

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 Benard Comrie

Professor Stefan Th. Gries

Professor William Croft (University of New Mexico)

June 2020

The dissertation of Daniel W. Hieber is approved.

Bernard Comrie

Stefan Th. Gries

William Croft

Marianne Mithun, Committee Chair

June 2020

Lexical flexibility in discourse:
A quantitative corpus-based approach

Copyright © 2020

by

Daniel W. Hieber

Typeset using L^AT_EX software and the Linux Libertine family of fonts.

Published under a Creative Commons Attribution 4.0 License (CC BY 4.0):

<https://creativecommons.org/licenses/by/4.0/>

This thesis may be downloaded at:

<https://files.danielhieber.com/publications/dissertation.pdf>

The source code, data, and accompanying scripts for this thesis are available on GitHub:

<https://github.com/dwhieb/dissertation>

Dedication (#469)

ACKNOWLEDGMENTS

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. ([#467](#))

CURRICULUM VITAE

Daniel W. Hieber

EDUCATION

| | |
|------------|---|
| June 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.), <i>Languages in Louisiana: Community & culture</i> (America's Third Coast Series). University Press of Mississippi. |
| 2019 | Semantic alignment in Chitimacha. <i>International Journal of American Linguistics</i> 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.), <i>Category change from a constructional perspective</i> (Constructional Approaches to Language 20), 15–46. John Benjamins. DOI: 10.1075/cal.20.02hie |
| 2016 | <i>The cohesive function of prosody in Ékegusií (Kisii) narratives: A functional-typological approach</i> . 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](https://doi.org/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

| | |
|---------------------------------|----------|
| Acknowledgments | v |
| Curriculum Vitae | vi |
| Abstract | viii |
| Table of Contents | ix |
| List of Figures | x |
| List of Tables | xi |
| List of Abbreviations | xii |
| Conventions | xiii |
| | |
| 1 Introduction | 1 |
| | |
| 2 Background | 2 |
| | |
| 3 Data & Methods | 3 |
| | |
| 4 Results | 4 |
| | |
| 5 Conclusion | 5 |
| | |
| References | 6 |
| | |
| Language Index | 7 |

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 |
| SUBJ | subject |

Conventions

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

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,b; 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](#) (p. xii) 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.

The source of each example is always provided after the example itself.

In all interlinear glossed examples, I follow the formatting conventions (but not necessarily the recommended abbreviations) of the Leipzig Glossing Rules (Bickel, Comrie & Haspel-

math 2015).

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 languages, the term is written in lowercase as *participle*. I follow these same capitalization conventions in this thesis.

The first mention of a language within each chapter is followed by its genealogical affiliation (following the format family > phylum) and the location where it is spoken. For example, Central Alaskan Yup'ik would appear as “Central Alaskan Yup'ik (Eskimo-Aleut > Eskimo; Alaska)”. Language information is taken from the Glottolog database (Hammarström, Forkel & Haspelmath 2019). Language names are given in English following Haspelmath (2017). A complete list of languages mentioned in this thesis, along with their ISO 639-3 codes and Glottolog codes, is in the [List of Languages](#).

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

After each graphical representation of data, I have included the file path within the accompanying GitHub repository for this thesis to the script which will generate that figure.

Chapter 1

Introduction

Chapter 2

Background

Chapter 3

Data & Methods

Chapter 4

Results

Chapter 5

Conclusion

References

SOURCES OF LITERATURE

The references listed in this section are literature on the topic of this thesis that have been cited in the text.

- 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>.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form* (Typological Studies in Language 9). Amsterdam: John Benjamins. DOI:[10.1075/tsl.9](https://doi.org/10.1075/tsl.9).
- Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Croft, William. 1995. Modern syntactic typology. In Masayoshi Shibatani & Theodora Bynon (eds.), *Approaches to language typology*, 85–144. Oxford: 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. Berlin: Mouton de Gruyter.
- Croft, William. 2003. *Typology and universals*. 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. DOI:[10.1017/CB09780511840579](https://doi.org/10.1017/CB09780511840579).
- 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](https://doi.org/10.1515/lingty-2014-0021).
- Dixon, R. M. W. 2010. *Basic Linguistic Theory*. Vol. 1: *Methodology*. Oxford: Oxford University Press.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2019. *Glottolog 4.0*. Max Planck Institute for the Science of Human History. <https://glottolog.org>.
- 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](https://doi.org/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](https://doi.org/10.1353/lan.2010.0021).
- Haspelmath, Martin. 2010b. The interplay between comparative concepts and descriptive categories (Reply to Newmeyer). *Language* 86(3). 696–699. DOI:[10.1353/lan.2010.0021](https://doi.org/10.1353/lan.2010.0021).
- Haspelmath, Martin. 2017. Some principles for language names. *Language Documentation & Conservation* 11. 81–93. DOI:[10125/24725](https://doi.org/10.10125/24725).
- 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](https://doi.org/10.1515/lity-2013-0013).
- 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](https://doi.org/10.1353/lan.2010.0000).
- Plank, Frans (ed.). 2016. *Linguistic Typology* 20(2): *Of categories: Language-particular – comparative – universal*.
- Song, Jae Jung. 2001. *Linguistic typology: Morphology and syntax* (Longman Linguistics Library). London: Routledge.
- Song, Jae Jung. 2018. *Linguistic typology* (Oxford Textbooks in Linguistics). Oxford: Oxford University 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: Oxford University Press. DOI:[10.1093/oxfordhb/9780199281251.013.0006](https://doi.org/10.1093/oxfordhb/9780199281251.013.0006).
- Whaley, Lindsay J. 1997. *Introduction to typology: The unity and diversity of language*. Thousand Oaks, CA: SAGE Publications.

Language Index

Central Alaskan Yup'ik, [xiv](#)

English, [viii](#)

Nuuchahnulth, [viii](#)

Yup'ik, *see* Central Alaskan Yup'ik

Todo list

| | |
|--|------|
| <div></div> add Dedication (#469) | iv |
| <div></div> add Acknowledgments (#467) | v |
| <div></div> add Abstract | viii |
| <div></div> cross reference | xiv |