

## Lexical flexibility in discourse

expected page count, excluding examples and figures: 115

### ABSTRACT (2 pages)

1. a brief statement of the problem of lexical flexibility
2. a description of the data and methods used to address this problem
3. a condensed summary of the findings

### 1 INTRODUCTION (20 pages)

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 Hook & Significance

Word classes (such as noun, verb, and adjective), which were once thought to be universal, easily identifiable, and easily understood, are today 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 (Haspelmath 2018; Hieber forthcoming: 1). Lexical categorization—how languages separate words into categories—is of central importance to theories of languages because it is tightly interconnected with linguistic categorization more generally, which in turn informs (and is informed by) our understanding of cognition. Categorization is a fundamental feature of human cognition (Taylor 2003: xi), and lexical categorization is perhaps the most foundational issue in linguistic theory (Vapnarsky & Veneziano 2017: 1).

#### 1.2 General Problem

Lexical flexibility presents an especially difficult challenge to our understanding of parts of speech.

1. provide a superficial definition of lexical flexibility
2. give examples of lexical flexibility from several languages
3. explain the problem this creates for traditional understandings of word classes

#### 1.3 State of the Field

##### 1.3.1 Overview

1. Lexical flexibility has only recently become an object of study in itself. We are only just beginning to understand the extent and kinds of lexical flexibility.
2. Historical research on lexical categories was of primarily two types:

- a. Descriptive studies arguing for or against the existence of a particular category in a particular language.
  - i. Iroquoian, Cayuga, Quechua, numerous others
- b. Theoretical studies concerned with fitting example facts from specific languages into the researcher's particular theory of categories.
3. These studies treat lexical flexibility as a problem to be solved or explained away, because flexible words do not fit tidily into the kinds of clearly-delineated, Aristotelian categories that were assumed to exist in all languages.
4. Recent work attempts to refine our theories of lexical categories in order to account for the existence of lexical flexibility.

### 1.3.2 What we know

#### 1.3.2.1 Summary of Lexicon-Based Studies

- Evans & Osada (2005)
- Creissels (2017)
- Mithun (2017: 163)

#### 1.3.2.2 Summary of Corpus-Based Studies

There are some studies which count the proportion of nouns vs. verbs vs. adjectives in texts. However, no corpus-based studies have to my knowledge focused explicitly on lexical flexibility.

- Hudson. 1994. About 37% of word-tokens are nouns
- Polinsky & Magar. 2020. Headedness and the lexicon: The case of verb-to-noun ratios

#### 1.3.2.3 Summary of Other Relevant Research

What we know from both cognitive science and historical linguistics suggests that it would be surprising for clear-cut categories to exist. Word meanings, lexical categories, and mental categories are all prototypical, and language change is gradual and gradient. Given these facts, we should expect lexical flexibility as the default. The fact to be explained is how discourse functions come to be grammaticalized over time, not why it is that some languages lack such distinctions in areas of their grammars.

### 1.3.3 What we don't know

1. How common is lexical flexibility? Is it overrepresented or underrepresented in the literature?
2. How does lexical flexibility vary from one language to another? "Are there significant cross-linguistic variations in the way parts of speech are realized?" (Vapnarsky & Veneziano 2017: 4)
3. With the few exceptions mentioned above, there are no studies documenting the empirical extent of flexibility within a lexicon or within a corpus.

## 1.4 Specific Problem

We don't have a good picture of what lexical flexibility looks like, or how it varies across languages. We have to understand the way categories are structured before we can begin to explain why they are structured that way. Once we have such empirical data, we can ask whether the evidence should incline us towards certain theories of lexical categories over others.

For example, if a language were discovered whose lexemes all occurred with equal frequency in referring constructions, predicative constructions, and modifying constructions, with no overt derivation, this would provide strong evidence for the existence of truly flexible categories. On the other hand, if even languages generally considered highly flexible turn out to be quite constrained in the extent of their flexibility, this suggests that maybe a few exceptional cases have caught the attention of linguists, but that lexical categories are otherwise fairly well-behaved and simply admit of the occasional exception.

Without making claims as to the categorical status or lexical unity/distinctiveness of lexical items, we should be able to provide a descriptive survey of flexible words and their behavior.

## 1.5 Purpose Statement

This study puts corpus-based data to the task of answering the above questions. This thesis is a quantitative corpus-based study of lexical flexibility in English (Indo-European) and Nuuchahnulth (Wakashan). *Lexical flexibility* is defined here as the capacity of a lexical item to serve in more than one discourse function—reference, predication, or modification—with no overt derivational marker of these functions. This is the first corpus-based examination of lexical flexibility.

## 1.6 Research Questions & Hypotheses

This thesis is primarily exploratory and descriptive, with some minimal testing of specific hypotheses.

### 1.6.1 Major Research Questions

1. How flexible are words within English and Nuuchahnulth?
2. How flexible are English and Nuuchahnulth overall?
3. Are higher frequency (or more evenly dispersed) lexemes more flexible than lower frequency (or less evenly dispersed) ones?

### 1.6.2 Minor Research Questions

1. English: How much does a strict vs. broad interpretation of flexibility affect the flexibility rating for English?
2. Nuuchahnulth: How does the flexibility of roots compare to that of stems?
3. [other research questions may be added, given time]

## 1.7 Theoretical Framework

This study aims to be framework-neutral in the sense of Haspelmath (2008). Its findings should be interpretable and of interest to researchers working in a range of linguistic theories and with different approaches to lexical categories. My own perspective on language is decidedly functional, with an orientation towards the interplay of cognition, discourse, and diachrony. I adopt a constructional approach to lexical categories and language structure generally, and I take word classes to be typological categories rather than categories of particular languages (see Section {{XX}}). These perspectives inform my interpretation of the data in this study, but do not influence my treatment and coding of the data themselves. The relevant factors in this study are operationalized in a theory-neutral way, and I expect that my coding decisions for individual data points will be found largely unobjectionable.

## 1.8 Delimitations & Assumptions

1. methodological limitations regarding how I coded the data
2. practical limitations regarding the size of the corpora (both large and small)

## 1.9 Definitions

Not all (perhaps none) of the definitions in this section will actually be introduced here. They'll be introduced in other places as they become relevant.

- word classes / parts of speech
- lexical vs. functional categories
- lexical flexibility / polycategoriality / polyfunctionality / heterosemy / hybridity
- acategoriality / precategoriality / underspecificity / vagueness
- lexical item / archlexemes
- pragmatic functions (reference, predication, modification) / propositional act functions
- categorial vs. categorical
- prototypical vs. prototypal
- conversion / zero-derivation / functional shift
- grammaticalization vs. grammaticization

## 1.10 Data & Methods

1. criteria for selecting languages
2. languages chosen and description of corpora used

## 1.11 Results

1. English
2. Nuuchahnulth
3. English vs. Nuuchahnulth
4. General Results

## 1.12 Contribution

- Methodological: Procedure for quantifying lexical flexibility that can be replicated with consistent results for other languages / corpora.
- Empirical: Introduces a new kind of data (quantitative corpus-based data) to research on lexical flexibility.
- Descriptive: Confirm / clarify the behavior of lexical flexibility in English and Nuuchahnulth specifically.
- General: Show that lexical flexibility is actually quite prevalent, and can vary drastically between languages.

## 1.13 Organization of this Study

## 2 BACKGROUND (47 pages)

The primary focus of this chapter is to explain the concept of lexical flexibility and consider its criticisms. I first briefly describe how flexible approaches to lexical categories developed as a response to weaknesses in traditional theories of parts of speech, and then survey the landmark studies and important findings on lexical flexibility. While lexical flexibility has recently become the focus of numerous descriptive studies, empirical coverage of the phenomenon is still limited.

In particular, there have been few quantitative approaches to the topic, and none which examine natural discourse data from corpora (that I am aware of). The focus of this study will be to extend the empirical coverage of lexical flexibility to these domains and give an analysis of the findings and their implications.

In the second half of this chapter, I present criticisms and difficulties for flexible approaches to lexical categories from various functional perspectives—construction grammar, typology, and cognitive linguistics. I conclude by offering a revised conceptualization of lexical flexibility which is more in line with this functional research.

## **2.1 Traditional Approaches to Parts of Speech (10 pages)**

### **2.1.1 Universalist**

Historically and still presently, categories like noun and verb were assumed to be basic and therefore not in need of explication or further examination (Bolinger & Sears 1981: 81; Payne 1997: 32; Schachter & Shopen 2007: 1-2; Stassen 2011: 95). These categories were thought to be basic in the sense of either a) being universally instantiated in all languages, or b) being universal categories available to all languages, but only instantiated in some (Hieber 2013: {{p. ??}}).

- classical descriptions of Greek, Latin, and Sanskrit
- colonial / missionary era (Basque, Celtic, American languages)
- generativist approaches
- Basic Linguistic Theory (“grab bag” approach)

### **2.1.2 Relativist**

- the Boasian turn (describing each language on its own terms)
- Sapir: "Each language has its own scheme." (1921: 125)

### **2.1.3 Problems & Critiques (Hieber forthcoming)**

## **2.2 Flexible Approaches to Parts of Speech (20 pages)**

Lexical flexibility is a longstanding issue in Americanist linguistics, going back to the earliest survey of North American languages (Gallatin 1836: 175–176). The concept of flexible word classes arose in response to the many debates and difficulties which linguists were consistently encountering in their attempts to describe the lexical categories of individual languages as well-defined and mutually exclusive.

### **2.2.1 Key Concepts & Findings**

- 2.2.1.1 Lexical Flexibility (Ross 1972; Broschart 1991; Hengeveld 1992; Launey 1994, 2004; Stassen 2004; Rijkhoff & Hengeveld 2005)
- 2.2.1.2 Monocategoriality (Gil 1994, 1995, 2005)
- 2.2.1.3 Criteria for Flexibility (Evans & Osada 2005)
- 2.2.1.4 Descriptive Studies (Lois & Vapnarsky 2003; van Lier 2006; Vogel & Comrie 2011; Rijkhoff 2011; Rijkhoff & van Lier 2013; Simone 2014; van Lier 2017; Vapnarsky & Veneziano 2017; Mithun 2019)
- 2.2.1.5 Locus of Categoriality

The locus of categoriality (root, stem, inflected word, discourse context) varies from language to language (and even word to word). (Not all researchers agree with this fact. Some claim that the

locus of categoriality lives at one level or the other: Baker 2003; 2015; Booij & Audring 2018; Distributed Morphology; Minimalism.)

## 2.2.2 Problems & Critiques

### 2.2.2.1 Item-Specific Knowledge / Semantic Shift

### 2.2.2.2 Lexicalization / Cognitive Processing Efficiency

## 2.3 Functional Approaches to Parts of Speech (15 pages)

### 2.3.1 Cognitive Approaches

- Prototype Theory (Rosch; Lakoff; Taylor)
- Cognitive Linguistics (Lakoff; Croft & Cruse; Evans)

### 2.3.2 Constructional Approaches

- Radical Construction Grammar (Croft)

### 2.3.3 Typological Approaches

- the problem of crosslinguistic comparability (Haspelmath)
- typological markedness theory (Croft)

### 2.3.4 Summary

The field of linguistics and typology especially is undergoing a radical shift in our understanding of lexical categories.

- universal > language-specific > typological
- categorical > prototypical > construction-specific / item-specific

## 2.4 Lexical Flexibility: A Conceptual Definition (2 pages)

How then is the notion of lexical flexibility still tenable or of interest, given the critiques of methodological opportunism presented by Croft?

- What are the range of phenomena that exist within Croft's "at least as marked as" formulation of categorial prototype theory?
- Flexibility can be thought of as the non-grammaticization of pragmatic functions—the areas of the grammar where pragmatic functions are not given dedicated constructions.
- Flexibility is best thought of as a historical process rather than a synchronic one.

## 3 DATA & METHODS (17 pages)

This chapter describes the data used and how those data were analyzed. It covers the selection criteria for both languages and lexemes, which corpora were used, and how the data were obtained and formatted. I also describe the methods used to annotate the data, and factors that influenced how the data were coded. I present and explain a measure of corpus dispersion that is used partly in place of, partly as a complement to, raw frequencies of lexemes. Lastly, I set forth a procedure for operationalizing and quantifying lexical flexibility in a crosslinguistically comparable way. The formulation of this lexical flexibility measure is a key methodological contribution of the thesis.

**3.1 Introduction** (3 pages)

## 3.1.1 Data Workflow

## 3.1.2 Principles

1. naturalistic discourse data
2. scientific accountability and reproducibility (data accessibility)

## 3.1.3 Steps for Replicating this Study (GitHub)

**3.2 Data** (4 pages)

## 3.2.1 Languages

## 3.2.1.1 English

## 3.2.1.2 Nuuchahnulth

## 3.2.2 Samples

## 3.2.2.1 100 Lexeme Sample

## 3.2.2.2 Small Corpus Sample (&lt; 10,000 tokens)

## 3.2.2.3 Criteria for Inclusion

**3.3 Methods** (4 pages)

## 3.3.1 Annotation Procedure

## 3.3.2 Annotation Decisions

**3.4 Data Analysis** (6 pages)

## 3.4.1 Frequency &amp; Dispersion

## 3.4.2 Measuring Lexical Flexibility

**4 RESULTS** (14 pages)

This chapter details the findings of the procedures described in Chapter 3: Data & Methods. I begin with a brief examination of several representative lexical items from English and Nuuchahnulth, and explain to the reader how to interpret the ternary plots used to present results in this thesis. I then take a broader look at the behavior of lexical flexibility in English and Nuuchahnulth, first individually and then in comparison. The major findings are as follows:

- Almost all lexemes of English exhibit a degree of flexibility. However, the degree of that flexibility is typically small.
- The behavior of lexemes in English mostly approximates what would be expected for prototypical nouns, verbs, and adjectives, with some marginal deviation from that prototype for most words.
- Lexemes in Nuuchahnulth show a great deal of noun-verb flexibility, but very little flexibility in the adjective direction.
- Lexemes in Nuuchahnulth are generally highly flexible.
- Roots and stems in Nuuchahnulth exhibit almost identical behavior in terms of their lexical flexibility.
- Nuuchahnulth is on average more flexible than English. (pending)
- English has more words which exhibit flexibility than Nuuchahnulth. (pending; caveats: sample size; potential correlation between flexibility and frequency)

- There is a small but robust correlation between lexical flexibility and frequency/dispersion in both English and Nuuchahnulth. (pending)

#### **4.1 Introduction** (3 pages)

#### **4.2 English** (4 pages)

#### **4.3 Nuuchahnulth** (4 pages)

#### **4.4 English vs. Nuuchahnulth** (1 page)

#### **4.5 General Results** (2 pages)

### **5 DISCUSSION & CONCLUSION** (15 pages)

This chapter summarizes the methods and main findings of this study, and then considers the implications of those results for theories of lexical categories. I argue that the data provide compelling evidence in favor of functional approaches to lexical categorization, most especially prototype theory and Croft's theory of lexical categories as typological markedness patterns. I also argue for a reversal of the canonical position on parts of speech: instead of working from the default assumption that all languages have clearly-defined or even loosely-defined parts of speech, we should begin from the understanding that dedicated referring, predicating, or modifying constructions must develop diachronically, and that even when they do, they do not do so for the entire lexicon, or in all areas of the grammar equally. Even languages like English whose lexemes pattern strongly with the standard prototypes of noun, verb, and adjective nonetheless exhibit varying degrees of flexibility for different lexemes. Lexical categories are not a given in grammar.

I conclude by discussion limitations of the present study and avenues for future research, followed by closing remarks.

#### **5.1 Summary** (3 pages)

#### **5.2 Discussion** (8 pages)

#### **5.3 Limitations** (1 page)

#### **5.4 Future Research** (1 page)

#### **5.5 Conclusion** (1 page)

### **APPENDIX**

The appendix contains lists of the 100 lexemes analyzed in English and Nuuchahnulth, and their accompanying statistics.