Dissertation prospectus:

Lexical flexibility in discourse

Daniel W. Hieber

University of California, Santa Barbara

1. Introduction

This dissertation investigates the discourse-functional motivations for *lexical flexibility*, i.e. the ability for a lexeme (or class of lexemes) to occur in multiple pragmatic functions (reference, predication, or modification, *inter alia*; Croft 1990) [VD] no overt coding (Hengeveld 1992:65; Croft 2001:66; van Lier 2016:197; ~~Evan~~ van Lier 2017:242). Two examples of flexible lexemes are shown in (1) and (2). The first set of examples is from English (Indo-European > Germanic), and the second is a much-debated set of examples from Mundari (Austroasiatic > Munda).

English (Indo-European > Germanic)

(1) a. We just put the **shoes** that were already made on a horse.
(DuBois et al. 2000:SBC001 251.37)

b. I would never ever ever trust myself to **shoe** a horse.
(DuBois et al. 2000:SBC001 220.41)

Mundari (Austroasiatic > Munda)

(2) a. **buru**=ko bai-ke-d-a
mountain=3PL.S make-COMPL-TR-IND
'They made the mountain.' (Evans & Osada 2005:354)

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

In (1), the word *shoe* is used for both reference (1a) and predication (1b) without any overt derivational morphology. In (2), *buru* 'mountain' is likewise used for both reference (2a) and predication (2b). This phenomenon is also frequently discussed as *conversion*, *zero-derivation*, or *functional shift*, depending on one's perspective (Crystal 2008:114).

When lexical flexibility is widespread in a language, it is often taken as evidence of *flexible word classes*, i.e. lexical categories which appear to subsume more than one traditional part of speech (Hengeveld 1992:65; Rijkhoff 2007:715; van Lier & Rijkhoff 2013:1; van Lier 2016 and

accompanying articles; Evan van Lier 2017:243; Vapnarsky & Veneziano 2017a).[1] Flexible categories have become a vibrant topic in recent years, prompting discussions on the existence of flexible categories in particular languages (Kinkade 1983; Van Eijk & Hess 1986; Broschart 1997; Sadock 1999; Evans & Osada 2005; Hengeveld & Rijkhoff 2005; Dorvlo 2009; Koch & Matthewson 2009; Floyd 2011; Chafe 2012; Chung 2012), the plausibility of flexible categories in general (Dixon 1977; Don 2004; Croft 2005; Evans & Osada 2005; Luuk 2010; Baker & Croft 2017; Palmer 2017), and detailed descriptive studies of flexible categories and the diversity of their expression across languages (Hengeveld 1992; Holton 1999; Hengeveld, Rijkhoff & Siewierska 2004; van Lier 2006; Luuk 2010; Rijkhoff & van Lier 2013; van Lier 2016; Cauchard 2017; Lichtenberk 2017; Vapnarsky & Veneziano 2017b). However, little attention has been paid to the functional motivations for lexical flexibility (though see Hopper & Thompson 1984, Thompson 1989, and Nakayama 2002). Why, in flexible languages, do speakers make the categorial choices they do? If a given lexeme can more-or-less freely alternate between, say, referential and predicative uses, what determines when a speaker uses one function over another? Since any choice between linguistic alternatives provides a means of conveying information, the presence of lexical flexibility in a language is yet another dimension of variation that speakers can manipulate to achieve their manifold discourse goals. How then is lexical flexibility deployed in discourse? This dissertation represents a first attempt to answer this question, investigating the discourse-functional correlates of lexical flexibility in a small but diverse sample of languages.

This focus on the role of lexical flexibility in discourse moves beyond the existing literature in that it aims to understand the functional underpinnings of lexical flexibility rather than debate its existence, the universality of lexical categories, or the existence of a particular lexical category in a language. Instead, I start from the premise that all languages have some lexemes (however few) that exhibit lexical flexibility to varying degrees and perhaps at different levels (root, stem, word, or construction), and that categorical distinctions between lexemes are more strongly and consistently expressed in some languages than others. A discourse-oriented

[1] Grammatical (as opposed to lexical), closed-class categories such as demonstratives may also exhibit flexibility (cf. François 2017; Lichtenberk 2017), but this phenomenon will not be discussed here.

approach is also of special interest because it has the potential to shed light on a recurring question in discussions of lexical flexibility: how much of the semantic shift that occurs when a lexeme changes function can be imputed to the discourse context, and how much to language- and lexeme-specific patterns that must be memorized by the speaker? In this dissertation I aim to assess the extent to which semantic and pragmatic/discourse properties contribute to the categoriality of lexemes.

This research potentially provides new insights into the emergence of lexical categories, in the sense of morphosyntactic constructions dedicated to specific pragmatic functions. If choice of lexical category in highly flexible languages is shown to be tied to discourse function in a way similar to that outlined by Hopper & Thompson (1984) and Thompson (1989), this suggests that categorial differences in language develop out of the gradual routinization and grammaticization[2] of discourse tendencies. Comparable developments of grammatical categories out of discourse tendencies are attested for other areas of grammar as well, including grammatical relations (Mithun 2012) and bound vs. free pronominal forms (Mithun 2013). Conversely, Nakayama (2002:3, 54, 57) argues that this model also explains the highly flexible nature of lexemes in Nuuchahnulth (a.k.a Nootka; Wakashan) as compared to more categorically rigid languages. Lexical categories in Nuuchahnulth are principally discourse tendencies rather than obligatory grammatical conventions, and so the language is flexible in virtue of the fact that it has not undergone this categorical grammaticization process. This dissertation provides support for this model of the emergence of lexical categories by showing the extent to which discourse and information status considerations contribute to choice of lexical category.

The specific research questions I look in this dissertation are as follows:

- Does the extent of lexical flexibility observed for a language correlate with size of corpus? Does it correlate with the token frequency of the lexeme? (Chapter 2)

[2] I use the term *grammaticization* in this dissertation in a somewhat atypical sense, to refer to the process whereby a construction becomes conventionalized and grammatically obligatory, and therefore "part of the grammar" (as opposed to just a discourse tendency). When necessary, the term *grammaticalization* will be used to distinguish between the above sense of grammaticization, and the standard definition wherein a lexical item becomes a grammatical one (Hopper & Traugott 2003:2). While this choice of terms is admittedly less than ideal, I have yet to find a better term for the notion I am calling *grammaticization*.

- Do certain semantic domains tend to exhibit more lexical flexibility than others? Is lexical flexibility sensitive to the animacy hierarchy? Are property concepts more flexible than time-stable items and/or events? (Chapter 3)
- Does the current and/or previous choice of grammatical role for a lexeme correlate with choice of lexical category? (Chapter 4)
- Does information status (given vs. new vs. activated) correlate with choice of lexical category? (Chapter 5)

I discuss my hypotheses and expected results for each of these research questions in my chapter outline below (§4).

2. Background

In this section, I outline at a high level the major approaches adopted by typologists in treating lexical categories generally, and flexible categories more specifically. I then advance the approach that will be adopted in this dissertation.

2.1. Approaches to Lexical Categorization

As is well known, the *classical* or *traditional* approach to parts of speech has its origins in the *Τέχνη Γραμματική / Tékhnē Grammatiké* ('The Art of Grammar') of the grammarian Dionysius Thrax in classical antiquity (2nd century B.C.E.). The *Tékhnē* synthesizes the work of Dionysius' predecessors, describing eight parts of speech for ancient Greek: noun, verb, participle, article, pronoun, preposition, adverb, and conjunction. These parts of speech were based largely on morphological (especially inflectional) criteria (Rauh 2010:17–20).

The *Tékhnē* was then translated and its model applied to Latin in the *Ars Grammatica* of Remnius Palaemon, initiating a tradition wherein the languages of Europe and eventually the world (see for example McDonald (2013)) were described using both Dionysius' eight categories (with some variation) and, importantly, his method of identifying those categories on the basis of primarily morphological criteria (Rauh 2010:20). Implicit in the classical approach is the assumption that parts of speech are universal, in the sense of being instantiated in all languages.

The American structuralists in the tradition of Franz Boas questioned this assumption in a programmatic and comprehensive way. Writing on grammatical rather than lexical 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" (Boas

1911:35). 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" (Boas 1911:43). Boas' student Edward Sapir applies this same language-particular approach to lexical categories: "[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." (Sapir 1921:125). Boas also strongly influenced Leonard Bloomfield, who treated language as a scientific object and, in applying Boasian methods, saw lexical categories as something to be empirically discovered in the different syntactic distributions of words, rather than imposed on a language *a priori* (Rauh 2010:33).

This structuralist approach to lexical categories, which came to be known as the *distributional method* (Harris 1951:5), constituted a major advance in the typological study of parts of speech, and essentially became the sole method of syntactic analysis in modern linguistics (Croft 2001:11). While a significant step forward, the distributional method for identifying word classes is however faced with one particularly potent problem: what to do when the distributional criteria for classifying lexemes yield conflicting results, or fail to yield consistent and well-defined categories.

A partial solution to this problem was the recognition, established in a series of studies by Eleanor Rosch (1973a; 1973b; 1975; Rosch & Mervis 1975; Rosch et al. 1976; Rosch 1978) and popularized among linguists by Lakoff (1987) and Taylor (1989 [2003]), that lexical categories are prototypal, and that members of a category do not necessarily exhibit all the properties associated with that category. This body of research collectively challenged the classical approach to lexical categories based on necessary and sufficient conditions cleanly delineating distinct categories. While linguists were generally quick to accept the existence of gradience and fuzzy boundaries for linguistic categories (Rauh 2010:7), the prototype approach did not really solve the essential problems of lexical categorization, namely, how to identify them, and their crosslinguistic status if any.

Recognizing this difficulty, Croft (2000; 2001:29–47) provides a detailed critique of the distributional method and its implications, and utilizes prototype theory in offering a typologically-oriented theory of lexical categories instead. Whenever distributional criteria conflict or fail to exclusively partition lexemes into distinct categories, he notes, typical practice

is that the linguist simply chooses whichever distributional criterion they believe to be the most important, and bases their categorization on that. This practice is what Croft calls *methodological opportunism*, and it is one replete with problems:

> There is no a priori way to decide which of several constructions with mismatching distributions, or which subset of constructions, should be chosen as criteria for identifying the category in question. Why should passivizability be the criterion for defining the Direct Object category? Why shouldn't the criterion be occurrence as the postverbal prepositionless Noun Phrase in the Active construction? The choice of criteria again looks suspiciously like serving a priori theoretical assumptions of the analyst, for example a priori assumptions about what should or should not be a Direct Object. Moreover, if one does choose one construction (or subset of constructions) to define a category, then one still has not accounted for the anomalous distribution pattern of the constructions that have been left out (in this case, occurrence as the postverbal prepositionless Noun Phrase in the Active construction).
>
> Language-internal methodological opportunism […] is unprincipled and ad hoc, and hence is not a rigorous scientific method for discovering the properties of the grammar of a language. (Croft 2001:41)

If one is consistent in the application of the distributional method, states Croft, then one must be prepared to accept a proliferation of minor categories for each language. Ultimately, every construction constitutes its own category, comprising the set of items that may appear in that construction. As a result, no language exhibits traditional major categories such as noun, verb, and adjective—only more narrow constructions such as, for example, Tense-Marked Intransitive Verb or Tense-Marked Transitive Verb, which may or may not share the same members.

For Croft, what exists in the grammar of particular languages is sets of constructions related in a taxonomic web rather than lexical categories per se. Parts of speech that approximate traditional categories exist only as crosslinguistic typological markedness tendencies. That is, when the semantic class of an item aligns with its pragmatic function (reference, predication, modification, etc.), that form will be unmarked. However, when an item is used in a non-prototypical manner, such as when an entity-denoting concept is used for predication, that use is structurally and/or behaviorally marked (Croft 2002:87–99). This theory of typological markedness is what "allows us to construct generalizations about categories *across* constructions" which otherwise do not share the same properties and members (Croft 2001:92). The typologically unmarked combinations of an object being used for reference, a property for

modification, and an action for predication form the prototypal core of the categories noun, adjective, and verb respectively (Croft 2001:89).

This dissertation utilizes Croft's typological markedness approach in exploring lexical flexibility. While the typological tendency is for non-prototypical uses of a lexeme to be structurally or behaviorally marked, lexical flexibility can be viewed as cases where non-prototypical uses of a lexeme are *not* marked in such a way. These cases do not however violate Croft's markedness tendencies, since the markedness principles are implicational in nature. That is, non-prototypical uses of a lexeme are *at least* as marked as prototypical ones; this does not preclude the possibility of both uses being equally marked, as in the case of conversion.

One final issue in research on lexical categories is whether they should be thought of as language-specific, and potentially incommensurable and uncomparable across languages, or as instantiations of crosslinguistically valid categories. This issue is hotly debated in the literature, and Croft's universal-typological approach is just one among many (Croft 2000; Pustet 2000; Croft 2005; Haspelmath 2007; Ramat 2009; Haspelmath 2010; Chung 2012; Croft & van Lier 2012; Haspelmath 2014; Beck 2016; Croft 2016; Rijkhoff 2016; Baker & Croft 2017). I do not aim to speak towards this debate in this dissertation. My focus instead is on comparing the ability of lexemes to appear in multiple pragmatic functions with no overt coding, across languages. I make no claims as to whether the constructions that these lexemes appear in constitute language-specific or universal categories.

2.2. Approaches to Lexical Flexibility

Lexical flexibility became a prominent topic of interest when early anthropological linguists investigated the structure of languages of the Americas in the 19th and 20th centuries, and found that it was difficult to reconcile classical categories with data from Native American languages (Boas 1911; Sapir 1921; Kuipers 1968; Jacobsen 1979; Kinkade 1983; Sadock 1999). Responses to this situation varied, and the positions adopted towards lexical flexibility have only multiplied in number with the more recent explosion of interest in the topic. This section briefly overviews these varied approaches toward lexical flexibility.

One common response to claims of lexical flexibility in a language is to show that the grammar does in fact show evidence for categorical distinctions, but that the evidence is simply

subtle (Dixon 2004; Floyd 2011; Palmer 2017).[3] In this approach, traditional categories are typically thought to be universally instantiated, to be found in a[illegible] the world's languages provided one looks hard enough. There are however two concerns with this approach: First, it would seem to engage in methodological opportunism (cf. Croft 2001). Criteria which highlight data suggestive of the category in question are privileged, while additional criteria that might suggest flexible membership or categorical subdivisions are ignored. Defining lexical flexibility in terms of shifts between language-specific categories like Noun and Verb is problematic if those categories are based on arbitrarily-chosen criteria in the first place.

Second, this response to lexical flexibility shifts the focus away from the very interesting ways in which categories differ across languages. Even when subtle evidence for categorical distinctions is found, there remain drastic and qualitative differences in the way that those categories are realized as compared to other languages with more clearly demarcated categories. Typologists should not be satisfied to gloss over these differences. Instead, differences in the strength of expression of lexical categories in a language should be taken as a dimension of variation to be mapped out and explored in a robust typological way. This approach has become more common in recent years (Rijkhoff & van Lier 2013; Eva van Lier 2017; Vapnarsky & Veneziano 2017b).

In stark contrast to this first approach, some have embraced the existence of flexible categories and argued extensively for their existence (Kuipers 1968; Kinkade 1983; Hengeveld 1992; Broschart 1997; Gil 2005; Hengeveld & Rijkhoff 2005; Luuk 2010; van Lier & Rijkhoff 2013). Some have even proposed that several new, flexible categories such as *non-verb* (Hengeveld 1992) or *noun/flexible* (Luuk 2010) be added to the classical typology of parts of speech. All these proposals have garnered heavy criticism. Since it is important for any study of lexical flexibility to address these criticisms, I briefly review them here, then discuss how a typological markedness approach to lexical flexibility avoids these problems.

Broadly speaking, the main argument leveled against lexical flexibility is that it ignores a great deal of item-specific knowledge speakers have about lexemes and their uses in different functions. Both Croft (2001:65–75) and Evans & Osada (2005), for example, criticize Hengeveld's notion of flexible categories (Hengeveld 1992; Hengeveld & Rijkhoff 2005) on the

[3] Note that Dixon's position on the universality of adjectives has shifted over time; cf. Dixon (1977).

basis that the meaning of a lexeme changes when it is used in different functions. Mithun also has in various studies (Mithun 1999; Mithun 2000; Mithun 2017) illustrated the impressive level of item-specific and idiosyncratic knowledge that speakers have about lexemes, their distributional contexts, and the semantic shifts they undergo in different constructions. Because the meaning that results from semantic shifts is conventional, often idiosyncratic, and language-specific, patterns of semantic shift constitute a basis for distinguishing between classes of lexemes. Even in cases where semantic shifts are patterned and non-idiosyncratic, the pattern of shift is still a language-specific fact that applies to a subset of the lexicon, thereby providing the basis for demarcating a lexical category. Researchers that emphasize the conventionalized and item-specific nature of lexical semantics thus tend to view cases of lexical flexibility as conversion or zero derivation, and languages purported to be highly flexible as ones in which such conversion is rampant.

Proponents of the existence of lexical flexibility have addressed these criticisms in two ways: First, many have argued that lexical items in flexible languages are *precategorial*, i.e. underspecified for lexical category (Hopper & Thompson 1984; Broschart 1997; Farrell 2001; Arad 2003; Don & van Lier 2003). In precategorial languages, lexical categorization is thought to be a property of the morphosyntactic construction that the item appears in, its pragmatics, or its discourse context, rather than the lexeme itself. The second response to lexical specificity is to argue that lexical items are semantically *vague*, i.e. they have a single, broad semantics which encompasses its use in various lexical categories (Farrell 2001; Hengeveld, Rijkhoff & Siewierska 2004; Hengeveld & Rijkhoff 2005; McGregor 2013). In this approach, the relevant component of the meaning of the lexeme is highlighted by its morphosyntactic context. What is common to both these approaches is that lexical categorization is not a property of the lexical item itself, but rather the result of a semantic coercion process whereby the lexical item receives its categorization from local context. Critics of lexical flexibility have not generally found these approaches to lexical specification satisfactory, and argue that even taking pragmatics, discourse, and local morphosyntactic context into account is insufficient to account for the semantic idiosyncrasies in the data (Croft 2001; Evans & Osada 2005; Mithun 2017).

If however lexical flexibility is understood in terms of the unmarked use of forms across different pragmatic functions (reference, predication, modification, etc.) rather than language-specific lexical categories, semantic shift need not be problematic for the study of lexical

flexibility. Rather, semantic shifts become a descriptive *desideratum*, i.e. a crucial part of what must be described when studying the use of a lexeme across different pragmatic functions. This dissertation helps satisfy this *desideratum* by investigating the interaction of discourse function, semantic shift, and lexeme-specific knowledge in a crosslinguistic sample.

2.3. Functional Motivations for Lexical Flexibility

This section briefly summarizes the relevant literature on the interaction of discourse and lexical categories, and especially lexical flexibility.

It has often been suggested that there is a semantic (or even logical; cf. the Port Royal Grammar) basis to the major lexical categories (Sapir 1921:117–119; Givón 1979:320–321; Lyons 1977:442–447), which are thought to have a prototypal structure. For example, prototypical nouns would be concrete, time-stable entities, while other nouns approximate this prototype to varying degrees. In an influential study, however, Hopper & Thompson (1984:708) argue that "the lexical semantic facts about N's and V's are secondary to their DISCOURSE ROLES; and that the semantic facts (perceptibility etc.) which are characteristic features of prototypical N's and V's are in fact derivative of (and perhaps even secondary to) their discourse roles." They demonstrate that lexemes show more nominal coding and behavior when they are used to introduce new referents into the discourse, but more verbal coding and behavior when they are used to assert the occurrence of an event. In a later article Thompson (1989) extends this framework to explain why adjectives crosslinguistically pattern as either verbs or nouns—when introducing a new referent into the discourse, adjectives tend to pattern nominally; when functioning as the discourse predicate, they tend to function verbally.

Hopper & Thompson also briefly touch on the issue of lexical flexibility in their conclusion, and it is worth providing an extensive excerpt here, because they directly anticipate some of the important conclusions of this dissertation:

> We should like to conclude, however, by suggesting that linguistic forms are in principle to be considered as LACKING CATEGORIALITY completely unless nounhood or verbhood is forced on them by their discourse functions. To the extent that forms can be said to have an a-priori existence outside of discourse, they are characterizable as ACATEGORIAL; i.e., their categorical classification is irrelevant. Categoriality—the realization of a form as either a N or a V—is imposed on the form by discourse. Yet we have also seen that the noun/verb distinction is apparently universal: there seem to be no languages in which all stems are indifferently capable

> of receiving all morphology appropriate for both N's and V's. This suggests that the continua which in principle begin with acategoriality, and which end with fully implemented nounhood or fully implemented verbhood, are already partly traversed for most forms. In other words, most forms begin with a propensity or predisposition to become N's or V's; and often this momentum can be reversed by only special morphology. It nonetheless remains true that this predisposition is only a latent one, which will not be manifested unless there is pressure from the discourse for this to occur.
>
> In other words, far from being 'given' aprioristically for us to build sentences out of, the categories of N and V actually manifest themselves only when the discourse requires it. (Hopper & Thompson 1984:747)

In essence, Hopper & Thompson acknowledge that lexemes are to a certain extent prespecified for category, and that this extent varies from lexeme to lexeme. However, to the extent that lexemes show flexibility between different traditional categories, the choice of category for a lexeme is determined primarily by its discourse function and information status. This is one of the primary claims that this dissertation aims to support.

A similar point is made by Nakayma (2002) for Nuuchahnulth (a.k.a. Nootka; Wakashan), which features prominently in debates on lexical flexibility. Nakayama concludes that word classes do exist in Nuuchahnulth, but that they are not strongly grammaticized: "word classes in Nuuchahnulth are not so much structural categories as behavioral categories: they represent groups of words defined by a set of regularities that are formed and maintained through repeated use in discourse rather than purely structural properties." (Nakayama 2002:57). Categorical choice in Nuuchahnulth thus appears to be driven primarily by discourse and information status considerations.

In this dissertation I intend to apply a discourse-oriented approach like those summarized above to a small but diverse sample of languages, with the expectation of providing empirical evidence of the following claims: a) that languages vary dramatically in the degree to which categorical distinctions have become grammaticized; and that b) in languages where categorical distinctions are not strongly grammaticized, choice of category is in large part determined by discourse function and information status rather than lexical prespecification.

3. Data & Methods

This section discusses the language sample that will be used for this research, and the way in which lexical flexibility will be operationalized quantitatively for comparison with other variables.

The language sample for this dissertation will consist in a small number of typologically diverse languages selected based on the following criteria:

- relevance to (and prominence in) the literature and debates on lexical flexibility
- geographic and typological diversity (e.g. both isolating and polysynthetic)
- availability of extensive lexical data and corpora
- differences in the purported degree of lexical flexibility (e.g. both rigid and flexible languages)

I currently plan to include at least the following languages in the sample. More may be added if time permits.

- Central Alaskan Yup'ik (Eskimo-Aleut > Eskimo)
- Chitimacha (isolate)
- English (Indo-European > Germanic)
- Nuuchahnulth (a.k.a. Nootka; Wakashan)
- Riau Indonesian (Austronesian > Malayo-Polynesian)
- Spanish (Indo-European > Romance)
- Swahili (Niger-Congo > Bantu)

Each of the research questions in this dissertation compares the degree of lexical flexibility of a lexeme to some other feature of the lexeme (semantic domain, inherent topicality, grammatical relation, and information status). Therefore it is important to operationalize the notion of lexical flexibility in a way that allows for comparison across lexemes and across languages. As stated in the introduction, I will identify instances of lexical flexibility as those in which the same form is used for two or more pragmatic functions with no overt derivational coding. My procedure for quantifying lexical flexibility will therefore be to take each word root in a corpus, and count the number of times that the root occurs in each of the three major pragmatic functions of reference, predication, and modification, without derivational morphology to mark that change (but allowing for inflectional morphology). This will give me the token frequency for each function

of that root. I will then repeat this procedure for stems and fully-inflected words. Table 1 shows a sample statistical summary of this procedure for the first 10 roots that appear in the Chitimacha (isolate) corpus.

Table 1. Sample statistical summary of lexical flexibility for Chitimacha (isolate)

Item	Gloss	Reference Count	Predication Count	Modification Count	Total Count
cuw-/t'ut-	'go'	9	581	1	591
hok-	'leave'	0	64	0	64
hus	3SG	79	0	437	516
kišut-	'swim'	0	13	0	13
kun	'some'	14	0	9	23
na:nčaʔa	'brother'	4	0	0	4
ničwa-	'go (VERT) to water'	0	9	0	9
še:ni	'pond'	10	0	2	12
te:t-	'say'	2	264	0	266
we(y)	DEM	553	7	1496	2056

As Table 1 shows, individual lexical items vary drastically in their overall frequency, and the frequency with which the item is used in each pragmatic function. Additionally, the size of the corpora for each of the languages in the sample also varies widely. These facts raise two methodological questions: 1) Does the degree of lexical flexibility for an item depend on its token frequency? 2) Does the degree of lexical flexibility exhibited in a corpus depend on the overall size of the corpus? If so, it will be important to control for these factors when operationalizing the degree of lexical flexibility for a lexeme. Chapter 3 of my dissertation, 'Data & Methods', will answer these methodological questions, and present a means of normalizing the data to account for these factors if necessary. The following section provides additional details about that and other chapters.

4. Outline

This section provides an outline of the dissertation. At a broad level, the dissertation will begin with a general introduction to the problem of lexical flexibility along with a review of the relevant literature (Chapter 1). The next chapter presents the data used in the dissertation, and the methods applied to measuring lexical flexibility; this chapter will be based in part on §3 above (Chapter 2). Once the degree of lexical flexibility has been assessed for each lexical item in the corpora, subsequent chapters then examine the extent to which that flexibility correlates with

various other features of the item or its immediate context (Chapters 3–6). Each chapter will introduce the research question and relevant literature, describe the way that relevant feature of interest will be operationalized, and then summarize the findings. Complete statistical summaries will be provided in an appendix as well as made available online. Finally, a concluding chapter will discuss the broader implications of the results in Chapters 3–6 (Chapter 7).

The dissertation outline is presented below.

4.1. Chapter 1: Introduction: The Challenge of Lexical Flexibility

The introductory chapter introduces the phenomenon of lexical flexibility and its importance for linguistic science. I frame the research question that is the focus of the dissertation, explain how I will answer that question, and preview my conclusions. After providing an outline of the plan of the dissertation, I proceed to review the literature and important concepts relating to lexical flexibility, to inform the reader of the issues involved in its study. This chapter will be based heavily on this prospectus.

4.2. Chapter 2: Data & Methods: Assessing Lexical Flexibility

This chapter provides an overview of the language sample, and the criteria for inclusion in that sample. For each language, I summarize its important typological characteristics, and the nature of the available data. I then operationalize the notion of lexical flexibility, similar to §3 above, and proceed to answer two methodological research questions presented in that section using the raw quantitative data. Methods for normalizing the data to account for these factors will then be discussed.

4.3. Chapter 3: Lexical Flexibility and Semantic Domains

This chapter answers the question of whether certain semantic concepts are more likely to participate in categorial alternations than others. Lexemes that have approximate semantic parallels in each of the languages are compared and ranked in terms of their overall flexibility. I then discuss any noticeable patterns or clusters in semantic domains for the most and least flexible items. I hypothesize that certain semantic domains, such as body part terms, will show significantly greater degrees of lexical flexibility than others, and that in the case of body part terms, the crucial determining semantic factor is not body parts per se, but rather the likelihood

of a semantic domain to undergo metaphorical extensions into the instrumental, positional, movement-oriented, or temporal domains.

I will also investigate whether lexical flexibility is sensitive to the *topicality hierarchy*, though to reflect the relative center of interest in discourse (DeLancey 1981; Mallinson & Blake 1981:86), though it has also been termed the agency hierarchy (Dixon 1979:85), animacy hierarchy (Comrie 1978:385–388; 1981:128), nominal hierarchy (Dixon 1994:85), noun phrase hierarchy (Filimonova 2005), person/animacy hierarchy (Blake 2004:137), and referential hierarchy (Bickel 2008). This hierarchy generally takes the form laid out in Figure 1, where more topical items are placed above less topical ones.

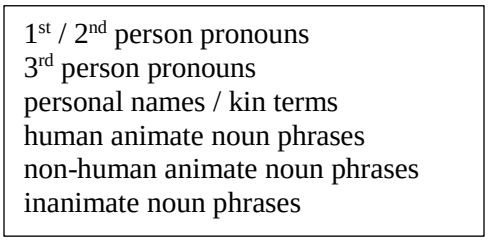

Figure 1. The topicality hierarchy

This scale of topicality has been shown to have relevance for many areas of the grammar across languages (Comrie 1978:385–388; 1981:128; Corbett 2000:56; Blake 2004:137; Siewierska 2004:148–161). I hypothesize that items typically placed higher on the topicality hierarchy are more likely to appear in nominal constructions, while those lower on the hierarchy are more likely to appear in verbal constructions, and to exhibit lexical flexibility generally.

Finally, this chapter will examine whether lexical flexibility among property concepts follows patterns similar to Dixon's (1977) crosslinguistic typology of adjectives. That is, are core property concepts (age, dimension, value, color) more or less flexible than more typologically peripheral property concepts (physical characteristics, shape, human propensity, speed)? I hypothesize that core property concepts will exhibit a greater degree of lexical flexibility than more peripheral ones.

4.4. Chapter 4: Lexical Flexibility and Grammatical Role

The focus of this chapter is to determine whether the current and/or previous choice of grammatical role for a lexeme in context correlates with choice of lexical category. According to Hopper & Thompson (1984), we should expect that continuing topics are more likely to be

construed with nominal constructions, since they are manipulable by the discourse. As such, I hypothesize that items which have previously been coded in a subject, ergative, or agent construction (depending on the language) will be more likely to be encoded using a nominal construction in their subsequent appearances.

4.5. Chapter 5: Lexical Flexibility and Information Status

This chapter is the core focus of the dissertation, and asks whether the choice of lexical category for an item corresponds to its information status (given, new, or activated), independent of its grammatical role. I hypothesize that given items are more likely to be coded using verbal constructions, while new items are more likely to be introduced into the discourse using nominal constructions. Moreover, Thompson (1989) finds that property concept words have two different functions in discourse: 1) to predicate a property of an established discourse referent, in which case they exhibit more verbal coding, and 2) to introduce a new discourse referent, in which case they exhibit more nominal coding. Accordingly, I hypothesize that the choice of lexical category for flexible property concept words will follow this same bifurcated pattern, depending on the function of the item at that point in the discourse.

4.6. Chapter 6: Conclusion – The Grammaticization of Categoriality

I conclude the dissertation by discussing the broader implications of the preceding chapters. I discuss the role of diachrony and lexicalization, and argue that speakers possess an immense amount of item-specific knowledge regarding what types of constructions are permissible for individual lexemes, and what meanings result from using roots in different pragmatic functions. However, I expect to find that the data also show various trends suggestive of a process whereby discourse tendencies—such as the tendency for new topics to appear in nominal constructions—become grammaticized as dedicated markers of pragmatic functions to varying degrees in different languages.

5. Timeline

I hope to complete and defend the dissertation by the end of the 2017–2018 academic year, with the possibility of extending the timeline to a second year. Meeting this goal requires following a rigorous timeline, laid out below.

Jan 15 - 26	Committee reads prospectus Finish Chapter 1
Jan 29 - Feb 9	Prospectus revisions Finish Chapter 2
Feb 12 - 23	Committee reads revised prospectus Finish Chapter 3
Feb 23 (?)	Prospectus defense
Feb 26 - April 6	6 weeks total
	Write Chapter 4 (2 weeks)
	Write Chapter 5 (2 weeks)
	Write Conclusion (2 weeks)
April 9 - May 4	Committee reads thesis
May 7 - 18	Thesis revisions
May 21 - 31	Committee reads revised thesis
June 1 (?)	Thesis defense

References

Arad, Maya. 2003. Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. *Natural Language & Linguistic Theory* 21: 737–778.

Baker, Mark & William Croft. 2017. Lexical categories: Legacy, lacuna, and opportunity for functionalists and formalists. *Annual Review of Linguistics* 3(2): 1–19.

Beck, David. 2016. Some language-particular terms are comparative concepts. *Linguistic Typology* 20(2): 395–402.

Bickel, Balthasar. 2008. On the scope of the referential hierarchy in the typology of grammatical relations. In Greville G. Corbett & Michael Noonan (eds.), *Case and grammatical relations: Papers in honor of Bernard Comrie*, pp. 191–200. (Typological Studies in Language 81). Amsterdam: John Benjamins.

Blake, Barry J. 2004. *Case*. 2nd ed. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Boas, Franz. 1911. Introduction. In Franz Boas (ed.), *Handbook of American Indian languages, Part 1*, pp. 1–84. (Bureau of American Ethnology Bulletins 40). Washington, D.C.: Smithsonian Institution.

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

Cauchard, Aurelia. 2017. Describing lexical flexibility in Caac (New Caledonia). *Studies in Language* 41(2): 521–542.

Chafe, Wallace L. 2012. Are adjectives universal? The case of Northern Iroquoian. *Linguistic Typology* 16(1): 1–39.

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

Comrie, Bernard. 1978. Ergativity. In Winfred P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, pp. 329–392. Austin: University of Texas Press.

Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. 2nd ed. Chicago: University of Chicago Press.

Corbett, Greville G. 2000. *Number*. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Croft, William. 1990. A conceptual framework for the grammatical categories (or: A taxonomy of propositional acts). *Journal of Semantics* 7(3): 245–280.

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

Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, William. 2002. *Typology and universals*. 2nd ed. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

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

Croft, William. 2016. Comparative concepts and language-specific categories: Theory and practice. *Linguistic Typology* 20(2): 337–393.

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

Crystal, David. 2008. *A dictionary of linguistics and phonetics*. 6th ed. (The Language Library). Malden, MA: Blackwell.

DeLancey, Scott. 1981. An intepretation of split ergativity and related patterns. *Language* 57(3): 626–657.

Dixon, Robert M. W. 1979. Ergativity. *Language* 55(1): 59–138.

Dixon, Robert M. W. 1994. *Ergativity*. (Cambridge Studies in Linguistics 69). Cambridge: Cambridge University Press.

Dixon, Robert M. W. 1997. Where have all the adjectives gone? *Studies in Language* 1(1): 19–80.

Dixon, Robert M. W. 2004. Adjective classes in typological perspective. In Robert M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Adjective classes: A cross-linguistic typology*, pp. 1–49. (Explorations in Linguistic Typology 1). Oxford: Oxford University Press.

Don, Jan. 2004. Categories in the lexicon. *Linguistics* 42(5): 931–956.

Don, Jan & Eva van Lier. 2003. Derivation and categorization in flexible and differentiated languages. In Jan Rijkhoff & Eva va Lier (eds.), *Flexible word classes: Typological studies of underspecified parts of speech*, pp. 56–88. Oxford: Oxford University Press.

Dorvlo, Kofi. 2009. Does Logba have an adjective class? In Masangu Deus Matondo, Fiona McLaughlin, & Eric Potsdam (eds.), *Selected proceedings of the 38th Annual Conference on African Linguistics: Linguistic theory and African language documentation*, pp. 95–105. Somerville, MA: Cascadilla Proceedings Project.

DuBois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson, & Mii Martey. 2000. *Santa Barbara Corpus of Spoken American English*. Philadelphia: Linguistic Data Consortium.

Eijk, Jan P. Van & Thom Hess. 1986. Noun and verb in Salish. *Lingua* 69(4): 319–331.

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

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

Filimonova, Elena. 2005. The noun phrase hierarchy and relational marking: Problems and counterevidence. *Linguistic Typology* 9(1): 77–113.

Floyd, Simeon. 2011. Re-discovering the Quechua adjective. *Linguistic Typology* 15(1): 25–63.

François, Alexandre. 2017. The economy of word classes in Hiw, Vanuatu. *Studies in Language* 41(2): 294–357.

Gil, David. 2005. Isolating-monocategorial-associational language. In Henri Cohen & Claire Lefebvre (eds.), *Handbook of categorization in cognitive science*, pp. 348–377. Amstedam: Elsevier.

Givón, Talmy. 1979. *On understanding grammar*. (Perspectives in Neurolinguistics & Psycholinguistics). New York: Academic Press.

Harris, Zellig. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.

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

Haspelmath, Martin. 2010. The interplay between comparative concepts and descriptive categories (Reply to Newmeyer). *Language* 86(3): 696–699.

Haspelmath, Martin. 2014. (Non-)universality of word-classes and words: The mid-20th century shift. *History & Philosophy of the Language Sciences*. Accessible at: https://hiphilangsci.net/2014/10/08/non-universality-of-word-classes-and-words-the-mid-20th-century-shift/.

Hengeveld, Kees. 1992. *Non-verbal predication: Theory, typology, diachrony*. (Functional Grammar Series 15). Berlin: Mouton de Gruyter.

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

Hengeveld, Kees, Jan Rijkhoff & Anna Siewierska. 2004. Parts-of-speech systems and word order. *Journal of Linguistics* 40(3): 527–570.

Holton, Gary. 1999. Categoriality of property words in a switch-adjective language. *Linguistic Typology* 3(3): 341–360.

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

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. 2nd ed. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Jacobsen, William H. 1979. Noun and verb in Nootkan. In Barbara S. Efrat (ed.), *The Victoria conference on northwestern languages*, pp. 83–155. Victoria, B.C.: British Columbia Provincial Museum.

Kinkade, M. Dale. 1983. Salish evidence against the universality of "noun" and "verb." *Lingua* 60(1): 25–39.

Koch, Karsten & Lisa Matthewson. 2009. The lexical category debate in Salish and its relevance for Tagalog. *Theoretical Linguistics* 35(1): 125–137.

Kuipers, Aert H. 1968. The categories verb-noun and transitive-intransitive in English and Squamish. *Lingua* 21: 610–626.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lichtenberk, Frank. 2017. Lexical and grammatical flexibility in Toqabaqita. *Studies in Language* 41(2): 496–501.

Lier, Eva van. 2006. *Parts-of-speech systems and dependent clauses: A typological study*. Utrecht: LOT.

Lier, Eva van. 2016. Lexical flexibility in Oceanic languages. *Linguistic Typology* 20(2): 197–232.

Lier, Evan van (ed.). 2017. *Lexical flexibility in Oceanic languages*. (Studies in Language 41). Amsterdam: John Benjamins.

Lier, Eva van & Jan Rijkhoff. 2013. Flexible word classes in linguistic typology and grammatical theory. In Eva van Lier & Jan Rijkhoff (eds.), *Flexible word classes: Typological studies of underspecified parts of speech*, pp. 1–30. Oxford: Oxford University Press.

Lier, Eva van. 2017. Introduction: Lexical flexibility in Oceanic languages. *Studies in Language* 41(2): 241–254.

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

Lyons, John. 1977. *Semantics, Vol. 2*. Cambridge: Cambridge University Press.

Mallinson, G. & B. J. Blake. 1981. *Language typology: Cross-linguistic studies in syntax*. (North-Holland Linguistic Series 46). Amsterdam: North-Holland.

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*. Accessible at: https://hiphilangsci.net/2013/06/12/the-creation-of-parts-of-speech-for-chinese-translingual-practice-across-graeco-roman-and-sinitic-traditions/.

McGregor, William B. 2013. Lexical categories in Gooniyandi, Kimberley, Western Australia. In Jan Rijkhoff & Eva van Lier (eds.), *Flexible word classes: Typological studies of underspecified parts of speech*, pp. 221–246. Oxford: Oxford University Press.

Mithun, Marianne. 1999. *The languages of Native North America*. (Cambridge Language Surveys). Cambridge: Cambridge University Press.

Mithun, Marianne. 2000. Noun and verb in Iroquoian languages: Multicategorisation from multiple criteria. In Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes*, pp. 397–420. (Empirical Approaches to Language Typology 23). Berlin: Mouton de Gruyter.

Mithun, Marianne. 2012. Core argument patterns and deep genetic relations: Hierarchical systems in Northern California. In Pirkko Suihkonen, Bernard Comrie & Valery Solovyev (eds.), *Argument structure and grammatical relations: A crosslinguistic typology*, pp. 257–294. (Studies in Language Companion Series 126). Amsterdam: John Benjamins.

Mithun, Marianne. 2013. Prosody and independence: Free and bound person marking. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, pp. 291–312. Berlin: Mouton de Gruyter.

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*, pp. 155–176. (Studies in Language Companion Series 182). Amsterdam: John Benjamins.

Nakayama, Toshihide. 2002. *Nuuchahnulth (Nootka) morphosyntax*. (University of California Publications in Linguistics 134). Berkeley: University of California Press.

Palmer, Bill. 2017. Categorial flexibility as an artefact of the analysis. *Studies in Language* 41(2): 408–444.

Pustet, Regina. 2000. How arbitrary is lexical categorization? Verbs vs. adjectives. *Linguistic Typology* 4(2): 175–212.

Ramat, Paolo. 2009. How universal are linguistic categories? *Universals of language today*, pp. 1–12. (Studies in Natural Language & Linguistic Theory 76). Dordrecht: Springer.

Rauh, Gisa. 2010. *Syntactic categories: Their identification and description in linguistic theories*. (Oxford Surveys in Syntax & Morphology 7). Oxford: Oxford University Press.

Rijkhoff, Jan. 2007. Word classes. *Language & Linguistics Compass* 1(6): 709–726.

Rijkhoff, Jan. 2016. Crosslinguistic categories in morphosyntactic typology: Problems and prospects. *Linguistic Typology* 20(2): 333–363.

Rijkhoff, Jan & Eva van Lier (eds.). 2013. *Flexible word classes: Typological studies of underspecified parts of speech*. Oxford: Oxford University Press.

Rosch, Eleanor H. 1973a. Natural categories. *Cognitive Psychology* 4(3): 328–350.

Rosch, Eleanor H. 1973b. On the internal structure of perceptual and semantic categories. *Cognitive development and the acquisition of language*, pp. 111–144. New York: Academic Press.

Rosch, Eleanor H. 1975. Cognitive representation of semantic categories. *Journal of Experimental Psychology* 104(3): 192–233.

Rosch, Eleanor H. 1978. Principles of categorization. In Eleanor Rosch & B. B. Lloyd (eds.), *Cognition and categorization*, pp. 27–48. Hillsdale, NJ: Lawrence Erlbaum.

Rosch, Eleanor H. & Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7(4): 573–605.

Rosch, Eleanor H., Carolyn B. Mervis, Wayne D. Gray, David M. Johnson & Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8(3): 382–439.

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.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace & Co.

Siewierska, Anna. 2004. *Person*. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Taylor, John R. [1989] 2003. *Linguistic categorization: Prototypes in linguistic theory*. 2nd ed. Oxford: Clarendon Press.

Thompson, Sandra A. 1989. A discourse approach to the cross-linguistic category ‘Adjective.’ In Roberta Corrigan, Fred R. Eckman & Michael Noonan (eds.), *Linguistic categorization*, pp. 245–266. (Current Issues in Linguistic Theory 61). Amsterdam: John Benjamins.

Vapnarsky, Valentina & Edy Veneziano. 2017a. 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*, pp. 1–34. (Studies in Language Companion Series 182). Amsterdam: John Benjamins.

Vapnarsky, Valentina & Edy Veneziano (eds.). 2017b. *Lexical polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches*. (Studies in Language Companion Series 182). Amsterdam: John Benjamins.