Chapter 1

Introduction

This chapter motivates the need for research on lexical flexibility by situating it within broader concerns regarding linguistic categories more generally, and categories in human cognition. The specific problem that this study seeks to address is our lack of understanding regarding what lexical flexibility looks like, and how it varies across languages. This thesis contributes to answering these questions via a quantitative corpus-based study of lexical flexibility in English (Indo-European) and Nuuchahnulth (Wakashan). It is the first study to examine lexical flexibility using natural discourse from corpus data. This chapter provides an overview of the thesis, including the specific research questions addressed, the data and methods used, a concise summary of the results, and a preview of the conclusions.

1.1 The "problem" of lexical flexibility

Word classes such as noun, verb, and adjective (traditionally called *parts of speech*) were once thought to be universal, easily identifiable, and easily understood. Today they are one of the most controversial and least understood aspects of language. While language scientists generally agree that word classes exist, there is much disagreement as to whether they are categories of individual languages, categories of language generally, categories of human cognition, categories of language science, or some combination of these possibilities (CITE: Mithun 2017: 166; Haspelmath 2018; Hieber forthcoming: 1). Lexical categorization—how languages

separate words into categories—is of central importance to theories of language because it is tightly interconnected with linguistic categorization generally, which in turn informs (and is informed by) our understanding of cognition. Categorization is a fundamental feature of human cognition (CITE: Taylor 2003: xi), and lexical categorization is perhaps the most foundational issue in linguistic theory (CITE: Croft 1991: 36; Vapnarsky & Veneziano 2017: 1).

One challenge for traditional theories of word classes is the existence of *lexical flexibility*—the use of a word in more than one discourse function, whether to refer (as a noun), to predicate (as a verb), or to modify (as an adjective). In traditional terms, flexible words are those which may be used for more than one part of speech. (A more precise definition of lexical flexibility is given in Chapter 2.) Examples of flexible words in several languages are shown below.

Give examples here. Discuss all the examples together in the lead out. Examples: English; Nuuchahnulth: Kingfisher 202

Flexible words like those in the examples above create an analytical problem for traditional theories of parts of speech. Traditional theories assume that words can be partitioned into mutually exclusive categories on the basis of a clear set of criteria, an approach that has its roots in the Aristotelian tradition of defining a category via its necessary and sufficient conditions. Flexible words would seem to violate this assumption because they appear to be members of more than one category at once, and the criteria for classification yield conflicting results.

Researchers have proposed numerous solutions to this problem. The most common response is to adjust the selectional criteria so that only certain features are considered definitional of the class, allowing these researchers to dismiss other, potentially contradictory evidence as irrelevant (CITE: Baker 2003; Dixon 2004; Floyd 2011 for Quechua; Chung 2012 for Chamorro; Palmer 2017). It is also common to analyze different uses of a putatively flexible word as instances of *heterosemy*—that is, entirely distinct words which share the same form

but belong to different word classes (CITE: Lichtenberk 1991). In this view, heterosemous words are related only historically, via a process of conversion or functional shift, in essence denying the existence of lexical flexibility (CITE: Evans & Osada 2005). Another approach is to claim that, while all words can be neatly categorized, some words in some languages may nonetheless be used for functions typically associated with other categories. A notable example of this is Launey's (CITE: 1994?, 2001?) analysis of Classical Nahuatl, which he calls an *omnipredicative* language. In this analysis, Classical Nahuatl has the traditional, clearly-delineated word classes of noun and verb, but allows for any word to function as a verb regardless of its category (hence the term *omnipredicative*). The reverse is not true however; only some verbs may function as nouns. This difference in behavior is taken as the basis for a categorical[1] distinction between nouns and verbs.

Make sure that this is an accurate description of the empirical facts of Classical Nahuatl.

Some researchers enthusiastically embrace the existence of lexical flexibility and abandon a commitment to the traditional categories of noun, verb, and adjective. Instead they analyze flexible lexemes as belonging to a broader, flexible word classes such as "flexibles", "contentives" or "non-verbs", etc. (CITE: Hengeveld & Rijkhoff 2005; Luuk 2010?). Other researchers abandon the commitment to word classes entirely. Mandarin, Tagalog, Tongan, and Riau Indonesian have each been analyzed as lacking parts of speech (CITE: Simon [1937], McDonald [2013], and Sun [2020] for discussions of early analyses of Mandarin; Gil [XXXX] for Tagalog; Broschart [XXXX] for Tongan; Gil [XXXX] for Riau Indonesian). Within generative linguistics, the Distributed Morphology framework takes it as an assumption that all word roots are category-neutral (CITE: Siddiqi 2018).

Note that these differences in perspective do not arise from disagreements about the empirical facts of each language. Researchers mostly agree on the empirical data, but disagree on the relative importance of various pieces of evidence, and on which criteria should be taken as diagnostic of a category (CITE: Croft & van Lier 2012: 58). It is rare that an argument for flexibility is refuted on the basis of the linguistic facts alone (CITE: cite Mithun's response to

Add example from Stassen [1997: 32] about two researchers coming to different conclusions about Sudanese. Also

[1] Throughout this thesis, I use the term *categorical* to mean 'without exception; unconditional' and the term *categorial* to mean 'having to do with categories'.

Sasse's analysis of Cayuga as a flexible language).

Since analyses of lexical flexibility depend more on the particular theoretical commitments of the researchers involved rather than any particular crucial pieces of evidence, this leads to an intractable problem: researchers cannot agree on the criteria that should be considered diagnostic for a given category in a specific language (let alone crosslinguistically). Instead they partake in *methodological opportunism* (CITE: Croft 2003?: ??), choosing the evidence and criteria which best support their theoretical commitments. Discussions in the literature about the existence of a particular category in a particular language are therefore often unproductive, and devolve into debates about theoretical assumptions or the relevance or importance of various pieces of evidence, which are ultimately unresolvable (CITE: Croft 2005: 435).

This is particularly unfortunate because lexical flexibility is by no means an isolated or minor phenomenon. Additional examples like those above could be provided for many or perhaps even all of the world's languages. Lexical flexibility is not as rare or marginal as traditional approaches to word classes lead one to believe. In a survey of word classes in 48 indigenous North American languages (CITE: Hieber forthcoming), every one of the languages surveyed exhibited lexical flexibility in at least some area of the grammar (although not all authors analyzed these cases as such). In my own experience researching lexical flexibility over the last decade, I have yet to encounter a language that does not exhibit a degree of flexibility in at least some words, however marginally. The prevalence with which different areas of the grammars of the world's languages lack sensitivity to the distinctions between reference (nouns), predication (verbs), and modification (adjectives) suggests that lexical flexibility may not be so much of a problem as it is a design feature of language (CITE: Hieber forthcoming). Indeed, as discussed in Chapter 2, existing research shows that lexical flexibility is constrained and shaped by the very principles that give rise to the crosslinguistic categories of noun, verb, and adjective in the first place (CITE: Croft 2000; Croft 2005; Croft & van Lier 2012).

It is only recently that lexical flexibility has become an object of study in itself, rather

than a problem to be solved. As explained above, most prior studies aim to advance a particular analysis rather than to expand empirical coverage of the phenomenon. While they often provide numerous examples, they are neither quantitative nor comprehensive. As yet, there are only a small number of empirical investigations into the extent and nature of lexical flexibility in individual languages (let alone crosslinguistically). The following section is a brief synopsis of the existing empirical studies.

1.2 Previous research

The existing studies on the empirical extent of lexical flexibility are of two types: lexicon-based studies which examine dictionaries to determine whether words may be used for multiple functions, and corpus-based studies which examine whether and how often words are used for multiple functions in discourse.

An early lexicon-based study, though not explicitly focused on lexical flexibility, is Croft's (CITE: 1984) study of categories of Russian word roots (summarized in (CITE: Croft 1991: 66)). Croft finds that Russian roots are unmarked, or among the least marked forms, when their semantic category (object, action, or property) aligns with their discourse function (reference, predication, or modification respectively). When roots are used for discourse functions that are atypical for their meaning—in other words, when they are used flexibly—they are marked in some way (or at least as marked as their prototypical uses). These data suggest that lexical flexibility is constrained in a principled way, by what Croft calls the *typological markedness of parts of speech* (explained in detail in Chapter 2).

In arguing that Mundari is *not* a flexible language, Evans & Osada (CITE: 2005) conduct a dictionary analysis using a focused 105-word sample as well as a larger 5,000 word-sample. In the 105-word sample, 74 words (72%) could be used as either noun or verb. In the larger sample, 1,953 words (52%) could be used as both noun and verb. The complete figures for the large sample are shown in Figure X. Evans & Osada argue on the basis of these data that, because

not all the words in the Mundari lexicon are flexible, Mundari cannot be considered a flexible language. As with any whole-language typology, however, this is an oversimplification. To overlook the flexibility of these words ignores the behavior of a vast portion of the lexicon. It is exactly this behavior which is of interest in this thesis. Evans & Osada's study constitutes an important contribution to our knowledge of the empirical extent of lexical flexibility across languages.

Table 1.1: Percentage of words used as nouns, verbs, or both in Mundari (Austroasiatic > Munda; India) (Evans & Osada 2005: 383)

noun only	772	20%
verb only	1,099	28%
noun and verb	1,953	52%
Total	3,824	100%

(CITE: Add Evans & Osada [2005: 383] citation to caption in this table.)

Creissels (CITE: 2017) is a careful lexicon-based of flexibility in Mandinka (Mande; West Africa). While Mandinka has nominal and verbal constructions that allow the predicative and referring functions of words to be distinguished unambiguously, it is not as easy to separate word stems themselves into similar classes, owing to the fact that no Mandinka lexemes are used exclusively in verbal constructions—all Mandinka lexemes may occur in nominal constructions as well. While Creissels does not dispute this fact, he shows that there is a crucial distinction to be made between two classes of word stems: 1) those whose nominal use is predictable and therefore analyzable as a case of "morphologically unmarked nominalization" (zero-marked conversion) from one category (verb) to another (noun)—these are always event nominalizations; and 2) those whose meaning in nominal constructions is idiosyncratic and therefore not predictable. Creissels calls the former *verbal* words and the latter *verbo-nominal*. He states that both word classes exhibit categorial flexibility, just of different natures. There is also a small set of nominal words used marginally as verbs. These cases are always semantically predictable. Even individual senses of a word can sometimes show varying behavior as to their flexibility. Although Creissels' study unfortunately does not provide counts of the

different stem types, it nonetheless adds to our understanding of lexical flexibility by showing how it may have varied realizations, within a single language or even a single word.

Mithun (CITE: 2017: 163) also conducts a lexicon-based analysis of words roots? stems? in Central Alaskan Yup'ik (Alaska; Eskaleaut) using Jacobson's (CITE: year?) exhaustive dictionary, and shows that only a small minority of roots? stems? (12%) exhibit flexibility and can be used as both nouns and verbs. The results of this study are shown in Table X. The words in these groups cannot be characterized in any general or semantic way (CITE: Mithun 2017: 163). Mithun's finding that flexibility in Yup'ik is rather marginal is surprising given that Yup'ik was the focus of an extensive debate about whether the language distinguished nouns and verbs (CITE: just cite Jacobson here). The fixation with these marginal cases in the literature seems disproportionate to their actual frequency of occurrence, again illustrating the disconnect between research advancing a particular analysis and research aiming to improve empirical coverage of the phenomenon. Just as with Mundari, however, it would be an oversight to simply ignore these flexible cases. What accounts for the large difference in the extent of flexibility in the lexicons of Mundari versus Yup'ik?

Table 1.2: Percentage of words used as nouns, verbs, or both in Central Alaskan Yup'ik (Eskaleaut > Eskimo; Alaska) (Mithun 2007: 163)

noun only	35%
verb only	53%
noun and verb	12%
Total	100%

(CITE: Add Mithun [2017: 163] citation to caption in this table.)

In summary, existing lexicon-based studies have yielded a range of results, each contribution to our understanding of lexical flexibility, but there are still too few such studies to draw any general conclusions as of yet.

Corpus-based studies of lexical flexibility are also scarce.

1.3 Overview of this study

Todo list