

Dissertation Prospectus:
Lexical flexibility in discourse

Daniel W. Hieber

University of California, Santa Barbara

1 Introduction

This dissertation presents a typology of flexible word classes, i.e. lexical categories which appear to subsume more than one traditional part of speech (Hengeveld 1992; Rijkhoff 2007; Luuk 2010; Rijkhoff & van Lier 2013; van Lier 2016 and accompanying articles; Vapnarsky & Veneziano 2017) and investigates the discourse-functional motivations for their existence. Flexible categories have become a vibrant topic in recent years, prompting discussions on the existence of flexible categories in particular languages (Broschart 1997; Chafe 2012; Chung 2012; Dorvlo 2009; Evans & Osada 2005a; Floyd 2011; Hengeveld & Rijkhoff 2005; Kinkade 1983; Koch & Matthewson 2009; Sadock 1999; Van Eijk & Hess 1986), the plausibility of flexible categories in general (Baker & Croft 2017; Croft 2005; Dixon 1982; Don 2004; Evans & Osada 2005a; Luuk 2010; Palmer 2017), and detailed typological studies of the nature of flexible categories and the diversity of their expression across languages (Hengeveld 1992; Hengeveld, Rijkhoff & Siewierska 2004; Holton 1999; Lichtenberk 2017; van Lier 2006; Luuk 2010; Rijkhoff & van Lier 2013; van Lier 2016; Cauchard 2017; Vapnarsky & Veneziano 2017). This dissertation contributes to this literature in two important ways:

First, the dissertation builds upon existing surveys of lexical flexibility (Croft 2001; Croft 2005; Evans & Osada 2005b; Rijkhoff & van Lier 2013; van Lier 2017), to propose ten criteria by which to assess the degree of lexical flexibility in a language. Subsequent chapters then apply those criteria to a small but diverse sample of languages to provide a quantitative look at the degree of lexical flexibility in those languages. It is hypothesized that languages will vary

drastically in both the degree and manner of expression of lexical flexibility, and that these differences will be borne out quantitatively by the data.

Second, while the existence and nature of lexical flexibility has been thoroughly debated, little attention has been paid to its discourse-functional motivations (though see Thompson (1989), Nakayama (1997), and Hopper & Thompson (1984)). *Why* are some languages more flexible than others? And why, in flexible languages, do speakers make the particular categorial choices in discourse that they do? If a given lexeme can more-or-less freely alternate between, say, nominal and verbal uses, what determines when a speaker uses one function over the other? The presence of lexical flexibility in a language provides a dimension of variation that speakers can manipulate to achieve their manifold discourse goals, since any choice between linguistic alternatives provides yet another means of conveying information. How then is lexical flexibility deployed in discourse? The final section of the dissertation is a first attempt at an answer to this question, summarizing the discourse-functional correlates of lexical flexibility among the languages surveyed.

This focus on the role of lexical flexibility in discourse diverges significantly from 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 particular language. Instead, I begin with the fact that all languages have some lexemes (however few) that exhibit lexical flexibility to varying degrees (however small), and that categorial distinctions between lexemes are more strongly and consistently expressed in some languages than others—facts which will be supported with empirical data in this dissertation. By first acknowledging the existence of lexical flexibility and its variability across languages, it becomes possible to make a first foray into investigating

the functions that lexical flexibility serves in discourse—a task which this dissertation sets out to accomplish.

The specific research questions asked by this dissertation are as follows:

- How do languages differ in their degree and expression of lexical flexibility?
- Crosslinguistically, do certain semantic concepts tend to exhibit greater lexical flexibility than others? If so, does this variation correspond to degree of inherent topicality? To grammatical role? To information status?
- In languages with flexible roots, stems, or wordforms, does the choice of lexical category depend on its information status (given vs. new, backgrounded vs. foregrounded, topic vs. focus)? Its inherent topicality? Its affectedness? Its grammatical role?
- Does lexical flexibility correlate with word order flexibility? In languages with both flexible word order and flexible lexemes, is there a correlation between choice of lexical category and choice of word order?

I expect the dissertation to conclude, based on the empirical data from the preceding chapters, that languages diverge significantly in the degree to which the pragmatic functions of reference, predication, and modification have become grammaticized into the structure of the language, and the degree to which those structures overlap or align to create fuzzy, emergent categories. Languages in which these pragmatic functions are strongly grammaticized in consistent and distinct ways show strong evidence of traditional lexical categories, while those in which the pragmatic functions are only weakly grammaticized or in non-distinct ways have much more flexible word classes. Ultimately, however, the grammaticization and distinctiveness of word classes in a language is a matter of degree, and it

is only through attentiveness to the dimensions of variation in lexical categories that we can begin to understand the functional motivations behind this diversity.

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 novel approach toward the study of flexible categories 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 European 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 ultimately 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). 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; 1978; Rosch & Mervis 1975; Rosch et al. 1976) and popularized among linguists by Lakoff (1987) and Taylor (1989 [2003]), that lexical categories are prototypical. Taylor in particular advances the thesis that lexical categories are prototypical, 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.¹ 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).

¹ See also DeLancey (1997) for a similar well-argued critique of the distributional method.

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 particular 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 propositional act function of either referring, predicating, or modifying, that form will be unmarked. However, when an item is used in a non-prototypical manner, such as an entity-denoting concept being used for predication, that use is structurally and/or behaviorally marked (Croft 2002). 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). This dissertation will utilize Croft’s typological markedness theory as a basis for investigating the diversity of lexical flexibility in particular languages.

It should be noted, however, that the issue of whether lexical categories should be thought of as language-specific, and potentially incommensurable and uncomparable across languages, or as instantiations of crosslinguistically valid categories, 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. Instead, my interest is in exactly the kinds of cross-constructural generalizations which Croft refers to—how specific constructions in specific languages work in tandem to express (via coding or behavioral tendencies) categorial distinctions between lexemes, more or less strongly. Put differently, the focus of this dissertation is investigating the extent to which constructions *align* to form emergent, gradient categories that have become grammaticized to varying degrees.

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 the data from Native American languages (Boas 1911; Jacobsen 1979; Kinkade 1983; Kuipers 1968; Sadock 1999; Sapir 1921). 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, and suggests that while none of these analyses is fully explanatory by themselves, the mechanisms they posit each contribute to a broader typology of 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 categorial distinctions, but that the evidence is simply subtle (Dixon 2004; Floyd 2011; Palmer 2017).² In this approach, traditional categories are typically thought to be universally instantiated, to be found in all of the world’s languages

² Note that Dixon’s position on the universality of adjectives has shifted over time; cf. Dixon (1977).

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. More germane to this dissertation, however, is the fact that 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, as this dissertation attempts to do.

In stark contrast to the first approach, some have embraced the existence of flexible categories and argued extensively for their existence (Broschart 1997; Gil 2005; Hengeveld 1992; Hengeveld & Rijkhoff 2005; Kinkade 1983; Kuipers 1968; Luuk 2010; Rijkhoff & van Lier 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 classic typology of parts of speech. All these proposals have garnered heavy criticism. It is important for any typology of lexical flexibility to incorporate these criticisms, so I briefly review them here.

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 (2005a), for example, criticize Hengeveld’s notion of flexible categories (Hengeveld 1992; Hengeveld & Rijkhoff 2005) on the basis that the meaning of a lexeme changes when it is used in different functions. Mithun also

has in various studies (1999:56–67; 2000; 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. Yet even in cases where semantic shifts are patterned and non-idiosyncratic, the particular 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 precategoryal, i.e. underspecified for lexical category (Arad 2003; Broschart 1997; Don & van Lier 2003; Farrell 2001; Hopper & Thompson 1984). In precategoryal languages, lexical categorization is thought to be a property of the particular morphosyntactic constructions 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 (Hengeveld & Rijkhoff 2005; Hengeveld, Rijkhoff & Siewierska 2004; Farrell 2001; McGregor 2013). In this approach, the relevant component of the meaning of the lexeme is highlighted by its particular 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 2005a).

The emergent approach put forward in this dissertation has the potential to reconcile these opposing perspectives. If one begins with the premise that all languages show flexibility to varying degrees (cf. Luuk (2010:362), who considers the possibility that all languages have at least some flexible lexemes), and that even individual lexical items may be more-or-less flexible, then it becomes entirely plausible that each of the mechanisms mentioned above are at work in the grammar of a language, each to varying degrees, and that it is the interplay of these mechanisms which give rise to the fuzzy prototypal categories that we see. What I hope to show in this dissertation is that the degree of association between a particular lexical item and a lexical category is a dimension of variability, and a matter of degree. Some lexemes are strongly specified for lexical category, while others are underspecified and receive some of their semantic interpretation from local context. Categoriality can be grammaticized at various levels of the grammar. Likewise, certain aspects of the semantic profile of a lexeme are highlighted more strongly in particular contexts. At the same time, context isn't everything. Speakers clearly have a vast store of item-specific lexical knowledge, and are aware of the range of constructions that an item can occur in. Understanding lexical categorization, then, requires an understanding of each of the dimensions of variability regarding exactly where categorization lives in the grammar.

2.3 Functional motivations for lexical flexibility

This section briefly summarizes the relevant literature on the interaction of discourse and lexical categories, and in particular 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 (Givón 1979:320–321; Lyons 1977:442–447; Sapir 1921:117–119), which are thought to have a prototypical 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 a lexeme tends to show a greater degree of nominal coding and behavior when it is used to introduce new referents into the discourse, but more verbal coding and behavior when being 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 (Nootka), 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." (p. 57). Categorical choice in Nuuchahnulth thus appears to be driven primarily by discourse and information-structural 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

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

- prominence in debates on lexical flexibility (e.g. Nuuchahnulth, Riau Indonesian, English)
- geographic and typological diversity: isolating (Riau Indonesian) vs. (poly)synthetic (Nuuchahnulth, Chitimacha); Africa (Swahili, Kisii) vs. North America (Nuuchahnulth, Chitimacha) vs. Austronesia (Riau Indonesian)
- availability of extensive lexical data and corpora (English, Swahili, Nuuchahnulth, and to a certain extent all the languages in the sample)
- differences in purported degree of lexical flexibility: highly flexible (Riau Indonesian, Nuuchahnulth) vs. highly rigid (French, Latin) vs. intermediate (Swahili)

The final list of languages to be included in the sample is not yet determined, and is expected to expand as time permits. However, at a minimum the following languages will be included: Chitimacha, English, Iñupiaq and/or Yup'ik, Nuuchahnulth, Riau Indonesian, Swahili and/or Kisii. Extensive corpora and lexicons are available for each.

For each case study in the dissertation, I will adopt different methods as appropriate to the research question. Chapters 5 & 6, on the degree and extent of flexibility, are intended to be largely quantitative, providing statistical analyses of variability in lexical flexibility across the

language sample. This will complement, and borrow methods from, existing quantitative work on lexical flexibility by van Lier (2016). Many of the counts in this section will need to be normalized to account for the different sizes of the corpora and lexicons. For example, it is likely that the number of distinct lexical categories that a lexeme may be used in is directly proportional to the size of the corpus, since the linguist is likely to continue discovering additional uses of a lexeme the larger the corpus becomes. Thus the reported number of lexical categories that a lexeme occurs in will be normalized to the size of the corpus, as well as perhaps other factors. I expect that, when summarized, the quantitative data on lexical flexibility will *not* show a stark bifurcation between strongly rigid and strongly flexible languages. Instead, I expect the degree of lexical flexibility to be largely scalar, suggesting that typologists should abandon the strict dichotomy between plainly rigid vs. plainly flexible languages.

Chapters 7 & 8, on the discourse-functional correlates of lexical flexibility, will still include quantitative components as appropriate, but will be structured as a typological survey and synthesis of existing grammatical descriptions as they relate to lexical flexibility and discourse. The language sample for these sections will be significantly larger, pulling from a large number of published grammars. This larger sample will take into account crosslinguistic sampling considerations (Bakker 2011; Daniel 2007; Hammarström 2009; Widmann 2001) to create an unbiased sample to the greatest extent possible.

4 Outline

The planned outline of the dissertation is as follows:

4.1 Section I: Preliminaries

4.1.1 Chapter 1: Introduction – The problem of lexical categories for typology

A brief introductory chapter overviewing the research question, its importance, and the plan of the dissertation. This chapter will likely be based heavily on this prospectus.

4.1.2 Chapter 2: Background – Approaches to lexical flexibility

This chapter surveys the prior literature on lexical flexibility, with a particular focus on its known functional motivations, especially discourse functions. Given that the literature on word classes and lexical flexibility is quite vast, this chapter does not aim to be an exhaustive survey of studies relating to parts of speech. Instead, it will focus primarily on typological surveys, and studies that explicitly treat the problem of lexical flexibility, or the high-level theoretical and methodological concerns relating to word classes. This chapter will make clear the need for studies, like this proposed dissertation, which explicate the functional motivations of lexical flexibility.

4.1.3 Chapter 3: Theoretical framework – A typology of lexical flexibility

This chapter briefly surveys, and then expands upon, existing definitions of lexical flexibility to propose a crosslinguistically valid definition that views lexical flexibility as resulting from the degree to which categorical distinctions have (or have not) become conventionalized into the grammar of a language. In some languages, categorical distinctions between lexemes are supported by copious pieces of grammatical evidence which strongly align, while in other languages the evidence for categorical distinctions is subtle and/or provides overlapping or conflicting results. The chapter then outlines ten criteria by which one might assess the degree of lexical flexibility in a lexicon. Though space does not permit

elaboration or precise definitions of each of the criteria here, I briefly list and explain them below:

1. *Structural coding*: Words which require no additional structural coding to change functions (i.e. conversion, zero-derivation) are more flexible than those that do.
2. *Inflectional category*: Words whose category can be inferred from their inflectional class are more rigid than words which show the same inflection across functions.
3. *Distributional potential*: Words which are limited in their distributional potential are more rigid than those that are not.
4. *Functional potential*: Words which are limited in the pragmatic functions they can fulfill are more rigid than those that are not.
5. *Featural potential*: Words which are limited in the morphological feature values they permit are more rigid than those which are not.
6. *Semantic shift*: Words that show little to no semantic shift, from one function to another are more flexible than those which are idiomatic.
7. *Semantic compositionality*: Words whose meaning is predictable when changing functions are more flexible than those whose meaning is idiosyncratic.
8. *Scope across functions*: Languages whose words generally show greater functional potential are more flexible than languages whose words show less functional potential.
9. *Scope across lexemes*: Languages which a greater number of flexible lexemes are more overall more flexible than those with fewer flexible lexemes.
10. *Scope across both lexemes and functions*: Languages with fewer strongly-grammaticized categories are more flexible than languages with multiple strongly-grammaticized categories.

4.1.4 Chapter 4: Data – The language sample

This chapter introduces the sample of languages, the motivations for selecting each language for inclusion in the study, and the nature of the data for each language. Data will consist of both lexical and textual material. Depending on the quality and type of the data, not all languages will be included in discussions of all ten criteria for lexical flexibility. A brief overview of the relevant features of the grammar of each language in the sample will also be provided.

4.2 Section II: Results

4.2.1 Chapter 5: Indistinguishability of categories

This chapter applies several of the criteria from Ch. 3 to a small sample of languages, with the hypothesis that languages will vary widely in their degree of lexical flexibility both overall and along each of the individual criteria. If this hypothesis is correct, it provides strong empirical evidence that languages differ drastically in the strength of the distinctions between lexical categories. This chapter looks particularly at how languages vary in regard to those criteria relating to indistinguishability of categories in a language, including: structural coding, inflection, distributional potential, feature values, lexical semantics, and semantic shift.

4.2.2 Chapter 6: Scope of flexibility

This chapter examines the remaining criteria for lexical flexibility, which pertain to the scope of that flexibility across lexemes and pragmatic functions.

4.3 Chapter 7: Lexical flexibility and categorical choice

This chapter examines the factors that contribute to choice of lexical category for flexible roots, stems, or words. If a given lexical item appears variously in morphosyntactic

constructions associated with different lexical categories, what determines when that item appears in one construction versus another? The possible determinants examined are grammatical role, affectedness, information status, and inherent topicality.

4.4 Chapter 8: Lexical flexibility and word order

This chapter investigates two questions: 1) Does a high level of overall lexical flexibility in a language correlate with a highly flexible word order? 2) Does choice of lexical category for a flexible stem correlate with choice of word order?

4.5 Chapter 9: Lexical flexibility and semantic domains

This chapter utilizes data on lexical flexibility from Chs. 4 & 5 to determine whether certain semantic concepts are more likely to participate in categorical alternations than others. Lexemes that have approximate parallels in each of the languages studied in Ch. 4 will be compared and ranked in terms of their overall flexibility. The chapter will then discuss any noticeable patterns or clusters in the semantic domains for the most flexible items.

4.6 Section III: Conclusion

4.6.1 Chapter 10: Conclusion

This chapter summarizes the empirical findings of the thesis and emphasizes the major theoretical implications, namely that a) lexical categories should be viewed as emergent and gradient rather than categorical; b) languages vary as to the strength with which they express categorial distinctions; and that c) lexical flexibility is functionally motivated and deployed by speakers towards various discourse ends.

4.7 Section IV: Appendices

The raw quantitative data for each of the case studies in this dissertation will be provided as a set of appendices, to enable independent confirmation of statistical tests, and future replicability.

5 Timeline

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

Oct 23 - Nov 3, 2017	Committee reviews prospectus
Nov 6 - 17, 2017	Prospectus revisions
Nov 20 - Dec 1, 2017	Committee reads revised prospectus
Dec 1 / 4	Prospectus defense
Dec 4 - 15	Chapters 1-2: Literature Review (2 weeks, already mostly written)
Dec 25, 2017	Christmas
Dec 26 - 29, 2017	Chapter 3: A typology of lexical flexibility (1 week, already mostly written)
Jan 1 - 5, 2018	LSA
Jan 8 - April 6, 2018	13 weeks total
	Chapters 5 & 6: Degree and scope of lexical flexibility (4 weeks)
	Chapter 7: Categorical choice (3 weeks)
	Chapter 8: Word order (2 weeks, or potentially remove)
	Chapter 9: Semantic domains (2 weeks, or potentially remove)
	Conclusion, Data & Methods, finishing touches (2 weeks)
April 9 - May 4, 2018	Committee reads thesis
May 7 - 18, 2018	Thesis revisions
May 21 - 31, 2018	Committee reads revised thesis
June 1, 2018	Thesis defense

6 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. doi:10.1146/annurev-linguistics-011516-034134. <http://www.annualreviews.org/doi/10.1146/annurev-linguistics-011516-034134>.
- Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 100–130. (Oxford Handbooks in Linguistics). Oxford: Oxford University Press.
- Beck, David. 2016. Some language-particular terms are comparative concepts. *Linguistic Typology* 20(2). 395–402. doi:10.1515/lingty-2016-0013.
- Boas, Franz. 1911. Introduction. *Handbook of American Indian Languages, Part 1*. (Bureau of American Ethnology Bulletin 40). Washington, D.C.: Smithsonian Institution.
- Broschart, Jürgen. 1997. Why Tongan does it differently: Categorical distinctions in a language without nouns and verbs. *Linguistic Typology* 1(1997). 123–165. doi:10.1515/lity.1997.1.2.123.
- Cauchard, Aurelie. 2017. Describing lexical flexibility in Caac (New Caledonia). *Studies in Language* 41(2). 521–542. doi:10.1075/sl.41.2.09cau. <http://www.jbe-platform.com/content/journals/10.1075/sl.41.2.09cau>.
- 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.
- 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*, 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. doi:10.1093/acprof:oso/9780198299554.001.0001.
<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198299554.001.0001/acprof-9780198299554>.
- 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. doi:10.1515/lity.2005.9.3.391.
- Croft, William. 2016. Comparative concepts and language-specific categories: Theory and practice. *Linguistic Typology* 20(2). 377–393. doi:10.1515/lingty-2016-0012.
- Croft, William & Eva van Lier. 2012. Language universals without universal categories. *Theoretical Linguistics* 38(1–2). 57–72. doi:10.1515/tl-2012-0002.
- Daniel, Michael. 2007. Representative sampling and typological explanation: A phenomenological lament. *Linguistic Typology* 11(1). 69–78. doi:10.1515/LINGTY.2007.006.
- DeLancey, Scott. 1997. Grammaticalization and the gradience of categories: Relator nouns and postpositions in Tibetan and Burmese. *Essays on language function and language type: Dedicated to T. Givón*.
- Dixon, Robert M. W. 1977. Where have all the adjectives gone? *Studies in Language* 1(1). 19–80.
- Dixon, Robert M. W. 1982. *Where have all the adjectives gone? and other essays in Semantics and Syntax*. *Studies in Language*. doi:10.1038/014265a0.
- 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*, 1–49. (Explorations in Linguistic Typology 1). Oxford: Oxford University Press.
- Don, Jan. 2004. Categories in the lexicon. *Linguistics* 42(5). 931–956. doi:10.1515/ling.2004.033.
- Don, Jan & Eva van Lier. 2003. Derivation and categorization in flexible and differentiated languages. In Jan Rijkhoff & Eva van Lier (eds.), *Flexible word classes: Typological studies of underspecified parts of speech*, 56–88. Oxford: Oxford University Press.
- Dorvlo, Kofi. 2009. Does Logba have an adjective class? . 95–105.
- Eijk, Jan P. Van & 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. 2005a. Mundari: The myth of a language without word classes. *Linguistic Typology* 9(2005). 351–390. doi:10.1515/lity.2005.9.3.351.
- Evans, Nicholas & Toshiki Osada. 2005b. Mundari and argumentation in word-class analysis. *Linguistic Typology* 9. 442–457.
- 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(1). 25–63. doi:10.1515/LITY.2011.003.
- Gil, David. 2005. Isolating-monocategorial-associational language. In Henri Cohen & Claire Lefebvre (eds.), *Handbook of categorization in cognitive science*, 348–377. Amsterdam: Elsevier.
- Givón, Talmy. 1979. *On understanding grammar*. (Perspectives in Neurolinguistics & Psycholinguistics). New York: Academic Press.
- Hammarström, Harald. 2009. Sampling and genealogical coverage in WALS. *Linguistic Typology* 13(1). 105–119. doi:10.1515/LITY.2009.006.
- 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. doi:10.1515/LINGTY.2007.011.
- Haspelmath, Martin. 2010. The interplay between comparative concepts and descriptive categories (Reply to Newmeyer). *Language* 86(3). 696–699. doi:10.1353/lan.2010.0021.
<http://muse.jhu.edu/content/crossref/journals/language/v086/86.3.haspelmath01.html>.
- Haspelmath, Martin. 2014. (Non-)universality of word-classes and words: The mid-20th century shift. *History & Philosophy of the Language Sciences*.
<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. doi:10.1515/lity.2005.9.3.391.
- Hengeveld, Kees, Jan Rijkhoff & Anna Siewierska. 2004. Parts-of-speech systems and word order. *Journal of Linguistics* 40(3). 527–570. doi:10.1017/S0022226704002762.
- Holton, Gary. 1999. Categoriality of property words in a switch-adjective language. *Linguistic Typology* 3(3). 341–360. doi:10.1515/lity.1999.3.3.341.
<http://www.degruyter.com/view/j/lity.1999.3.issue-3/lity.1999.3.3.341/lity.1999.3.3.341.xml>.
- Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in Universal Grammar. *Language* 60(4). 703–752. doi:10.1371/journal.pone.0005772.
<http://www.jstor.org/stable/413797>.
- Jacobsen, William H. 1979. Noun and verb in Nootkan. In Barbara S. Efrat (ed.), *The Victoria conference on northwestern languages*, 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. doi:10.1016/0024-3841(83)90045-1.

Koch, Karsten & Lisa Matthewson. 2009. The lexical category debate in Salish and its relevance for Tagalog. *Theoretical Linguistics* 35(1). 125–137. doi:10.1515/THLI.2009.007.

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.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. *Mind & Language*. Chicago: University of Chicago Press. doi:10.1111/j.1468-0017.1989.tb00245.x.

Lichtenberk, Frank. 2017. Lexical and grammatical flexibility in Toqabaqita. *Studies in Language* 41(2). 496–501. doi:10.1075/sl.41.2.07lic. <http://www.jbe-platform.com/content/journals/10.1075/sl.41.2.07lic>.

Lier, Eva van. 2006. *Parts-of-speech systems and dependent clauses: A typological study*. *Folia Linguistica*. Vol. 40. doi:10.1515/flin.40.3-4.239.

Lier, Eva van. 2016. Lexical flexibility in Oceanic languages. *Linguistic Typology* 20(2). 197–232. doi:10.1515/lingty-2016-0005.

Lier, Evan van. 2017. Introduction. *Studies in Language* 41(2). 241–254. doi:10.1075/sl.41.2.01van. <http://www.jbe-platform.com/content/journals/10.1075/sl.41.2.01van>.

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

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

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/>.

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*, 221–246. Oxford: Oxford University Press.

Mithun, Marianne. 1999. *The languages of Native North America*. 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*, 397–420. (Empirical Approaches to Language Typology 23). Berlin: Walter de Gruyter.

Mithun, Marianne. 2017. Polycategoriality and zero derivation: Insights from Central Alaskan Yup'ik Eskimo. (Ed.) Valentina Vapnarsky & Edy Veneziano. *Lexical polycategoriality: Cross-linguistic, cross-theoretical, and language acquisition approaches*. Amsterdam: John Benjamins. 155–174.

Nakayama, Toshihide. 1997. Discourse-pragmatic dynamism in Nuuchahnulth (Nootka) morphosyntax. University of California, Santa Barbara. doi:10.16953/deusbed.74839.

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

Palmer, Bill. 2017. Categorical flexibility as an artefact of the analysis. *Studies in Language* 41(2). 408–444. doi:10.1075/sl.41.2.05pal. <http://www.jbe-platform.com/content/journals/10.1075/sl.41.2.05pal>.

Pustet, Regina. 2000. How arbitrary is lexical categorization? Verbs vs. adjectives. *Linguistic Typology* 4(2). 175–212. doi:10.1515/lity.2000.4.2.175.

<http://www.degruyter.com/view/j/lity.2000.4.issue-2/lity.2000.4.2.175/lity.2000.4.2.175.xml>.

Ramat, Paolo. 2009. How universal are linguistic categories? *Universals of language today*, 1–12.

(Studies in Natural Language & Linguistic Theory 76). 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.

doi:10.1111/j.1749-818X.2007.00030.x. <http://doi.wiley.com/10.1111/j.1749-818X.2007.00030.x>.

Rijkhoff, Jan. 2016. Crosslinguistic categories in morphosyntactic typology: Problems and prospects. *Linguistic Typology* 20(2). 333–363. doi:10.1515/lingty-2016-0010.

Rijkhoff, Jan & Eva van Lier. 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.

doi:10.1016/0010-0285(73)90017-0.

Rosch, Eleanor H. 1973b. On the internal structure of perceptual and semantic categories.

Cognitive development and the acquisition of language, 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*, 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. doi:10.1016/0010-0285(75)90024-9.

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.

doi:10.1016/0010-0285(76)90013-X.

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.

doi:10.2307/3713880. <https://books.google.com/books?id=ofgrAAAAYAAJ>.

Taylor, John R. 1989. *Linguistic categorization: Prototypes in linguistic theory*. 1st 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*, 245–266. (Current Issues in Linguistic Theory 61). Amsterdam: John Benjamins.

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

Widmann, Thomas Martin. 2001. *Language sampling for typological studies*. University of Aarhus.