/P	p/P	/S	/
Ngiti	/	s/P	/	/S	/	/P	/P	/P	/	/P	/	/	s/P	p/P	/S	/
MABAN																
Maba	s/	s/	/	/S	p/	/S	/S	/S	/	s/	p/P	s/	/S	p/P	/	/
Masalit	s/	s/	/	s/	/	s/	s/	/	s/	s/	/P	s/	s/	p/P	s/	/
Runga	/	s/	/	/S	/	/S	/S	/S	/S	s/	p/P	s/	p/S	p/P	/	/
MANGBETU																
Mangbetu	/	s/P	/	/	/	/	/	s/	/	/	/S	p/	s/	p/P	/P	/
MANGBUTU-EFE																
Lese	/	p/	/	/	/	/	/	/S	/	p/	/	s/	/	p/P	/	/
Mamvu	/	/S	/	/	/	/	/	/S	/	p/	/S	s/	p/P	p/P	/	/
MORU-MA'DI																
Avokaya	/	s/	/	/S	/	/	/	/S	/S	/	/	s/	/	/P	/	/
Logoti	/S	/	/	/S	/	/	s/	/S	/	/S	/	/	/S	/P	/	/
Lugbara	/S	p/P	/P	/S	/	/	/	/S	/	/S	/	s/	/P	/P	/	/
Ma'di	/S	/	/	/S	/P	/	/S	/S	/	/S	/	/S	/	p/P	/	/P
Moru	/S	s/	/	/S	/	/	/S	/S	/	/S	/	s/	s/S	p/P	/	/
NARA																
Nara (in Ethiopia)	/	s/S	/	/S	s/	/	/S	/S	/	/P	/P	s/	/P	s/P	/	/
NILOTIC																
Acholi	/	p/	/	/P	/P	/P	s/	/	/	/	s/S	/	s/	p/P	/P	/
Anywa	/	p/P	/	/P	/	/	/	s/	/	/P	/	s/	/	/	/	/
Bari	/	p/P	/	/P	/	/P	/	/P	/	/P	/S	s/	/S	/P	/P	/
Camus	/	/	/	/P	/	/	/	/P	/	/	/S	s/	/	p/S	/	/
Dinka	/	/P	/	/	/	/P	s/	s/	/	/P	/	/	s/	/P	/P	/
Jur Luwo	/	/	/	/P	/	/	/	/	/	/P	/	s/	s/	/	/	/
Karimojong	/S	s/P	/	/P	p/	/	/	/S	/	/	/S	p/	p/S	p/S	/P	/
Lango	/	/P	/	/P	/	/P	/	/S	s/	/P	s/S	s/	/S	p/P	/	/
Luo	/	p/P	/	/P	/	/P	/S	/S	/S	/P	s/S	s/	s/S	p/P	/	/P
Maasai	/P	/P	/P	/P	/	/P	/	/P	/	/	p/S	s/	/S	p/S	/P	/P
Nandi	/S	/	/	/P	/	/	s/	s/	/	p/	s/S	s/	s/	p/S	/P	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Nuer	/	/	/	/P	/	/P	/	/S	/	/P	/	/	s/	s/P	/P	/
Pokot	/	p/	/	s/P	/	/P	s/	s/	/	/P	s/S	s/	s/	/S	/P	/P
Päri	/	p/	/	s/P	/	/P	/	/	/	/P	p/P	s/	s/	s/S	/P	/
Sebei	/	/	/	/P	/	/	s/	s/	/	/	/S	s/	s/	/S	/	/
Shilluk	/	p/P	/	/P	/	/P	/	/	/	/P	/S	s/	s/S	/P	/P	/
Teso	/	/P	/P	p/P	p/	/	/	/S	/	/P	p/S	/	/	p/S	/P	/P
Toposa	/	s/	/	/P	/	/	/	/S	/	/	/S	/	/	/S	/	/
Turkana	/S	p/P	/	p/P	p/	/P	/	/S	/	p/	p/S	/	/S	p/S	/P	/P
NUBIAN																
Ghulfan	/	s/	/	s/	/	/	/	/	/	/	/P	s/	/	s/P	/	/
Midob	s/	s/	/	s/S	/	/	/	/P	/	s/	/P	s/	/P	s/P	/	/S
Nubian (Dongolese)	s/	s/P	/	/S	s/	/	/	/P	/S	s/	s/P	s/	p/P	s/P	s/	/
Nubian (Kunuz)	s/	s/	/	/S	s/	/S	/	/P	s/	s/	/P	s/	p/	s/P	s/	/
NYIMANG																
Nyimang	/	s/P	/	s/S	/	/S	/S	/S	/	/P	s/P	/	/P	s/P	/P	/
SONGHAY																
Koyra Chiini	/	/P	/	/S	/	/	/S	/S	/	/P	/S	/S	/P	/P	/P	/
Koyraboro Senni	/P	/P	/	/S	/	/	/S	/S	/	/P	/P	/	/P	/P	/P	/
Zarma	/	/P	/	/S	/	/	/S	/S	/S	/P	/	/	/P	/P	/	/
SURMIC																
Baale	/	s/P	/	s/	/	s/	/	/S	/	/	s/	s/	/	/	/	/
Chai	/	/	/	s/S	/	/	/	/	/	/	/	s/	s/	/P	/	/
Didinga	/	/P	/	s/P	/	/P	/	/S	/S	/P	s/S	s/	s/S	s/S	/P	/P
Koegu	s/	s/	/	/P	/	/P	/	/S	/	/	/S	/	/S	p/P	/P	/
Majang	/S	/S	/	s/S	/	/P	/	/S	/	/P	/S	s/	/S	s/S	/	/
Me'en	/S	s/	/	s/	/	/P	/	s/	/	s/	s/S	s/	s/	/P	/P	/
Murle	/	/	/	s/	/P	/P	/	/S	/	/P	s/S	s/	s/S	/S	/	/P
Mursi	/	/	/	s/	/	/P	p/	s/	/	/P	s/S	s/	s/	/P	/	/
Tennet	/	p/S	/	s/	/P	/P	/	/S	/	/P	s/S	s/	/S	s/S	/P	/
Tirmaga	/	/S	/	s/S	/	/	/	/	/	/	s/S	s/	s/	/P	/	/
TAMAN																
Miisiirii	/	/	/	s/	/	/	/	/	/	s/	/P	/	/	/P	/	/
Sungor	/	s/	/	/	/	/	/	/	/	s/	/	s/	/S	/	/	/
Tama	/	s/	/	s/	/	/	s/	/	/	s/	/P	s/	/P	p/P	/	/
TEMEIN																
Temein	/	p/P	/	s/	/	/P	/	/S	/	/P	/S	/	/S	p/P	/	/
WESTERN SAHARAN																
Kanembu	/	s/	/	s/	/	/	/	/	/	/	s/P	s/	s/	s/P	/	/
Kanuri	s/	s/	/	/S	p/	/	/S	/S	/S	s/	/P	s/	s/	s/P	/	/
Tubu	/S	s/	/	/S	/	/	/S	/S	/	s/	/P	s/	s/	/P	/	/
Shabo																
SHABO																
Shabo	/	/P	/	s/S	/	/	/	/	/	/S	/P	/S	/	/P	/	/
Eurasia																
Ainu																
AINU																
Ainu	/S	s/S	/S	/S	s/	/S	/	/P	/P	/P	p/P	s/	p/	p/P	/S	/S

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Altaic																
MONGOLIC																
Bao'an	s/	s/S	/	/S	s/	/S	/	/P	/S	/P	/P	/	/	/P	s/	/
Buriat	/	s/	/	s/S	s/	/S	/	/P	/	s/	/P	s/	s/	s/P	/	/
Dagur	/S	s/S	/	s/S	s/	/S	/	/P	/	/P	/P	/	s/	/P	/S	/S
Kalmyk	s/	s/S	/	s/S	s/	/S	/	/	/	/	/P	s/	s/P	s/P	/	/
Khalkha	/S	s/S	/	s/S	s/	/S	/	/P	/	s/	/P	/S	s/	/P	/S	/
Mangghuer	s/	s/S	/	/S	s/	/S	/	/P	/S	/P	/P	/	/	/P	/S	/
Moghol	/	s/	/	s/	/	/	/	/	/	/P	/P	s/	/	s/P	/P	/
Mongol (Khamnigan)	/S	s/	/	s/	/	/	/	/	/	/	/P	s/	s/	s/P	/	/
Oirat	/	/	/	s/	/	/	/	/	/	/P	/P	s/	s/	/P	/	/
Ordos	/	s/	/	s/	/	/	/	/	/	/	/P	s/	s/	/P	/	/
Santa	s/	s/	/	s/S	/	/	/	/	/	/P	/P	s/	s/	/P	/	/
Shira Yughur	/	s/	/	s/S	/	/	/	/	/	/S	/	s/	/	/	/	/
TUNGUSIC																
Even	/	s/	/	s/S	/	/	/	/	/	/	/P	s/	s/	s/P	/	/
Evenki	/S	s/S	/	s/S	s/	/S	/	/P	/	/	/P	s/	s/P	s/P	s/	/
Manchu	s/	s/	/	s/S	s/	/S	/	/	/	s/	/P	s/	/	/P	/S	/
Orok	/	s/	/	s/S	s/	/	/	/	/	p/	/P	s/	s/	s/P	/	s/
Udihe	/	/	/	/S	/	/	/	/	/	/P	/P	/	/	/P	/	/
TURKIC																
Azari (Iranian)	/	s/	/	s/	s/	/	/	/	/P	s/	/P	s/	s/	s/P	/	/
Azerbaijani	/S	s/	/	s/	s/	s/	/	/	/	s/	/P	s/	s/	s/P	/	/
Bashkir	/	s/	/	s/S	/	/	/	/P	/	s/	/P	s/	s/P	s/P	/P	/
Chuvash	s/	s/	/	s/S	s/	/	/	/P	/	s/	/P	s/	s/P	s/P	/S	/
Karachay-Balkar	s/	s/	/	s/S	/	/	/	/P	/P	/	/P	s/	s/	s/P	/	/
Karakalpak	s/	s/S	/	s/S	/	/	/	/P	/P	s/	/P	s/	s/	s/P	/	/
Khalaj	/	s/	/	s/S	s/	s/	/	/P	/P	s/	/P	s/	s/	s/P	/P	/
Kirghiz	/S	s/S	/	s/S	/	/	/	/	/P	/	/	s/	s/	s/	/	/
Noghay	/	s/	/	s/	/	/	/	/	/	/	/P	s/	s/	s/P	/	/
Tatar	/	s/S	/	s/S	/	/	/	/P	/	/	/P	s/	s/P	s/P	/P	/
Turkish	/S	s/S	s/	s/S	s/	/	/	/P	/P	s/	/P	s/	s/P	s/P	/	/
Turkmen	/	/	/	/S	/	/	/	/P	/	/	/P	/	/	s/P	/	/
Tuvan	/S	s/S	/	s/S	s/	/	/	/P	/	s/	/P	s/	s/	s/P	/	/
Uyghur	/	s/	/	s/	/	/	/	/	/	s/	/P	s/	s/	s/P	/	/
Uzbek	s/	s/S	/S	s/	s/	/	/	/P	/	s/	/P	s/	s/	s/P	/P	/
Yakut	/	s/	/	s/S	s/	/	/	/P	/	s/	/P	s/	s/	s/P	/	/
Austro-Asiatic																
MUNDA																
Bhumij	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Gorum	/	s/S	/	/S	p/	/	/	/P	/	p/	s/P	s/	s/	p/P	/S	/
Gta'	/S	/	/	/	/	/	/	/P	/	p/	/P	/	/	/P	/	/
Gutob	s/	s/S	/	/	p/	/	/	/P	/	p/	/P	/	/	/P	/	/
Ho	/S	s/	/	s/S	/	/	/	/P	/	/P	s/P	s/	/	/P	/S	/
Juang	s/	s/	/	s/S	/	/	s/	/	/	/	s/P	s/	s/	/P	/	/
Kharia	/S	/S	/	/S	p/	/	/	/P	/	/P	/P	/	/	/P	/	/
Korku	/	s/S	/	s/S	/	/S	/	/P	/	/	s/P	s/	/P	/P	/S	/S
Mundari	/S	s/	s/	/S	/S	/	/	/P	/	/P	s/P	s/	s/	/P	/	/
Remo	s/	s/	s/	s/S	p/	/S	/	/P	/S	p/	/P	s/	/	s/P	s/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Santali	s/	s/S	/	s/S	p/	/	/	/P	/P	/P	s/P	s/	s/P	s/P	/	/
Sora	/	s/	/	s/S	/	/	/	/P	/P	p/	s/P	s/	s/	s/P	/	/
Basque																
BASQUE																
Basque	/	/S	/	/S	p/	/S	/S	/S	/S	/P	p/P	/	/P	s/P	/S	/
Burushaski																
BURUSHASKI																
Burushaski	s/	s/S	/	s/S	s/	/S	/P	/P	/	p/	p/P	s/	p/P	s/P	/	/
Chukotko-Kamchatkan																
NORTHERN CHUKOTKO-KAMCHATKAN																
Chukchi	/	/S	/	s/S	/	/S	/	/P	/	/P	s/P	s/	/P	p/P	/P	/
Koryak	/	/	/	s/S	/	/	/	/P	/	/	s/P	s/	/	p/P	/P	/
SOUTHERN CHUKOTKO-KAMCHATKAN																
Itelmen	/	s/	/	s/	/	/S	/	/P	/	/	s/P	s/	/P	/P	/P	s/
Dravidian																
CENTRAL DRAVIDIAN																
Kolami	s/	s/	/	s/S	s/	/	/	/P	/	/	/P	s/	/	s/P	s/	/
NORTHERN DRAVIDIAN																
Brahui	/	s/S	/	s/S	s/	/S	/	/P	/	s/	s/	s/	s/	s/	/	/
Kurukh	/	s/	/	/S	/	/	/	/P	/	/	/P	s/	/P	s/P	/	/
Malto	/	/	/	/	/	/	/	/P	/	/	/P	/	/	/P	/	/
SOUTH-CENTRAL DRAVIDIAN																
Gondi	/S	s/S	/S	s/S	s/	/	/	/P	/	s/	/P	s/	/P	s/P	/	/
Koya	/	s/	/	s/	s/	/	/	/P	/	s/	/P	s/	/	s/P	/	/
Kuvi	/S	s/S	/	s/S	s/	/S	/	/P	/	s/	s/P	s/	/	s/P	/S	/
Telugu	s/	s/S	/	s/S	s/	/S	/	/P	/	s/	/P	s/	/	s/P	/	s/
SOUTHERN DRAVIDIAN																
Betta Kurumba	/	s/	/S	s/S	/	/	/	/P	/	/	/P	s/	/	s/P	/S	/
Kannada	/S	s/	s/	s/S	s/	/S	/	/P	/P	s/	/P	s/	/P	s/P	/	/S
Malayalam	/S	s/	/	s/S	s/	/S	/	/P	/P	s/	/P	s/	/P	/P	/	/
Tamil	s/	s/S	/	s/S	/S	/S	/	/P	/	s/	/P	s/	/	s/P	s/	/
Tulu	/	/	/	s/S	s/	/	/	/P	/	s/	/P	s/	/	s/P	/	/
Indo-European																
ALBANIAN																
Albanian	/P	s/P	/P	s/P	/	/P	s/	/P	/P	/P	/S	s/	/S	s/P	/P	/P
ARMENIAN																
Armenian (Eastern)	/	s/	/	s/S	/	/S	s/	/P	/P	/	/	s/	s/P	s/P	/P	/
Armenian (Western)	/	s/	/	s/S	s/	/S	s/	/P	/S	/	/P	s/	s/P	s/P	/P	/
BALTIC																
Latvian	/P	/P	/	s/P	/	/P	/P	/P	/P	p/	/S	s/	/P	s/P	/P	/P
Lithuanian	/P	s/P	/	s/P	/	/P	/	/P	/	p/	/S	s/	/	s/P	/P	/P
CELTIC																
Breton	/P	s/	/	/P	/	/	/P	/S	/P	/	p/S	s/	s/P	/S	/P	/
Cornish	/P	s/P	/	/P	/	/	/P	/S	/	/P	/S	s/	/P	s/P	/P	/
Gaelic (Scots)	p/	/P	/P	/P	/	/P	/P	/S	/	/P	/S	s/	/P	s/S	/P	/
Irish	/P	/P	/	p/P	/	/P	/P	/S	/	/P	/S	s/	/P	s/S	/P	/P
Welsh	/P	s/P	/P	/P	/	/P	/P	/S	/	/P	/S	s/	/P	s/S	/P	/
Welsh (Colloquial)	/	/	/	/	/	/	/	/	/	/S	/S	/	/	/S	/	/
GERMANIC																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Danish	/	s/P	/P	s/P	/	/P	s/	/P	/P	/	/S	s/	/P	/P	/P	/P
Dutch	/	s/P	/	/P	/	/P	/P	/P	/P	/	/	s/	/	s/P	/P	/P
English	/	s/P	/	/P	/P	/P	/P	/P	/P	/P	/S	s/	/P	s/P	/P	/P
Frisian	/	/	/	/P	/	/P	/P	/P	/P	/	/	/	/	s/P	/	/
German	/	s/P	/P	s/P	/	/	/P	/P	/P	/	/	s/	/P	s/P	/P	/P
Icelandic	/	s/P	/	s/P	/	/P	s/	/P	/	/	/S	s/	/S	s/P	/	/
Luxemburgeois	/	/	/	/P	/	/	/P	/	/P	/	/	s/	/	/P	/	/
Norwegian	/	s/P	/P	s/P	/	/P	s/	/P	/P	/	/S	s/	/	/P	/P	/P
Swedish	/	s/P	/P	/P	/P	/P	s/	/P	/P	/	/S	s/	/P	/P	/P	/P
GREEK																
Greek (Modern)	/P	s/P	/	s/P	/	/	/P	/P	/P	/P	/S	s/	/	s/	/P	/P
INDIC																
Assamese	s/	s/	/	s/	/	/	s/	/P	/	p/	/P	s/	s/	s/P	/	/
Bengali	/	s/	/	s/	/	/	/	/P	/	/S	/P	/	/P	s/P	/	/
Bhojpuri	/P	s/S	/	s/S	s/	/	/	/P	/	/P	/P	s/	/P	s/P	/P	/
Brokskat	/	/	/	/	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Darai	/	/	/	/	/	/	/	/P	/	/	/P	/	/	/P	/	/
Dhivehi	/S	/	/	/	/	/	/	/P	/	p/	/	/	/	s/	/	/
Domari	/	s/P	/	s/P	s/	s/	/	/P	/	/	s/S	s/	s/	s/	/P	/
Dumaki	/S	s/S	/	s/S	/	/S	/	/	/	/P	/P	s/	/	s/P	/	/
Gujarati	/	s/S	/S	/S	s/	/S	/	/P	/	/P	/P	/	/P	/P	/P	/
Hindi	/P	s/S	/S	s/S	s/	/S	/	/P	/	/P	s/P	s/	/P	/P	/P	/S
Kalami	/S	s/	/	s/S	s/	/S	/	/P	/P	/P	/P	s/	/	s/P	/	/
Kashmiri	/	s/P	/	s/S	s/	/P	/	/P	/	s/	s/	s/	/P	s/P	/P	/
Khowar	/	s/	/	s/S	/	/	/	/	/	/	/P	s/	/	s/P	/	/
Kumauni	/P	s/S	/S	s/S	s/	/	/	/P	/	/P	/P	s/	/	/P	/	/S
Lamani	/S	s/S	/	s/S	s/	/S	/	/P	/	/S	/P	s/	/P	s/P	/S	/
Magahi	/	/	/	s/S	/	/	/	/	/	/	s/P	s/	/	s/P	/	/
Maithili	/	s/S	/	s/S	s/	/S	/	/P	/	/P	s/P	/S	/	s/P	/P	/
Marathi	/S	s/S	/	s/S	/S	/S	/	/P	/	/S	/P	/	/P	s/P	s/	/
Nepali	/	s/S	/	s/S	/S	/S	/	/P	/	s/	/P	s/	/	s/P	/S	/S
Oriya	/S	s/	/	s/	/	/	s/	/P	/	s/	/P	s/	/	s/P	/	/
Palula	/S	s/S	/	s/S	/	/	/P	/P	/	/P	s/	s/	/P	s/	/	/
Panjabi	/P	s/S	/S	s/S	s/	/S	/	/P	/	/P	s/P	s/	/P	s/P	/P	/
Prasuni	/	/	/	p/S	/	/	/	/P	/P	/	/P	/	/P	s/P	/	/
Romani (Welsh)	/P	s/P	/	s/P	s/	/	/P	/P	/	/	/S	s/	/P	s/	/P	/
Savi	/	s/	/	s/S	/	/	/	/P	/	/	/P	s/	/P	s/P	/	/
Shekhawati	/	/	/	/S	/	/	/	/	/	/P	/P	/	/	/P	/	/
Shina	s/	s/	/	s/S	/	/	/	/P	/	/P	/P	s/	/P	s/P	/	/
Sindhi	/	s/	/	s/S	/	/	/	/	/	/	/P	s/	/	s/P	/	/
Sinhala	/S	s/S	/	s/S	s/	/	/	/P	s/	/S	/P	s/	/	s/P	/	/
Torwali	/	s/	/	/S	/	/	/	/P	/	/	/P	s/	/	/P	/S	/
Urdu	/P	s/S	/	/S	s/	/S	/	/P	/	/P	/P	s/	/P	s/P	/P	/S
IRANIAN																
Baluchi	/	/	/	s/	s/	/	/	/	s/	p/	/P	s/	s/	s/P	/P	/
Gilaki	/	/	/	s/	/	/	/P	/	/P	/	/	s/	/	/	/	/
Hawrami	/	/	/	/P	/	/	/	/P	/	/	/	/	/	s/	/	/
Kabatei	/	/	/	/S	/	/	/	/P	/	/	/P	/	/	/	/	/
Koluri	/	/	/	/S	/	/	/	/P	/	/	/P	/	/	/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Kurdish (Central)	/	/	/	/P	s/	s/	s/	/	s/	p/	p/P	s/	s/S	s/P	/P	/P
Mazanderani	/	/	/	/S	/	/	/	/	/P	/	/	s/	/	/	/	/
Ormuri	/	s/	/	/	/S	/S	/P	/	/	/	/P	s/	s/	s/P	/	/
Ossetic	/	s/S	/	s/S	/S	/	/	/P	/	/P	/P	s/	/P	s/P	/	/P
Parachi	/	/	/	/	/	/	/	/	/	/P	/P	/	/	s/P	/	/
Pashto	/	s/	/S	s/S	s/	/S	/	/P	/	p/	/P	s/	/	s/P	/P	/P
Persian	/P	/S	/P	/P	s/	/	/	/P	/S	p/	s/P	s/	/S	s/P	/P	/P
Shughni	/	/	/	/	/	/	/P	/P	/	/	/P	/	/	/P	/	/
Tajik	/P	/S	/	/P	s/	/	/	/P	/S	p/	/P	s/	/S	s/P	/P	/P
Talysh (Southern)	/	/	/	/S	/	/	/	/P	/	/	/P	/	/	/	/	/
Tati (Southern)	/	/	/	/S	/	/	/	/	/P	p/	/	s/	/	/	/	/
Vafsi	/	/	/	/	/	/	/	/P	/	/	/P	/	/	/	/	/
Wakhi	s/	s/P	/S	s/	s/	/S	/P	/P	/	/P	/P	s/	/P	s/P	/	/
Zazaki	/	/P	/	s/	/	/S	/	/P	/P	p/	/P	s/	/	s/P	/P	/
ROMANCE																
Catalan	/	/	/	/P	/	/	/P	/P	/P	/P	/S	/	/	/P	/	/
French	/P	s/P	/	/P	/P	/P	/P	/P	/P	/S	p/S	s/	/P	s/P	/P	/P
Italian	/	s/P	/	/P	/	/	/P	/P	/P	/P	/S	s/	/P	s/P	/P	/
Portuguese	/P	s/P	/	/P	/	/	/P	/P	/P	/P	/S	s/	/P	s/P	/	/
Romanian	/	s/P	/	s/P	/P	/P	s/	/	/P	/P	/S	s/	/S	s/P	/P	/
Romansch (Sursilvan)	/	/	/	/	/	/	/	/P	/	/S	/S	/	/P	/P	/P	/
Sardinian	/P	s/P	/	/	/	/	/P	/	/	/P	/S	s/	/S	s/P	/P	/
Spanish	/	s/P	/	/P	/P	/P	/P	/P	/P	/P	p/S	s/	/P	s/P	/P	/P
SLAVIC																
Belorussian	/P	/	/	s/	/	/	/	/P	/	/P	/	/	/	s/	/	/
Bulgarian	/	s/P	/	/P	/	/	/	/P	/P	/P	/S	s/	/	s/P	/P	/
Czech	/	/S	/	s/P	/	/P	/	/P	/	p/	/S	s/	/P	s/P	/P	/
Macedonian	/	s/	/	/P	/	/	/S	/P	/	/P	/S	/	/	s/P	/	/
Polish	/P	s/P	/P	s/P	/P	/P	/	/P	/	/P	/S	s/	/P	s/P	/P	/P
Russian	/	/P	/	s/P	/P	/P	/	/P	/	/P	/S	s/	/P	s/P	/P	/P
Serbian-Croatian	/P	/S	/	s/P	/	/P	/	/P	/	/P	/S	s/	/P	s/P	/P	/
Slovene	/	/	/	s/P	/	/	/	/P	/	/P	/S	s/	/P	s/P	/	/
Sorbian	/	/P	/	s/P	/	/	/	/	/	p/	/P	s/	/	s/P	/P	/
Ukrainian	/P	s/P	/P	s/P	/	/	/	/P	/	/P	/S	s/	/	s/P	/P	/P
Japonic																
JAPANESE																
Japanese	/S	s/S	/	/S	s/	/S	/	/P	/P	s/	/P	s/	/P	/P	/S	/
Shuri	s/	/	/	/S	/	/	/	/	/	/	/P	/	/	/P	/	/
Kartvelian																
KARTVELIAN																
Georgian	/	s/	/	s/S	/	/	/	/P	/	/P	p/P	s/	/	p/P	/P	/
Korean																
KOREAN																
Korean	s/	s/S	/	s/S	s/	/	/	/P	/	/P	/P	s/	/	/P	/S	/S
Nahali																
NAHALI																
Nahali	/	s/	/	s/	/	/	/	/	/	/P	/	/	/	/P	/	/
Nakh-Daghestanian																
AVAR-ANDIC-TSEZIC																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Andalal	/	/	/	/	/	/	/	/P	/	/	/	/	/	/	/	/
Avar	/	s/S	/	s/S	/	/S	/	/P	/	/S	p/P	s/	/	/P	/	/
Godoberi	/	/	/	s/	/	/	/	/P	/	s/	/P	s/	/	/P	/	/
Hunzib	s/	s/S	/S	s/S	s/	/S	/	/P	/	s/	p/P	s/	/P	/P	s/	/
Khvarshi	s/	s/S	/S	s/S	s/	/S	s/	/P	/	s/	p/P	s/	/P	/P	s/	/
Tsez	s/	/	/S	s/S	/	/S	/	/P	/	/S	/P	s/	/	p/P	/	/
LAK-DARGWA																
Dargwa	/	/	/	s/S	s/	/	/	/	/	/	/P	s/	/P	/P	/	/
Lak	/	/	/	s/S	/	/S	/	/	/	/	/P	/	/P	/P	/	/
LEZGIC																
Archi	s/	s/S	/	s/S	/S	/	/P	/P	/P	s/	p/P	s/	/	/P	/	/S
Khinalug	s/	s/	/	s/	/	s/	/	/	/	/	p/	s/	/	/	/	/
Kryz	/	/	/	/	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Lezgian	s/	s/S	/S	s/S	s/	/	/	/P	/P	s/	/P	s/	/	/P	s/	/S
Rutul	/	s/	/	s/S	/	/	/	/P	/	/	/P	s/	/P	p/P	/	/
Tsakhur	/	/	/	s/S	/	/	/	/P	/	/	/P	s/	/	/P	/	/
Udi	/	s/	/	s/	s/	/	/	/P	/	p/	/P	s/	/P	s/P	/P	/
NAKH																
Chechen	s/	s/S	/	s/S	/	/S	/	/P	/	/P	p/P	s/	/P	/P	/	/
Ingush	s/	s/S	/	s/S	s/	/S	/	/P	/	s/	p/	s/	/P	/	s/	/
Tsova-Tush	/	s/	s/	s/S	s/	/S	/	/P	/	/P	p/P	s/	/P	s/P	/	/
Nivkh																
NIVKH																
Nivkh	/S	s/	/	s/S	s/	/S	/	/P	/	s/	p/P	s/	p/	/P	/	s/
Northwest Caucasian																
NORTHWEST CAUCASIAN																
Abaza	s/	s/	/	/S	p/	/	p/	/P	s/	/	p/P	s/	p/	p/P	/	/
Abkhaz	s/	s/	/	/S	p/	/	/P	/P	/S	/	p/P	s/	p/P	p/P	s/	/
Adyghe (Abzakh)	/	/	/	s/S	/	/	s/	p/	/	/	/P	/	p/	/P	/	/
Kabardian	/	/	/	s/S	p/	/	s/	/P	/	s/	p/P	s/	/	p/P	/	/
Ubykh	s/	s/	s/	/S	/	/	p/	p/	p/	/	p/P	s/	p/	p/P	s/	/
Uralic																
FINNIC																
Estonian	/P	s/P	/P	s/S	/	/P	/	/P	/	/P	/S	s/	/P	s/P	/P	/
Finnish	/	s/P	/	s/S	s/	/P	/	/P	/	/P	/S	s/	s/P	s/P	/P	/
Liv	/	/	/	/	/	/	/	/	/	/P	/S	/	/	/P	/P	/
Votic	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
MARI																
Mari (Meadow)	/	s/S	/	s/S	/	/	/	/	/	/P	/P	s/	s/	s/	/	s/
MORDVIN																
Mordvin (Erzya)	/	s/	/	s/S	/	/	s/	/	/	/P	s/S	s/	s/	s/P	/	/
PERMIC																
Komi-Permyak	/	s/	/	s/S	s/	/	s/	/	/	/P	/P	s/	s/	s/	/	/
Komi-Zyrian	/	/	/	s/S	/	/	/	/	/	/P	/S	/	/	/P	/	/
Udmurt	/	s/	/	s/S	s/	/	/	/P	/	/S	/P	s/	s/	s/P	/S	/
SAAMI																
Saami (Northern)	/	s/P	/	s/S	/	/	/	/	/	/P	/S	s/	s/	s/P	/	/
SAMOYEDIC																
Enets	s/	s/	/	s/S	/	/	/	/	/	/P	/P	s/	s/	s/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Kamass	/	s/	/	s/S	/	/	/P	/P	/	/P	/P	s/	s/	s/P	/	/
Nenets	/	s/	/	s/S	/	/	/	/P	/	/P	/P	s/	s/	s/P	/	/
Nganasan	s/	s/	/	s/	/	/	/	/	/	/P	/P	/	s/	s/P	/	/
Selkup	/	s/	/	s/	/	/	/	/P	/	/	/P	s/	s/	s/P	/	/
UGRIC																
Hungarian	/S	s/	/	s/S	/	/S	/P	/P	/P	/P	/S	s/	s/	s/P	/P	/P
Khanty	/	s/	/	s/S	/	/	/	/P	/	/P	/P	s/	s/P	s/P	/P	/
Mansi	/	s/	/	s/S	/	/S	/	/P	/	/P	s/P	s/	s/P	s/P	/	/
Yeniseian																
YENISEIAN																
Ket	/	p/	/	s/S	/	/	/	/P	/	/P	p/P	s/	p/P	p/P	/	/
Yukaghir																
YUKAGHIR																
Yukaghir (Kolyma)	/	s/S	/	s/S	s/	/S	/	/P	/	p/	/P	s/	s/P	s/P	/S	/
Yukaghir (Tundra)	/	s/	/	s/S	s/	s/	/	p/	/	p/	/P	s/	/	s/P	/	s/
SE Asia & Oceania																
Austro-Asiatic																
ASLIAN																
Jahai	/P	p/	/P	/P	p/	/	/	/S	/	/P	/S	/	/S	/P	/P	/
Mah Meri	/	/	/	/P	/	/	/	/S	/	/S	/S	/	/	/	/	/
Semelai	/	/	/	/P	/	/	/	/S	/	/P	/S	/	/S	p/S	/	/
Temiar	/	p/P	/	p/P	p/	/	/	/S	/	/P	/S	/	/S	p/P	/	/
BAHNARIC																
Brao	/S	/P	/	/P	/	/	/	/S	/	/P	/S	/	/	/P	/	/P
Chrau	/S	/P	/S	/P	p/	/P	/	/S	/	/P	/S	/	/S	/P	/P	/P
Cua	/	/	/	p/	/	/	/	/	/	/	/S	/	/	/P	/	/
Loven	/S	/P	/	/P	/	/	/	/S	/	/P	/S	/	/S	/P	/P	/
Nyaheun	/S	/	/	/P	/	/	/	/S	/	/P	/S	/	/	/P	/	/P
Sapuan	/S	/	/	/	/	/	/	/	/	/P	/S	/	/	/P	/	/
Sedang	/	/	/	/P	p/	/	/	/S	/	/P	/S	/P	/S	/P	/	/
Sre	/S	/P	/	/P	p/	/P	/	/S	/	/P	/S	p/	/	/P	/	/
Stieng	/S	/P	/	/P	/	/	/	/S	/	/P	/S	/P	/S	/P	/P	/P
KATUIC																
Katu	/	/	/	/P	/	/	/	/S	/	/	/	/	/	/	/	/
Pacoh	/S	s/	/P	/P	p/	/P	/	/S	/	/P	/S	/	/S	/P	/	/
KHASIAN																
Khasi	/	/P	/P	/P	p/	/P	/	/P	/	/P	/S	/P	/S	/P	/P	/P
KHMER																
Khmer	/S	/P	/P	/P	/	/P	/S	/S	/S	/P	/S	/S	/S	/P	/P	/P
MONIC																
Mon	/S	/S	/S	/P	/	/	/S	/S	/	/P	/S	/S	/S	/P	/P	/
NICOBARESE																
Nancowry	/P	/	/	/P	/	/	/	/P	/	/	/S	/	/S	/S	/	/
Nicobarese (Car)	/	s/	/	/P	p/	/	/P	/P	/P	/P	s/S	/	/S	/S	/P	/P
PALAUNG-KHMUIC																
Khmu’	/S	/P	/	/P	/P	/P	/	/S	/	/P	/S	/	/S	/P	/	/P
Mlabri (Minor)	/	/	/	/P	/	/	/P	/S	/	/	/S	/	/	/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Palaung	/	/P	/	/P	/	/	/	/S	/	/P	/S	/	/	/P	/P	/
PEARIC																
Kasong	/	/	/	/	/	/	/	/S	/	/P	/S	/	/	/P	/P	/
VIET-MUONG																
Vietnamese	/S	/P	/	/P	/P	/P	/P	/S	/	/P	/S	/P	/	/P	/P	/P
Austronesian																
ATAYALIC																
Atayal	/S	p/	/	/P	p/	/	/P	/S	/	/P	/S	/	/S	/S	/P	/
Seediq	/	p/P	/	/P	p/	/	/	/S	/	/P	/S	/	/S	/S	/	/
BARITO																
Ma'anyan	/	/P	/	/	/	/	/S	/S	/P	/	/S	/	/S	/P	/	/
Malagasy	/P	p/	/	/P	/	/	/P	/	/	/P	/S	/	/S	/S	/P	/
BILIC																
Tboli	/	p/	/	/P	p/	/	/	/S	/	/P	/S	/P	s/	/S	/P	/
CELEBIC																
Muna	/	p/P	/P	/P	p/	/	/	/S	/	/P	s/S	s/	s/S	p/P	/	/P
Pendau	/P	p/	/P	/P	/	/	/	/S	/	/P	/	/	/S	/	/P	/P
Tukang Besi	/S	p/P	/	/P	p/	/	/S	/S	/	/P	s/S	/	/S	p/S	/P	/P
Wolio	/	/	/	/P	p/	/	/	/P	/	/	s/S	/P	s/	p/S	/P	/
CENTRAL LUZON																
Kapampangan	/	p/	/	/P	p/	/	/	/	/	/	/S	/	/S	/S	/	/P
CENTRAL MALAYO-POLYNESIAN																
Alune	/	/	/	/P	/	/	/	/S	/	/S	s/S	s/	/	p/P	/P	/
Buru	/	/	/	/P	/	/	/	/S	/	/	/S	/	/	/P	/	/
Kambera	/	/	/	/P	/	/	/P	/	/	/	/S	/	s/	/S	/P	/
Keo	/	/P	/	/P	/P	/	/P	/S	/	/P	/S	/	/	/P	/P	/P
Lamaholot	/	/	/	/	/P	/	/S	/S	/S	/S	/S	/	s/S	p/P	/P	/P
Leti	/	/	/	/P	/	/	/S	/S	/	p/	/S	/	/S	p/P	/	/
Manggarai	/	/	/	/P	/	/	/	/S	/	/	/S	/	s/S	/P	/P	/
Ngad'a	/	/	/	/P	/	/	/	/	/	/P	/S	/	/	/P	/P	/
Nuaulu	/	/	/	/P	/	/P	/	/S	/S	/S	/S	s/	/	p/P	/P	/
Paulohi	/	/	/	/P	/	/	/	/S	/	/P	s/S	s/	/	p/P	/P	/
Sawu	/	/	/	/	/	/	/	/	/	/S	/	/	/	/	/	/
Tetun	/S	/P	/P	/P	/	/	/S	/S	/S	/P	/S	/S	/P	p/P	/P	/
Tetun Dili	/	/	/	/P	/	/	/S	/S	/S	/	/S	/S	/P	/P	/	/
Tugun	/	/	/	/	/	/	/	/S	/	/	/S	/S	p/	p/P	/	/
CHAMORRO																
Chamorro	/P	/P	/	/P	p/	/	/P	/P	/	/P	/S	/S	s/	/S	/P	/P
ENGGANO																
Enggano	/	/P	/	p/P	/	/	/	/S	/	/P	/S	p/	s/	p/P	/P	/
GREATER CENTRAL PHILIPPINE																
Agta (Central)	/P	/	/	/P	p/	/	s/	s/	/	/P	/S	/S	s/	s/S	/	/P
Agta (Dupaningan)	/	/	/	/	/	/	/	/	/	/P	/S	/S	/	/S	/	/
Bikol	/	/	/	/P	p/	/	/	/	/	/	/S	/P	/	/S	/P	/
Buol	/	/	/	/	/	/	/	/	/	/	/	/	/	/S	/	/
Cebuano	/	p/	/	/P	/	/	/	/	/P	/	/S	/	/	/S	/	/
Hiligaynon	/	/	/	/P	/P	/	/	/P	/	/P	/S	/	/	/S	/	/P
Mamanwa	/	/P	/	/P	p/	/	/	/P	/	/P	/S	/P	/	/S	/P	/
Manobo (Western Bukidnon)																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
	/S	p/	/	/P	p/	/	/P	/P	/P	/P	/S	/P	/	p/S	/P	/
Tagalog	/	p/	/	/	p/	/	/	/	/	/P	/S	/P	/	/S	/P	/
Tausug	/	/	/	/P	/	/	/	/	/	/P	/S	/P	/	/S	/	/
LAMPUNG																
Lampung	/P	/	/	/P	s/	/	/S	/S	/	p/	/S	/	/	/P	/P	/P
MALAYO-SUMBAWAN																
Acehnese	/P	/P	/P	/P	p/	/	/	/S	/	/P	/	/	/S	p/	/P	/P
Cham (Eastern)	/S	/	/	/P	p/	/	/	/S	/	/	/S	/	/S	/P	/	/
Cham (Western)	/S	/P	/	/	/	/	/	/	/	/S	/S	/	/	/P	/P	/
Iban	/P	/P	/P	/P	/	/	/	/S	/	/P	/S	/	/S	/P	/P	/
Indonesian	/P	s/P	/P	/P	/	/P	/S	/S	/	/P	s/S	/	/S	/P	/P	/P
Manadonese	/	/P	/	/P	/	/	/	/P	/	/	/S	/	/	/P	/P	/
Minangkabau	/	/P	/	/P	/	/	/	/S	/	/P	/	/	/	/	/P	/
Mualang	/	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/P	/
Salako	/	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/	/
Sundanese	/P	/P	/	/P	/	/P	/S	/S	/	/P	/S	/	/S	/P	/P	/P
Urak Lawoi’	/S	/	/	/P	/	/	/P	/S	/	/P	/S	/	/S	/P	/	/
MINAHASAN																
Tondano	/	p/	/	/P	p/	/	/	/S	/	/P	/S	/P	s/	/	/P	/
MOKLEN-MOKEN																
Moklen	/	/	/	/P	/	/P	/	/	/P	/S	/S	/	/S	/P	/	/
NORTH BORNEO																
Banggi	/	/	/	/	/	/	/	/	/	/	/S	/	/	/P	/	/
Belait	/	/	/	/P	/	/	/	/S	/	/P	/	/	/	/	/	/
Ida’an (Begak)	/	p/	/	/P	p/	/	/	/S	/P	/P	/S	/	/S	/	/P	/
Kadazan	/	/	/	/P	/	/	/	/	/	/P	/S	/	/S	/S	/	/
Kimaragang	/	/	/	/	/	/	/	/	/	/	/S	/	/S	/S	/	/
Tatana’	/	/	/	/	/	/	/	/S	/	/	/S	/	/S	/S	/	/
Timugon	/	/	/	/P	/	/	/S	/S	/	/P	/S	/	/S	/S	/P	/
NORTHERN LUZON																
Balangao	/	/	/	/	p/	/	/	/P	/	/	/S	p/	s/	/	/	/
Bontok	/P	p/P	/P	/P	/	/	/P	/	/	/P	/S	p/	s/	s/	/P	/
Ifugao (Batad)	/P	p/P	/P	/P	/	/	/	/P	/	/P	/S	/	/S	s/S	/P	/
Ilocano	/	p/	/	/P	/	/	/	/P	/	/P	/	p/	/	/	/	/
Pangasinan	/	/	/	/P	p/	/	/	/	/	/P	/S	/	/	/S	/P	/
NORTHWEST SUMATRA-BARRIER ISLANDS																
Batak (Karo)	/	s/	/	/P	/	/	s/	/S	/	/P	s/S	/	s/	/	/P	/
Batak (Toba)	/	/	/	/P	p/	/	/	/S	/	/P	s/S	/P	s/	p/S	/P	/P
Mentawai	/	/	/	/	/	/	/	/S	/	/P	/	/	s/	/S	/	/
Nias	/P	p/P	/P	p/P	p/	/P	/	/S	/P	/P	/S	p/	s/	p/S	/P	/
Simeulue	/	/	/	/P	/	/	/	/S	/	/	/S	/	/S	p/S	/P	/
OCEANIC																
Adzera	/	p/	/P	/P	/	/	/	/S	/S	p/	/S	/	s/	/P	/	/P
Ala’ala	/P	p/	/	s/	p/	/	/	/S	/	/P	s/	/	/	p/P	/P	/
Amara	/	/	/	/P	/	/	/	/S	/	/	/S	/	s/	p/P	/	/
Ambae (Lolovoli Northeast)	/	/P	/	/P	/	/	/	/S	/	/	s/S	/P	s/P	/P	/P	/
Anejom	/S	/P	/	/P	/	/	p/	/S	/P	/P	s/S	p/	s/S	/S	/	/
Araki	/	/P	/	/P	/	/	/	/S	/	/P	s/S	/P	s/	/P	/P	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Aribwatsa	/	p/	/	/	/	/	/	/	/	/	/S	/	s/	p/P	/	/
Arop-Lokep	/	/	/	/P	/	/	/	/S	/	/S	/S	/P	s/S	p/P	/P	/
Arosi	/S	/P	/	/P	p/	/	/	/S	/P	/P	s/S	/	s/	/P	/	/
Asengseng	/S	/	/	/P	/	/	/	/S	/	/	/S	/P	s/	/P	/	/
Avava	/S	/S	/	/P	/	/P	/	/S	/S	/	/S	/S	s/S	p/P	/	/
Ayiwo	/	/	/	/P	/	/	/	/S	/	/	/S	/	/S	/P	/P	/
Bali-Vitu	/	/	/	/P	/	/	/S	/S	/P	/P	s/S	/P	s/	/P	/P	/
Banoni	/	p/P	/P	/P	p/	/	/	/P	/P	/P	/S	/P	s/	/P	/	/
Bariai	/	/P	/	/P	/	/	/S	/S	/S	/S	/S	/P	s/P	p/P	/	/
Buang (Mapos)	/	/	/	/	/	/	/	/S	/	/	/S	/	/	/P	/	/
Buma	/	/	/	/P	/	/	/	/S	/	/S	/S	/	/	p/P	/P	/
Cèmuhî	/P	/	/	/P	/	/	/P	/S	/P	/P	s/S	/P	s/	/S	/P	/
Dawawa	/S	s/	/	s/S	/	/	s/	/P	/	/P	s/P	s/	s/P	p/P	/P	/P
Drehu	/P	/P	/	/P	/	/P	/P	/S	/	/P	/S	/P	s/	/	/P	/P
Efate (South)	/	/P	/P	p/P	/	/	/	/S	/	/	/S	/	s/S	/P	/P	/
Erromangan	/P	/	/	/P	/	/	/	/S	/P	p/	/S	/P	/S	p/P	/	/
Fijian	/	/P	/P	/P	p/	/	/	/S	/	/P	s/S	/	s/P	/S	/P	/P
Futuna-Aniwa	/S	/P	/	/P	p/	/	/P	/S	/P	/P	/S	/P	s/P	/P	/P	/P
Gapapaiwa	/S	/	/	p/S	/	/	/	/	/	/P	/P	/	/	/P	/	/
Gela	/S	/	/	/P	/	/	/P	/S	/	/P	s/S	/P	s/	/S	/	/
Gumawana	/	p/	/	/S	p/	/	/S	/S	/P	/P	s/P	s/	s/P	p/P	/P	/
Halia	/	s/S	/	/P	/	/P	/	/S	/	/	s/S	/P	s/	s/P	/P	/P
Hawaiian	/	/P	/	/P	p/	/	/P	/P	/P	/P	/S	/P	/P	/S	/P	/
Hoava	/	/	/	/P	/	/	/	/S	/P	/P	/S	/P	s/	/S	/P	/
Iaai	/P	/P	/	/P	/	/P	/P	/S	/P	/P	/S	/P	s/P	/P	/P	/P
Iduna	/	p/S	/	/S	p/	/	/	/P	/	/P	s/P	s/	/	p/P	/	/
Ifira-Mele	/	/	/	/	/	/	/	/S	/	/S	/S	/	/	/P	/	/
Iwal	/S	p/	/	/P	/	/	/	/	/	/S	s/S	/	/	p/P	/	/
Jabêm	/S	p/	/	s/	/	/P	/	/S	/	/S	/S	/	s/P	p/P	/	/
Kairiru	/	/P	/	/	/	/	/S	/S	/	/S	s/P	/P	s/S	p/P	/P	/P
Kaliai-Kove	/	p/S	/	/P	p/	/P	/	/S	/	/S	s/S	p/	s/	p/P	/P	/
Kaulong	/S	/	/	/P	/	/	/S	/S	/S	/S	/S	/	s/	p/P	/P	/
Kele	/S	p/	/	/P	/	/	/P	/S	/	/S	/S	/	/S	p/P	/	/
Kilivila	/	p/	/	/P	/	/	/	/S	/	/P	s/S	/	s/P	p/	/P	/
Kiribati	/	p/P	/P	/P	p/	/	/	/S	/	/P	s/S	/P	s/P	/S	/P	/P
Kokota	/	/	/	/P	/	/	/	/P	/P	/P	/S	/P	s/	/S	/P	/
Koluwawa	/	/	/	/	/	/	/	/S	/	/	/P	s/	/P	/P	/	/
Kosraean	/P	/P	/	/P	/	/	/S	/S	/	/P	s/S	/	s/	/P	/P	/
Kwaio	/	/	/	/P	/	/	/	/S	/	/P	/S	/P	/	/P	/	/
Labu	/	/	/	/P	/	/	/	/S	/	/S	/S	/	/	p/P	/	/
Lamen	/	/	/	/	/	/	/	/	/S	/	/S	/S	/	/P	/	/
Lenakel	/S	p/	/	/P	/	/	/	/S	/S	/	/S	/S	s/S	p/P	/P	/
Lewo	/S	p/	/	/P	/	/	/	/S	/S	/S	s/S	/S	s/P	p/P	/	/
Longgu	/	/	/	/P	/	/	/S	/S	/	/P	s/S	/	s/	/S	/P	/
Loniu	/	/P	/	/P	/	/	/P	/S	/P	/S	/S	/P	s/S	p/P	/P	/
Lou	/	/P	/	/P	/	/	/	/S	/	/	/S	/P	s/P	/P	/P	/
Madak	/	/	/	/	/	/	/	/P	/	/	/	/	/	/	/P	/
Mae	/S	/P	/P	/P	p/	/	/P	/S	/P	/P	/S	/P	/P	/P	/	/P
Maisin	/S	p/S	/	/S	/	/S	/	/S	/	/	s/P	/	s/	p/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Maleu	/	/P	/	/P	p/	/	/	/S	/	/S	s/S	/	/	p/P	/P	/
Manam	/	s/	/P	/S	/	/S	/	/S	/	/P	s/P	/	s/S	p/P	/	/
Mangap-Mbula	/S	/	/	/P	/	/	/	/S	/	/S	/S	/	/	p/P	/	/
Maori	/	/P	/P	/P	/	/	/P	/	/P	/P	/S	/P	/P	/S	/P	/
Marquesan	/P	/	/	/P	p/	/	/P	/	/P	/P	/S	/P	/P	/S	/P	/
Matukar	/	s/	/	/	/	/	/	/	/	/	s/	/	/	p/	/	/
Mavea	/	p/	/P	/P	p/	/	/S	/S	/S	p/	s/S	/	s/S	p/P	/P	/P
Mekeo	/S	/	/	/S	/	/	/	/	/	p/	/P	/	s/	/P	/	/
Minaveha	/	p/	/	/S	/	/	/	s/	/P	/P	s/P	/	s/	p/P	/	/
Misima	/	/	/	/P	/	/	/	/	/	/	/P	s/	s/	/P	/	/
Mokilese	/P	/P	/	/P	p/	/	/S	/S	/S	/P	/S	/	s/P	/P	/P	/P
Mono-Alu	/	/	/	/	/	/	/P	/P	/P	/	/	/	s/P	/	/	/
Motu	/S	/P	/	/S	p/	/	/P	/	/S	/	s/P	/	s/	/P	/	/
Musom	/	p/	/	/P	/	/	/S	/S	/S	/S	/S	/	s/	p/P	/	/
Mussau	/S	/P	/	/P	/	/	/S	/S	/	/P	s/S	/S	s/P	/P	/P	/
Mwotlap	/	p/	/	p/	/	/	/	/S	/	/	/S	/P	s/	/P	/	/
Nadroga	/	/	/	/	/	/	/S	/S	/	/P	/S	/	/	/S	/	/
Nakanai	/	/P	/	/P	/	/	/	/S	/	/P	/S	/P	s/S	/P	/	/
Nalik	/	/	/	/P	p/	/	/	/S	/	/P	/S	/P	/	/P	/	/
Nauruan	/	/	/	/	/	/	/	/	/	/	/S	/	/P	/	/	/
Nehan	/	/P	/	/P	p/	/	/	/S	/	/P	/S	/	s/	/P	/	/P
Nengone	/	/	/	/	/	/	/P	/S	/P	/P	/S	/	/	/P	/	/
Nguna	/	/P	/	/P	/	/	/	/S	/	/P	/S	/	/	/P	/	/
Niuafo'ou	/	/	/	/P	/	/	/P	/S	/P	/P	/S	/P	/	/S	/	/
Niuean	/	/P	/	/P	/	/P	/	/	/P	/P	/S	/P	/	/S	/P	/P
Numbami	/	/	/	/P	/	/	/	/	/	/S	/S	/	/	p/P	/	/
Paamese	/	p/S	/	/P	/	/	/S	/	/	p/	s/S	/S	s/	p/P	/P	/
Patep	/	/P	/	/P	/P	/P	/	/S	/	/S	/S	/	/	/P	/P	/P
Patpatar	/	/	/	/	/	/	/	/P	/	/	/S	/P	/	/P	/	/
Pohnpeian	/	/P	/	/P	p/	/	/	/S	/S	/P	s/S	/	s/P	/P	/P	/P
Port Sandwich	/	/	/	/P	/	/	/	/S	/	p/	/S	/S	s/S	p/P	/	/
Puluwat	/	/P	/	/P	/	/	/	/S	/	/P	s/S	/	s/	/P	/P	/P
Raga	/	/	/	/P	/	/	/	/S	/	/	/S	/P	s/P	/P	/	/
Ramoaaina	/	/	/	/P	/	/	/	/S	/	/	/S	/P	/	/P	/	/
Rapanui	/P	/P	/	/P	p/	/	/P	/S	/P	/P	/S	/P	/	/S	/P	/P
Roro	/	/	/	/S	/	/	/	/S	/	/	/P	/	/P	/P	/	/
Rotuman	/P	s/	/	/P	/	/	/S	/S	/	/	/S	/	/S	/P	/P	/P
Roviana	/	/	/	/	/	/	/P	/S	/P	/P	s/S	/	s/	/S	/P	/
Sa'a	/	/	/	/P	p/	/	/	/	/	/	s/	/P	s/	/	/	/
Sakao	/	/P	/	/P	/	/	/	/S	/	p/	/S	p/	s/	p/P	/P	/
Saliba (in Papua New Guinea)	/S	s/	/	/S	/	/	/S	/P	/S	/P	s/P	s/	s/	p/P	/	/
Samoan	/P	p/P	/P	/P	/	/P	/P	/	/P	/P	/S	/P	/P	/S	/P	/P
Seimat	/	/	/	/	/	/	/	/S	/	/	/S	/	s/	/P	/	/
Siar	/	/	/	/P	/	/	/P	/	/P	/P	/S	/	/	/P	/	/
Sinaugoro	/S	p/	/	/S	p/	/P	/	/P	/	/P	s/P	/	s/P	p/P	/	/P
Sio	/S	/	/	/P	/	/P	/	/S	/S	/S	s/S	/	s/P	p/P	/P	/
Sisiqa	/	/	/	/P	/	/	/S	/S	/S	/P	/S	/P	s/	/P	/	/
Sobei	/S	/	/	/P	/	/	/S	/S	/	/P	s/S	s/	s/P	p/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Sonsorol-Tobi	/P	/P	/	/P	/	/	/	/S	/	/P	s/S	/	s/	/P	/P	/
Southeast Ambrym	/	/	/	/	/	/	/	/S	/	/	/S	/S	/	/P	/	/
Suau	/	/	/	/	p/	/	/	/	/	/	/P	/	/	/P	/	/
Sudest	/	/	/	/	/	/	/S	/S	/	/P	/S	p/	/	/P	/P	/
Sursurunga	/	/	/	/	/	/	/	/S	/	/	/S	/	/	/	/	/
Tahitian	/	/P	/P	/P	p/	/P	/	/P	/P	/P	/S	/P	/P	/S	/P	/P
Taiof	/	/	/	/	/	/	/	/S	/P	/P	/S	/	/	/P	/	/
Takia	/S	/S	/	/S	/	/	/S	/S	/S	/P	s/P	/S	s/	p/P	/	/
Tamabo	/	/P	/	/P	/	/	/	/S	/P	/P	/S	p/	s/	/P	/P	/
Tawala	/	p/	/	/	/	/	/S	/P	/	/P	s/P	/	/	p/P	/	/
Teop	/	/	/	/P	p/	/	/	/S	/P	/	/S	/P	s/	/P	/P	/
Tigak	/S	/P	/	/P	/P	/	/P	/S	/P	/P	/S	/P	s/	/P	/P	/P
Tinrin	/	/P	/	/P	/	/	/P	/S	/	/P	/S	/P	s/S	/	/P	/P
Tobati	/	/	/	/S	/	/	/	/S	/	/	s/P	/	/	p/P	/P	/
Tokelauan	/	/	/	/P	/	/	/P	/S	/P	/P	/S	/	/P	/S	/	/
Tolai	/P	/P	/	/P	/	/	/P	/P	/P	/P	/S	/P	s/P	/P	/P	/
Tongan	/	/P	/	/P	/	/P	/P	/S	/P	/P	/S	/P	/	/S	/P	/P
Toqabaqita	/	/P	/	/P	/	/P	/	/S	/	/P	s/S	/S	s/	/P	/P	/
Tuam	/	/	/	/P	/	/	/	/	/	/	/S	/	/	p/P	/	/
Tumleo	/	/	/	/	/	/	/	/	/	/	/S	/	s/P	p/P	/	/
Tungak	/	/	/	/P	/	/	/P	/S	/P	/	/S	/P	/P	/P	/	/
Tuvaluan	/	/	/	/P	/	/	/P	/S	/P	/P	/P	/P	/P	/S	/	/
Ulithian	/	/P	/	/P	/	/	/	/S	/	/P	s/S	/	s/P	/P	/	/
Unua	/	/	/	/	/	/	/	/S	/	/	/	/	/	/	/	/
Ura	/	/	/	/	/	/	/	/	/P	/	/S	/P	/	p/P	/	/
Vinitiri	/	/	/	/P	p/	/	/P	/P	/	/P	/S	/P	s/P	/P	/P	/
Vinmavis	/	/	/	/P	/	/	/	/S	/	/	/S	/S	s/S	p/P	/P	/
Wala	/	/	/P	/P	p/	/	/S	/S	/P	/P	s/S	/S	s/S	/P	/P	/P
Wampar	/S	/	/	/P	/	/P	/S	/S	/S	/P	/S	/S	s/	p/P	/	/P
Wedau	/	/P	/	s/	/	/	/	/P	/	/P	s/P	p/	s/	/P	/	/
Woleaian	/	/P	/	/P	p/	/	/S	/S	/P	/P	s/S	/	s/P	/P	/P	/
Wuvulu	/P	/	/	/P	/	/	/P	/P	/P	/	/S	/	s/	/P	/	/
Xârâcùù	/	/	/	/P	p/	/	/	/S	/	/P	/S	/P	s/	/P	/	/
Ùa Pou	/	/	/	/P	/	/	/P	/P	/P	/	/	/P	/	/	/	/
PAIWANIC																
Amis	/	/	/	/P	/	/	/	/P	/	/	/S	/	/	/S	/	/
Paiwan	/	/	/	/P	/	/	/	/P	/	/P	/S	/	/S	/S	/P	/
PALAUAN																
Palauan	/	p/P	/	/P	p/	/	/	/P	/	/P	s/S	p/	s/	p/P	/P	/
PUYUMA																
Puyuma	/	/	/P	/P	p/	/	/P	/P	/P	/P	/S	/P	s/P	p/S	/P	/P
RUKAI																
Rukai (Mantauran)	/	/	/P	/	p/	/P	/P	/P	/	s/	/S	p/	s/	s/S	/	/P
Rukai (Tanan)	/	p/	/	/P	p/	/	/	/P	/	/P	/S	p/	s/P	s/S	/	/
SAMA-BAJAW																
Bajau (Sama)	/	/	/	/P	/	/	/S	/S	/P	/P	/S	/	/	/	/P	/
Bajau (West Coast)	/	/P	/P	/P	/	/	/S	/S	/P	/P	/S	/P	s/	/	/P	/
SOUTH HALMAHERA - WEST NEW GUINEA																
Ambai	/S	/S	/P	/P	/P	/	/S	/S	/S	/S	s/S	/	/P	p/P	/	/P

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Biak	/	/P	/	/P	/	/S	/S	/S	/S	/S	/S	/	s/S	p/P	/P	/P
Irarutu	/	p/	/	/P	/	/	/S	/S	/	/S	/S	/	p/	p/P	/	/
Mor	/S	/	/	/P	/	/	/	/S	/	/S	s/S	s/	s/	p/P	/	/
Taba	/	p/	/S	/P	p/	/	/	/S	/	/S	/S	/	/P	/P	/P	/P
Warembori	/	/	/	/P	/	/	/S	/S	/	/S	/S	/	/	p/P	/	/
SOUTH SULAWESI																
Selayar	/	p/	/	/P	p/	/	/S	/S	/	/	/S	/	s/	p/S	/	/
SULAWESI																
Bugis	/	/P	/	/	/	/	/S	/S	/	/P	/S	/	/	/P	/P	/
TSOUIC																
Tsou	/	/P	/	/	/	/P	/	/	/	/P	/S	/	/	/S	/	/
WESTERN PLAINS																
Thao	/	/	/	/P	p/	/	/	/	/	/P	/S	/	/	s/	/	/
YAPESE																
Yapese	/	/P	/	/P	/	/	/P	/S	/P	/P	s/S	/P	s/	/S	/P	/P
Great Andamanese																
GREAT ANDAMANESE																
Great Andamanese	/	s/	/	s/S	p/	/	/	/	/	/	/P	/	p/	p/P	/	/
Kusunda																
KUSUNDA																
Kusunda	/	s/	/	/S	/	/S	/	/	/	s/	/P	/	/	s/P	/	/
Miao-Yao																
MIAO-YAO																
Hmong Njua	/	/P	/S	/P	/P	/P	/	/S	/P	/P	/S	/	/P	/P	/P	/P
Mien	/S	/	/	/	/	/	/	/P	/	/P	/S	/	/P	/P	/	/
Punu	/	/	/	/	/	/	/	/S	/	/P	/S	/	/P	/P	/	/
Shompen																
SHOMPEN																
Shompen	/	/	/	/	/	/	/	/	/	/	/S	/	/S	/S	/	/
Sino-Tibetan																
BAI																
Bai	/S	/	/	/	/	/	/	/S	/	/P	/S	/S	/P	/P	/	/
BODIC																
Amdo (Themchen)	/	/	/	/S	/	/	/S	/	/	p/	/P	/	/	/P	/	/
Balti	s/	s/P	/	/S	/	/	s/	/P	/S	/	/P	s/	/	/P	/S	/
Baragaunle	/	/	/	/	/	/	/	/P	/	p/	/P	/	/	/P	/	/
Byansi	/P	s/S	/S	s/S	s/	/S	/	/P	/P	p/	/P	s/	/P	s/P	/	/
Chantyal	/	s/	/	/S	/	/	/	/P	/	p/	/P	s/	/	/P	/	/
Chaudangsi	/	/	/	/	/	/	/	/P	/	/P	/P	/	/	/P	/	/
Darma	/P	s/S	/S	/S	s/	/S	/	/P	/	p/	/P	/S	/P	s/P	/	/S
Gahri	/S	/	/	/	/	/	/	/	/	/P	/P	/	/P	/P	/	/
Gurung	/S	s/	/S	/S	/S	/S	/	/P	/	p/	/P	/S	/P	/P	s/	/S
Jad	/	/	/	/	/	/S	/	/	/	/	/P	/	/	/P	/	/
Johari	/	/	/	/S	/	/	/	/P	/	/P	/P	/	/	/P	/	/
Kanashi	/	/	/	/S	/	/	/	/	/	/P	/P	/	/	/P	/	/
Kham (Dege)	/	/	/	/S	/	/	/S	/S	/S	p/	/P	/	/P	/P	/	/
Kham (Nangchen)	/	/	/	/S	/	/	/	/S	/	p/	/P	/	/	/P	/	/
Kinnauri	s/	s/S	/S	s/S	/	/S	/	/P	/	/P	s/P	s/	/P	s/P	/	/S
Kurtöp	/	s/	/S	/S	/	/S	/S	/P	/	p/	/P	/S	/P	/P	/S	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Kyirong	/	/	/	/S	/	/	/S	/P	/S	p/	/P	/	/	/P	/	/
Ladakhi	s/	s/	/	s/S	s/	/S	/	/P	s/	/	/P	s/	/	/P	/	s/
Lhomi	s/	s/S	/	s/S	/	/S	/	/S	/	/	/P	/	/P	s/P	/	/
Manange	/	/	/	/	/	/	/	/	/	p/	/P	/	/	/P	/	/
Marchha	/	/	/	/	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Nar-Phu	s/	s/	/	/S	/	/S	/	/P	/S	p/	/P	/	/	/P	/	/
Nyamkad	s/	/	/	/	/	/	/	/P	/	/P	/P	/	/P	/P	/	/
Pattani	s/	/	/	s/S	/	/	/	/P	/	p/	/P	s/	/	/P	/	/
Purki	s/	s/	/S	s/S	/	/S	/	/P	/	s/	/P	s/	/P	s/P	/	/
Rang Pas	/	/	/	/	/	/	/	/	/	/P	/P	/	/	/P	/	/
Sherpa	/S	s/S	/	/S	/	/	/	/P	/	p/	/P	/	/	/P	/	/
Shigatse	/	/	/	/	/	/	/	/	/	p/	/P	/	/	/P	/	/
Sikkimese	/S	/S	/S	/S	/	/S	/S	/	/S	/S	/P	/S	/P	/P	/S	/
Spitian	/	/	/	/	/	/	/	/P	/	/	/P	s/	/P	s/P	/	/
Tamang (Eastern)	/S	s/S	/S	/S	/S	/S	/P	/P	/S	/P	/P	/S	/P	/P	s/	/
Tamang (Western)	/	s/S	/	/	/S	/S	/	/	/	/	/	/	/	/	/	/S
Tangbe	s/	s/	/	/S	/	/	/	/	/	/P	/	/	/	/	/	/
Thakali	/	/	/	/	/	/	/	/P	/	/	/P	s/	/	/P	/	/
Tibetan (Dingri)	/	/	/	/	/	/S	/	/	/	/	/P	/	/	/P	/	/
Tibetan (Drokpa)	/	/	/	/	/	/	/	/S	/	p/	/P	/	/	/P	/	/
Tibetan (Modern Literary)	s/	s/S	/S	/S	/S	/S	/S	/S	/S	/	/P	/S	/P	s/P	/	/S
Tibetan (Standard Spoken)	/S	/	/	/	/	/	/	/S	/	p/	/P	/	/	/P	/S	/
Tinani	s/	/	/	/	/	/	/	/P	/	/	/P	/	/	/P	/	/
Tod	/	/	/	/	/	/	/	/	/	p/	/P	/	/	/P	/	/
Tshangla	/S	s/S	/S	/S	/	/	/	/P	/S	/	/P	/	/P	/P	s/	/S
Zhang-Zhung	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
BODO-GARO																
Atong	/S	s/	/S	/S	s/	/S	/	/P	/S	s/	/P	/	/	/P	/S	s/
Bodo	s/	s/	/	s/	/	/	/	/P	/	/	/P	s/	p/P	/P	/	/
Deuri	s/	s/	s/	/	s/	/	/	/	/	s/	/P	s/	/	s/P	/	/
Dimasa	/	s/	/S	/S	/	/	/	/P	/	s/	/P	s/	/P	/P	/S	/
Garo	s/	s/	/S	/S	s/	/S	/	/P	/	s/	/P	/	/P	/P	s/	/
Kachari	/	s/	/	s/S	/S	/	/	/P	/	s/	/P	s/	/	/P	/	/
Kokborok	s/	s/	/	s/S	s/	/	/	/P	/	s/	/P	s/	p/	/P	/	/
BURMESE-LOLO																
Achang	/S	/S	/S	/S	/	/	/	/S	/	/	/P	/	/P	/P	/	/S
Akha	/S	/	/S	/S	/P	/S	/	/S	/	/P	/P	/	/P	/P	/S	/S
Bisu	s/	/	/	s/S	/	/	/	/S	/	p/	/P	/	/P	/P	/	s/
Burmese	s/	s/S	s/	s/S	s/	/S	/	/P	/	/	/P	/S	/	/P	s/	s/
Hani	/	/	/S	/S	/	/S	/S	/S	/	/P	/P	/	/P	/P	/	/S
Jino	/	/S	/	/S	/	/P	/	/	/	/	/P	/	/	/P	/	/
Lahu	/S	/S	/S	/S	/S	/	/S	/S	/S	/P	/P	s/	/P	/P	/S	/S
Lalo	/S	/S	/S	/S	/S	/	/	/S	/	/P	/P	/	/	/P	/S	/S
Lisu	/S	/S	/S	/S	/	/S	/	/S	/	/P	/P	s/	/P	/P	/	/S
Maru	/S	/S	/	/S	/S	/	/	/P	/	/P	/P	/S	/P	/P	/S	/
Nasu	/	/	/	/S	/	/	/	/	/	/P	/P	/	/P	/P	/	/
Nuosu	/	/	/	/S	/	/S	/	/S	/	/P	/P	/	/P	/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Nusu	/	/	/S	/	/	/	/	/S	/	/	/P	/	/	/P	/	/
CHINESE																
Cantonese	/	s/S	/	/	/P	/P	/	/P	/P	/P	/S	/	/	/P	/	/P
Hakka	/	s/P	/	/	/	/	/	/P	/	/P	/S	/	/	/P	/	/P
Mandarin	/S	s/P	/	/	/P	/P	/	/P	/	/P	/S	s/	/	/P	/	/P
Wu (Shanghai)	/S	/	/P	/P	/	/P	/	/P	/	/P	/S	/	/P	/P	/P	/
Wutun	/S	s/	s/	s/	/	/	/	/	/	/	/P	s/	/	/P	/	/
DHIMALISH																
Dhimal	/P	s/	/	s/S	s/	/S	/	/P	s/	p/	s/P	/	/P	s/P	/	/
DIGAROAN																
Digaro	/	s/	s/	/S	s/	/S	/	/	/	s/	/P	/	/P	s/P	/	/
Idu	s/	/	/	/S	/	/	/	/P	/	s/	/P	/	/	/P	/	/
HRUSO																
Miji	/	s/	/	s/	/	/	/	/P	/	/	/	s/	/	/	/	/
JINGPHO																
Jingpho	/S	/S	/S	/S	/	/S	/	/	/	/P	/P	/S	/P	/P	/S	/
KAREN																
Karen (Bwe)	/S	/	/S	/P	/	/P	/S	/S	/S	/	s/S	/	p/	p/P	/	/
Karen (Pwo)	/S	/	/	/P	/	/	/	/S	/	/S	/S	/	/	/P	/	/
Karen (Sgaw)	/S	/S	/S	/P	/	/P	/	/S	/	/	/S	/S	/P	/P	/	/P
Kayah Li (Eastern)	/S	/	/S	/	/	/	/	/S	/	/S	/S	/	/P	/P	/	/
KUKI-CHIN																
Angami	/S	s/S	/S	/S	p/	/	/S	/S	/S	/S	/P	/	/P	/P	/S	/S
Ao	/S	s/	/	/S	s/	/S	/S	/S	/S	p/	/P	s/	p/	/P	s/	/
Bawm	/S	/S	/S	/S	/S	/S	/	/S	/	/S	/P	/S	/P	/P	/S	/
Chin (Mara)	/S	/	/S	/S	/P	/S	/	/	/	/S	/P	/S	/P	/P	/S	/
Chin (Siyin)	/S	/S	/S	/S	/	/S	/	/P	/S	/S	/P	s/	/P	/P	/S	/S
Chin (Tiddim)	/	/S	/S	/S	/	/S	/	/P	/	/S	/P	/	p/	p/P	/S	/
Hmar	/S	/S	/	/S	/	/	/	/	/	/S	/P	s/	/	/P	/	/
Koireng	/	/	/	/	/	/	/	/S	/	/	/P	/	/	/P	/	/
Lai	/S	/S	/	/S	/S	/S	/	/	/	/S	/P	/S	/P	/P	/S	/S
Lotha	s/	s/	/	/S	s/	/	/S	/S	/	p/	/P	/	/P	/P	/	/
Meithei	/S	s/	/	/S	s/	/	/S	/S	/S	s/	/P	/	p/P	/P	/	s/
Mikir	s/	s/	/S	s/S	p/	/	/	/P	/	s/	/P	/S	p/P	/P	s/	/
Mizo	/S	/S	/S	/S	/S	/S	/	/	/	/S	/P	/S	/P	/P	/S	/S
Naga (Mao)	/S	/S	/	/	/	s/	/	/	/	/S	/P	/	/P	/P	s/	/
Naga (Tangkhul)	s/	s/	/	s/	/	/S	/	/S	/	p/	/P	/	/P	/P	/	/
Naga (Zeme)	/S	s/	/	/S	/	/S	/	/P	/	/S	/P	/S	/P	/P	/	/
Ralte	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Sema	/S	s/S	/	/S	/	/S	/	/S	/	s/	p/P	/	p/P	s/P	/S	/
Tarao	/S	s/	/	s/	p/	/	/	s/	/	s/	/P	s/	/	p/P	/P	/
Thadou	/	/	/	/S	/	/	/	/P	/	/	/	/	/	/	/	/
LEPCHA																
Lepcha	/S	s/S	/S	/S	/S	/S	/S	/P	/S	/	/P	/S	/P	/P	/S	/S
MAHAKIRANTI																
Athpare	s/	s/	/	s/	/	/	/	/P	/	s/	s/P	s/	p/	/P	/S	/
Belhare	/S	s/	/	s/	/	/	/	/	/	s/	/P	/	p/	/P	/	/
Camling	/	/	/	s/	/	/	/	/P	/	/	/P	/	p/P	/P	/	/
Chepang	/	/	/	/	/	/	/	/P	/	s/	/P	/	/P	/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Dumi	/	s/	/	/S	/S	/	/	/P	/	/	s/P	/	p/	s/P	/	/
Hayu	/	s/S	/	/S	/	/S	/	/P	/	/P	s/P	s/	/	s/P	s/	/
Khaling	/S	s/S	/	s/	/	/	/	/P	/	p/	s/P	s/	p/	s/P	s/	/S
Kham	p/	s/S	/	s/S	/S	/	/	/P	/	p/	/P	s/	p/P	/P	s/	/S
Kulung	/	/	/	s/	/	/	/	/	/	/	/P	/	p/	/P	/	/
Limbu	/	s/	/	s/	/	/S	/	/P	s/	/	s/P	/	/	/P	s/	/
Magar	/	s/	/S	/S	/	/S	/	/P	/P	p/	/P	s/	/P	/P	/P	/
Newar (Dolakha)	/S	s/S	/	/S	/	/S	/	/P	/	p/	/P	/	/P	s/P	/	/
Newari (Kathmandu)	/S	s/S	/S	s/S	s/	/S	/	/P	/	p/	/P	/	/P	s/P	/	/
Thangmi	/	s/	/	s/S	/	/S	/	/P	/	p/	/P	s/	/	s/P	/	/
Thulung	/	s/S	/	s/S	s/	/	/	/P	/	/	s/P	/	/P	s/P	/	/
Wambule	/	/S	/S	/S	/S	/	/	/P	/P	p/	/P	/S	/P	/P	/P	/
Yamphu	/	/	/	/	/	/	/	/P	/	p/	/P	/	/	/P	/	/
MIJU																
Miju	/	s/	/	/	/	/	/	/	/	/	/P	s/	/	s/P	/	/
MRU																
Hkongso	/S	/	/P	/S	/	/S	/P	/P	/S	/P	/S	/S	/	/P	/S	/
NAXI																
Naxi	/	/	/S	/	/	/	/	/P	/	/P	/	/	/	/	/	/
NORTHERN NAGA																
Chang	/	s/S	/	/S	/	/	/	/P	/	p/	/P	s/	p/P	/P	/	/
Jugli	s/	/	/	/S	/	/	/	/	/	/	/P	s/	/P	/	/	/
Lungchang	/S	s/	/	/S	/	/	/	/P	/	/	/P	s/	/P	s/	/	/
Nocte	s/	s/	p/	s/S	/	/	/	/P	/	s/	s/P	s/	/P	s/P	/	s/
Tutsa	/S	s/	/P	/S	/	/	/	/P	/	/	/P	/	/P	s/P	/	/S
NUNGISH																
Anong	/	/	/	/	/	/	/	/S	/	/	/P	/	/	/P	/	/
Dulong	/S	/	/	/S	/	/P	/	/P	/	p/	/P	/S	/	/P	/S	/
Rawang	/S	s/S	/S	/S	p/	/S	/	/P	/	/P	/P	/S	/P	/P	/S	/S
QIANGIC																
Pumi	/	s/	/	/S	/S	/S	/	/P	/	/	/	/	/P	s/P	/S	/
Qiang	s/	/	/	/S	/	/	/S	/S	/S	p/	s/	/	/	s/	/	/
SULUNG																
Sulung	/S	s/	/	/	/	/	/	/	/	/	/P	/	/P	/	/	/
TANI																
Apatani	/S	s/	/	/S	s/	/	/	/P	/	s/	/P	/S	/P	/P	/S	/
Bokar	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Bori	/	s/	s/	/S	/	/	/	/	/	s/	/P	/	/P	/	/	s/
Galo	/S	s/	s/	s/S	/	/	/	/S	/	s/	/P	/	/P	/P	/	/
Milang	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Miri (Hill):	/	/	/	s/	/	/	/	/	/	s/	/	/	/	/	/	/
Mising	s/	s/	/	/S	s/	/S	/S	/P	/	s/	/P	/	/P	s/P	/	/
Nishi	/	s/	/	s/S	/	/S	/	/	/S	s/	/P	/S	/P	/P	/	/
Tagin	s/	s/	s/	s/S	s/	/	/	/	/	s/	/P	/S	/	/P	/	/
TUJIA																
Tujia (Northern)	/	/	/	/	/	/	/	/P	/	/	/P	/	/	/P	/	/
WESTERN ARUNACHAL																
Bugun	/S	s/	s/	/	/	/	/	/	/	p/	/P	s/	/	/	/	/
Sherdukpen	/	s/	/	/	/	s/	/	/	/	/	/	s/	/	/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
rGYALRONG																
Gyarong (Cogtse)	p/	p/	/	/S	/	/	/	/P	/	p/	/P	s/	p/P	/P	/	/
rGyalrong (Caodeng)	/	/	/	/	/	/	/	/P	/	p/	/P	s/	p/	/P	/	/
South Andamanese																
SOUTH ANDAMANESE																
Jarawa (in Andamans)	/	/	/	/	/	/	/	/	/	/	/P	/	p/	/P	/	/
Onge	/	s/S	/	s/	/	/	s/	p/	/	/S	/P	s/	s/P	p/P	/	/
Tai-Kadai																
HLAI																
Hlai (Baoding)	/	/	/	/P	/	/	/	/S	/	/P	/S	/	/S	/P	/	/
KADAI																
Gelao	/	/	/	/P	/	/	/	/S	/	/S	/S	/	/S	/P	/	/
KAM-TAI																
Dong	/S	/S	/	/P	/	/	/	/S	/	/	/S	/	/S	/P	/	/
Khün	/S	/	/	/	/	/	/	/S	/	/P	/S	/	/S	/P	/	/
Lao	/S	/	/	/P	/P	/P	/	/S	/	/P	/S	/	/S	/P	/P	/P
Nung (in Vietnam)	/S	/P	/	/P	/	/	/	/S	/	/P	/S	/P	/S	/P	/	/P
Shan	/	/	/	/	/	/	/	/	/	/	/S	/	/	/P	/	/
Thai	/S	/P	/	/P	/P	/P	/	/S	/S	/P	/S	/	/	/P	/P	/P
Yay	/	/	/	/	/	/	/	/S	/	/	/S	/	/S	/P	/P	/
Yong	/	/	/	/P	/	/	/	/S	/	/	/	/	/	/	/	/
Zhuang (Northern)	/S	/	/	/P	/	/	/	/	/	/	/S	/	/	/P	/	/

Australia-New Guinea																
Abun																
ABUN																
Abun	/S	/S	/	/P	/	/	/S	/S	/S	/	/S	/	/P	/P	/	/
Anêm																
ANêM																
Anêm	/	/S	/	/P	/	/	/	/S	/	/S	s/S	/	s/	p/P	/	/
Australian																
ANINDILYAKWA																
Anindilyakwa	/	s/	/	/	/	/	/	/P	/	/P	p/S	p/	s/	p/	/	/
ANSON BAY																
Bachamal	/	s/	/	/S	/	/	/	/	/	/P	p/	/	/	p/	/	/
Pungupungu	/	/S	/	/S	/	/	/	/	/	/P	/P	/	/S	/P	/	/
BUNABAN																
Bunuba	/	/	/	/	/	/	/	/P	/	/	/	/	/P	/P	/	/
Gooniyandi	/	p/S	/	/	/	/S	/P	/P	/	/P	/P	/	s/P	/P	/	/
BURARRAN																
Burarra	/	s/	/	p/S	/	/	/	/	/	/P	p/P	/	p/	p/P	/	/
Gurr-Goni	/	/	/	p/	/	/	/	/P	/	/	/	/	/	p/P	/	/
DJERAGAN																
Miriwung	/	/	/	s/	/	/	/	/	/	/P	/	/	/P	p/	/	/
EASTERN DALY																
Kamu	/	/	/	/S	/	/	/	/S	/	/P	/	/	/	/P	/	/
Madngele	/	/	/	/S	/	/	/	/	/	/P	/P	/	/S	/P	/	/
GAAGUDJU																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Gaagudju	/	s/	/	/S	/	/	/	/P	/	/P	p/	/	/	p/	/	/
GARRWAN																
Garrwa	/	/S	/	s/P	/	/	/P	/P	/	/P	/S	s/	/P	/S	/P	s/
GUNWINYGIC																
Bininj Gun-Wok	/P	/	/	s/	s/	/	/	/	/P	/P	p/P	p/	/	p/P	/P	/
Gunbalang	/	/	/	/P	/	/	/	/P	/	/	p/S	/	/S	p/P	/	/
Ngalkbun	/P	s/	/	s/	/	/	/	/	/	/	p/P	s/	/	p/P	/	/
IWAIDJAN																
Iwaidja	/P	s/	/	/	p/	/	/	/P	/	/	p/S	/	p/	p/P	/	/
Maung	/P	s/	/	/P	/	/	/	/P	/	/P	p/S	/P	p/S	p/P	/P	/
JAMINJUNGAN																
Jaminjung	/	s/	/	/	/	/	/	/	/	/	/	/	/	p/	/	/
LARAGIYAN																
Laragia	/	/	/	/S	/	/	p/	/	/	/P	p/P	s/	s/S	p/P	/	/
LIMILNGAN																
Limilngan	/	/	/	s/	/	/	/	/	/	/	/	/	/	p/	/	/
MANGARRAYI																
Mangarrayi	/S	/	/	/P	/	/	/P	/P	/	/P	p/P	s/	s/	p/S	/P	s/
MARAN																
Alawa	/	s/S	/	s/P	/	/S	/	/P	/	/P	/	p/	s/	/	/P	/
Mara	/	s/	/	p/P	/	/	/	/P	/	/	p/	p/	s/S	p/	/P	/
Warndarang	/	s/	/	s/	/	/	/	/	/	p/	p/S	p/	p/	p/P	/	/
MURRINH-PATHA																
Murrinh-Patha	/S	/S	/	/	/	/	/	/S	/	/P	p/P	/	/	p/P	/	/
NAKKARA																
Nakkara	/	/	/	/	/	/	/	/	/	/P	/	/	p/	/	/	/
NDJéBBANA																
Ndjébbana	/	/S	/	s/S	/	/	/	/P	/	/P	p/P	/	/S	p/P	/	/
NGALAKAN																
Ngalakan	/	s/	/	s/	s/	s/	/	/P	/	s/	p/	/	s/	p/	/	/
NGANDI																
Ngandi	/P	s/	/	s/	s/	s/	/	/P	/	s/	p/P	p/	s/	p/P	/	/
NORTHERN DALY																
Malakmalak	/	/S	/	s/	/	/	/	/S	/	/P	/P	/	/	/P	/	/P
Tyeraity	/	/S	/	/	/	/	/	/	/	/	/P	/	/S	/P	/	/
NUNGGUBUYU																
Nunggubuyu	/P	s/	/	s/	s/	/	/	/P	/	/P	p/	p/	/	p/	/	/
NYULNYULAN																
Nyigina	/	/	/	/	/	/	/	/P	/	/P	s/S	/	/	p/	/	/
Nyulnyul	/P	/	/	/	/	/	/P	/P	/	/P	s/	/	p/	p/	/	/
Warrwa	/	/	/	/	/	/	/P	/	/	/	s/	/	p/	p/	/	/
Yawuru	/	/	/	/	/	/	/	/	/	/P	/	/	/	/S	/	/
PAMA-NYUNGAN																
Adynyamathanha	/	s/	/	s/S	/	/	/	/	/	/	s/	/	/	s/P	/P	/
Alyawarra	/	s/	/	s/S	s/	/	/	/S	/	s/	/P	/	s/S	/P	/	s/
Anguthimri	/	s/	/	s/	/	/	/	/S	/	/	/P	s/	/	/P	/	/
Arabana	/	s/	/	s/	/	/	/	/	/	/P	/P	s/	/P	/P	/	/
Arrernte (Mparntwe)	/	s/S	/	/S	/	/	/S	/S	/	s/	/P	/	s/	/P	/	/
Arrernte (Western)	/	s/	/	s/S	s/	/	/	/S	/	s/	/	s/	/	s/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Badimaya	/	s/	/	/	/	/	/	/P	/	/P	/P	s/	/P	/P	/	/
Bandjalang	/P	s/	/	s/	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Bilinara	/	s/	/	s/	/	/	/	/P	/	/	/	/	/	/	/P	/
Biri	/	s/	/	s/	s/	/	/	/P	/	/	s/S	/	/	s/P	/	/
Bularnu	/P	s/	/	s/	/	/	/	/	/	/P	/P	s/	/	/P	/P	/
Bunganditj	/	s/	/	s/	/	/	/	/	/	/P	s/	s/	/	/	/	/
Dharawal	/	s/	/	s/	/	/	/	/	/	/P	s/	s/	/	s/	/	/
Dhargari	/	s/	/	s/	/	/	/	/	/	/	/	s/	/	/	/P	/
Dharumbal	/	s/	/	s/	s/	/	/	/	/	/	/P	/	/	/P	/	/
Diyari	/	s/S	/	s/	/	/S	/P	/P	/	/P	/P	/	/	/P	/	/
Djabugay	/	s/	/	s/	/	/	/	/	/	/P	/P	/	/	/P	/	/
Djambarrpuyngu	/	s/P	/	s/	/	/	/	/P	/	/P	/	/	/P	/P	/	/
Djapu	/	s/	/	s/	s/	/	/	/P	/	/P	/P	/S	/P	/P	/P	/
Djaru	/	s/	/	s/	/	/	/	/P	/	/	/	/	/	/P	/	/
Djinang	/	s/S	/	s/	/	/	/	/P	/	/P	/	s/	/P	/	/	/
Dyirbal	/	s/	/	s/	/	/	/	/P	/	/P	/P	/	/P	/P	/	/
Gidabal	/	s/	/	s/	/	/	/	/P	/	/	/	/	/	/	/	/
Gugada	/	s/P	/	/S	/	/	/	/S	/	/	/P	/	/	/P	/	s/
Gumbaynggir	/	s/	/	s/	/	/	/	/	/	/	/	s/	/	/	/	/
Gureng Gureng	/	s/	/	s/	/	/	/	/	/	/P	/	s/	/	/P	/P	/
Guugu Yimidhirr	/	s/	/	s/	s/	/	/	/P	/	/P	/P	s/	/	/P	/	/
Innamincka	/	s/	/	/S	/	/	/	/S	/	/P	/P	s/	/S	/P	/	/
Kalkatungu	/	s/	/	s/	/	/	/	/	/	/	s/P	/	/P	s/P	/	/
Karadjeri	/P	s/	/	s/S	/	/	/	/	/	/P	s/	/	/	s/	/	/
Kugu Nganhcara	/	s/P	/	/S	s/	/	/	/S	/	/P	/P	/	/S	s/P	/	/
Kuku-Yalanji	/P	s/	/	s/P	/	/	/	/	/	/P	/P	/	/P	/P	/P	/
Kuuk Thaayorre	/	s/P	/	/	/	/S	/	/S	/	/P	/P	/	/S	/P	/P	/
Kuuku Ya'u	/	s/	/	/	/	/	/	/S	/	/	/P	/	/S	/P	/	/
Madimadi	/	s/	/	s/P	/	/	/	/P	/	/P	/S	/	s/	/	/	/
Margany	/P	s/	/	s/	/	/	/	/	/	/P	/P	s/	/P	/P	/	/
Martuthunira	/	s/	/	s/	/	/	/	/P	/	/P	/S	s/	s/	/P	/	/
Muruwari	/P	s/	/	s/	/	/	/	/	/	/P	s/	s/	/S	s/P	/	/
Ngaanyatjarra	/	s/	/	s/S	/	/	/	/S	/	s/	/P	/	s/	/P	/	/
Ngadjumaja	/	s/	/	s/	s/	/S	/	/	/	/P	/P	s/	/	/P	/	/
Ngarinyeri	/	/	/	/	/	/	/	/	/	/	/	/	/	/S	/	/
Ngawun	/	s/	/	s/	/	/	/	/P	/	/P	/S	/	/	/P	/	/
Ngiyambaa	/	s/	/	s/	s/	/S	/	/P	/	/P	/P	s/	/S	/P	/	s/
Nhanda	/	s/	/	s/	/	/	/	/	/	/P	s/	s/	s/	/	/	/
Nyangumarta	/	/	/	/	/	/	/	/P	/	/P	/	/	/	/	/	/
Nyawaygi	/	/	/	s/	/	/	/	/	/	/	/P	/	/	/P	/	/
Paakantyi	/	s/	/	s/	/	/	/	/	/	/P	s/S	/	s/P	s/P	/	/
Panyjima	/	s/	/	s/	/	/	/	/P	/	/	/	/	/	/P	/	/
Pitjantjatjara	/	/	/	/S	/	/	/	/S	/	/	/P	/	/	/P	/	/
Pitta Pitta	/	s/	/	s/	s/	/	/	/	/	/P	/	/	/P	/	/	/
Ritharngu	/S	s/	/	s/	s/	/	/	/	/	s/	/	s/	s/	/	/	/
Uradhi	/	s/	/	s/	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Walmatjari	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Wangkumara	/	/	/	s/	/	/	/	/	/	/	/	/	/P	/P	/	/
Warlpiri	/	/	/	s/	/	/	/	/	/	/P	/	s/	/	/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Warluwara	/	/	/	/	/	/	/	/	/	/P	/P	/	/	/P	/	/
Warrgamay	/	s/	/	s/	/	/	/	/	/	/P	/	/	/	/	/	/
Warrnambool	/	s/	/	s/	/	/	/	/	/	/P	/S	/	s/	/S	/	/
Wathawurrung	/P	/	/	s/	/	/	/	/	/	/	/S	/	s/	/S	/	/
Watjarri	/	s/	/	/S	/	/	/	/S	/	/	/	/	s/	/	/	/
Wembawemba	/	s/	/	s/	/	/S	/	/P	/	/P	s/S	/	s/	/S	/	/P
Wik Munkan	/	s/	/	s/	/	/	/S	/S	/	/P	s/P	/	/S	s/P	/	/
Wik Ngathana	/S	/	/	s/	/	/	/	/S	/	/	/P	/	/	/P	/	/
Wirangu	/	s/	/	/S	/	/	/	/	/	/	/	/	/	/	/	/
Wulguru	/	s/	/	s/	/	/	/	/	/	/	/S	/	/	s/S	/	/
Yalarnnga	/	s/	/	s/	s/	/	/	/P	/	/P	/	/	s/	/	/	/
Yanyuwa	/	/	/	/	/	/	/	/P	/	/	/	p/	/P	/	/	/
Yaygir	/	s/	/	s/	/	/	/	/	/	/	/	/	/	/	/	/
Yidiny	/	s/	/	s/	/	/	/	/	/	/P	/P	/	/	/P	/	/
Yindjibarndi	/S	s/	/	s/	s/	/	s/	/P	/	/P	/S	s/	/P	/P	s/	/
Yingkarta	/	s/	/	s/	/	/	s/	/P	/	/P	/S	/	s/	/P	/	/
Yuwaalaraay	/P	s/	/	s/	/	/	/P	/P	/	/P	/P	s/	/	/P	/	/
REMBARNGA																
Rembarnga	/	s/	/	s/	/	/	/	/P	/	/	p/P	/	/	p/P	/	/
SOUTHERN DALY																
Ngan'gityemerri	/	/P	/	/S	/	/	/	/S	/	/	/P	/	/S	/	/	/
Ngankikurungkurr	/	/P	/	/S	/	/	/	/S	/	/P	/	/	/S	/P	/	/
TANGKIC																
Kayardild	/	s/	/	s/	s/	/	/P	/P	/P	s/	/	/	/P	/P	/	/
Yukulta	/P	/	/	s/	/	/	/	/P	/	/	/P	s/	/P	/P	/	/
TIWIAN																
Tiwi	/S	p/S	/	/P	s/	/	/	/P	/	/P	p/S	s/	/P	p/P	/P	/
WAGIMAN																
Wagiman	/	/	/	s/	/	/	/	/	/	/P	p/P	/	/	p/P	/	/
WARAY																
Waray (in Australia)	/	p/	/	/S	/	/	/	/	/	/	p/S	/	/	p/P	/	/
WEST BARKLY																
Djingili	/P	s/	/	s/S	/	/P	/	/	/	/P	/P	s/	/	/P	/	/
Wambaya	/	s/	/	s/	/	/	/	/P	/	/P	/S	/	/	/	/	/
WESTERN DALY																
Emmi	/	/	/	/S	/	/	/	/	/	/	s/	/	/	p/P	/	/
Maranungku	/	/	/	/S	/	/	/	/S	/	/P	/P	/	s/S	/P	/	/
Maringarr	/	s/	/	/	/	/	/	/	/	/P	p/P	/	/S	p/P	/	/
Marrithiyel	/	s/	/	/S	/	/	/	/	/	/P	p/P	/	/S	p/P	/	/
WORRORRAN																
Gunin	/	s/	/	/S	/	/	/S	/S	/	/	p/	/	p/P	p/	/	/
Ungarinjin	/	/S	/	/S	/S	/S	/S	/S	/	/P	/P	/	/S	/S	/P	/
Worora	/	/	/	/S	/	/	/	/	/	/P	/	/	p/	/	/	/
YANGMANIC																
Wardaman	/	/	/	s/S	/	/	s/	/P	/	/P	p/	s/	/P	p/	/	/
Awin-Pare																
AWIN-PARE																
Aekyom	/	s/	/	/	/	/	/	/S	/	/	/P	/	p/P	/P	/	/
Baibai-Fas																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
BAIBAI-FAS																
Fas	/	s/	/	/S	/	/	/	/	/	/	p/P	/	/S	s/P	/	/
FAS-BAIBAI																
Baibai	/	/	/	s/	/	/	/	/	/	/	/P	/	/	s/P	/	/
Border																
BORDER																
Amanab	/S	s/P	/	s/	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Awyi	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Imonda	/	s/S	/	/S	/	/S	/	/	/	/P	/P	/	/P	/P	/	/
Manem	/	s/	/	/	/	/	/	/	/	/	/P	/	/	s/P	/	/
Waris	s/	s/	/	s/	/	/	/	/S	/	/	/P	/	/	p/P	/	/
Bosavi																
BOSAVI																
Edolo	/	/	/	/S	/	/	/	/S	/	/	/P	/	/P	/P	/	/
Kaluli	/	/	/	/S	/	/	/	/	/	/	/P	/	/	s/P	/	/
Bulaka River																
BULAKA RIVER																
Maklew	/	p/	/	/S	/	/	/	/	/	/P	p/P	/	s/	s/P	/	/
Yelmek	/	s/P	/	/S	/	/	/	/	/	/P	/P	/	/	s/P	/	/
Dagan																
DAGAN																
Daga	/	s/	/	s/S	/	/	/	/S	/	/P	s/P	/	s/S	s/P	/	/S
East Bird's Head																
EAST BIRD'S HEAD																
Meyah	/S	p/	/	/P	/	/P	/S	/S	/S	/S	s/S	s/	p/P	p/P	/P	/
Moskona	/	/	/	/P	/	/	/	/S	/	/	/S	/	/	/P	/	/
Sougb	/	/	/	/P	/	/	/	/S	/	/S	s/S	s/	p/	p/P	/	/
East Geelvink Bay																
EAST GEELVINK BAY																
Bauzi	/	s/	/	/S	/	/	/	/	/	/S	/P	/	/	/P	/	/
East Strickland																
EAST STRICKLAND																
Nomad	/	s/	/	/	s/	/	/	/	/	/	/P	/	/P	/P	/	/
Eastern Trans-Fly																
EASTERN TRANS-FLY																
Meryam Mir	/	/P	/	s/	/	/	/P	/	/	/P	p/P	s/	/	p/P	/P	/
Eleman																
TATE																
Kaki Ae	s/	/	/	/	/	/	/	/P	/	/P	/P	/	/	s/P	/	/
Gapun																
GAPUN																
Gapun	/	s/	/	/S	/	/	/	/	/	/	s/P	/	/	/P	/	/
Gogodala-Suki																
GOGODALA																
Gogodala	/	s/	/	s/	/	/	/	/	/	/	s/P	/	/	s/P	/	/
SUKI																
Suki	/	s/	/	s/	s/	/	/	/P	/	/	s/P	s/	/P	s/P	/	/
Guriaso																
GURIASO																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Guriaso	/	/	/	s/	/	/	/	/	/	/	/P	/	/	p/P	/	/
Hatam																
HATAM-MANSIM																
Hatam	/S	/	/P	/P	/	/S	/	/S	/	/S	/S	/	p/P	p/P	/	/
Kayagar																
KAYAGAR																
Tamagario	/	s/	/	/	/	/	/	/	/	/	/P	/	/	s/P	/	/
Kiwaian																
KIWAIAN																
Kiwai	/	s/	/	s/S	/	/	/	/P	/	/	/P	s/	/	p/P	/	/P
Koiarian																
KOIARIAN																
Barai	/P	s/	/	s/S	/	/	/	/S	/	/P	s/P	s/	/	/P	s/	s/
Koiali (Mountain)	/	s/	/	s/S	/	/	/	/S	/	s/	s/P	s/	s/	s/P	s/	/
Koiari	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Koita	/S	s/P	/	s/	/	/	/	/S	/	s/	s/P	s/	/	s/P	/	/
Ömie	/	s/S	/	s/S	/	/	/	/P	/	/P	s/P	/	/	s/P	/	/
Kolopom																
KOLOPOM																
Kimaghama	/S	/S	/	/S	/	/	/	/	/	/S	/P	/S	/	/P	/	/
Riantana	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Kuot																
KUOT																
Kuot	/	/	/	/P	/	/	/	/P	/	/P	/S	/P	/S	/S	/	/
Kwerba																
KWERBA																
Isirawa	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Kwerba	/	/	/	/S	/	/	/	/	/	/P	/P	/	/P	/	/	/
Kwomtari																
KWOMTARI																
Kwomtari	/	s/	/	/S	/	/	/P	/P	/	/P	s/P	s/	/	s/P	/	/
Nai	/	s/	/	s/	/	/	/	/	/	s/	s/P	/	/P	s/P	/	/
Lakes Plain																
LAKES PLAIN																
Iau	/	/	/	/S	/	/	/	/S	/	/S	/P	/	/S	/P	/	/
Sikaritai	/	s/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Left May																
LEFT MAY																
Owininga	/	/	/	/S	/	/	/	/	/	/	/P	/	/	/P	/	/
Mairasi																
MAIRASI																
Mairasi	/	s/	/	/	/	/	/	/	/	s/	s/P	/	p/	s/P	/	/
Marind																
MARIND																
Boazi	/	/	/	/S	/	/	/	/	/	s/	/P	/	p/	p/P	/	/
Marind	/	p/	/	/S	/	/	/	/	/	/P	p/P	/	p/	p/P	/	/
Yaqay	/P	p/	/	/S	/	/	/	/	/	/	p/P	/	p/	p/P	/	/
Zimakani	/	p/	/	/	/	/	/	/	/	/	p/P	/	/	p/P	/	/
Maybrat																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
MAYBRAT																
Maybrat	/S	/	/	p/P	/	/	/S	/S	/	/S	/S	/	p/S	p/P	/P	/
Mombum																
MOMBUM																
Mombum	/	s/	/	/S	/	/	/	/	/	/	/P	/	/	s/P	/	/
Moraori																
MORAORI																
Moraori	/S	s/	/	/S	/	/	/	/	/	/P	p/P	/	/	s/P	/	/
Morehead-Upper Maro																
MOREHEAD-UPPER MARO																
Kanum (Bädi)	/	s/	/	/S	/	/	/	/	/	/P	p/P	/	/	s/P	/	/
Yei	/	/	/	/S	/	/	/	/	/	/P	/P	/	/	s/P	/	/
Mpur																
MPUR																
Mpur	/S	/	/	/P	/	/	/	/S	/	/S	/S	/	p/P	p/P	/	/
Nimboran																
NIMBORAN																
Nimboran	/	s/	/	/S	/	/	/	/P	/	/P	s/P	/	p/P	s/P	/	/
North Bougainville																
NORTH BOUGAINVILLE																
Rotokas	/	s/	/	/S	s/	/	/	/P	s/	/P	/P	s/	s/S	s/P	/P	/
Pauwasi																
EASTERN PAUWASI																
Karkar-Yuri	/	/	/	/S	/	/S	/	/P	/	/	/P	/	/	/P	/	/
Pawaian																
PAWAIAN																
Pawaian	s/	s/	/	/S	/	/P	/	/P	/	/	/P	/	/P	s/P	/	/
Piawi																
PIAWI																
Haruai	/	s/	/	/S	/	/	/	/	/	s/	/P	/	/S	s/P	/	/
Ramu-Lower Sepik																
ANNABERG																
Rao	/	s/P	/	/	/	/	/	/	/	s/	/P	/	/P	/P	/	/
BOTIN																
Kambot	/	s/	/	/S	/	/	/	/	/	s/	/P	/	/P	/P	/	/
LOWER RAMU																
Kaian	/	/	/	/	/	/	/	/	/	/	/	s/	/P	/	/	/
Watam	s/	s/	/	/S	/	/	/	/	/	p/	/P	/	/	/P	/	/
LOWER SEPIK																
Chambri	/	s/	/	/	/	/	/	/	/	/	/P	s/	/	p/P	/	/
Yimas	/	s/	p/	s/S	/	s/	/	/S	/	p/	p/P	s/	/	p/P	/	/
MIKAREW																
Kire	/	/	/	/S	/	/	/	p/	/	/S	/P	s/	/	/P	/	/
Mikarew	/	s/	/	/	/	/	/	/	/	/	/	s/	/	/	/	/
Senagi																
SENAGI																
Anggor	s/	/S	/	/S	/	s/	/	/S	/	s/	s/P	/	/P	s/P	/	/
Dla (Menggwa)	/	s/S	/	/S	/	/S	/	/P	/	/	s/P	/	/P	s/P	s/	/
Dla (Proper)	/	s/	/	/S	/	/	/	/	/	/	/P	/	/P	/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Sentani																
SENTANI																
Sentani	/	s/	/	/S	/	/	/	/P	/	p/	s/P	/	/	s/P	/	/
Tabla	/	s/	/	/S	/	/	/	/P	/	/	s/P	/	/P	s/P	/	/
Sepik																
ABAU																
Abau	/	/	/	/S	/	/	/	/S	/	/S	/P	/	/P	/P	/	/
IWAM																
Iwam	/S	p/	/	/	/	/	/	/	/	/S	/P	/	/	/P	/	/
NDU																
Ambulas	/P	s/	/	/S	p/	/S	/	/P	/	/	/P	/	/P	/P	/	/S
Boiken	/	s/	/	s/	/	/	/	/	/	/	/	/	/	/	/	/
Hanga Hundi	/	/	/	/S	/	/	/	/P	/S	/	/P	s/	s/	s/P	/	/
Iatmul	/	s/	/	/	/	/	/	/P	/	/	/P	/	/P	s/P	/	/S
NUKUMA																
Kwanga	/S	s/	/	/S	/	/	/P	/P	/	/S	/P	/S	/P	/P	/	/
Kwoma	/	s/P	/	/S	/	/S	/S	/P	/	s/	/P	/S	/P	/P	s/	/S
Mende (Papua New Guinea)	/	/	/	/S	/	/	/	/S	/	/	/P	/	/	/P	/	/
RAM																
Awtuw	/S	s/	/	/S	/	/	/P	/P	/S	p/	/P	s/	/P	/P	/S	/
SEPIK HILL																
Alamblak	/	s/	/	/S	p/	/S	/P	/P	/	/P	s/P	/	/P	s/P	/	/
Sare	/	/S	/	/S	/	/	/	/P	/	/	/P	/	/P	/P	/P	/
TAMA SEPIK																
Mehek	/	s/	/	s/	/	/	/	/	/	/	/	/	/	s/	/	/
WOGAMUSIN-CHENAPIAN																
Wogamusin	/	s/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
YELLOW RIVER																
Namia	/S	/	/	/S	/	/	/S	/S	/S	p/	s/P	/	/P	/P	/S	/P
Skou																
KRISA																
I'saka	/	/S	/	/	/	/	/	/S	/	/S	/P	/	/S	p/P	/	/
SERRA HILLS																
Poko-Rawo	/S	/	/	/	/	/	/	/S	/	/	/	/	/	/P	/	/
Womo	/	/	/	s/	/	/	/	/S	/	/S	s/P	/	/P	p/P	/	/
WARAPU																
Barupu	/P	/	/	/	/	/P	/S	/S	/	/P	s/P	/	/S	p/P	/	/
WESTERN SKOU																
Dumo	/S	/	/	/	/	/P	/	/S	/	/S	/P	/	/S	p/P	/	/
Skou	/S	/S	/	/S	/	/	/S	/S	/	/S	p/P	p/	s/	p/P	/	/
Wutung	/S	/	/	/P	/P	/P	/	/S	/S	/S	/P	/	/S	p/P	/	/
Solomons East Papuan																
BILUA																
Bilua	/	s/P	/	/S	/S	/	/P	/P	/P	/P	s/S	/S	p/	/	/	/P
LAVUKALEVE																
Lavukaleve	/S	s/S	s/	s/S	/	/	/S	/	/S	/	p/P	s/	p/	/P	/	/
SAVOSAVO																
Savosavo	/	s/	/	/S	/	s/	/P	/P	/	/P	/P	/	/P	/P	/S	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
South Bird's Head																
SOUTH BIRD'S HEAD																
Inanwatan	/S	s/	/	/S	/	/	/S	/	/	s/	p/	s/	p/P	p/P	/	/
South Bougainville																
SOUTH BOUGAINVILLE																
Buin	/	s/	/	s/	/	/	/	/	/	/P	/P	s/	p/	s/	/	/
Motuna	/	s/	/	/	s/	/	/P	/P	/	s/	s/P	s/	p/P	s/P	/	/
Nasioi	/	s/	/	s/S	/S	/S	/	/P	/	/	s/P	/	p/	s/P	/	/S
Sulka																
SULKA																
Sulka	/	p/S	/	/P	/	/	/	/S	/	p/	/S	p/	p/S	p/P	/P	/
Tasmanian																
TASMANIAN																
Tasmanian	/	/	/	/S	/	/	/	/	/	/	/	/	/	/	/	/
Teberan																
TEBERAN																
Folopa	s/	s/	/	/S	/	/	/P	/P	/P	s/	/P	s/	/P	/P	/	/
Timor-Alor-Pantar																
EAST PANTAR																
Teiwa	/	/	/S	/S	/	/S	/S	/S	/	/S	p/P	/S	p/	/P	/	/P
GREATER ALOR																
Abui	/	s/	/	/	/	s/	/S	/P	/	/S	p/P	/S	p/	/P	/	/
Adang	/	/S	/	/S	/	/P	/S	/S	/	/S	p/P	/	p/P	/P	/	/
Klon	/S	/	/S	/	/	/	/	/S	/	/S	p/P	/	p/	/P	/	/
Woisika	/	/	/	/	/	/	/	/S	/	s/	/P	/	/	/P	/	/
MAKASAE-FATALUKU-OIRATA																
Fataluku	/	/	/	/	/	/	/	/S	/	/	/	/	/	/	/	/
Makasae	/S	/P	/P	/S	/	/	/	/S	/	/P	/P	/	/P	/P	/P	/P
Tor-Orya																
ORYA																
Orya	/	/P	/	/S	p/	/	/S	/S	/	/	/P	/	/P	s/P	/	/
Torricelli																
KOMBIO-ARAPESH																
Arapesh (Abu)	/S	/	/	/	/	/	/	/S	/P	/S	/S	/	/	p/P	/	/
Arapesh (Mountain)	/S	p/P	/	/P	/	/	/	/P	/P	/P	/S	s/	/P	p/P	/	/
Kombio	/	/P	/	s/P	/	s/	/	/S	/S	/P	/S	/	/S	/P	/	/
Mufian	/S	p/	/	/	/	/	/	/S	/P	/	s/S	/P	/S	p/P	/	/
Urat	/S	/P	/	/P	/	/	/	/S	/S	/	/S	/P	/S	p/P	/	/
MARIENBERG																
Kamasau	/	/S	/	/S	/	/	/	/S	/	/S	s/P	/	/P	p/P	/	/
MONUMBO																
Monumbo	/	/	/	/S	/	/	/	/	/	/	s/P	/	/	p/P	/	/
WAPEI-PALEI																
Au	/	/	/	/P	/	/	/	/S	/	/	s/S	/	/S	p/P	/P	/P
Olo	/S	s/	/P	/P	/	/	/	/S	/	/S	s/S	s/	/S	p/P	/P	/
Walman	/S	/	/	/S	/	/	/	/S	/S	/P	p/S	s/	/	p/P	/P	/
WEST WAPEI																
One	/	s/	/	s/	/	/	/	/	/	/S	/S	s/	/	/P	/	/
Trans-New Guinea																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
ANGAN																
Baruya	/	s/	/	s/S	/	s/	/S	/S	/S	/	p/P	/	p/P	s/P	/	/
Hamtai	/S	/S	/	/S	/	/	/	/S	/	p/	p/P	/	p/	s/P	/	/
Menya	/S	/	/	/S	/	/	/	/S	/	/	/P	/	/P	/P	/	/
ASMAT-KAMORO																
Asmat	/S	s/S	/	/	s/	/	/P	/P	/	/S	s/P	s/	/P	s/P	/S	/
Kamoro	p/	/	/	s/	/S	/	/	/	/	s/	s/P	/	p/	s/P	/	/
AWJU-DUMUT																
Kombai	/S	s/S	/	/S	/	/	/	/S	/	s/	/P	/	p/	s/P	/S	/
Korowai	/S	s/S	/	s/	/	/S	/	/P	s/	/	/P	s/	p/	s/P	/	/
Pisa	/S	s/	/	/S	/	/	/	/	/	/	/P	/	/	s/P	/	/
Wambon	/	s/S	/	/S	/	/	/	/P	/	s/	/P	s/	p/	s/P	/S	/
BINANDEREAN																
Binandere	/	s/	/	/S	/	/	/S	/S	/S	/	/P	/	/	s/P	/	/
Guhu-Samane	/S	s/	/	s/	/	/	/	/	/	/	/P	s/	/	/P	/	/
Orokaiva	/	s/	/	/S	/	/S	/	/S	/	/	/P	/	/	s/P	/	/
Suena	s/	s/	/	s/S	/	/	/	/S	/	/P	/P	/	/S	s/P	/P	/
CHIMBU																
Dom	/	s/S	/	/S	/	/S	/	/S	/	/	/P	/	s/	s/P	/	/
Golin	s/	s/	/	/S	/	/S	/	/S	/	s/	/P	/S	s/	s/P	/	/S
Kuman	s/	s/	/	/S	/	/	/	/P	/	s/	/P	/	s/P	s/P	/	/
Salt-Yui	s/	s/	/	/S	/	/S	/	/	/	s/	/P	/	s/	s/P	/	/
Tabare	/	s/	/	/	/	/	/S	/	/S	s/	/P	/	s/	s/P	/	/
Umbu Ungu	/	s/	/	/S	/	/	/	/	/	/	/P	/	/P	s/P	/	/
Wahgi	/S	s/	/	s/S	/	/S	/	/S	/S	/P	/P	/	s/P	s/P	/	/
DANI																
Dani (Lower Grand Valley)																
	s/	s/	/	s/S	/	/	/	/S	/	/S	/P	/	p/	s/P	/	/
DUNA																
Duna	/S	s/	/	/S	/	/	/	/S	/	/	/P	/	/P	/P	/	/
EAST KUTUBU																
Foe	s/	/	/	/	/	s/	/	/S	/	/	/P	/	/	/P	/	/
ENGAN																
Kewa	/S	s/	/	/S	p/	/S	/	/P	/	p/	/P	s/	/P	s/P	/S	s/
Kyaka	s/	s/	/	/S	/	/	/S	/S	/S	/	/P	/S	/	s/P	/	/
FASU																
Fasu	s/	s/	/	/S	/	/	/S	/S	/	/	/P	/	/P	/P	/	/
FINISTERRE-HUON																
Borong	/	/	/	/	/	/	/	/S	/	/	/P	/	/	/P	/	/
Burum	/	s/	/	s/	/	/	/	/	/	/	p/	/	s/	s/	/	/
Iyo	/	/	/	/S	/	/	/	/S	/	/	/P	/	/P	/P	/	/
Komba	/	/	/	/	/	/	/	/S	/	/	/P	/	/P	/P	/	/
Mungkip	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Nabak	/	s/	/	/S	p/	/	/	/S	/S	/P	/P	s/	s/P	s/P	/	/
Nankina	/	/	/	/S	/	/	/	/	/	/S	/P	/	s/	/P	/	/
Ono	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Rawa	/	s/	/	/	/	/	/	/	/	/P	/P	/	s/	s/P	/	/
Selepet	/	s/S	/	s/S	/	/	/	/S	/	/	s/	s/	s/	s/	/	/S
GOILALAN																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Kunimaipa	/	s/	/	s/	/	/	/	/S	/	p/	p/P	/	/	s/P	s/	/
KAINANTU-GOROKA																
Agarabi	s/	s/	/	s/	/	/	/	/	/	/	p/P	/	p/	s/P	/	/
Awa	/	s/	/	s/	/	/	/	/P	/	/P	/P	s/	p/	s/P	/	/
Fore	/	s/	/	/S	/	/	/	/P	/	/	p/P	s/	/P	s/P	/	/
Gadsup	s/	s/	/	s/	/	/	/	/	/	/P	p/P	s/	p/	s/P	/	s/
Gahuku	s/	s/	/	s/	/	/	s/	/	s/	s/	p/P	s/	/	s/P	s/	/
Hua	s/	s/	/	s/S	/	/	/	/P	/	p/	p/P	/	/	s/P	/	/
Kanite	/	/	/	/	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Kosena	/	/	/	/	/	/	/	/P	/	/	/	/	/	/	/	/
Siane	s/	s/	/	s/	/	/S	/	/	/	s/	/P	/	s/P	s/P	s/	/
Tairora	/	s/S	/	s/	/	/	/	/P	/	/P	s/P	s/	p/	s/P	s/	/
Usarufa	s/	s/S	/	s/	/	/S	/	/P	/	p/	p/P	s/	/	s/P	/	/S
Yagaria	/S	/	/	/S	/	/S	/	/	/S	p/	p/P	/	/P	s/P	/	/
MADANG																
Amele	/S	s/S	/	/S	/P	/S	/	/S	/S	/	s/P	/	s/P	s/P	/S	/
Anamuxra	/	s/	/	/	/	s/	/	/S	/S	/	p/P	s/	p/P	s/P	/	/
Bau	/	s/	/	/	/	/	/	/	/	/	s/	/	s/	s/	/	/
Bongu	/S	/	/	/S	/	/	/	/	/	/	/P	/	/P	/P	/	/
Garus	/	s/	/	/	/	/	/	/	/	/P	s/	/	s/	s/	/	/
Girawa	/	/	/	/S	/	/	/	/P	/	/	/P	/	s/P	/P	/	/
Kalam	/	s/	/	/S	/	/	/	/S	/	/	/P	/	/S	s/P	/	/
Kobon	/S	s/	/	/S	/	/	/	/S	/S	s/	/P	s/	p/S	s/P	/	/
Maia	/	/	/	/S	/	/	/	/S	/	/	/P	/	/	/P	/	/
Mauwake	/S	s/S	/S	/S	/	/S	/S	/S	/	/P	/P	/	p/P	s/P	/	/S
Mugil	/	/	/	/	/	/	/	/	/	p/	p/	/	p/	s/	/	/
Nend	s/	/	/	/S	/	/	/	/S	/	/	/P	/	/	s/P	/	/
Siroi	/S	s/S	/S	/S	/S	/S	/	/S	/	/S	s/P	/S	/	s/P	/S	/S
Tauya	/S	s/	/	/S	/	/	/	/P	/S	/P	p/P	/	p/S	s/P	/	/
Usan	/	s/	/	/S	/	/	/S	/S	/	/P	p/P	/	/P	s/P	/	/
Waskia	/	s/	/	/S	/	/	/S	/S	/S	/P	s/P	/	p/	s/P	/	/S
MAILUAN																
Magi	/	s/	/	/S	p/	/	/	/P	/	/P	s/P	/	/	s/P	/	/
MEK																
Eipo	s/	s/	/	/S	/	/S	/P	/	/	/	s/P	s/	/P	s/P	/	/
Una	s/	s/	/	/S	/	/	/	/P	/	/P	/P	/	p/	s/P	/	/
Yale (Kosarek)	s/	s/	/	/S	/	/	/	/S	/	/	s/P	/	/P	s/P	/	/
OK																
Bimin	/	/	/	/S	/	/	/S	/S	/S	/S	/P	/	/	/P	/	/
Indonesia)	/	s/	/	/S	/	/	/	/	/	/S	s/P	/	/	s/P	/S	/
Kati (in West Papua	/	s/	/	/S	/	/	/	/	/	/S	s/P	/	/	s/P	/S	/
Mian	/S	s/	/	/S	/	/S	/S	/S	/	/	/P	/	/P	s/P	/S	/
Telefol	/	/	/	/S	/	/	/	/	/S	/	p/P	s/	/P	s/P	/	/
Tifal	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
OKSAPMIN																
Oksapmin	/	s/	/	/S	p/	/	/S	/S	/S	/P	p/P	s/	/P	/P	/S	/
PANIAI LAKES																
Ekari	/	s/	/	/	/	/	/	/	/	/	p/P	/	/	s/P	/	/
Moni	s/	s/	/	/	/	/	/	/	/	/P	/P	/	/	s/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
SOMAHAI																
Momuna	/	s/	/	s/S	/	/	/	/	/	/	/P	/	/P	s/P	/	/
WIRU																
Wiru	/S	s/	/	/S	s/	/	/	/	/	s/	/P	/	/P	s/P	/	/
Turama-Kikorian																
TURAMA-KIKORIAN																
Rumu	/S	s/	/	/S	s/	/S	/	/S	/	s/	/P	/	/	/P	/	/
West Bomberai																
WEST BOMBERAI																
Baham	/	/	/	/	/	/	/	/S	/	/	/P	/	/	/P	/	/
Iha	/	s/	/	/	/	/	/	/	/	/	/P	/	/	s/P	/	/
West Papuan																
NORTH HALMAHERAN																
Pagu	/	s/	/	/	p/	/	/P	/S	/P	s/	p/P	/	/P	p/P	/	/
Sahu	/	/	/	/P	/	/	/	/S	/	/S	p/S	/	/P	p/P	/P	/
Ternate	/	/	/	/P	/	/	/	/	/	/	/	/	/P	p/	/	/
Tidore	/	/P	/S	/P	p/	/	/P	/S	/S	/S	/S	/	p/P	p/P	/	/S
Tobelo	/	/	/	s/P	/	/	/	/S	/	s/	p/P	/	p/	p/P	/P	/
West Makian	/	/	/	/P	/	/	/	/S	/	/S	/S	/	p/	p/P	/P	/P
Yale																
YALE																
Nagatman	/S	s/	/	/S	/	/	/	/	/	/	s/P	s/	p/	s/P	/	/
Yareban																
YAREBAN																
Yareba	/S	s/	/	s/S	/P	/	/	/	/	s/	/P	s/	/P	s/P	/	s/
Yawa																
YAWA																
Yawa	/S	/	/	/	/	/	/	/S	/	/	p/P	/	p/	/P	/	/
Yelî Dnye																
YELî DNYE																
Yelî Dnye	/	p/	/	/S	/	/	/	/S	/	/	/P	/S	p/P	p/P	/	/

North America																
Algic																
ALGONQUIAN																
Arapaho	/	/	/	/	/	/	/	/P	/	/	/S	/	/	/S	/	/
Blackfoot	p/	p/	/	/	s/	/	s/	/P	s/	p/	/S	s/	p/	/	/	/P
Cree (Eastern Swampy)	/	/	/	s/	/	/	/	/	/	/P	/	/	/	/	/	/
Cree (Plains)	/	p/	/	s/	/	/	/	/P	/	/P	/S	s/	p/	/	/	p/
Massachusett	s/	/	/	/P	/	/	/	/P	/	/	/	s/	p/	/	/	/
Menomini	s/	p/	/	s/	/	/	/	/P	/	/	/	s/	p/P	/	/P	/P
Ojibwa (Eastern)	/	/P	/	s/	/	/S	/P	/P	/	/P	/S	s/	/	/S	/P	/
Passamaquoddy-Maliseet	/	/P	/	s/P	/	/	/P	/P	/	/	/S	s/	/P	/	/P	/
WIYOT																
Wiyot	/	/P	/	s/	s/	/	/P	/P	/	/P	s/	/	p/	s/P	/	/
YUROK																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Yurok	/	/P	/	s/	/	/	/P	/P	/	/P	s/	s/	p/P	/P	/P	/P
Atakapa																
ATAKAPA																
Atakapa	/	s/	/	/S	/	/	/	/P	/	/	p/P	/	/	s/P	/	/
Caddoan																
CADDOAN																
Pawnee	/	s/	/	s/	s/	s/	/	p/	/	p/	p/	s/	p/	p/	/	/
Wichita	/P	/	/	s/	s/	/S	/S	/P	/	/P	p/P	/	p/	p/	/P	/P
Chimakuan																
CHIMAKUAN																
Quileute	s/	/S	/	s/	s/	/	/P	/P	/P	/P	/S	p/	s/	/S	/P	s/
Chitimacha																
CHITIMACHA																
Chitimacha	/S	s/P	/	s/S	s/	/	/	/	/	s/	s/P	s/	/P	s/P	/S	/
Chumash																
CHUMASH																
Chumash (Barbareño)	s/	s/P	/	/P	p/	/	/	p/	p/	p/	s/S	/	p/	p/S	/	p/
Coahuiltecan																
COAHUILTECAN																
Coahuilteco	s/	s/	/	s/	/	/	/	/	/	s/	p/P	/	p/	p/P	/	/
Eskimo-Aleut																
ALEUT																
Aleut	/	s/	/	s/	/	/	/	/	/	/	/P	s/	s/	s/P	/	/
ESKIMO																
Greenlandic (West)	s/	s/	/	s/S	s/	/	/	/	/	s/	s/P	s/	s/	s/P	/	s/
Iñupiaq	/	s/	/	s/	/	/	/	/	/	/	s/P	s/	s/	s/P	/	/
Yup'ik (Central)	/	s/	/	s/	s/	/	/	/	/	s/	s/	s/	s/	s/	s/	s/
Yupik (Siberian)	/	/	/	/	/	/	/	/S	/	/	/P	/	/	/P	/	/
Esselen																
ESSELEN																
Esselen	/	/	/	s/	/	/	/	/	/	/	/	/	p/	/	/	/
Haida																
HAIDA																
Haida	/	s/S	/S	/S	s/	/	s/	/P	/	/	/P	s/	/	/P	/S	/S
Hokan																
CHIMARIKO																
Chimariko	s/	s/	/	s/S	s/	/	s/	/	/	s/	p/P	/	/	p/P	/	/S
PALAIHNIHAN																
Achumawi	/	s/P	/	s/P	/	/	/	/	/	/P	/S	/	/	p/P	/	/
Atsugewi	/	/	/	s/	/	/	p/	/P	/	/	/	/	/	p/	/	/
POMOAN																
Pomo (Eastern)	/	s/	/	s/	s/	s/	/	/	/	s/	/P	s/	p/	s/P	s/	/
Pomo (Northern)	/	/	/	/S	/	/S	/	/	/	/	/P	/	/	/P	/	/S
Pomo (Southeastern)	s/	s/	/	s/S	/	/	/	/P	/	/	/P	s/	/	/P	s/	/
SERI																
Seri	/S	p/S	/	/S	/	/	/S	/S	/S	p/	p/P	s/	p/	p/P	/S	/
SHASTA																
Shasta	/	/	/	/	/	/S	/	/	/	/	/S	/	/	p/	/	/
YUMAN																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Cocopa	/	s/	/	s/	p/	/	s/	/	/	/	p/P	s/	p/	p/P	/	/
Diegueño (Mesa Grande)	s/	/S	/	/S	/	/	/S	/S	/	/S	p/P	/	p/	p/P	/	/S
Hualapai	/	s/S	/	/S	/	/	/	/	/	/S	p/P	s/	p/P	/P	s/	/S
Kiliwa	s/	p/S	/	s/	/S	/S	/	s/	s/	/S	p/P	s/	p/S	p/P	/	/
Maricopa	s/	s/S	/	/S	p/	/S	/S	/S	/	/	p/P	/	p/S	p/P	/	/S
Mojave	/S	s/S	/	/S	/	/	/	/	/	s/	/P	/	p/	p/P	/	/
Tiipay (Jamul)	/S	s/S	/	/S	p/	/S	/S	/S	/	/S	p/P	/	p/	p/P	/S	/
Huavean																
HUAVEAN																
Huave (San Mateo Del Mar)	/	p/P	/	/P	/	/	/S	/S	/	/P	s/S	p/	p/S	p/P	/P	/
Iroquoian																
NORTHERN IROQUOIAN																
Mohawk	/	/	/	s/	s/	/	/	/P	/	/P	p/	p/	p/	p/	/P	/
Oneida	/	/	/	s/	/	/	/P	/P	/	/P	p/S	/	p/	p/S	/P	/
Seneca	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Tuscarora	/	/	/	s/	s/	/	/	/	/	/P	p/	p/	p/	p/	/	/
Wyandot	/	/	/	/	/	/	/	/	/	/P	p/S	/	p/	p/	/	/
SOUTHERN IROQUOIAN																
Cherokee	/	s/	/	/S	/	/S	/	/P	/	/P	p/P	p/	p/	p/	/	/
Karankawa																
KARANKAWA																
Karankawa	/	/P	/	/	/	/	/	/P	/	/	/	/	/P	/	/	/P
Karok																
KAROK																
Karok	/	s/	/	/S	/	/	/P	/P	/	/	p/	s/	p/	p/P	/	/
Keresan																
KERESAN																
Acoma	/	s/P	/	/S	/	s/	/	/P	/	/	p/	s/	p/	/P	/P	/
Keresan (Santa Ana)	/P	s/S	/	/S	/	/	/	/P	/	/P	p/P	/	p/	p/P	/	/
Kiowa-Tanoan																
KIOWA-TANOAN																
Kiowa	/P	s/P	/	s/	s/	/	/	/P	/	/	p/P	s/	p/	p/P	s/	/
Tiwa (Northern)	/	s/	/	s/	/	/	/	/	/	p/	p/	s/	p/	p/P	/	/
Kutenai																
KUTENAI																
Kutenai	/P	/P	/P	/	s/	/P	/P	/P	/	/P	s/S	p/	s/	s/S	/P	/P
Mayan																
MAYAN																
Cakchiquel	/	/	/	/P	/	/	/	/P	/	/	/S	/	/	/S	/	/
Chontal Maya	/	s/P	/	/P	/	/	/P	/	/	/P	s/S	s/	s/P	/P	/P	/
Chuj	/	p/P	/	/P	/	/	/	/S	/	/P	p/S	/	p/	p/S	/	/
Huastec	/	/P	/P	/P	s/	/P	/P	/P	/P	/P	/S	s/	/P	/P	/P	/P
Ixil	/	p/	/	/P	/	/	/	/	/	/	/S	/	p/	p/S	/	/
Jakaltek	/	/P	/	/P	/P	/	/	/S	/P	/P	p/S	/P	p/	p/S	/P	/P
Lacandón	/S	s/P	/	/P	/	/	/	/P	/	/P	s/S	/	/P	s/	/P	/
Mam	/	p/P	/	/P	s/	/	/	/P	/P	/P	/S	/	p/	/S	/P	/P
Sipakapense	/P	p/	/	/P	/	/	/P	/P	/P	/P	p/S	/P	p/	p/S	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Tzotzil	/P	/	/	/P	/	/	/	/	/	/P	/S	s/	p/	/S	/P	/
Tzutujil	/P	/	/P	/P	s/	/P	/P	/	/P	/P	p/S	/P	p/	p/	/P	/P
Mixe-Zoque																
MIXE-ZOQUE																
Mixe (Ayutla)	/S	/	/	/	/	/	/	/P	/	/	/	/	/	/	/	/
Mixe (Coatlán)	/P	s/	/	/P	/	/	/	/P	/	/P	/	/S	p/	/	/P	/
Zoque (Chimalapa)	/	s/	/	/S	/	/	/P	/P	/	/P	/S	/	/	/P	/	/
Zoque (Copainalá)	s/	s/P	/	s/	/	/	/P	/P	/	/	p/S	s/	p/P	p/S	/	/P
Zoque (Ostuacan)	/	s/P	/	s/	p/	/	/P	/	/	/P	p/S	/	p/	p/	/P	p/
Zoque (Soteapan)	/	s/P	/	s/P	p/	/	/P	/P	/	/P	p/S	/	p/	p/P	/	/
Muskogean																
MUSKOGEAN																
Alabama	/	/	/	/S	/	/	/	/	/	/	/P	s/	/	/P	/	/
Chickasaw	s/	s/	/	/S	/	/	/	/	/	/	/P	/	p/	/P	/	/
Choctaw	/S	s/S	/	/S	s/	/S	/	/S	/	s/	p/P	/	p/	p/P	s/	/
Creek	s/	s/S	/S	/S	/	/	/	/P	/	s/	p/P	s/	p/	s/P	/	/
Koasati	/	/S	/	/S	/	/	/S	/P	/	s/	p/P	/	p/	s/P	/	/
Mikasuki	/	s/	/	s/	/	/	/	/	/	s/	p/P	/	p/	s/P	s/	/
Seminole	s/	s/	/	/S	s/	/	/	/P	/	s/	p/P	s/	p/	s/P	/	/S
Na-Dene																
ATHAPASKAN																
Apache (Western)	s/	p/	/	s/	/	/S	s/	/P	/	/	p/P	/	/	p/P	/	/
Chipewyan	/S	p/S	/	/S	/	/S	/	/P	/	/S	p/P	/	p/	p/P	/S	/
Hupa	/	/	/	s/S	p/	/	/P	/P	/	/P	p/	/	p/	p/	/	/
Koyukon	/	/	/	/S	/	/	/	/P	/	/	/P	/	/P	/P	/S	/
Navajo	/	p/S	p/	/S	/	/S	/S	/P	/S	/P	p/P	s/	p/	p/P	/P	/
Sarcee	s/	/	/	/S	/	/S	/	/P	/	/P	p/P	s/	p/	p/P	/	/
Slave	/P	p/S	/P	/S	p/	/S	/	/P	/	/S	p/P	s/	p/	p/P	/	/S
Tanacross	/	p/	/	/S	p/	/	/	/P	/	/	p/P	s/	p/	p/P	/S	/
TLINGIT																
Tlingit	/	/S	/	s/S	/	/	/	/P	/	/P	p/P	s/	p/	p/P	/S	/
Natchez																
NATCHEZ																
Natchez	/	/	/	/S	/	/	/	/	/	/	/P	/	/	/P	/	/
Oregon Coast																
ALSEA																
Alsea	/	/	/	/P	/	/	/	/P	/	/P	s/S	/	/	s/S	/	/P
COOSAN																
Coos (Hanis)	/S	s/	/	/S	s/	/	/P	/P	/	/P	p/	s/	/P	/S	/P	/
SIUSLAWAN																
Siuslaw	/	s/	/	/	/	/	/	/P	/	/P	s/S	s/	s/	s/	/P	/
Oto-Manguean																
CHINANTECAN																
Chinantec (Comaltepec)	/P	p/S	/	/P	p/	/	/P	/S	/P	p/	/S	/P	/S	/S	/P	/
Chinantec (Lealao)	/P	p/	/	/P	/	/	/	/S	/	/P	/S	/P	/S	/S	/P	/
Chinantec (Palantla)	/	p/	/	/P	/	/	/S	/S	/	p/	s/S	/	/S	s/S	/P	/
Chinantec (Quiotepec)	/P	p/	/	/P	/	/S	/	/	/	/P	/S	/	/S	s/S	/P	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
MIXTECAN																
Mixtec (Atatlahuca)	/	/	/	/P	/	/	/	/	/	/	/	/	/	/	/	/
Mixtec (Chalcatongo)	/	p/	/	/P	p/	/P	/	/S	/	/P	/S	/S	/S	p/S	/P	/
Mixtec (Jicaltepec)	/S	p/P	/P	/P	/	/S	/	/S	/	/P	/S	/	s/	s/S	/P	/
Mixtec (Ocotepec)	/P	p/P	/	/P	p/	/S	/	/S	/	/P	/S	/P	/S	/S	/P	/P
Mixtec (Peñoles)	/S	p/P	/P	/P	p/	/P	/	s/	/	/P	/S	/P	s/	/S	/P	/P
Mixtec (Yosondúa)	/S	/	/	/P	/	/	/	/S	/P	/	/S	/P	/S	/S	/	/
Trique (Copala)	/S	p/	/	/P	/	/	/	/S	/P	/P	/S	/	/P	s/S	/P	/
OTOMIAN																
Ocuilteco	/P	p/	/	p/P	s/	/	/	/P	/P	p/	s/S	p/	p/	p/P	/P	/
Otomí (Mezquital)	/P	p/P	/	/P	/P	/P	/	/P	/	/P	s/S	/	/P	p/S	/P	/P
PAMEAN																
Pame	/	p/P	/	/	/	/	/	/P	/	/	s/S	/	/P	p/P	/	/
POPOLOCAN																
Mazatec (Chiquihuitlán)	/	/	/	/	/	/	/	/	/P	s/	/S	/	s/	/S	/	/
Mazatec (Huautla)	/P	p/	/	/P	/	/S	/	/	/	/	s/	/	s/	s/	/P	/
Popoloca (San Juan Atzingo)	/	/	/	/	/	/	/	/	/	s/	/	/	/	/	/	/
SUBTIABA-TLAPANEC																
Tlapanec	/P	p/S	/	s/P	/	/S	/	/S	/	p/	s/S	/	s/	/S	/	/S
ZAPOTECAN																
Chatino (Sierra Occidental)	/S	p/P	/	/P	/	/S	/P	/S	/	/P	/S	/	/S	/S	/P	/
Chatino (Yaitepec)	/	p/	/	/P	/	/	/	/S	/	/P	/S	/	/S	/S	/P	/
Zapotec (Isthmus)	/S	p/	/	/P	p/	/P	/	/S	/P	/P	/S	/P	s/S	s/S	/P	/
Zapotec (Mitla)	/P	p/	/P	p/P	p/	/	/	/S	/	p/	s/S	p/	s/	s/S	/	/
Zapotec (Yatzachi)	/P	p/	/	/P	p/	/P	s/	/S	/	/	s/S	/	/	s/S	/	/P
Zapotec (Zoogocho)	/	p/	/	/P	/P	/P	/S	/S	/P	/P	s/S	/	s/S	s/S	/P	/
Penutian																
CHINOOKAN																
Chinook (Lower)	/	/S	/	/P	s/	/	/	/P	/	/P	p/S	s/	p/	p/S	/P	/
Chinook (Upper)	/	/	/	/	/	/	/	/P	/	/	p/	/	p/	p/S	/	/
COSTANOAN																
Costanoan	/	s/P	/	s/	/	/S	/P	/	/	/P	/S	/	/P	/S	/	/
Mutsun	/P	s/	/P	s/	s/	/	/	/P	/	/P	/S	s/	p/P	/P	/P	/P
KLAMATH-MODOC																
Klamath	/P	s/	s/	s/S	p/	/S	/	/P	/	/P	/	/	/P	/	/P	/
MAIDUAN																
Konkow	s/	s/	/	s/	/	/S	p/	p/	/	s/	/P	s/	p/P	/P	s/	/
Maidu (Northeast)	s/	s/	s/	s/	s/	/P	/	/P	/	s/	/P	s/	p/	s/P	/	/
Nisenan	/	s/	/	s/	s/	/	/	/P	/	s/	/P	/	p/	/P	/	/
MIWOK																
Miwok (Southern Sierra)	/	s/	/	/	s/	/	/	/P	/	/P	s/	s/	s/	s/P	/	s/
SAHAPTIAN																
Klikitat	/	s/	/	s/	p/	/	/	/P	/	/P	/S	s/	p/	p/S	/	s/
Nez Perce	/P	s/	/	s/	p/	/P	/	/P	/	/P	p/	/	p/P	p/	/P	/
Yakima	/P	s/	/	s/	p/	/	/	/P	/	/P	/S	s/	/	p/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
TSIMSHIANIC																
Gitksan	/S	/	/	/P	s/	/	/	/	/	/	/	p/	/	/	/	/
Nisgha	/S	/P	/	/P	s/	/	/	/S	/	/P	s/S	/	s/	/S	/P	/
Tsimshian (Coast)	/	/P	/	/P	/P	/	/S	/S	/	/	s/S	p/	s/	/S	/P	/
WINTUAN																
Patwin	/	s/	/	s/S	/	/	/	/	/	s/	/P	s/	/P	/P	/	/
Wintu	/	s/S	/	s/	/	/	/	/	/	/	/	/	/	s/	s/	/S
YOKUTS																
Wikchamni	/P	s/	/	s/	/	/	/	/P	/	/P	/	s/	/P	/	/P	/
Yawelmani	/P	s/	/	s/	s/	/	/	/	/	/P	/S	s/	/	/	/	s/
Yokuts (Yaudanchi)	/P	s/P	/	s/P	/	/	/	/P	/	/P	/S	s/	/	/S	/	s/
Salinan																
SALINAN																
Salinan	p/	/S	/	/P	s/	/P	/P	/P	/	/	s/S	s/	/	p/P	/	s/
Salishan																
BELLA COOLA																
Bella Coola	s/	s/	/	/P	/	/	/S	/S	/	/P	s/S	p/	s/	s/S	/	/
CENTRAL SALISH																
Comox	/	/	/	/P	/	/	p/	p/	/	/	/S	/	p/	/S	/	/
Halkomelem (Upriver)	/	/	/	/P	/	/	/	/P	/P	/P	s/S	/	s/	s/S	/	/
Musqueam	/	/	/P	/P	/	/	/	/P	/	/P	/S	/	/	/S	/P	/P
Squamish	/	/P	/	/P	s/	/P	/P	/P	/P	/P	s/S	/	/	s/S	/P	/
INTERIOR SALISH																
Coeur D'alene	/	p/	/	/P	/	/	/	/S	/	/	s/S	/P	/	s/S	/	/
Kalispel	/	/	/	p/	/	/	/P	/P	/	/P	/S	/	/	p/S	/P	/
Lillooet	/	/P	/	/P	/	/	/	/P	/	/	s/S	/	s/	s/S	/	/
Shuswap	/	/P	/	p/P	s/	/	/	/	/	/P	s/S	/	/	s/S	/	/
Thompson	/P	/	/	/P	/	/	/	/	/	/P	/S	/	/	/S	/	/
Siouan																
SIOUAN																
Assiniboine	/S	/	/	/	/	/	/S	/S	/S	/	/P	/	/	/P	/	/
Biloxi	/S	s/S	/	s/S	/	/	/	/S	/	/S	p/P	/	p/	p/P	/S	/S
Catawba	/	/	/	/S	/	/	/S	/	/S	/	/P	/	/	/P	/S	/
Crow	/S	s/	/	/S	/	/S	/S	/P	/S	s/	p/P	s/	p/	p/P	s/	/
Hidatsa	/	/	/	/S	s/	/	/S	/S	/S	s/	/P	s/	/	p/P	/S	s/
Lakhota	/S	s/S	/	/S	s/	/S	/S	/S	/S	/S	p/P	/S	p/	p/P	/	/S
Mandan	s/	s/S	/	s/S	s/	/	s/	s/	s/	/	p/P	s/	p/	p/P	s/	/
Omaha	/S	s/S	/	/	/	/	/S	/P	/	s/	/P	/	/	/P	/	/S
Osage	/	s/	/	/S	/	/	/	/	/S	/S	p/P	/S	/S	p/P	/S	/
Tutelo	s/	s/S	/	/S	/	/	s/	/S	s/	/	p/P	/	p/	p/P	/	/
Takelma																
TAKELMA																
Takelma	/	s/	/	p/S	s/	/	/	/P	/	/P	s/P	s/	s/	s/P	/	/
Tarascan																
TARASCAN																
Purépecha	/	s/	/	s/	s/	/	/	/P	/	/P	s/S	s/	s/	s/P	/	s/
Tequistlatecan																
TEQUISTLATECAN																
Chontal (Huamelultec Oaxaca)																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
	/P	s/P	/	p/	/	/	/	/	/	/P	s/S	s/	p/	s/S	/P	/
Timucua																
TIMUCUA																
Timucua	/	s/	/	/	/	/	/	/	/	/	p/P	s/	s/	p/P	/	/
Tol																
TOL																
Tol	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Tonkawa																
TONKAWA																
Tonkawa	s/	s/	/	/S	p/	/	/	s/	/	s/	p/P	s/	p/	s/P	s/	/
Totonacan																
TOTONACAN																
Tepehua (Huehuetla)	/P	/	/P	p/P	/	/S	/P	/P	/P	/P	/S	/	p/	/	/P	/
Tepehua (Tlachichilco)	/	/P	/	p/P	/	/S	/P	/	/P	/P	/S	s/	p/	/P	/P	s/
Totonac (Misantla)	/	/	/	/	/	/	/P	/P	/	/	/	/	p/	/	/	/
Totonac (Papantla)	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Totonac (Xicotepec De Juárez)	/P	/	/	/P	/	/S	/P	/P	/P	/	s/S	/	p/	/P	/P	/
Tunica																
TUNICA																
Tunica	s/	s/	/	s/S	/	/	/	/	/	s/	p/P	s/	p/	s/P	s/	/
Uto-Aztecan																
AZTECAN																
Nahuatl (Huasteca)	/	s/	/P	/P	s/	/P	/	/P	/	/P	p/S	s/	p/	p/S	/P	/
Nahuatl (Michoacán)	/	s/	/	/P	s/	/P	/	/P	/	/P	p/S	/	p/	p/P	/P	/P
Nahuatl (North Puebla)	/	s/	/	/	s/	/	/	/P	/	/	p/S	s/	p/	p/P	/	s/
Nahuatl (Tetelcingo)	/	s/	/P	/P	s/	/P	/	/P	/	/P	p/S	s/	p/	p/P	/P	/P
Pipil	/	s/	/	/P	s/	/	/P	/	/P	/P	p/S	s/	p/	p/S	/P	s/
CAHITA																
Eudeve	/P	s/	/	s/S	s/	s/	/	/P	/	/P	/P	p/	p/P	s/P	/	/S
Yaqui	/P	s/	/	s/S	s/	/	/	/P	/P	/P	/P	s/	/P	/P	/	s/
CORACHOL																
Cora	/	/	/	/S	/	/	/P	/P	/	/	/S	/	p/	/S	/	/
Huichol	/	s/	/	s/	s/	/	/	/	/	p/	p/P	s/	p/	p/P	/	s/
HOPI																
Hopi	/P	s/	/	s/S	/	/	/	/P	/	/P	/P	s/	/	s/P	/	/
NUMIC																
Chemehuevi	/	s/	/	s/S	s/	s/	/	/	/	s/	s/	s/	s/P	/P	/	s/
Comanche	/	s/	/	s/S	s/	/	/P	/P	/	/P	/P	s/	/P	/P	/	/S
Kawaiisu	/	/	/	s/S	/	/	/	/	/	/P	s/	s/	s/	/	/	/
Mono (in United States)	/	s/	/	s/S	/	s/	s/	s/	/	/P	s/P	s/	p/P	s/P	/	/
Paiute (Northern)	/	s/	/	/	p/	s/	/	/P	/	/P	/P	/	/P	s/P	/	p/
Shoshone	/	s/	/	s/	/	/	/	/P	/	/	/P	s/	/	s/P	/	s/
Tümpisa Shoshone	/	s/	/	s/	/	/S	/	/	/	/P	/P	s/	/P	/P	/	s/
Ute	/	s/	/	s/S	s/	/S	/S	/P	/	/	/P	s/	s/P	/	/	/
TAKIC																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Cahuilla	/	s/	/	s/S	s/	/	/	/	/	/P	p/P	s/	p/	p/P	/	/
Cupeño	/	s/	/	s/S	s/	/S	/	/P	/	/P	p/P	s/	p/P	p/P	/	s/
Luiseño	/	s/P	/	s/S	s/	/P	/	/P	/	/P	/	s/	p/	/P	/	s/
TARAHUMARAN																
Guarijío	/	s/	/	/S	/	/	/	/P	/	/P	/	/	/	/	/	/
Tarahumara (Central)	/	s/	/	s/S	/	/	/P	/P	/P	/P	/P	/	/	/P	/	/
Tarahumara (Western)	/	s/S	/	s/	s/	/S	/P	/P	/P	/P	/P	/	/P	/P	/	s/
TEPIMAN																
Nevome	/P	s/P	/	/S	s/	/	/	/P	/P	/P	p/P	p/	p/	/P	/	s/
O'odham	/P	s/	/	/	s/	/	/P	/P	/	/P	p/S	p/	p/	/S	/	s/
Pima Bajo	/	s/	/	s/	/	/	/	/P	/	/P	/P	p/	p/	/P	/	/
Tepehuan (Northern)	/	s/	/	s/S	/	/S	p/	/P	/	/P	p/S	p/	p/	p/S	/	s/
Tepehuan (Southeastern)	/	s/	/	/S	/	/	/	/P	/P	/	/S	p/	p/	s/S	/	/
Wakashan																
NORTHERN WAKASHAN																
Heiltsuk	/	/	/	/P	/	/	/	/P	/	/P	/S	/	/	/S	/	/
Kwakw'ala	/	s/	/	/P	s/	/	/	/	/	/	s/S	/	s/	s/	/	/
SOUTHERN WAKASHAN																
Kyuquot	/	s/	/	/P	/	/	s/	/	/	/P	/S	s/	s/	/S	/	/
Makah	/	s/S	/	/	/	/	/	/P	/	/P	/S	/	/S	/S	/	/
Nuuchahnulth	/	/	/	/	/	/	/	/	/	/	/S	/	/	/S	/	/
Wappo-Yukian																
WAPPO																
Wappo	s/	s/	/	s/	/	/	/P	/P	/	s/	/P	s/	p/	/P	/	/
Washo																
WASHO																
Washo	/S	s/	/	s/S	/	/	/	/P	/	s/	p/P	/	p/P	p/P	/	s/
Xincan																
XINCAN																
Xinca (Guazacapán)	/	/	/	/P	/	/S	/P	/S	/P	/P	/S	s/	/	/S	/P	/
Yuchi																
YUCHI																
Yuchi	/	s/S	/S	s/S	/	/	/S	/P	/	/P	p/P	/S	p/	p/P	/S	/
Zuni																
ZUNI																
Zuni	/	s/	/	s/S	/	/	/	/P	/P	/	/P	s/	/P	/P	/	/

South America																
Aikaná																
AIKANá																
Aikaná	/	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/
Alacalufan																
ALACALUFAN																
Qawasqar	s/	s/S	/	/S	/	/S	/	/	/	/S	/P	/	/P	/P	/	/
Andoke																
ANDOKE																
Andoke	/	/	/	s/S	/	/	/	/P	/	/	/	/	/	/	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Arauan																
ARAUAN																
Jarawara	/	s/	/	/	/	/	/	/	/	/	/P	/	/	p/P	/	/
Paumarí	s/	s/S	/	/S	p/	/	/	/P	/	/	p/	/	/	p/P	/	/S
Araucanian																
ARAUCANIAN																
Mapudungun	/	s/	/	/S	/	/	/P	/P	/	s/	/S	/P	/P	s/S	/	/
Arawakan																
ARAWAKAN																
Achagua	/	s/	/	/S	/	/	/	/	/	/P	/S	s/	p/	p/P	/	/
Arawak	/	s/	/	/S	s/	/	/P	/P	/P	/S	s/S	s/	p/P	p/P	/	s/
Baré	/	s/	/	s/S	/	/	/	/P	/	/P	s/S	s/	p/	p/P	/	/
Garífuna	/	s/S	/	/P	s/	/	/	/S	/	/P	s/S	s/	p/	/S	/P	/
Goajiro	/P	s/	/	s/P	/	/	/P	/S	/	/P	s/S	s/	p/	p/S	/	/
Maipure	/	s/	/	/S	/	/	/	/	/	/	/S	s/	p/	p/P	/	/
Piapoco	/	/	/	/	/	/	/	/P	/	/	/	/	/	/	/	/
Resígaro	/P	s/P	/	/S	s/	/	/	/P	/	/P	/P	s/	p/P	p/P	/S	s/
Tariana	/	/	/S	/S	/	/S	/P	/P	/P	s/	/P	s/	p/P	p/P	/	/P
Warekena	/	s/	/	s/P	/	/	/	/P	/	/	s/S	s/	p/	p/P	/	/
Yucuna	/	s/	/	/S	/	/	/	/P	/	/	/S	s/	/	p/P	/	/
BOLIVIA-PARANA																
Ignaciano	/	s/	/	/P	p/	/	/	/	/	/P	s/S	s/	p/	p/S	/	/P
EASTERN ARAWAKAN																
Palikur	/P	s/	p/	/P	s/	/	/	/P	/	/	s/S	s/	p/	/P	/	s/
NORTHERN ARAWAKAN																
Baure	/	s/P	/	s/	p/	/	/P	/P	/	/P	s/S	s/	p/	p/S	/P	/
PRE-ANDINEARAWAKAN																
Campa (Axininca)	/	s/	/	s/	s/	/P	/	/	/	/P	s/S	s/	p/	s/	/	/
Machiguenga	/	/	/	s/	/	/	/	/	/	/P	s/S	/	/	p/S	/	/
PURUS																
Apurinã	/	s/	/	/S	s/	/	/	/P	/	/P	s/	s/	p/P	p/P	/P	s/
Piro	/	s/	/	s/	s/	/	/P	/P	/	/	s/P	/	p/	p/P	/	/P
WESTERN ARAWAKAN																
Amuesha	/	s/P	/	s/	/	/	/	/P	/	/	s/S	s/	p/	p/	/	/
Aymaran																
AYMARAN																
Aymara	/S	s/	s/	/S	s/	s/	/	/P	/P	/P	s/P	s/	s/	s/P	/	/S
Jaqaru	/	s/	/	s/	/	/	/	/P	/	/P	s/	s/	s/	s/	s/	/
Barbacoan																
BARBACOAN																
Awa Pit	/S	s/S	/	/S	s/	/S	/	/P	/	/	s/P	s/	/P	s/P	/	s/
Cayapa	/	s/	/	s/	/	/	/	/P	/	s/	/P	/	/	s/P	/S	/
Tsafiki	s/	s/	/	/S	/	/	/	/P	/	s/	/P	/	/	/P	/	/
Betoi																
BETOI																
Betoi	s/	/	/	s/S	/	/	/	/	/	s/	s/P	s/	p/	/P	/	/
Bora-Huitotoan																
BORAN																
Muinane	/	s/	/	s/	s/	/	p/	/	/	s/	/P	s/	p/	/P	/	s/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
HUITOTO																
Huitoto (Minica)	/	s/P	/	s/	/	/	/	/P	/	s/	/P	s/	/P	s/P	/S	s/
Huitoto (Murui)	/	s/	/	s/S	s/	/	/	/	/	s/	/P	s/	/	s/P	/S	s/
Bororan																
BORORO																
Bororo	/	s/P	/	s/	s/	/	/	/	/	s/	/S	/	p/	p/	/	/
Botocudoan																
BOTOCUDOAN																
Krenak	/	/	/	/	/	/	/	/S	/	/	/P	/	/	/P	/	/
Cahuapanan																
CAHUAPANAN																
Jebero	/	s/	/	s/	p/	/	/	/P	/	s/	s/	s/	s/	s/	/	p/
Camsá																
CAMSá																
Camsá	/	p/	/	s/S	/	/	/	/	/	p/	p/	s/	/	p/P	/	/
Candoshi																
CANDOSHI																
Candoshi	s/	/	/	/	/	/	/	/	/	s/	/P	/	/	/P	/	/
Cariban																
CARIBAN																
Akawaio	/	s/S	/	s/S	/	/	/	/	/	/S	/P	s/	p/	/P	/	/
Apalaí	/	s/	/	/S	s/	/S	/	/	/	s/	p/P	/S	p/	p/	s/	/S
Carib	s/	s/S	/	/S	s/	/S	/	/P	/P	s/	p/P	s/	p/P	p/P	/S	/S
Carib (De'kwana)	/	s/S	/	/S	/	/S	/	/P	/	s/	p/P	s/	p/	p/P	/	/
Hixkaryana	/	s/S	/	/S	s/	/S	/	/	/	s/	p/P	/S	p/	p/S	/S	/S
Kuikúro	/	/	/	/S	/	/	/	/	/	/	/P	/	/	p/S	/	/
Macushi	/	s/S	/S	s/S	/	/	/	/P	/	/S	p/P	s/	p/P	s/	/S	s/
Tiriyo	/	s/	/	/S	s/	/	/	/	/	s/	p/P	s/	p/	p/S	/	/
Wai Wai	/	s/	/	/S	/	/	/	/	/	s/	p/P	/	/	p/	/	/
Cayuvava																
CAYUVAVA																
Cayuvava	/	/	/	/P	s/	/	/	/P	/	p/	s/S	/	/	p/S	/P	/
Chapacura-Wanhan																
CHAPACURA-WANHAN																
Wari'	/	/	/	/P	/	/P	/	/S	/	/P	/S	/P	s/S	/S	/P	/
Chibchan																
ARUAK																
Damana	/	/	/	/	/	/	/	/	/	/	/	/	/	/P	/	/
Ika	/	s/S	/	/S	/	/S	/	/	/	s/	p/P	/S	p/P	/P	s/	/S
CHIBCHAN PROPER																
Muisca	s/	s/	/	s/	/	/	/	/P	/	s/	/P	/	/	p/P	/	/
Tunebo	/	/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
GUAYMI																
Ngäbere	/	s/P	/	s/S	/	/	/P	/S	/	/P	/P	s/	/P	/P	/	/
KUNA																
Kuna	/	/	/	/S	/	/	/	/	/	s/	/P	/	/	/P	/	/
PAYA																
Pech	/	/	/	/	/	/	/	/	/	s/	/	/	/	/	/	/
RAMA																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Rama	/	s/	/	/S	/	/	/	/P	/S	/	/P	/	p/P	p/P	/	s/
TALAMANCA																
Bribri	/	s/P	/	/S	/	/P	/	/S	/	/P	/P	s/	/P	/P	/	/P
Teribe	/S	s/	/	/S	/P	/P	/S	/S	/	/S	/P	s/	/P	s/P	/	/
Chimúan																
CHIMúAN																
Mochica	/	/	/	s/S	/	/	/	/S	/	/	/	/	/P	/	/	/
Chiquitano																
CHIQUITANO																
Chiquitano	/	p/	/	/P	/	/	/	/	/	s/	/S	s/	p/	p/	/	/
Choco																
CHOCO																
Emberá (Northern)	s/	s/S	/	/	/	/S	/	/P	/	/S	/P	s/	/P	s/P	s/	/
Epena Pedee	s/	s/S	/	/S	s/	/S	/P	/P	/	s/	/P	/	/P	/P	/	/
Waunana	/	s/	/	s/	/	/	/	/	/	s/	/P	s/	/	/P	/	/
Chonan																
CHONAN																
Selknam	/	/S	/	/S	/	/	/	/	/	/S	/P	/	/	/S	/	/
Cofán																
COFáN																
Cofán	s/	/	/	/	/	/	/	/P	/	s/	/P	/	/	/P	/	/
Creole																
CREOLE																
Ndyuka	/	/P	/	/P	/	/P	/P	/S	/P	/P	/S	/P	/	/P	/P	/
Saramaccan	/S	/P	/	/P	/	/	/P	/S	/P	/	/P	/	/	/P	/	/
Fulniô																
FULNIô																
Fulniô	s/	/	/	/S	s/	s/	/	/P	/	s/	p/P	s/	p/	p/P	s/	/
Guahiban																
GUAHIBAN																
Cuiba	/	s/	/	/	/	/	/	/	/	/	/P	/	/	/P	/	/
Sikuani	/	/	/	/S	/	/	/	/	/	/	/P	/	/	/P	/	/
Guaicuruan																
GUAICURUAN																
Abipón	/	s/	/	/P	/	/	/	/P	/	/P	s/S	s/	/	/P	/P	/
Kadiwéu	/	/	/	/	/	/	/	/	/	p/	p/S	s/	p/	p/P	/	/
Mocoví	/	s/	/	/	/	/	/	/P	/	p/	/S	s/	p/	/P	/P	/
Pilagá	/P	s/	/	/	/	/	/	/P	/	p/	p/S	s/	p/	p/P	/	/
Toba	/	s/	/	/	/	/	/	/P	/	p/	s/S	s/	p/	p/	/P	/
Guató																
GUATó																
Guató	/	/	/	/P	/	/	/	/P	/	/	/S	/	p/	/S	/	/
Harakmbet																
HARAKMBET																
Amarakaeri	/	/	/	/S	/	/	/	/P	/	/	/	/	/	/	/	/
Hibito-Cholón																
HIBITO-CHOLóN																
Cholón	/	s/	/	s/S	/	/	/	/P	/	s/	p/P	s/	p/	p/P	s/	/
Huarpe																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
HUARPE																
Allentiac	/	s/	/	/	/	/	/	/	/	/	/	/	/	s/	/	/
Iranxe																
IRANXE																
Iranxe	/	s/	/	/	/	/	/	/	/	s/	/P	/	/	/P	/	/
Itonama																
ITONAMA																
Itonama	p/	/P	/	/	/	/	/	/P	/	/	/S	s/	p/	/P	/	/
Je-Jabutí																
GE-KAINGANG																
Apinayé	/P	/P	/	/S	/	s/	/	/S	/	/	/P	/S	p/P	p/P	/	/
Canela	/P	/P	/	/S	/S	/	/	/S	/	/S	p/P	/	p/	/P	/	/
Kaingang	/	/S	/	/S	/	/	/	/	/	/S	p/P	/	/P	/P	/S	/
Kayapó	/P	/	/	s/	/	/	/	/	/	/S	/P	/	p/	/P	s/	/S
Krahô	/	s/	/	/S	/	/	/	/	/	s/	p/P	/	p/	p/P	/	/
Panará	/P	/	/	/S	/	/	/	/	/	/	/	/	/	/P	/	/
Xaványte	/P	/P	/	/S	/	/S	/	/	/	/S	p/	/	p/	/P	/	/
Xerénte	/	/	/	/	/	/	/	/P	/	/	/P	/	p/	/P	/	/
JABUTí																
Jabutí	/	/	/	/S	/	/	/	/P	/	/	/P	/	p/	p/P	/	/
Jirajaran																
JIRAJARAN																
Ayomán	/	/	/	s/	/	/	/	/	/	/	/S	/	p/	/	/	/
Jivaroan																
JIVAROAN																
Achuar	s/	s/	/	s/	/	/	/	/	/	s/	s/P	/	s/P	s/	/	/
Aguaruna	/S	s/S	/	/S	/	/	/	/P	/P	s/	/P	/	s/	s/P	s/	s/
Jivaro	/	s/	/	s/	/	/	/	/	/	s/	/P	s/	/P	s/P	/	/
Kapixana																
KAPIXANA																
Kanoê	/S	/	/	s/	/	/S	/	/P	/	/	/P	/	/P	/P	/P	/
Karajá																
KARAJá																
Javaé	/	s/	/	/S	/	/	/	/P	/	/	/P	s/	p/	p/P	/	/
Kariri																
KARIRI																
Kipea	/	s/	/	/P	/	/	/	/P	/	s/	/S	s/	p/	p/S	/	/P
Katukinan																
KATUKINAN																
Canamarí	/	s/	/	/S	s/	/	/	/P	/	s/	/	/S	p/P	p/	/	/
Kunza																
KUNZA																
Atacameño	/	/	/	/	/	/	/	/	/	/	/P	/	p/	/P	/	/
Kwazá																
KWAZá																
Kwazá	s/	s/	/	s/	/	/	/	/	/	s/	s/P	/	/	s/P	/	/
Leko																
LEKO																
Leko	/	s/	s/	s/	s/	/	/	/P	/	s/	p/	s/	p/	s/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Lule																
LULE																
Lule	/	/	/	/	/	/	/	/	/	/	/	s/	s/	s/	/	/
Macro-Ge																
RIKBAKTSA																
Rikbaktsa	/	/	/	/S	p/	/	/	/	/	/P	/P	/	p/	/P	/	/
Mascoian																
MASCOIAN																
Lengua	/	s/	/	/	s/	/	/	/	/	/	/	s/	p/	p/	/	/
Matacoan																
MATACOAN																
Chorote	/	/	/	/	/	/	/	/	/	/P	/	/	/	/	/	/
Wichí	/P	s/P	/	/P	s/	/P	/P	/S	/	/	s/S	s/	p/P	p/P	/P	/S
Maxakalían																
MAXAKALíAN																
Maxakalí	/P	/	/	/S	/	/	/	/	/	/	/P	/	p/	/P	/	/
Misumalpan																
MISUMALPAN																
Miskito	/	s/S	/	/S	/	/	/	/P	/S	s/	/P	/S	s/P	s/P	/S	/
Mosetenan																
MOSETENAN																
Mosetén	/	s/	/	/	/	/	/	/	/	/P	/S	s/	/P	/P	/	/
Movima																
MOVIMA																
Movima	/	/	/	/P	s/	/	/	/	/	/P	/S	/	/S	/S	/P	s/
Muniche																
MUNICHE																
Muniche	s/	s/	/	/	s/	/	/	/S	/	/P	/S	s/	/	/S	/	s/
Mura																
MURA																
Pirahã	/	s/	/	s/S	/	/S	/	/S	/	s/	/P	/	/P	/P	/	s/
Nadahup																
NADAHUP																
Dâw	/	s/	/	/S	/	/	/	/P	/	/	/P	/	/P	/P	/	/
Hup	/	s/	/	/S	/	/	/	/P	/	s/	/P	/	/P	/P	/	/
Nadëb	/	/	/	/S	/	/	/	/	/	/	/P	/	/	/P	/	/
Nambikuaran																
NAMBIKUARAN																
Nambikuára	s/	s/	/	/	/	s/	s/	/	s/	s/	s/P	/	p/	s/P	/	s/
Ofayé																
OFAYé																
Ofayé	/	/	/	/S	/	/	/	/P	/	/	/P	/	/	/P	/	/
Panoan																
PANOAN																
Amahuaca	/	s/S	/	/S	/	/	/	/P	/	s/	/P	/	p/P	s/	/	/
Cashibo	/	s/	/	s/S	/	/	/P	/P	/S	s/	/P	s/	/	/P	/	/
Chácobo	/S	s/S	/	/S	s/	/S	/	/	/	s/	s/P	/S	/P	s/	s/	s/
Katukina Pano	/	/	/	/	/	/	/	/	/	/	/P	/	p/	/P	/	/
Matis	s/	s/	/	s/	/	/	/	/	/	s/	/P	s/	/	/P	/	/

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Matsés	/	/	/	/S	/	/	/	/	/	/	/P	/	/	/P	/	/
Sharanahua	/S	s/	/	s/S	/	/	/	/P	/	s/	/P	/	/	/P	/	/S
Shipibo-Konibo	/	s/S	/	/S	s/	/	/	/P	/	s/	/P	s/	/P	/P	/	s/
Yaminahua	/	/	/	s/	/	/	/	/	/	s/	/P	s/	/	/P	/	/
Peba-Yaguan																
PEBA-YAGUAN																
Yagua	/	s/P	/	s/S	s/	/S	/	/P	/	/P	/S	s/	p/	/	/	/P
Puelche																
PUELCHE																
Gününa Küne	/	/	/	s/P	/	/	/	/	/	p/	s/	/	s/	/	s/	/
Puinave																
PUINAVE																
Puinave	s/	s/	/	s/	/	/	/	/	/	/	p/	s/	p/	p/	/	/
Puquina																
PUQUINA																
Puquina	/	/	/	s/S	/	/	/	/	/	/	s/	/	/P	s/	/	/
Páezan																
PáEZAN																
Páez	s/	s/S	/	/S	/	/S	/	/P	/	s/	s/P	s/	/P	s/P	s/	/S
Quechuan																
QUECHUAN																
Quechua (Huallaga)	s/	s/S	/	/S	/	/	/	/P	s/	/	s/P	s/	s/	s/P	/	/
Quechua (Imbabura)	/S	s/S	/S	s/S	s/	/S	/	/P	/	/P	s/P	s/	/P	s/P	s/	/
Sáliban																
PIAROA																
Piaroa	/	s/	/	s/	/	/	/	/P	/	s/	s/P	s/	p/P	/P	/	/
SáLIBA																
Sáliba (in Colombia)	/	s/	/	/S	/	/	/	/	/	s/	/	/	/	s/	/	/
Tacame																
TACAME																
Esmeraldeño	/	/	/	s/P	/	/	/	/	/	/	s/S	/	s/	s/	/	/
Tacanan																
TACANAN																
Araona	p/	s/S	/	s/S	/	/	/	/P	/	/	/P	/S	/P	s/P	/S	s/
Cavineña	/	/	/	/S	/	/	/	/P	/	/	/	/	/	/	/	/
Ese Ejja	/P	s/S	/	/S	/S	/	/	/	/	/S	/P	/S	/P	/P	/S	/S
Tacana	/P	s/S	/	/S	/	/	/	/P	/	/	/P	/S	/P	s/P	/S	s/
Taushiro																
TAUSHIRO																
Taushiro	/	/	/	/S	/	/	/	/	/	/	/S	/	/	/S	/	/
Ticuna																
TICUNA																
Ticuna	/	/	/	s/	s/	/	/	/P	/	/	p/	s/	p/	/	/P	s/
Timote-Cuica																
TIMOTE-CUICA																
Cuica	/	/	/	/P	/	/	/	/	/	/	/S	/	/	/P	/	/
Trumai																
TRUMAI																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
Trumai	/	s/S	/	s/S	/	/	/	/P	/	/S	/	/S	/	/P	s/	s/
Tucanoan																
TUCANOAN																
Barasano	s/	s/S	/	s/S	s/	/S	/	/P	/	s/	/P	s/	/P	s/	s/	s/
Carapana	s/	s/S	/	s/	/	/S	/	/P	/	s/	/	s/	/P	s/	s/	s/
Cubeo	s/	s/	/	s/S	/	/	/	/	/	s/	/P	/	/P	s/S	/	/
Desano	s/	s/S	/	/S	/	/S	/	/P	/	s/	/P	s/	/P	s/P	/S	/
Koreguaje	s/	/S	/	s/S	/	/	/	/	/	/	/S	/	/	/S	/	/
Retuarã	/P	s/	/	/S	/	/	/	/P	/	s/	/P	/	/	p/P	/	/
Secoya	/	/	/	/	/	/	/	/P	/	/	/	s/	/P	/	/	/
Siona	/	s/	/	s/	/	/	/	/P	/	s/	/P	s/	/	s/P	/	/
Tucano	s/	s/S	/	/S	s/	/	/	/P	/	s/	/P	s/	/P	s/P	/S	s/
Tuyuca	s/	/	/	/S	/	/	/	/P	/	s/	/P	/	/P	/P	/	/
Wanano	/	s/S	/	s/	/	/	s/	/P	/	s/	/P	s/	p/	s/P	/	s/
Tupian																
ARIKEM																
Karitiana	/	/	/	/S	/	/P	/	/	/	/S	/S	/	/	/P	/	/
MONDE																
Gavião	/P	/	/	/S	/	/	/	/P	/	/P	/P	/	p/	/P	/S	/
Suruí	/P	/	/	/S	/	/	/	/	/	/	/P	/	/	/P	/S	/
MUNDURUKU																
Mundurukú	/	/	/	/S	/	/	/	/	/	/	p/P	s/	p/	p/P	/	/
RAMARAMA																
Karó (Arára)	/S	/S	/	/S	p/	/S	/	/P	/	/S	p/P	/	p/P	/P	/	/
TUPARI																
Makurap	/	s/	/	/S	p/	/	/	/	/	s/	/P	/	p/	/P	/S	/
Mekens	/S	/	/	/S	/	/S	/	/P	/	/S	p/P	/	p/	/P	/P	/
TUPI-GUARANí																
Asuriní	/	/	/	/S	/	/	/	/	/	/	/P	/	p/	/S	/	/
Cocama	/	s/S	/	/S	s/	/	/	/P	/	/P	/S	/	/P	/P	/	/S
Guajajara	/	/S	/	/S	/S	/	/	/P	/	p/	p/S	/	p/	p/S	/S	/S
Guaraní	/	/S	/	/S	s/	/	/P	/P	/	/	p/S	/S	p/	p/P	/	s/
Kamaiurá	/	s/	/	s/S	/	/	/	/P	/	s/	p/P	s/	p/	p/P	/	/
Kayabí	/	s/	/	/S	s/	/	/	/	/	s/	/P	/	p/	p/P	/	/
Sirionó	/S	s/	/	/S	/	/	/	/P	/	/	p/P	/	p/	p/P	/S	/
Tapieté	/	s/S	/	s/P	/	/	/	/P	/	s/	p/P	s/	p/	p/P	/	/
Urubú-Kaapor	/S	/S	/	/S	p/	/S	/	/P	/	/S	/P	s/	p/	p/P	/S	/S
Wayampi	/	s/	/	/S	/	/	/	/	/	/	p/P	/	p/	p/P	/S	/
Émérillon	/	s/	/	/S	/	/	/	/P	/	/	p/P	/	p/	p/P	s/	/
YURUNA																
Juruna	/	/	/	/	/	/	/	/P	/	/	/	/	p/P	/	/	/
Urarina																
URARINA																
Urarina	/	/	/	/S	/	/	/P	/P	/	s/	/P	s/	/	s/S	/S	/
Uru-Chipaya																
URU-CHIPAYA																
Chipaya	s/	s/	/	s/	s/	/	/	/P	/	/P	/P	s/	/P	s/P	/	/
Uru	/	/	/	s/	/	/S	/	/	/	/P	/P	/	/	/P	s/	/
Waorani																

	Q	TA	able	case	caus	cop	def	dem	indef	neg	obj	pl	poss	subj	subor	want
WAORANI																
Waorani	s/	s/	/	/S	/	/S	/	/P	/	s/	/P	s/	/P	s/P	/	/
Warao																
WARAO																
Warao	s/	s/S	/	/S	/	/S	/	/P	/	s/	p/P	s/	p/	s/P	/	/
Yanomam																
YANOMAM																
Sanuma	/	/S	/	/S	/S	/S	/	/P	/	/S	/P	/S	/P	/P	/S	/
Shiriana	/S	s/	/	s/S	s/	/	/	/P	/	/S	/P	/S	s/P	/P	/S	/
Yaruro																
YARURO																
Yaruro	/	/	/	/	/	/	/	/	/	s/	/	/	/	/	/	/
Yurakare																
YURAKARE																
Yuracare	/	s/	/	/S	/	/	/P	/P	/	/P	p/	/	p/	s/S	/	/
Yurumanguí																
YURUMANGUí																
Yurimanguí	/	s/	/	/	/	/	/	/	/	s/	/	/	/	/	/	/
Yámana																
YáMANA																
Yahgan	/	s/	/	s/S	/	/	/	/	/	/	/P	/	/P	p/P	/	/
Zamucoan																
ZAMUCOAN																
Ayoreo	/	/	/	/	/	/	/	/	/	/P	/S	/	p/	p/P	/	/
Zaparoan																
ZAPAROAN																
Iquito	/	s/	/	s/S	/	/S	/	/P	/	/P	p/S	/	p/	p/P	/	/
Zaparo	/	s/P	/	s/	s/	/S	/	/	/	/	/S	s/	p/	p/P	/	/