University of California, Santa Barbara

Lexical flexibility in discourse:

A quantitative corpus-based approach

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of

Philosophy in Linguistics

by

Daniel W. Hieber

Committee in Charge:

Professor Marianne Mithun, Chair

Professor Bernard Comrie

Professor Stefan Th. Gries

Professor William Croft (University of New Mexico)

January 2021

The dissertation of Daniel W. Hieber is approved.

_____

Bernard Comrie

_____

Stefan Th. Gries

_____

William Croft

_____

Marianne Mithun, Committee Chair

January 2021

Lexical flexibility in discourse:

A quantitative corpus-based approach

Copyright © 2020

by

Daniel W. Hieber

Dedication

ACKNOWLEDGMENTS

Curriculum Vitae
Daniel W. Hieber

Education

December 2020	Ph.D. in Linguistics, University of California, Santa Barbara
March 2016	M.A. in Linguistics, University of California, Santa Barbara
June 2008	B.A. in Linguistics & Philosophy, The College of William & Mary (magna cum laude)

Professional Experience

2018–2019	Editor, Custom Language Products, Rosetta Stone
2015–2017	Teaching Assistant, Department of Linguistics, University of California, Santa Barbara
2014–2015	Research Assistant (under Prof. Carol Genetti), Department of Linguistics, University of California, Santa Barbara
2011–2013	Associate Researcher, Research Labs, Rosetta Stone
2008–2011	Editor, Endangered Languages Program, Rosetta Stone
2007–2008	Intern, Endangered Languages Program, Rosetta Stone
2006	Spanish Instructor, Nielsen Builders
2004–2006	Lab Assistant, Language Lab, The College of William & Mary
2003–2004	Latin Instructor, Bridgewater Home School Unit

Publications

2019	The Chitimacha language: A history. In Nathalie Dajko & Shana Walton (eds.), *Languages in Louisiana: Community & culture* (America's Third Coast Series). University Press of Mississippi.
2019	Semantic alignment in Chitimacha. *International Journal of American Linguistics* 85(3): 313–363. doi:10.1086/703239.
2018	Category genesis in Chitimacha: A constructional approach. In Kristel Van Goethem, Muriel Norde, Evie Coussé, & Gudrun Vanderbauwhede (eds.), *Category change from a constructional perspective* (Constructional Approaches to Language 20), 15–46. John Benjamins. doi:10.1075/cal.20.02hie.
2016	*The cohesive function of prosody in Ékegusií (Kisii) narratives: A functional-typological approach.* M.A. thesis, University of California, Santa Barbara.

2013	On linguistics, language, and our times: A linguist's narrative reviewed. *Linguistic Typology* 17(2): 291–321. Review article of *I am a linguist* by R. M. W. Dixon (Brill, 2010). doi:10.13140/RG.2.2.13238.96329.
2013	(with Sharon Hargus & Edward Vajda, eds.) *Working papers in Athabaskan (Dene) languages 2012.* Alaska Native Language Center Working Papers 11. ANLC.

Awards

2019	SSILA Best Student Presentation Award
2015	National Science Foundation (NSF) Graduate Student Research Fellowship (GRFP)
2015	2nd place, University of California Grad Slam
2015	Winner, University of California, Santa Barbara Grad Slam
2013	Chancellor's Fellowship, University of California, Santa Barbara
2006	Boren Scholarship, National Security Education Program (NSEP)

Fields of Study

Major Fields: Linguistic Typology, Language Documentation & Description, Language Revitalization, Prosody, Discourse, Language Change, Language Contact, Digital Linguistics, Corpus Linguistics

Linguistic Typology with Professor Bernard Comrie & Professor Marianne Mithun

Language Documentation & Description with Professor Eric Campbell, Professor Carol Genetti, & Professor Marianne Mithun

Language Revitalization with Professor Carol Genetti

Prosody with Professor Carol Genetti, Professor Matthew Gordon, & Professor Marianne Mithun

Discourse with Professor Patricia Clancy, Professor John W. DuBois, Professor Carol Genetti, & Professor Marianne Mithun

Language Change with Professor Marianne Mithun

Language Contact with Professor Marianne Mithun

Digital Linguistics with Professor Eric Campbell & Professor Stefan Th. Gries

Corpus Linguistics with Professor Eric Campbell, Professor John W. DuBois, & Professor Stefan Th. Gries

Abstract

Lexical flexibility in discourse:

A quantitative corpus-based approach

by

Daniel W. Hieber

This thesis is a quantitative corpus-based study of lexical flexibility in English (Indo-European) and Nuuchahnulth (Wakashan). *Lexical flexibility* is the capacity of lexical items to serve in more than one discourse function—reference, predication, or modification (or more traditionally, noun, verb, or adjective). In this thesis I develop a procedure and metric for quantifying the lexical flexibility of words in a corpus, and apply that metric to English and Nuuchahnulth. I find that the two languages differ drastically in not only their degree of lexical flexibility, but the way in which that flexibility is realized. This study advances the discussion of lexical flexibility—as well as parts of speech more generally—by adding a new kind of empirical evidence to the discussion (quantitative corpus-based data), and in doing so provides answers to several longstanding and much-debated questions about how lexical categories operate in English and Nuuchahnulth.

The abstract should include 1) a brief statement of the problem; 2) a description of the methods and procedures used to gather data or study the problem; 3) a condensed summary of the findings. The abstract should be double-spaced. The recommended length is 1–2 pages. (add Abstract)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

The following table provides the meaning of each abbreviation used in interlinear glossed examples throughout this thesis.

1	first person
2	second person
3	third person
ACC	accusative
CAUS	causative
COMPL	completive
DEF	definite
FIN	finite
HUM	human
INCEP	inceptive
IND	indicative
MOM	momentaneous
NEG	negative
PAST	past
PFV	perfective
PL	plural
SG	singular
SPEC	specific
SUBJ	subject
TR	transitive

List of Languages

The following table provides information about each language mentioned in this thesis: the name of the language in English, the International Standards Organization (ISO) 693-3 language code, and the Glottolog code (Hammarström, Forkel & Haspelmath 2019). Genealogical information follows the format `family > phylum`.

Language Name (English)	ISO 639-3	Glottocode	Genetic Affiliation
Ancient Greek	grc	anci1242	Indo-European > Hellenic
Cayuga	cay	cayu1261	Iroquoian > Northern Iroquoian
Central Alaskan Yup'ik	esu	cent2127	Eskimo-Aleut > Yupik
Chamorro	cha	cham1312	Austronesian > Malayo-Polynesian
Classical Nahuatl	nci	clas1250	Uto-Aztecan > Nahuan
English	eng	stan1293	Indo-European > Germanic
Latin	lat	lati1261	Indo-European > Italic
Kutenai	kut	kute1249	isolate
Mandarin	cmn	mand1415	Sino-Tibetan > Sinitic
Mandinka	mnk	mand1436	Mande > Manding
Mundari	unr	mund1320	Austroasiatic > Munda
North Efate	llp	nort2836	Austronesian > Oceanic
Nuuchahnulth	nuk	nuuc1236	Wakashan > Southern Wakashan
Quechua	qwe	quec1387	Quechuan
Quiché Maya	quc	kich1262	Mayan > Quichean

Riau Indonesian	ind	indo1316	Austronesian > Malayan
Russian	rus	russ1263	Indo-European > Balto-Slavic
Sundanese	sun	sund1251	Austronesian > Malayo-Polynesian
Tagalog	tgl	taga1280	Austronesian > Philippine
Tongan	ton	tong1325	Austronesian > Polynesian
Ute	ute	utee1244	Austronesian > Polynesian

Conventions

This note documents the conventions I have adopted regarding linguistic data, terminology, and presentation of data throughout this thesis.

Interlinear Examples

It is well known that the world's languages realize widely different sets of morphosyntactic categories (Whaley 1997: 58; Haspelmath 2007). Moreover, even when these categories bear the same name, they may differ drastically in their behavior (Dixon 2010: 9). It is the subject of much debate whether these language-specific categories can be mapped onto each other or compared in any useful way (Croft 1995; Song 2001: 10–15; Croft 2003: 13–19; Haspelmath 2010a,c; Newmeyer 2010; Stassen 2011; Hieber 2013: 308–310; Croft 2014; Plank 2016; Song 2018: 44–58). Recognizing these difficulties, I have made no attempt to standardize the linguistic terminology used in examples from different languages. I have, however, standardized the abbreviations used to refer to those terms. For example, even though one researcher may abbreviate Subject as SUBJ and another researcher abbreviate it as SUB, I nonetheless gloss all Subject morphemes as SUBJ. See the List of Abbreviations for a complete list of glossing abbreviations.

I have not attempted to standardize the transcription systems and orthographies used in examples. All examples are given as transcribed in their original source. The reader should consult those original sources for further details regarding orthography.

In all interlinear glossed examples, I follow the formatting conventions (but not necessarily the recommended abbreviations) of the Leipzig Glossing Rules (Bickel, Comrie & Haspelmath 2015). The source of each example is always provided after the example itself.

Prose

It is increasingly common in typological studies to write language-particular terms and categories with an initial capital letter, and to write terms that refer to language-general or semantic/functional concepts (e.g. the crosslinguistic notion of subject) in lowercase (Comrie 1976: 10; Bybee 1985: 47 (fn. 3), 141; Croft 2000: 66; Haspelmath 2010a: 674; Croft 2014: 535). For example, the English Participle suffix *-ing* is, obviously, specific to English, and does not exist in any other language; therefore it is capitalized and written as *Participle*. If, however, a writer is discussing the category of participles generally and crosslinguistically, not specific to any particular language, the term is written in lowercase as *participle*. I follow these same capitalization conventions in this thesis.

When a language is discussed in the text, its genealogical information is provided in parentheses, following the format `family > phylum`. For example, Central Alaskan Yup'ik would be listed as "Central Alaskan Yup'ik (Eskimo-Aleut > Yupik)". A complete list of languages mentioned in this thesis, along with their ISO 639-3 codes, Glottolog codes, and genetic information is in the List of Languages. Each language is also included in the Language Index. Language names are given in English following Haspelmath (2017). Language information is taken from the Glottolog database (Hammarström, Forkel & Haspelmath 2019).

Quotations

Within quotations, *italics* indicate emphasis in the original, while **boldface** indicates my emphasis.

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 addressed 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 > Germanic) and Nuuchahnulth (Wakashan > Southern Wakashan). It is the first study to examine lexical flexibility using natural discourse data from corpora. 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 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 (Mithun 2017: 166; Haspelmath 2019; Hieber forthcoming). Lexical categorization—how languages assign words to categories—is of central importance to theories of language because it is tightly intercon-

nected with linguistic categorization 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 (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 with no overt derivational morphology, whether it is used to refer (like a noun), to predicate (like a verb), or to modify (like 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 §2.5.) Examples of flexible words in several languages are shown below. In the examples, **N** stands for a word being used nominally, **V** for a word being used verbally, and **A** for a word being used adjectivally. The flexible word in each set of examples is shown with **emphasis**.

(1) English (Indo-European > Germanic)
 N: And the spots of **paint** would change every hundred degrees.
<div align="right">(Ide & Suderman 2005: FrancisClem)</div>

 V: One story does come to my mind though where you **painted** the foundation coating on the house and got tar all over you.
<div align="right">(Ide & Suderman 2005: BorelRaymondHydellII)</div>

 A: And it happened to be one of the rare **paint** jobs.
<div align="right">(Ide & Suderman 2005: sw2236)</div>

(2) Mandinka (Mande > Manding)
 N: **Kuuráŋ**-o   mâŋ   díyaa.
   sick-DEF   PFV.NEG   pleasant
   'Sickness is not pleasant.' <div align="right">(Creissels 2017: 46)</div>

 V: Díndíŋ-o   máŋ   **kuraŋ**.
   child-DEF   PFV.NEG   **sick**
   'The child is not sick.' <div align="right">(Creissels 2017: 46)</div>

(3) Mundari (Austroasiatic > Munda)
 N: **buru**=ko   bai-ke-d-a.
   mountain=3PL.SUBJ   make-COMPL-TR-IND
   'They made the mountain.' <div align="right">(Evans & Osada 2005: 354)</div>

V: saan=ko   **buru**-ke-d-a.
firewood=3PL.SUBJ   mountain-COMPL-TR-IND
'They heaped up the firewood.'   (Evans & Osada 2005: 355)

(4)   <u>Nuuchahnulth (Wakashan > Southern Wakashan)</u>

N: watqšiƛ   ʔaƛimt   ...
watq-ši(ƛ)   **ʔaƛa**-imt   ...
swallow-MOM   **two**-PAST   ...
completely.swallowed   two   ...
'He swallowed two of them [...]'   (Louie 2003: Qawiqaalth 57)

V: wiḱaƛ   haʔukšiƛ   ʔaƛiičiƛ
wik-'aƛ   haʔuk-ši(ƛ)   **ʔaƛa**-'i·čiƛ
not-FIN   eat-MOM   **two**-INCEP
didn't   ate   became.two
'He (Mink) didn't eat them and the crabs became two.'   (Louie 2003: Mink 266)

A: hiiɬtqyaaṗup   ʔaƛa   qʷayaḋiik
hiɬ-tqya·ṗi-up   **ʔaƛa**   qʷayaḋi:k
there-back-MOM.CAUS   **two**   wolf
put.on.the.back   two   wolf
'Two wolves put (the dead wolf) on their back.'   (Louie 2003: FoodThief 46)

(5)   <u>Quechua (Quechuan)</u>

N: rikaška: **hatun**-(kuna)-ta
I.saw   **big**-(PL)-ACC
'I saw the big one(s)'   (Schachter & Shopen 2007: 17)

V: chay   runa   **hatun**   (kaykan)
that   man   **big**   is
'that man is big'   (Schachter & Shopen 2007: 17)

A: chay   **hatun**   runa
that   **big**   man
'that big man'   (Schachter & Shopen 2007: 17)

(6)   <u>Tongan (Austronesian > Polynesian)</u>

N: na'e   lele   e   kau   **fefiné**
PAST   run   SPEC   PL.HUM   **woman**.DEF
'The women were running.'   (Broschart 1997: 134)

V: na'e   **fefine**   kotoa   e   kau   lelé
PAST   **woman**   all   SPEC   PL.HUM   run.DEF
'The ones running were all female.'   (Broschart 1997: 134)

(7) Central Alaskan Yup'ik (Eskimo-Aleut > Yup'ik)

	a.	*iqa-*	'dirt'; 'be dirty'
		-ngtak	'very'
	N:	***iqa-ngtak***	'one that is very dirty'
	V:	***iqa-ngtaq-***	'be very dirty'

(Mithun 2017: 159)

	b.	*tangerr-*	'see'
		-uaq	'imitation, inauthentic'; 'pretend to, without serious purpose'
	N:	***tangerr-uaq***	'movie, vision, hallucination'
	V:	***tangerr-uar-***	'hallucinate, watch a movie'

(Mithun 2017: 159)

	c.	*iqeq-*	'corner of mouth'
		-mik	'thing held in one's mouth'; 'to put in one's'
	N:	***iq-mik***	'chewing tobacco'
	V:	***iq-mig-***	'put in one's mouth'

(Mithun 2017: 160)

In the English example in (1), the predicative use of *paint* takes the English Past Tense suffix *-ed* like any prototypical verb in English, but there is no morpheme present that explicitly converts the word from noun to verb (or vice versa). The remaining examples illustrate the same situation for a variety of language families around the world. Even though in some cases there is inflectional morphology indicating the function of the word, none of these examples have explicit derivational morphology converting the target words from one function to another.

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 based on 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 classifying them 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 (Baker [2003]; Dixon & Aikhenvald [2004]; Palmer [2017]; Floyd [2011] for Quechua; Chung [2012] for Chamorro). 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 (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 (Evans & Osada 2005). Another approach is to say that languages exhibiting flexibility have only some of the traditional categories. A notable example of this is Launey's (Launey 1994; 2004) analysis of Classical Nahuatl (Uto-Aztecan), which he calls an *omnipredicative* language. In this analysis, all lexical words are predicates, so there is just one giant class of verbs.

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 class such as "flexibles", "contentives" or "non-verbs", etc. (Hengeveld & Rijkhoff 2005; Luuk 2010). Other researchers abandon the commitment to word classes entirely. Mandarin (Sino-Tibetan > Sinitic), Tagalog (Austronesian > Philippine), Tongan (Austronesian > Polynesian), Riau Indonesian (Austronesian > Malayo-Polynesian), and Proto-Indo-European have each been analyzed as lacking parts of speech by some researchers (see Simon [1937], McDonald [2013], and Sun [2020] for discussions of early analyses of Mandarin; Gil [1993] for Tagalog; Broschart [1997] for Tongan; Gil [1994] for Riau Indonesian; Kastovsky [1996] for Proto-Indo-European). Within generative linguistics, the Distributed Morphology framework takes it as an assumption that all word roots are category-neutral (Siddiqi 2018). In a more functionalist orientation, Farrell (2001) argues that *all* instances of flexible words (which he describes as cases of "functional shift") involve roots underspecified for category.

Note that these differences in perspective do not arise from disagreements about the empirical facts. 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 (Wetzer 1992: 235; Stassen 1997: 32; Croft & van Lier 2012: 58). Examples of contested languages include Iroquoian (Chafe 2012), Mundari (Evans & Osada 2005; Hengeveld & Rijkhoff 2005), Quechua (Schachter & Shopen 2007: 17; Floyd 2011), and Sundanese (Robins 1968: 352; Hardjadibrata 1985: 62–63), with many others that could be cited as well. It is rare that an argument for flexibility is refuted by linguistic facts alone (though see Mithun's [2000] response to Sasse [1988] regarding Cayuga).

Since analyses of lexical flexibility depend more on the theoretical commitments of the researchers involved rather than any 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* (Croft 2001: 30), 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 (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 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 (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 the existence of lexical categories in a language is not necessarily a given (Hieber forthcoming).

Indeed, given what we know from both cognitive science and diachronic linguistics, it would be surprising if clear-cut categories *did* exist. Word meanings, lexical categories, and mental categories are all prototypal[1] (Taylor 2003), and language change is both gradual and gradient (Hopper & Traugott 2003; Traugott & Trousdale 2010) There will be more or less central members of any given category, and at any given point in time a word might be in a stage of transition or expansion from one category into another, meaning that it will show attributes of both. Given these facts, the real curiosity is how discourse functions come to be grammaticalized in language over time, not why it is that some languages lack such distinctions in certain areas of their grammars. Lexical flexibility is not so much a problem as it is a design feature of language. It is precisely the liminal categorial[2] status of flexible words that makes them interesting:

> In the functionalist view, linguists should recognize the boundary status of the cases in question and try to understand why they are boundary cases. The major empirical fact that has led to concrete results for typology is the discovery that the cross-linguistic variation in such things as the basic grammatical distinctions is patterned. (Croft 1991: 23)

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). What follows is a brief synopsis of the existing studies of this latter type.

---

[1]In this thesis, I use the term *prototypical* to mean 'having the properties of the prototype, exemplar, or central member of a category' and the term *prototypal* to mean 'having a prototype structure, with central and less central members'. The term *prototypal* is borrowed from the programming community, where it is used to describe programming languages (such as JavaScript) in which objects inherit properties from shared prototypes. Word classes may be described as prototypal, and their members as prototypical or non-prototypical.

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

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 (1984) study of categories of Russian (Indo-European > Balto-Slavic) word roots (summarized in 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 §2.4).

A study of Mundari (Austroasiatic > Munda), Evans & Osada (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 Table 1.1. Evans & Osada argue on the basis of these data that, because not all the words in the Mundari lexicon are flexible, Mundari is *not* 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 flexible 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.

Mithun (2017: 163) also conducts a lexicon-based analysis of words roots in Central Alaskan Yup'ik (Eskimo-Aleut > Yupik) using Jacobson's (2012) exhaustive dictionary, and shows that

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

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

only a small minority of roots (12%) are flexible, and can be used as both nouns and verbs. The results of this study are shown in Table 1.2. Mithun reports that the words in these groups cannot be characterized in any general or semantic way. 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 (Sadock 1999). 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. Instead we should ask 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 (Eskimo-Aleut > Yupik) (Mithun 2017: 163)

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

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

Corpus-based studies of lexical flexibility are also scarce. In a study of the discourse functions of property words in English and Mandarin (Sino-Tibetan > Sinitic), Thompson (1989)

reports that predicative uses of adjectives are in fact more common than attributive (modifying) uses of adjectives in conversation. The resulting figures from this study are shown in Table 1.3. Some of the attributive adjectives reported in Table 1.3 have "anaphoric head nouns" (Thompson 1989: 258), meaning that they are adjectives functioning to refer, so the figures presented are not entirely representative of the pragmatic functions of these words. The study also does not discuss the extent to which *individual* words exhibit this predicate-modifier flexibility—we only have the data in aggregate—and it also excludes any prototypical nouns being used to modify. These methodological choices are appropriate for a study of the discourse uses of prototypical adjectives, but the result is that we cannot infer much about the extent of lexical flexibility in English or Mandarin from this study.

Table 1.3: Distribution of functions of property words in English (Indo-European > Germanic) and Mandarin (Sino-Tibetan > Sinitic) (Thompson 1989: 253, 257)

	English		Mandarin	
predicative adjectives	209	86%	243	71%
attributive adjectives	34	14%	97	29%

Nonetheless, Thompson's study suggests a functional underpinning to the observed flexibility in prototypical property words. She finds that property words have primarily two functions in discourse: 1) to introduce new referents; and 2) to predicate an attribute about a referent. It is therefore no surprise that property words in some languages have their own specialized constructions, since they represent a unique mix of referring and predicating functions. However it is equally unsurprising that some languages encode property concepts using either referring or predicating constructions, since prototypical adjectives exhibit behavior related to both functions.

A similar study to Thompson's is Croft's (1991: §2.5) investigation of *textual markedness*, which refers to the fact that prototypical uses of a word are more frequent than non-prototypical uses of a word in texts (as might be expected by prototype theory; Taylor 2003: 56). Croft counts the frequency with which object, action, and property words are used for

each of the pragmatic functions of reference, predication, and modification in four languages: Quiché Maya (Mayan > Quichean), North Efate (a.k.a. Nguna; Austronesian > Malayo-Polynesian > Oceanic), and Ute (Uto-Aztecan > Numic). The resulting counts give confirmation to textual markedness theory. In all four languages, the most frequent use of words is in their prototypical function. Object words are most frequently used to refer, action words are most frequently used to predicate, and property words are most frequently used to modify. Like Thompson's (1989) study, however, we do not know these distributions for individual words. Additionally, Croft's data include cases of overtly marked uses of words in non-prototypical functions, which would not be considered instances of lexical flexibility.

Finally, there are some studies which count the proportion of nouns vs. verbs. vs. adjectives in English texts (Hudson 1994; Polinsky & Magyar 2020). Again, the data are not disaggregated to the word level, so no firm conclusions can be draw about the extent of lexical flexibility.

In sum, no existing studies examine the distribution of pragmatic functions for individual words, and limit themselves to only flexible (morphologically unmarked) cases. To my knowledge, the studies just reviewed exhaust those that take an empirical approach to determining the extent of lexical flexibility in or across languages. There are numerous additional studies of lexical flexibility, but these either a) focus on particular analyses or theories of flexible words rather than attempt to expand the empirical coverage of lexical flexibility, as mentioned earlier; or b) focus on various dimensions of the *behavior* of flexible words rather than studying the overall *prevalence* of flexibility. This point is not a criticism, but simply a recognition of a lacuna in existing research. The emergent literature which treats lexical flexibility as a phenomenon of interest in its own right and applies empirical data to the task of understanding its behavior has advanced our knowledge of the various ways lexical flexibility can be realized, and what the constraints on that variation are. Existing research shows, for example, 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 (Croft 2000;

2005; [Croft & van Lier 2012]). This literature and its many findings are reviewed in §2.3.

There is however still much to discover about lexical flexibility. Most significantly, we do not yet know the overall prevalence of the phenomenon. Most grammatical descriptions of flexibility present a relatively small set of handpicked examples, so that we do not know how representative these examples are. [Croft (2001]: 70) makes this point nicely:

> How do we know that when we read a grammar of an obscure "flexible" language X that the author of the grammar has systematically surveyed the vocabulary in order to identify what proportion is flexible? If English were spoken by a small tribe in the Kordofan hills, and all we had was a 150 page grammar written fifty years ago, might it look like a highly flexible language? ([Croft 2001]: 70)

Equally significant (and equally unknown) is whether there are any commonalities among words or languages which exhibit greater flexibility than others. These questions are relevant even if one adopts the position that flexible uses of words are truly heterosemous, related only historically. There remains the question of how such rampant heterosemy arises in the first place. Are there patterns or principles that guide the emergence of heterosemous forms? Whether one prefers to analyze this phenomenon as conversion, zero derivation, functional shift, polycategoriality, heterosemy, acategoriality, or something else, the fact is we do not yet have a strong empirical grasp of just how this phenomenon is realized in the world's languages. This thesis is a first foray into filling that empirical gap. The following section describes the contribution made by this thesis to addressing this gap, and gives an overview of the present study.

1.3   Overview of this study

This thesis is a quantitative corpus-based study of lexical flexibility in English (Indo-European > Germanic) and Nuuchahnulth (Wakashan > Southern Wakashan). It is exploratory and descriptive, with the primary goal of describing the prevalence of lexical flexibility within and across languages. The specific research questions investigated are as follows:

R1: How flexible are words in English and Nuuchahnulth?

R2: Is there a correlation between degree of lexical flexibility for a word and frequency (or corpus dispersion)?

R3: How do the semantic properties of words pattern with respect to their flexibility?

I explore each of these questions from several angles. R1, "How flexible are words in English and Nuuchahnulth?" is the core focus of this thesis. To answer it, I count the frequency with which lexical words are used for each of the three functions of reference, predication, and modification in a corpus of spoken texts for each language. Each word is given a flexibility rating from 0 to 1 based on how evenly its uses are distributed across the three functions. A rating of 0 indicates that the word is highly inflexible, with all its occurrences being used for a single function; a rating of 1 indicates that the word is maximally flexible, with its occurrences evenly distributed across the three functions. By quantifying the flexibility of each word in this way, it then becomes possible to look for statistical correlations between the flexibility of a word and other factors, such as those addressed by the other two research questions. It also enables us to answer the question of just how pervasive flexibility is in the two languages.

R2, "Is there a correlation between degree of lexical flexibility for a word and frequency (or corpus dispersion)?", uses the flexibility ratings calculated in R1 to consider whether the flexibility of a word correlates with either its overall frequency or with its corpus dispersion. *Corpus dispersion* refers to how evenly/regularly the word appears in a corpus, a measure which is thought to more accurately capture the notion of frequency of exposure (Gries 2008; forthcoming). This question has three motivations: First, some researchers have claimed or implied that all words may exhibit flexibility if you examine enough tokens of the word (Mosel & Hovdhaugen 1992: 77). If true, this would lend some empirical support to the claim that all words are to some degree flexible, or perhaps even acategorial. Second, higher-frequency words often preserve irregular or atypical forms or functions (Bybee 2007: Ch. 13), such that words with higher frequencies might be more likely to retain their non-prototypical, flexible

uses. Third, the fact that a word is flexible means that there is a wider range of constructions it can appear in. This could reasonably result in a higher overall frequency for flexible words. Each of these potential factors invite inquiry into the relationship between frequency and flexibility.

R3, "How do the semantic properties of words pattern with respect to their flexibility?", is investigated using a mix of quantitative and qualitative methods. Unlike the other two research questions, which are intended to capture the extent of flexibility in and across languages, R3 is an inquiry into the semantic *behavior* of flexible (and inflexible) words. This research question is directly motivated by Croft's (1991; 2000; 2001; forthcoming) typological markedness theory of lexical categories, which claims among other things that words used in non-prototypical functions (for example, a property word being used to refer, as a noun) will always show a semantic shift in the direction of the meaning typically associated with that function. So, if a property word is used to refer, its meaning should be more object-like than property-like; that is, it should mean something like 'an entity with the property X' rather than 'the abstract property X'. Croft's (1991) seminal work in this area provides strong empirical evidence for this semantic markedness principle, but is nonetheless somewhat preliminary. Croft himself has in various places implored linguists to investigate the lexical semantics of these functional shifts further (Croft 2005: 440; Croft & van Lier 2012: 70), but as yet little research has responded to this call (though see Rogers [2016] and Mithun [2017]). Investigating the semantic patterns that appear in cases of lexical flexibility is therefore another contribution of this thesis, addressed by question R3.

The preceding notes are simply a high-level summary of the principal research questions investigated in this thesis. A complete description of the methods used in answering each question is given in Chapter 3.

This study aims to be framework neutral in the sense of Haspelmath (2010b). Its findings should be interpretable and of interest to researchers working in a range of linguistic theories and with different approaches to lexical categories. As mentioned in §1.2, the results of this

study do not depend on whether one analyzes lexical flexibility as polycategoriality, conversion, or something else. While my own perspective on language is decidedly functional, this is of little relevance to how I coded the data, the procedures for which are described in detail in Chapter 3. The relevant factors in this study are operationalized in a theory-neutral way (to the extent such a thing is possible), and I expect that my coding decisions for individual data points will be found largely unobjectionable.

Several principles guided the choice of data used for this study. First, a self-imposed requirement for this project is that of empirical accountability and replicability. It should be possible for other researchers to apply the measure of lexical flexibility defined in Chapter 3 to new corpora, or to replicate the results of the present study on the existing dataset. As such, I only used data that were publicly available and, if possible, open access. Second, since the aim of this study is to investigate lexical flexibility in actual language *use*, I rely solely on naturalistic data from spoken texts. This has the additional advantage of abetting comparison between other, less well documented languages, since the majority of corpora of minority languages consist mainly of spoken texts. Third, I sought to examine data from languages that have featured prominently in discussions of lexical flexibility in the literature, with the intention of offering a more expansive empirical foundation for future discussions. With these principles in mind, I chose to focus this study on English and Nuuchahnulth.

English has at various times been described as both a highly flexible language with fluid category membership (Crystal 1967: 47–48; Vonen 1994; Farrell 2001: 111; Cannon 1985) and a fairly rigid language with clearly-delineated categories (Rijkhoff 2007: 710; Schachter & Shopen 2007: 4, 11, 12; Velupillai 2012: 122, 126). It is used as a point of comparison for nearly every discussion of lexical flexibility, but we do not have a clear idea of just how flexible English words are. Its inclusion in this study is therefore well justified. The data for English are from the Open American National Corpus (OANC), a 15-million word corpus of American English comprising numerous genres of both spoken and written data, all of which is open access (Ide & Suderman 2005). This study uses just the spoken portion of the corpus, consisting

of approximately 3.2 million words, which is itself composed of two distinct subcorpora—the Charlotte Narrative & Conversation Collection (or simply "the Charlotte corpus") and the Switchboard Corpus.

Nuuchahnulth (formerly referred to in the literature as Nootka) is a Wakashan language presently spoken by a hundred or so people on and around Vancouver Island, British Columbia, in the Pacific Northwest. Nuuchahnulth, together with the other members of the Wakashan family (especially Makah and Kwak'wala / Kwakiutl) is one of the widely discussed languages in the literature on lexical flexibility (Swadesh 1939; Jacobsen, Jr. 1979; Braithwaite 2015). This is due largely to the following examples of flexible words from Swadesh (1939).

<u>Nuuchahnulth (Wakashan > Southern Wakashan)</u>

(8) a. qo·ʔas-ma   ʔi·ḥ-ʔi
         man-3SG.IND   large-DEF
         'The large one is a man.'   (Swadesh 1939: 78)

      b. ʔi·ḥ-ma   ʔo·ʔas-ʔi
         large-3SG.IND   man-DEF
         'The man is large.'   (Swadesh 1939: 78)

(9) a. mamo·k-ma   ʔo·ʔas-ʔi
         work-3SG.IND   man-DEF
         'The man is working.'   (Swadesh 1939: 78)

      b. ʔo·ʔas-ma   mamo·k-ʔi
         man-3SG.IND   work-DEF
         'The working one is a man.'   (Swadesh 1939: 78)

Hardly a single typological survey of lexical categories or study of lexical flexibility has failed to include these examples since. Yet we still do not know how representative these examples are of Nuuchahnulth in general. What is more, lexical flexibility is an areal feature of the entire Pacific Northwest. The nearby Chimakuan, Chinookan, Coosan, Sahaptian, Salishan, and Tsimshianic families as well as the isolate Kutenai each exhibit lexical flexibility to a presumably strong degree, since they have caught the attention of so many researchers in this regard (Chimakuan: Andrade [1933: 179]; Chinookan: Duncan, Switzler & Zenk

[2023]; Coosan: Frachtenberg [1922: 318]; Sahaptian: Wetzer [1996: 142]; Salishan: Kuipers [1968], Hébert [1983], Kinkade [1983], van Eijk & Hess [1986], Jelinek & Demers [1994], Mattina [1996], Beck [1999: 135–169], Montler [2003], Beck [2013], Davis, Gillon & Matthewson [2014]; Tsimshianic: Davis, Gillon & Matthewson [2014]; Kutenai: Morgan [1991]). Again, we do not actually know whether this literature is truly representative of the pervasiveness of the phenomenon, or whether its "exotic" nature as compared to Indo-European languages has simply garnered undue attention to the topic in this geographic region. Nuuchahnulth, being the most discussed of these languages, is therefore nearly obligatory to include in a study such as this one.

The data used for the investigation of Nuuchahnulth come from a corpus of texts collected and edited by Toshihide Nakayama and published in Little (2003) and Louie (2003). The corpus consists of 24 texts dictated by speakers Caroline Little and George Louie, containing 2,081 utterances and 8,366 tokens (comprising 4,216 types). The texts cover a variety of genres, including procedural texts, personal narratives, and traditional stories. I manually retyped these texts as scription files for analysis. Scription is a simple text format for representing interlinear glosses in a way that is both familiar to linguists and computationally parseable (Hieber 2020a). The resulting digitally-searchable corpus is available on GitHub at `https://github.com/dwhieb/Nuuchahnulth`.

Other languages that would have been obvious choices for inclusion in this study are Riau Indonesian (Austronesian > Malayo-Polynesian) (Gil 1994), Mundari (Austroasiatic > Munda) (Evans & Osada 2005; Hengeveld & Rijkhoff 2005), Classical Nahuatl (Uto-Aztecan) (Launey 1994; 2004), and Central Alaskan Yup'ik (Eskimo-Aleut > Yupik) (Thalbitzer 1922; Sadock 1999; Mithun 2017). Each of these has generated contested claims about their flexibility and the existence of flexibility more generally. However, practicalities have limited me to examining just English and Nuuchahnulth for the time being. I leave investigations of other languages to future research and researchers.

Both the English and Nuuchahnulth corpora were converted to the Data Format for Digital

Linguistics (DaFoDiL) (a JSON format for representing linguistic data; Hieber [2020b]) for tagging and scripting purposes. This made it possible to use the Digital Linguistics (DLx) ecosystem of tools and software to more quickly tag and analyze the data. More information about Digital Linguistics may be found at https://digitallinguistics.io.

All of the datasets, scripts, and source files for this thesis are publicly available on GitHub at https://github.com/dwhieb/dissertation.

Turning now to results:

Regarding R1, "How flexible are words in English and Nuuchahnulth?", I find that English and Nuuchahnulth differ significantly not only in their overall degree of flexibility, but also in how that flexibility is realized. In English, the majority of words surveyed are flexible, but only to a small degree. Most lexical words of English can be used as nouns, verbs, or adjectives, but there is a strong tendency for each word to be used for primarily one function. English thus shows a consistent but somewhat marginal degree of flexibility. In contrast, most words in Nuuchahnulth are highly flexible, but primarily along the noun-verb axis; Nuuchahnulth words are very freely used as both nouns and verbs, but only infrequently used as adjectives. Nuuchahnulth thus shows a consistently high degree of flexibility, but primarily in just one dimension.

For R2, "Is there a correlation between degree of lexical flexibility for a word and frequency (or corpus dispersion)?", I find that higher frequency words are more flexible than lower frequency words, but that the effect is very small. The same facts hold when comparing degree of lexical flexibility with corpus dispersion. Words that are more evenly dispersed in a corpus have a slight tendency to be more flexible than those that are less evenly dispersed. These findings suggest that the degree of flexibility exhibited by a word does depend in part on how regularly speakers use it.

Lastly, R3 asks "How do the semantic properties of words pattern with respect to their flexibility?". With respect to Nuuchahnulth, I find that property words, especially numerals and quantifiers, are the most flexible semantic class of words. Nearly all of the most flexible

words denote property concepts. Deictic expressions such as *this, that, here, there* also rank very highly in their flexibility. I also find that there are strong correlations between morphologically marked aspect (durative, continuative, inceptive, etc.) and discourse function. In Nuuchahnulth, aspect markers may be used with either predicates or referents; they are not an exclusively verbal category. However, I find that the presence of any aspect marker does correlate strongly with predication, lending additional empirical evidence to Hopper & Thompson's (1984) claim that items used in their prototypical function will show the inflectional behaviors typical of that function. The momentaneous and telic aspect markers are the only ones in Nuuchahnulth which show any sort of tendency towards use with referents, while the durative was the only aspect marker to show any sort of tendency towards use with modifiers. Since aspect is a grammatical category that expresses how speakers construe the temporal structure of an event, these data suggest that flexibility has a great deal to do with how speakers conceptualize or construe words—as an action, object, or property—as has been suggested by Croft (1991: 99; 2001: 104).

Nuuchahnulth also has a definite suffix *-ʔi:* used with referents. Nakayama (2001: 48) states that this suffix is used with action words being construed as objects. This observation suggests that the definite suffix may have a clarifying function, appearing whenever an action word is used for the atypical role of reference (as predicted by Croft's structural coding hypothesis; see §2.4 for more details). One hypothesis that arises from applying typological markedness theory to Nuuchahnulth is that aspect markers which correspond to more object-like construals of a word (durative, telic, momentaneous) are more likely to be marked with the definite suffix. This turns out to be true, but only trivially so—only a tiny percentage (7.98%) of words with definite markers also had aspect markers. However, this leads to the far more interesting observation that the definite marker and the aspect markers in Nuuchahnulth are *almost* entirely mutually exclusive. They only rarely co-occur. These facts demonstrate that even in a language with rampant flexibility, as this study shows Nuuchahnulth to be, flexibility is nonetheless bound by universal typological constraints.

To summarize, this thesis makes contributions in several areas. The first is methodological: this thesis lays out a procedure for quantifying lexical flexibility for individual words in a corpus that can be replicated for other languages and corpora (Chapter 3). The second is empirical and descriptive: I describe the extent of lexical flexibility and the manner in which it operates in English and Nuuchahnulth (Chapter 4). The final contribution is analytical and theoretical: I argue that the data and statistical analysis presented in this thesis support Croft's typological markedness theory of word classes, in which lexical categories such as noun, verb, and adjective are not in fact categories of particular languages as has been historically assumed, but instead are emergent patterns that arise from how speakers use object, action, and property words for different functions in discourse (reference, predication, and modification). Words used for functions that are not prototypical of their meaning *tend* to be more marked (morphologically, behaviorally, semantically, and/or frequentially) than prototypical uses, but this is not an absolute universal. Lexical flexibility is the natural and expected result of the fact that these non-prototypical uses are *not* always *morphologically* marked, even when they are marked in other ways (Chapter 5).

The remainder of this thesis is organized as follows: Chapter 2: Background summarizes previous definitions of lexical flexibility and discusses their shortcomings. I propose an alternative, functionally-oriented definition that is consistent with cognitive and typological approaches to word classes instead. Chapter 3: Data & Methods describes in detail how the data were coded and analyzed for each of the major research questions (and contributing subquestions) in this study. I discuss factors that influenced how the data were coded, and outline the various coding decisions that were made. I present and explain a measure of corpus dispersion that is used partly in place of, and partly as a complement to, raw frequencies of words. Lastly, I set forth a procedure for operationalizing and quantifying lexical flexibility in a crosslinguistically comparable way. Chapter 4: Results presents the empirical findings from this study. I demonstrate how the methodological techniques from Chapter 3 are applied to individual words, and then present aggregated views of the data for English and

Nuuchahnulth respectively. Chapter 5: Discussion & Conclusion considers the implications of the results in Chapter 4 for theories of lexical categories. I argue that the data support a typological-universal theory of word classes, and that lexical flexibility should be viewed as a natural result of the cognitive and diachronic processes at work in language, rather than as an exceptional phenomenon. I conclude by discussing some limitations of the present study and avenues for future research, followed by closing remarks.

Chapter 2

Background

The focus of this chapter is to explain the concept of lexical flexibility, consider its criticisms, and offer a more robust, functionally-grounded definition instead. I first briefly describe how flexible approaches to lexical categories developed as a response to weaknesses in traditional theories of parts of speech. I then survey the landmark studies and important findings on lexical flexibility, along with criticisms of this research. Following that, I summarize approaches to lexical categories from several functionalist perspectives—cognitive linguistics, typology, and construction grammar. I conclude by offering a revised formulation of lexical flexibility which is more in line with this functional research.

2.1 Introduction: Approaches to lexical flexibility

The field of linguistics as a whole, and the subfield of typology in particular, is undergoing a radical shift in how we understand lexical categories, along primarily two dimensions. The first dimension is our understanding of what lexical categories are a property *of*. Early researchers viewed categories as universal properties of both language and languages (CITE: Haspelmath on g-language vs. p-language; add these terms in parentheses). I call this the *universalist* position. After Boas, many researchers then came to view categories as language-specific, with patterned similarities across languages. I call this the *relativist* approach. Most recently, some researchers view categories as typological patterns rather than properties of any particular language. This is the *typological* position, and the one I adopt here.

The second dimension of historical change in linguistic theories of categories is in the *nature* of the categories themselves. In the Classical tradition, categories were thought to be categorical and well-defined by a set of necessary and sufficient conditions (in the tradition of Aristotle). After the cognitive turn in the 1980s, many linguists came to view categories as prototypal, with some members of a category being more central, or better exemplars, than others. Cognitive research into the nature of idioms then led to the development of construction grammar, which sees language as consisting of a network of constructions rather than monolithic categories. I adopt a constructional approach to categories in this thesis.

These theoretical paradigm shifts are summarized in (10). At each stage of development, there has not been a wholesale displacement of previous theories (CITE: Kuhn). There are still many who regard word classes as universal and categorical, and the typological-constructional approach is still nascent.

(10) a. universal > language-specific > typological

    b. categorical > prototypal > constructional

Section 2.2 gives a synopsis of these theoretical positions and shows how research on lexical flexibility developed in recognition of the shortcomings of traditional approaches. Section 2.3 summarizes the key concepts and findings that have arisen from the research on lexical flexibility. Such research, however, is not without its own shortcomings. Section 2.3 also presents the main criticisms that have been leveled against flexible analyses of word classes. Section 2.4 then presents an alternate, functionally-oriented approach—the typological-constructional perspective. The final section of this chapter (§2.5) then applies this functional perspective to formulate an improved definition of lexical flexibility.

2.2 Traditional approaches

This section is a necessarily brief history of approaches to lexical categories up until the cognitive turn of the 1980s. It covers the universalist position that developed in the Classical

tradition, the relativist position that developed as a result of Boas' cultural relativism, and the structuralist (or "distributionalist") position that developed in the tradition of Saussure. Depending on how one understands and applies these different perspectives, none of them are mutually exclusive. It is especially common for linguists to simultaneously hold that lexical categories must be identified on the basis of language-internal evidence alone (the relativist position) and that lexical categories are universal in some sense or another (the universalist position).

2.2.1 Universalism

Historically and still presently, many researchers assumed that a small set of lexical categories are basic and universal to all languages (**BolingerSears1981**Croft 1991: 2**Payne1997**; Stassen 2011: 95). The set typically consists of some variation of the following: Noun, Verb, Adjective, Adverb, Pronoun, Adposition, Conjunction, Numeral, and Interjection (**Haspelmath2001**). This set has its origins in the *Τέχνη Γραμματική* / *Tékhnē Grammatiké* ('The art of grammar') of the the 2nd century B.C.E. grammarian Dionysius Thrax. The *Tékhnē* synthesizes the work of Dionysius' predecessors, describing eight parts of speech for Ancient Greek. These parts of speech were based largely on morphological (especially inflectional) criteria (**Rauh2010**). The *Tékhnē* was then translated and its model applied to Latin in the *Ars Grammatica* of Remnius Palaemon. The *Ars Grammatica* initiated a tradition wherein the languages of Europe and eventually the world (e.g. Mandarin [McDonald 2013]) were described using both Dionysius' categories (with occasionally additions / subtractions) as well as his method of identifying those categories on the basis of morphological criteria (**Rauh2010**).

Implicit in the Classical method is the assumption that lexical categories are universal in the sense of being instantiated in all languages. However, as European scholars began to encounter non-Indo-European languages (or even non-Romance languages) in both Europe and abroad, this assumption was challenged, as early as the first grammatical descriptions of Irish in the 7th century. At first, these languages either had Classical grammar imposed upon them

or were deemed grammatically deficient (**Suarez1983**). Nonetheless, missionary linguists in the early colonial era were indeed aware of the significant grammatical differences between these languages and Latin, and made their best attempts at describing them (**Suarez1983**). It is also important to realize that the project of describing the languages in the Americas and other zones of colonial influence was partially contemporaneous with the publication of the first grammars of the vernacular languages of Europe, as illustrated in Figure 2.1 (the data for which are shown in Table 2.1). Between 1524 and 1572, over 100 catechisms, manuals for confession, collections of sermons, grammars, and vocabularies were written in or about ten languages within the Viceroyalty of New Spain alone (an area smaller than present-day Mexico), mostly by Spanish Franciscan and Jesuit missionaries (**Suarez1983**). The task of converting the indigenous peoples to Christianity via the medium of their own languages was so important to the Spanish crown that the first bishop of Mexico, Francisco de Zumárraga, brought a printing press to Mexico in 1534 (just 15 years after the arrival of the first Spaniards in Mexico in 1519). The first book printed in Mexico was a Spanish-Nahuatl catechism by Alonso de Molina (**Suarez1983**). All this is merely to illustrate that language scholars in the colonial era were wrestling with the lexical categories of non-Indo-European languages—and therefore aware of the challenges these languages posed to Classical theories—at a very early stage.

Table 2.1: Some first grammatical descriptions of European vs. American languages

Language	Year	Title	Author
Irish	600s	*Auraicept na n-Éces* 'The scholars' primer'	Longarad
Occitan	1327	*Leys d'amors* 'Laws of love'	Guilhèm Molinièr
Welsh	1382–1410	*Llyfr Coch Hergest* 'Red book of Hergest'	unknown
Tuscan	1437–1441	*Grammatica della lingua toscana* 'Grammar of the Tuscan language'	Leon Battista Alberti

Table 2.1: Some first grammatical descriptions of European vs. American languages

| Language	Year	Title	Author
Castilian	1492	*Gramática de la lengua castellana* 'Grammar of the Castilian language'	Antonio de Nebrija
French	1530	*L'Éclaircissement de la langue francoyse* 'Explication of the French language'	John Palsgrave
German	1534	*Ein Teutsche Grammatica* 'A German grammar'	Valentin Ickelsamer
Basque	1545	*Linguæ Vasconum Primitiæ* 'First fruits of the Basque language'	Bernard Etxepare
Totonac	1539–1554	*Arte de la lengua totonaca* 'Grammar of the Totonac language'	Andrés de Olmos
Nahuatl	1547	*Arte para aprender la lengua mexicana* 'Grammar for learning the Mexican language'	Andrés de Olmos
Tarascan	1558	*Arte de la lengua tarasca de Michoacán* 'Grammar of the Tarascan language of Michoacán'	Maturino Gilberti
Dutch	1559	*Den schat der Duytsscher Talen* 'The treasure of the Dutch language'	John III van de Werve
Quechua	1560	*Grammatica o arte de la lengua general de los Indios de los Reynos del Peru* 'Grammar or Art of the General Language of the Indians of the Royalty of Peru'	Domingo de Santo Tomás
Tzeltal Maya	1571	*Ars Tzeldaica* 'Tzeltal Grammar'	Fray Domingo de Hara
Zapotec	1578	*Arte en lengua Zapoteca* 'Grammar in the Zapotec language'	Juan de Córdova
English	1586	*Pamphlet for Grammar*	William Bullokar
Mixtec	1593	*Arte de lengua Mixteca* 'Grammar of the Mixtec language'	Antonio de los Reyes
Timucua	1614	*Gramatica de la lengua Timuquana de Florida* 'Grammar of the Timucua language of Florida'	Francisco Pareja
Narragansett	1643	*A key into the language of America*	Roger Williams
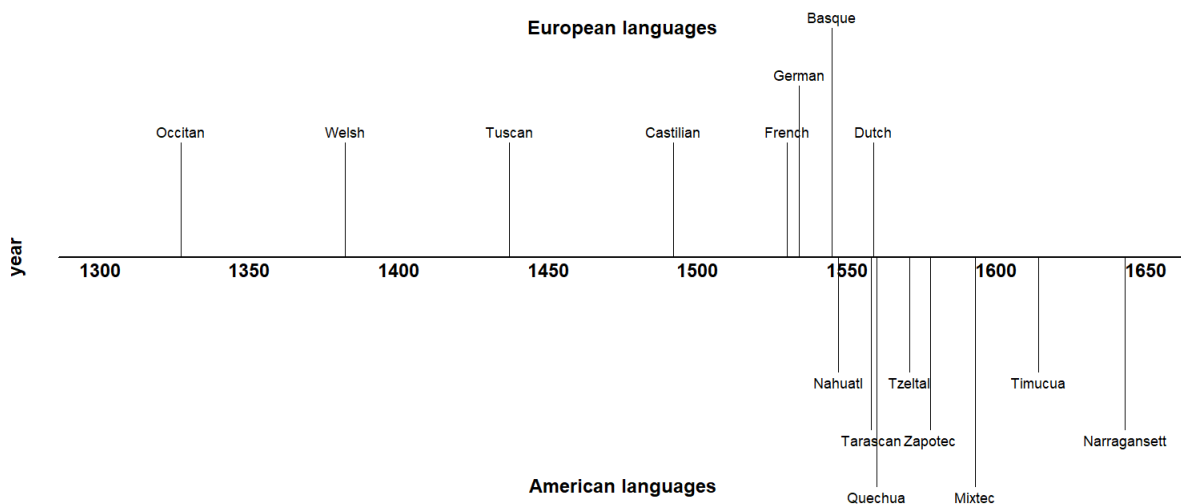
Figure 2.1: Approximate date of some of the first grammatical descriptions of European vs. American languages

As documentary linguistics turned its attention to North American (as opposed to Mesoamerican) languages, lexical flexibility in particular became a more prominent issue. In fact, even the first comprehensive survey of North American languages contains an entire section on "Conversion of nouns into verbs" (**Gallatin1836**), in which Gallatin depicts lexical flexibility as a rampant feature of all languages on the continent:

> It is the substantive [i.e. copula / auxiliary] verb which we [speakers of Indo-European languages] conjugate; whilst the [Native American] conjugates what we call the adjective and even the noun itself, in the same manner as [s/he] does other intransitive verbs. [...] I believe it must appear sufficiently obvious, that this general if not universal character of the [Native American] languages, the conversion into verbs and the conjugation, through all the persons, tense, and moods, of almost all the adjectives and of every noun which, without a palpable absurdity, is suspectible of it, is entirely due to the absence of the substantive verb. (**Gallatin1836**)

As evidenced by the above passage, increasing familiarity with non-Indo-European languages prompted some writers to abandon the universalist commitment. However, categorial universalism is still a widely-held position today, either in the sense of a) being universally instantiated in all languages (commonly assumed by most generative frameworks; although see Culicover [**Culicover1999**]), or b) being available to all languages, but only instantiated in

some (sometimes called the "grab bag" approach, as exemplified by Dixon's Basic Linguistic Theory framework [**Dixon2010a**; **b**; **c**]; [Hieber 2013: ???**Croft**]).

2.2.2 Relativism

American ethnographers in the tradition of Franz Boas questioned the univeralist assumption in a programmatic and comprehensive way. Writing on grammatical categories, Boas states, "Grammarians who have studied the languages of Europe and western Asia have developed a system of categories which we are inclined to look for in every language" (**Boas1911**). He concludes that this endeavor is a folly, and that "in a discussion of the characteristics of various languages **different fundamental categories** will be found" (**Boas1911**). Boas' students all adopted his grammatical relativism, and it became a foundational principle of the American linguistics tradition. His student Edward Sapir, writing on lexical categories specifically, makes one of the best-known and strongest statements of this position in his influential textbook *Language*: "[N]o logical scheme of the parts of speech—their number, nature, and necessary confines—is of the slightest interest to the linguist. Each language has its own scheme. Everything depends on the formal demarcations which it recognizes." (**Sapir1921**).

Many linguists today hold to Boas' grammatical relativism in some fashion or another. Textbooks and typological surveys commonly state that languages have varying numbers of lexical categories, though usually with the caveat that all languages seem to differentiate at least Noun and Verb (e.g. Velupillai 2012: §6.2). Some researchers, especially those working in typology, argue that linguists are still not rigorous *enough* in their application of grammatical relativism; they criticize certain kinds of crosslinguistic comparisons for imposing the categories of one language onto another (Croft 2001**Haspelmath2010LaPolla2016**) (CITE: You could probably cite your 2013 article for this too.). This position is discussed further in §2.4.

2.2.3 Structuralism

Developing alongside the early anthropological linguistics of Boas was the linguistic structuralism of Ferdinand de Saussure. His work informed both the Prague school under Nikolay Trubetzkoy and Roman Jakobson, and the distributional method of Leonard Bloomfield. The term *structuralism* has any number of uses (**Matthews2001**); here I refer to the idea that "language is a [...] self-contained, self-regulating system, whose elements are defined by their relationship to other elements" (**Matthews2014**). In particular, I am referring to the positivistic flavor of structuralism as practiced by Bloomfield, which focused on the structural relations between elements and establishing a set of rigorous scientific discovery procedures for linguistic structures (**Bloomfield1933**). Bloomfield saw lexical categories as something to be empirically discovered in the different syntactic distributions of words, rather than imposed on a language a priori (**Rauh2010**). Zellig Harris later refined and expanded on this methodology (**Harris1951**), which in turn was incorporated into Noam Chomsky's ealy Phrase Structure Grammar.

The signature methodological feature of this form of structuralism is the *distributional method*, a procedure for defining categories in terms of the set of contexts in which its words can appear—that is, their distributions (**Harris1951**2001: 11). As an illustration of distributional analysis applied to lexical categories, Croft (1991: 11–12) considers the distributions of the English words *cold*, *happy*, *dance*, and *sing* in two constructions: in the Predicate construction after *be*, and in the 3rd Person Singular Present Tense (*-s*) construction. Example data are shown below.

(11) a. Jack is cold.
      b. * Jack colds.

(12) a. Jack is happy.
      b. * Jack happies.

(13) a. * Jack is dance.
      b. Jack dances.

(14) a. * Jack is sing.
      b. Jack sings.

We can see that *cold* and *happy* have the same distributions in these tests (both may appear in the Predicate construction but not the Person-Tense inflection construction), while *dance* and *sing* have the same distribution (the inverse situation as *cold* and *happy*). The results of these two distributional tests are summarized in Table 2.2.

Table 2.2: Distribution of English Verbs and Adjectives (adapted from Croft [2001: 12])

	Predicate Construction	Inflectional Construction
Adjective: *cold*, *happy*, etc.	✓	✗
Verb: *sing*, *dance*, etc.	✗	✓

As applied in practice, however, the distributional method suffers from one serious drawback when used to argue for large, traditional categories like noun, verb, and adjective: distributional tests yield conflicting and overlapping results. Perhaps no two lexical items behave the same in every distributional test. Each new test that is introduced therefore partitions the lexicon into smaller and smaller classes. This fact has been demonstrated empirically for French verbs (**Gross1979**), Russian numerals (**Corbett1978**), and English adjectives vs. adverbs (Crystal 1967: 54). Distributional tables like that in Table 2.2 from each of these studies are reproduced in Table ??, Table ??, and Table ?? respectively. It is clear from these studies that distributional analysis does *not* lead to large, unified categories like noun, verb, and adjective, but rather a myriad of small constructions (Croft 2005: 434). This fact is the motivation underlying constructional approaches to language.

add distributional tables

Many scholars nonetheless choose to retain lexical categories as a necessary component of their linguistic theories or descriptions, at the expense of consistent application of the distributional method. Rather than considering all possible distributional contexts for a word, these scholars instead treat certain constructions as definitional of the category. Other distributional tests which yield cross-cutting results are either ignored, or treated as evidence

of subcategories instead of categories. Many researchers even prefer the term *syntactic categories* over *lexical categories* for this reason, focusing on just the syntactic evidence for categories (CITE: Rauh 2010). A severe methodological problem for this approach is that there are no generally agreed-upon principles for determining which distributional tests should be considered definitional. In this regard, Schachter & Shopen (2007: 4) note, "there may be considerable arbitrariness in the identification of distinct parts of speech rather than subclasses". Different scholars choose or prioritize different kinds of evidence for lexical categories over others on the basis of their theoretical commitments. This is the reason, as stated in Section 1.1, that disagreements about the existence of particular lexical categories in particular languages are typically *not* about the empirical facts. The results of a given distributional analysis are not usually controversial; the choice of distributional tests used to support one's analysis are. Unsurprisingly, then, debates over how to analyze lexical categories in various languages have been largely unproductive and unresolved (Croft 2005: 435). The problem only worsens when scholars attempt to apply the same criteria across languages. Distributions of words with similar meanings vary drastically across languages (Croft 2001: §1.4.1).

The real methodological problem here is *not* that we have yet to ascertain the correct principles for selecting the right distributional tests. The problem is being selective regarding which tests to apply in the first place. If we take the distributional method seriously, then we must apply it consistently, without ignoring distributional evidence that contradicts our theoretical or pretheoretical assumptions. To do otherwise is a kind of *methodological opportunism* (CITE: Croft 1991: 10–15, 41 (check these pages); Croft 2001: ???).

Partly in response to these problems, a growing cadre of linguists in the last 30 years have adopted one of various *flexible* approaches to word classes. Flexible analyses of word classes come in many flavors, some of which arguably still commit methodological opportunism, and others of which introduce new difficulties. These flexible approaches are reviewed in the following section.

2.3   Flexible approaches

2.3.1   Key concepts

2.3.2   Key findings

2.3.3   Problems & critiques

2.4   Functional approaches

2.4.1   Prototype theory

2.4.2   Typological markedness theory

2.5   Lexical flexibility: A functional definition

Chapter 3

Data & Methods

This chapter describes the data used for this study, and how those data were analyzed. It covers the selection criteria for 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, and 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 this thesis.

Chapter 4

Results

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 an aggregate look at the behavior of lexical flexibility in English and Nuuchahnulth, first individually and then in comparison (R1). Next, I present the results of the investigation of the relationship between degree of lexical flexibility and frequency / dispersion (R2). Finally, I discuss the behavior of flexible items with respect to their semantics (R3).

Chapter 5

Conclusion

This chapter summarizes the methods and main findings of this study, and the considers the impliations 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 cognitive 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 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 adjectives nonetheless exhibit varying degrees of flexibility for different lexemes. Lexical categories are not a given in grammar. I conclude by discussing some limitations of the present study and avenues for future research, followed by closing remarks.

References

Andrade, Manuel J. 1933. Quileute. In Franz Boas (ed.), *Handbook of American Indian languages*, vol. 3, 151–292. Columbia University Press.

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns, and adjectives* (Cambridge Studies in Linguistics 102). Cambridge University Press. doi:10.1017/CBO9780511615047.

Beck, David. 1999. *The typology of parts of speech systems: The markedness of adjectives*. University of Toronto. (Ph.D. thesis).

Beck, David. 2013. Unidirectional flexibility and the noun-verb distinction in Lushootseed. In Jan Rijkhoff & Eva van Lier (eds.), *Flexible word classes: typological studies of underspecified parts of speech*, 185–220. Oxford University Press. doi:10.1093/acprof:oso/9780199668441.003.0007.

Bickel, Balthasar, Bernard Comrie & Martin Haspelmath. 2015. *The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses*. Max Planck Institute for Evolutionary Anthropology. Leipzig: Department of Linguistics. https://www.eva.mpg.de/lingua/resources/glossing-rules.php.

Braithwaite, Ben. 2015. Nuu-chah-nulth nouns and verbs revisited: Root allomorphy and the structure of nominal predicates. In Joanna Błaszczak, Dorota Klimek-Jankowska & Krzysztof Migdalski (eds.), *How categorical are categories?: New approaches to the old questions of noun, verb, and adjective* (Studies in Generative Grammar 122), 47–74. Mouton de Gruyter. doi:10.1515/9781614514510-003.

Broschart, Jürgen. 1997. Why Tongan does it differently: Categorial distinctions in a language without nouns and verbs. *Linguistic Typology* 1(2). 123–165. doi:10.1515/lity.1997.1.2.123.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form* (Typological Studies in Language 9). John Benjamins. doi:10.1075/tsl.9.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford University Press. doi:10.1093/acprof:oso/9780195301571.001.0001.

Cannon, Garland. 1985. Functional shift in English. *Linguistics* 23(3). 411–432. doi:10.1515/ling.1985.23.3.411.

Chafe, Wallace. 2012. Are adjectives universal?: The case of Northern Iroquoian. *Linguistic Typology* 16(1). 1–39. doi:10.1515/lingty-2012-0001.

Chung, Sandra. 2012. Are lexical categories universal?: The view from Chamorro. *Theoretical Linguistics* 38(1-2). 1–56. doi:10.1515/tl-2012-0001.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems* (Cambridge Textbooks in Linguistics). Cambridge University Press.

Creissels, Denis. 2017. The flexibility of the noun/verb distinction in the lexicon of Mandinka. In Valentina Vapnarsky & Edy Veneziano (eds.), *Lexical polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches* (Studies in Language Companion Series 182). John Benjamins. doi:10.1075/slcs.182.02cre.

Croft, William. 1984. Semantic and pragmatic correlates to syntactic categories. *Chicago Linguistic Society* 20. 53–71.

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information.* University of Chicago Press.

Croft, William. 1995. Modern syntactic typology. In Masayoshi Shibatani & Theodora Bynon (eds.), *Approaches to language typology*, 85–144. Oxford University Press.

Croft, William. 2000. Parts of speech as typological universals and language particular categories. In Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes* (Empirical Approaches to Language Typology 23), 65–102. Mouton de Gruyter.

Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective.* Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001.

Croft, William. 2003. *Typology and universals.* 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge University Press. doi:10.1017/CBO9780511840579.

Croft, William. 2005. Word classes, parts of speech, and syntactic argumentation. *Linguistic Typology* 9(3). 431–441. doi:10.1515/lity.2005.9.3.391.

Croft, William. 2014. Comparing categories and constructions crosslinguistically (again): The diversity of ditransitives. *Linguistic Typology* 18(3). 533–551. doi:10.1515/lingty-2014-0021.

Croft, William. Forthcoming. Word classes in Radical Construction Grammar. In Eva van Lier (ed.), *The Oxford handbook of word classes.* Oxford University Press.

Croft, William & Eva van Lier. 2012. Language universals without universal categories. *Theoretical Linguistics* 38(1-2). 57–72. doi:10.1515/tl-2012-0002.

Crystal, David. 1967. English. *Lingua* 17(3-4). 24–56. doi:10.1016/0024-3841(66)90003-9.

Davis, Henry, Carrie Gillon & Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4). e180–e226. doi:10.1353/lan.2014.0076.

Dixon, R. M. W. 2010. *Basic Linguistic Theory, Vol. 1: Methodology.* Oxford University Press.

Dixon, R. M. W. & Alexandra Y. Aikhenvald (eds.). 2004. *Adjective classes: A cross-linguistic typology* (Explorations in Linguistic Typology 1). Oxford University Press.

Duncan, Philip T., Valerie (Lamxayat) Switzler & Henry B. Zenk. 2023. Chinookan. In Carmen Jany, Marianne Mithun & Keren Rice (eds.), *The languages and linguistics of indigenous North America: A comprehensive guide* (The World of Linguistics 13). Mouton de Gruyter.

van Eijk, Jan P. & Thom Hess. 1986. Noun and verb in Salish. *Lingua* 69(4). 319–331. doi:`10.1016/0024-3841(86)90061-6`.

Evans, Nicholas & Toshiki Osada. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology* 9(3). 351–390. doi:`10.1515/lity.2005.9.3.351`.

Farrell, Patrick. 2001. Functional shift as category underspecification. *English Language & Linguistics* 5(1). 109–130. doi:`10.1017/S1360674301000156`.

Floyd, Simeon. 2011. Re-discovering the Quechua adjective. *Linguistic Typology* 15(2011). 25–63. doi:`10.1515/LITY.2011.003`.

Frachtenberg, Leo J. 1922. Coos. In Franz Boas (ed.), *Handbook of American Indian languages*, vol. 2 (Bureau of American Ethnology Bulletins 40), 297–430. Smithsonian Institution.

Gil, David. 1993. Parts of speech in Tagalog. *Southeast Asian Linguistics Society* 3. 67–90.

Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17(2). 179–200. doi:`10.1017/S0332586500003000`.

Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4). 403–437. doi:`10.1075/ijcl.13.4.02gri`.

Gries, Stefan Th. Forthcoming. On, or against?, (just) frequency. In Hans C. Boas (ed.), *Applications of cognitive linguistics*. Mouton de Gruyter.

Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2019. *Glottolog 4.0*. Max Planck Institute for the Science of Human History. `https://glottolog.org`.

Hardjadibrata, R. R. 1985. *Sundanese: A syntactical analysis* (Pacific Linguistics Series D 65). Australian National University.

Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11(1). 119–132. doi:`10.1515/LINGTY.2007.011`.

Haspelmath, Martin. 2010a. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. doi:`10.1353/lan.2010.0021`.

Haspelmath, Martin. 2010b. Framework-free grammatical theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis* (Oxford Handbooks in Linguistics), 341–366. Oxford University Press. doi:`10.1093/oxfordhb/9780199544004.013.0014`.

Haspelmath, Martin. 2010c. The interplay between comparative concepts and descriptive categories (Reply to Newmeyer). *Language* 86(3). 696–699. doi:`10.1353/lan.2010.0021`.

Haspelmath, Martin. 2017. Some principles for language names. *Language Documentation & Conservation* 11. 81–93. doi:`10125/24725`.

Haspelmath, Martin. 2019. How comparative concepts and descriptive linguistic categories are different. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation* (Trends in Linguistics: Studies & Monographs 324). Mouton de Gruyter. doi:10.1515/9783110607963-004.

Hébert, Yvonne M. 1983. Noun and Verb in a Salishan language. *Kansas Working Papers in Linguistics* 8(2). 31–82.

Hengeveld, Kees & Jan Rijkhoff. 2005. Mundari as a flexible language. *Linguistic Typology* 9(3). 406–431. doi:10.1515/lity.2005.9.3.391.

Hieber, Daniel W. 2013. On linguistics, linguists, and our times: A linguist's personal narrative reviewed. *Linguistic Typology* 17(2). 291–321. doi:10.1515/lity-2013-0013.

Hieber, Daniel W. 2020a. *Scription.* doi:10.5281/zenodo.2595548.

Hieber, Daniel W. 2020b. *The Data Format for Digital Linguistics (DaFoDiL).* doi:10.5281/zenodo.1438589.

Hieber, Daniel W. Forthcoming. Word classes. In Carmen Jany, Marianne Mithun & Keren Rice (eds.), *The languages and linguistics of indigenous North America: A comprehensive guide* (The World of Linguistics 13). Mouton de Gruyter.

Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in Universal Grammar. *Language* 60(4). 703–752. doi:10.2307/413797.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization.* 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge University Press. doi:10.1017/CBO9781139165525.

Hudson, Richard. 1994. About 37% of word-tokens are nouns. *Language* 70(2). 331–339. doi:10.2307/415831.

Ide, Nancy & Keith Suderman. 2005. *Open American National Corpus.* http://www.anc.org/.

Jacobsen, Jr., William H. 1979. Noun and verb in Nootkan. In Barbara S. Efrat (ed.), *The Victoria conference on northwestern languages* (Heritage Record 4), 83–153. British Columbia Provincial Museum.

Jacobson, Steven A. 2012. *Yup'ik Eskimo dictionary (2 vols.)* 2nd edn. Alaska Native Language Center.

Jelinek, Eloise & Richard A. Demers. 1994. Predicates and prenominal arguments in Straits Salish. 70(4). 68–111. doi:10.4324/9780203068236.

Kastovsky, Dieter. 1996. Verbal derivation in English: A historical survey, or, much ado about nothing. In Derek Britton (ed.), *English historical linguistics 1994* (Current Issues in Linguistic Theory 135), 93–118. John Benjamins.

Kinkade, M. Dale. 1983. Salish evidence against the universality of 'noun' and 'verb'. *Lingua* 60(1). 25–39. doi:10.1016/0024-3841(83)90045-1.

Kuipers, Aert H. 1968. The categories verb-noun and transitive-intransitive in English and Squamish. *Lingua* 21. 610–626. doi:10.1016/0024-3841(68)90080-6.

Launey, Michel. 1994. *Une grammaire omniprédicative: Essai sur la morphosyntaxe du nahuatl classique* (Sciences du Langage). CNRS.

Launey, Michel. 2004. The features of omnipredicativity in Classical Nahuatl. *STUF* 57(1). 49–69. doi:10.1524/stuf.2004.57.1.49.

Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67(3). 475–509. doi:10.1353/lan.1991.0009.

Little, Caroline. 2003. *Caroline Little's Nuu-chah-nulth (Ahousaht) texts with grammatical analysis.* Toshihide Nakayama (ed.) (Endangered Languages of the Pacific Rim A2-27). Nakanishi Press.

Louie, George. 2003. *George Louie's Nuu-chah-nulth (Ahousaht) texts with grammatical analysis.* Toshihide Nakayama (ed.) (Endangered Languages of the Pacific Rim A2-028). Nakanishi Press.

Luuk, Erkki. 2010. Nouns, verbs and flexibles: Implications for typologies of word classes. *Language Sciences* 32(3). 349–365. doi:10.1016/j.langsci.2009.02.001.

Mattina, Nancy J. 1996. *Aspect and category in Okanagan word formation.* Simon Fraser University. (Ph.D. thesis).

McDonald, Edward. 2013. The creation of 'parts of speech' for Chinese: 'translingual practice' across Graeco-Roman and Sinitic traditions. *History & Philosophy of the Language Sciences.* https://hiphilangsci.net/2013/06/12/the-creation-of-parts-of-speech-for-chinese-translingual-practice-across-graeco-roman-and-sinitic-traditions/.

Mithun, Marianne. 2000. Noun and verb in Iroquoian: Multicategorisation from multiple criteria. In Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes*, 397–420. Mouton de Gruyter. doi:10.1515/9783110806120.397.

Mithun, Marianne. 2017. Polycategoriality and zero derivation: Insights from Central Alaskan Yup'ik Eskimo. In Valentina Vapnarsky & Edy Veneziano (eds.), *Lexical polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches* (Studies in Language Companion Series 182), 155–176. John Benjamins. doi:10.1075/slcs.182.06mit.

Montler, Timothy. 2003. Auxiliaries and other categories in Straits Salishan. *International Journal of American Linguistics* 69(2). 103–134. doi:10.1086/379680.

Morgan, Lawrence. 1991. *A description of the Kutenai language.* University of California, Berkeley. (Ph.D. thesis).

Mosel, Ulrike & Even Hovdhaugen. 1992. *Samoan reference grammar.* Scandanavian University Press.

Nakayama, Toshihide. 2001. *Nuuchahnulth (Nootka) morphosyntax* (University of California Publications in Linguistics 134). University of California Press.

Newmeyer, Frederick J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86(3). 688–695. doi:10.1353/lan.2010.0000.

Palmer, Bill. 2017. Categorial flexibility as an artefact of the analysis: Pronouns, articles, and the DP in Hoava and Standard Fijian. *Studies in Language* 41(2). 408–444. doi:`10.1075/sl.41.2.05pal`.

Plank, Frans (ed.). 2016. *Linguistic Typology* 20(2): *Of categories: Language-particular – comparative – universal.*

Polinsky, Maria & Lilla Magyar. 2020. Headedness and the lexicon: The case of verb-to-noun ratios. *Languages* 5(1). 1–25. doi:`10.3390/languages5010009`.

Rijkhoff, Jan. 2007. Word classes. *Language & Linguistics Compass* 1(6). 709–726. doi:`10.1111/j.1749-818x.2007.00030.xWord`.

Robins, R. H. 1968. Basic sentence structures in Sundanese. *Lingua* 21(C). 351–358. doi:`10.1016/0024-3841(68)90061-2`.

Rogers, Phillip. 2016. *Illustrating the prototype structures of parts of speech: A multidimensional scaling analysis.* University of New Mexico. (M.A. thesis).

Sadock, Jerrold M. 1999. The nominalist theory of Eskimo: A case study in scientific self-deception. *International Journal of American Linguistics* 65(4). 383–406. doi:`10.1086/466400`.

Sasse, Hans Jürgen. 1988. Der irokesische Sprachtyp. *Zeitschrift für Sprachwissenschaft* 7. 173–213. doi:`10.1515/ZFSW.1988.7.2.173`.

Schachter, Paul & Timothy Shopen. 2007. Parts-of-speech systems. In Timothy Shopen (ed.), *Language typology and syntactic description, Vol. 1: Clause structure*, 2nd edn., 1–60. Cambridge University Press. doi:`10.1017/CBO9780511619427.001`.

Siddiqi, Daniel. 2018. Distributed Morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory* (Oxford Handbooks in Linguistics), 143–165. Oxford University Press. doi:`10.1093/oxfordhb/9780199668984.013.15`.

Simon, Walter. 1937. Has the Chinese language parts of speech? *Transactions of the Philological Society* 36(1). 99–119. doi:`10.1111/j.1467-968X.1937.tb00675.x`.

Song, Jae Jung. 2001. *Linguistic typology: Morphology and syntax* (Longman Linguistics Library). Routledge. doi:`10.4324/9781315840628`.

Song, Jae Jung. 2018. *Linguistic typology* (Oxford Textbooks in Linguistics). Oxford University Press.

Stassen, Leon. 1997. *Intransitive predication* (Oxford Studies in Typology & Linguistic Theory). Clarendon Press.

Stassen, Leon. 2011. The problem of cross-linguistic identification. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology* (Oxford Handbooks in Linguistics), 90–99. Oxford University Press. doi:`10.1093/oxfordhb/9780199281251.013.0006`.

Sun, Linlin. 2020. *Flexibility in the parts-of-speech system of Classical Chinese* (Trends in Linguistics: Studies & Monographs 334). Mouton de Gruyter. doi:`10.1515/9783110660791.`

Swadesh, Morris. 1939. Nootka internal syntax. *International Journal of American Linguistics* 9(2). 77–102. doi:10.1086/463820.

Taylor, John R. 2003. *Linguistic categorization*. 3rd edn. (Oxford Textbooks in Linguistics). Oxford University Press.

Thalbitzer, William. 1922. Eskimo. In Franz Boas (ed.), *Handbook of American Indian languages*, vol. 2 (Bureau of American Ethnology Bulletins 40), 971–1069. Smithsonian Institution.

Thompson, Sandra A. 1989. A discourse approach to the cross-linguistic category 'Adjective'. In Roberta Corrigan, Fred R. Eckman & Michael Noonan (eds.), *Linguistic categorization* (Current Issues in Linguistic Theory 61), 245–266. John Benjamins. doi:10.1075/cilt.61.16tho.

Traugott, Elizabeth Closs & Graeme Trousdale (eds.). 2010. *Gradience, gradualness and grammaticalization* (Typological Studies in Language 90). John Benjamins. doi:10.1075/tsl.90.

Vapnarsky, Valentina & Edy Veneziano. 2017. Lexical polycategoriality – Cross-linguistic, cross-theoretical and language acquisition approaches: An introduction. In Valentina Vapnarsky & Edy Veneziano (eds.), *Lexical polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches* (Studies in Language Companion Series 182), 1–33. John Benjamins. doi:10.1075/slcs.182.01val.

Velupillai, Viveka. 2012. *An introduction to linguistic typology*. John Benjamins. doi:10.1075/z.176.

Vonen, Arnfinn M. 1994. Multifunctionality and morphology in Tokelau and English. *Nordic Journal of Linguistics* 17(2). 155–178. doi:10.1017/S0332586500002997.

Wetzer, Harrie. 1992. ''Nouny" and "verby" adjectivals: A typology of predicative adjectival constructions. In Michel Kefer & Johan van der Auwera (eds.), *Meaning and grammar: Cross-linguistic perspectives* (Empirical Approaches to Language Typology 10), 223–263. Mouton de Gruyter. doi:10.1515/9783110851656.223.

Wetzer, Harrie. 1996. *The typology of adjectival predication* (Empirical Approaches to Language Typology 17). Mouton de Gruyter. doi:10.1515/9783110813586.

Whaley, Lindsay J. 1997. *Introduction to typology: The unity and diversity of language*. SAGE Publications.

Language Index

Ancient Greek, 24

Cayuga, 6
Central Alaskan Yup'ik, *see* Yup'ik
Chamorro, 5
Chimakuan, 16
Chinookan, 16
Classical Nahuatl, *see* Nahuatl
Coosan, 16

English, xvii, 2, 4, 9–11, 15, 18, 29

Greek, *see* Ancient Greek

Indo-European, 5, 17
Indonesian, 5, 17
Iroquoian, 6

Kutenai, 16

Latin, 24

Mandarin, 5, 9, 10, 24
Mandinka, 2
Maya, 11

Mundari, 2, 6, 8, 9, 17

Nahuatl, 5, 17
Nguna, *see* North Efate
Nootka, *see* Nuuchahnulth
North Efate, 11
Nuuchahnulth, 3, 16, 18, 19

Proto-Indo-European, *see* Indo-European

Quechua, 3, 5, 6
Quiché Maya, *see* Maya

Riau Indonesian, *see* Indonesian
Russian, 8

Sahaptian, 16
Salishan, 16
Sundanese, 6

Tagalog, 5
Tongan, 3, 5
Tsimshianic, 16

Yup'ik, xvii, 4, 8, 9, 17

To Do