

Contents

List of Figures	xiii
List of Tables	xvii
1 Introduction	1
1.1 The role of lexicality in linguistic organization	1
1.2 Approach	6
1.3 Structure of the Thesis	8
1.4 Contributions	12
2 Background	15
2.1 Typology and Comparative Concepts	15
2.2 Approaches to Comparative Concepts in Computational Typology	17
2.2.1 Comparative concepts in typological databases	18
2.2.2 Phonological typology	19
2.2.3 Semantic category systems	22
2.2.4 Multidimensional scaling	25
2.2.5 Deep learning models of comparative concepts	28
2.3 The Distinctness of Lexicon and Syntax	30
2.3.1 Splitting the dimensions of the distinction	31
2.3.2 Relevant phenomena at different formal levels	34
2.3.3 Theories that reify the distinction	34

2.3.4	Doing without the distinction	34
2.3.5	Evidence from psychology and neuroscience	34
2.3.6	Computational operationalization	34
2.4	The Gradient Nature of Lexicality	35
2.5	Finding Meaning in Computational Models	35
I	Inflection and Derivation	37
3	Corpus-based Measures for Inflection and Derivation	39
3.1	Introduction	39
3.2	Motivation for our measures	44
3.3	Method	48
3.3.1	Orthography-based measures	49
3.3.2	Distributional-embedding-based measures	50
3.4	Data	54
3.4.1	Data selection and summary	56
3.5	Distribution of the individual measures	58
3.5.1	Effects of Frequency	60
3.6	The role of syntactic information	61
4	Predicting Inflection and Derivation Cross-linguistically	65
4.1	Predicting inflection and derivation	65
4.2	Classification of Linguistic Types of Inflection	69
4.2.1	Categories of inflectional meaning	70
4.2.2	Inherent vs. contextual inflection and transpositions	73
4.2.3	Summary	75
4.3	Discussion	76
4.3.1	The role of our individual measures	76
4.3.2	Language generality	79

4.3.3	The classification approach	81
4.3.4	Inflection and derivation: gradient or categorical?	83
4.3.5	Are inflection and derivation identifiable from the statistics of language?	84
4.3.6	Classification and syntactic change	87
4.3.7	Future work	88
4.4	Conclusion	90
II	Word Classes	93
5	Groundedness and the Lexical–Functional Distinction	95
5.1	Introduction	95
5.2	Background	98
5.2.1	Contentfulness and word class	99
5.2.2	Measuring contentfulness	99
5.3	Method	101
5.4	Experimental setup	103
5.5	Results	106
5.5.1	Which word classes are grounded?	106
5.5.2	Which word classes are more grounded?	108
5.5.3	How consistent is word class groundedness across languages?	110
5.5.4	Semantic dimension of the measure	112
5.6	Discussion and Conclusion	114
6	Splitting and lumping: Visual groundedness as an organizing factor among lexical classes	119
6.1	Introduction	119
6.2	Continua among lexical word classes	122

6.3	Japanese adjectives	126
6.3.1	Method	128
6.3.2	Results	130
6.3.3	Discussion	132
6.4	The Tensedness Correlation	135
6.4.1	The typological finding	136
6.4.2	Theoretical explanation of the finding	138
6.4.3	Methodological background	139
6.4.4	Results	146
6.4.5	Discussion	147
6.5	Conclusion	151
	Bibliography	153

Chapter 1

Introduction

1.1 The role of lexicality in linguistic organization

The distinction between LEXICAL and FUNCTIONAL linguistic units has played a role in the theory and analysis of language for millennia. This is to say, a distinction can be drawn between two poles for the role that linguistic units play in communication. On one end, we have the LEXICAL: linguistic signs which carry specific meanings, often referring out to objects or events in the world. On the other end, we have the FUNCTIONAL: signs which do not so much carry specific meanings, but rather serve to organize and clarify the relationships between lexical elements.

Because the roles served by FUNCTIONAL units are similar to roles that, in some languages, are expressed not through independent units but through structural patterns or rules (that is, through *grammar*), such units are sometimes referred to as GRAMMATICAL units. For example, in English, whether a word serves as the subject or object of a verb is indicated through word order alone, while in some languages, there are linguistic signs (called, variably *case markers*, *adpositions*, or *flags*) which explicitly mark these relationships. As such, for theories which treat grammar as a separate system from the lexicon, functional

units present a key challenge, as they straddle the boundary between these two systems, having the realized form of a linguistic sign, but the organizational function of grammar.

Yet to say the characteristic role of functional units is clarifying the relationship between lexical elements is an oversimplification. Many types of functional elements rather clarify or elaborate on the meanings of individual lexical elements. Adding to the puzzle, however, is a surprisingly high degree of cross-linguistic consistency in the roles of functional elements. For example, while many languages mark number, definiteness, or tense through functional elements, it has been claimed that no languages mark *color* or *shape* through functional elements (?). Linguists have handled this observation of cross-linguistic consistency in different ways. Many generative linguists have hypothesized a rich, innate, universal set of functional heads which are realized in different ways across languages (??). In contrast, cognitive linguists have attempted to find common cognitive motivations for the cross-linguistic consistency of functional elements.

However, while there is some consistency in what concepts can be expressed as obligatory, paradigmatic, bound markings across languages, there are also serious definitional issues around where the boundary between lexical and functional units lies. To claim that “only certain concepts can be expressed functionally” presupposes a consistent definition of what it means to be functional; for this to avoid circularity, the definition of functional must not rely on the concepts themselves, but rather on something about linguistic *distribution*.

Yet the formal and distributional properties of functional expression are far from clear-cut. Typically, these are identified as (I) being CLOSED-CLASS, (II) being BOUND, and (III) being OBLIGATORY. While functional categories are typically closed-class (resisting new members), prototypically lexical categories can also be closed class, like Bemba adjectives (Dixon, 1977) or Jaminjung verbs (Pawley,

2006). Further, closedness is not a binary property, with closed classes varying substantially both in their size and their resistance to admitting new members. Purported functional elements can vary significantly in their degree of boundness, and indeed “boundness” is a complex property of dubious categorical status (??), with no general consensus on how to define or measure it. Finally, obligatoriness is also gradient, with some functional elements being optional in some or even many contexts.

The problem grows only more complex when we consider that the lexical status of a linguistic unit can vary contextually and diachronically. For example, in Mandarin Chinese, the intensifier *hěn* (“very”) is undergoing semantic bleaching into a copular function for adjectives (?, pp. 508–509):

- (1.1) *tā hěn gāo*
 3SG very tall
 “She is tall.”

Here, *hěn* has lost much of its lexical meaning of “very” and serves primarily to link the subject to the adjective. Grammaticalization processes like this are universal cross-linguistically, and tend to produce boundary cases between lexical and functional units. Examples of such boundary cases include: “light nouns” like quantificational nouns (e.g. *lot*) or taxonomic nouns (e.g. *type*); “light verbs” like *have* in *have a rest*; and adpositions (??). Typically, the loss of semantic force in these cases is accompanied by changes in formal and distributional properties (decategorization).

While typical discussions of the distinction between lexical and functional units tend to focus on a contrast between lexical words or roots and functional words or affixes, a closely related distinction is drawn *within* the domain of morphology between DERIVATION and INFLECTION. Morphology described as *inflectional* typically have all the prototypical properties of functional units: inflection is typically closed-class, bound, and obligatory, and align closely with

“possible grammatical concepts.” In contrast, *derivational* morphology may express rich and perhaps unconstrained meaning, interacting ideosyncratically with the meaning of roots, and is typically optional. However, there are a few key differences with the lexical–functional distinction at the level of words. First, there are strong cross-linguistic tendencies for derivations to occur closer to the root than inflections do (Greenberg’s Universal 28) (Greenberg, 1966a; ?). Second, simple conversions or transpositions of word class (e.g. converting an adjective to a noun: *happy*→*happiness*) tend to pattern more similarly to derivations than inflection (e.g. scoping inside inflections), despite their apparent lack of semantic weight and highly obligatory and productive nature. This has led to substantial debate over whether such morphological constructions are better considered inflection (?) or derivation (ten Hacken, 1994). Further, where exactly the boundary between “transpositions of word class” and more semantically rich derivations lies is also unclear. For example, the *-er* nominalizer in English forms agentive nouns from verbs (*teach*→*teacher*). Again, we have encountered a distinction between two poles that seem to be related to semantic weight, but with unclear boundaries. While derivational morphology tends to be more semantically rich than inflectional morphology, it is typically less semantically rich than lexical roots; however, in highly agglutinative languages like Inuktitut, derivational morphemes can carry semantic content comparable to roots in other languages.

This thesis concerns itself with these divisions between more lexical and more functional linguistic units, at multiple levels of linguistic structure. I treat both the lexical–functional distinction at the level of words and the inflection–derivation distinction at the level of morphology, and connect them to prototype phenomena among the major lexical classes of nouns, verbs, and adjectives. *In this thesis, I refer to this gradient of semantic weight at different levels of linguistic structure as the spectrum of LEXICALITY.*

Evidence for the importance of lexicality to human language processing comes not just from the structure of language, but from the psychological and neurological study of language processing itself. Differential psychological processing for inflection and derivation has been observed by Laudanna et al. (1992) and Kirkici and Clahsen (2013), and while neurological evidence is somewhat mixed, it does point to differential processing between inflection and derivation (?). At the level of the traditional lexical–functional distinction, differences have been documented in access speed (?), brain region activation (?), and lateralization (?). Famously, there is also evidence for the individuation of the lexical and functional from aphasia: on the one hand, we have agrammatic aphasia¹, where lexical access seems intact, but the use of grammatical words and complex structures is impaired; on the other, there are fluent aphasia like Wernicke’s aphasia, where speakers produce fluent sentences with rich syntactic structure, but fail to produce sentences that convey content clearly, struggling to access relevant contentful words. These evidence bases have even shed some light on boundary cases, with aphasic evidence supporting the “grammatical” status of light verbs (in Dutch; ?), but aphasia and neurological data providing mixed evidence about lexical vs. functional adpositions (in English and German; ???). Aphasic evidence also supports differential processing among lexical classes (specifically between nouns and verbs; ??Bird et al., 2003).

Despite this rich base of evidence for differential representation and processing across the lexicality spectrum, the direct study of semantic weight/force/contentfulness on linguistic structure remains largely pre-theoretical in linguistics, especially in large-scale cross-linguistic study. A major cause of this theoretical lacuna is the difficulty of specifying *semantic contentfulness* in a principled, cross-linguistically applicable way. This has led many linguists to avoid this notion entirely, focusing instead on how notions like frequency shape grammatical

¹This is sometimes known as Broca’s aphasia.

expression (?). In this thesis, I seek to address this gap by developing measures of the *semantic* dimensions of lexicality distinctions and investigating *how they relate* to traditional “problematic” cross-linguistic grammatical distinctions.

1.2 Approach

Multiple Levels of Linguistic Structure This thesis spans a *wide range* of levels of linguistic structure. While previous work has largely treated these distinctions at different formal and semantic levels as different, unrelated problems (e.g. the inflection–derivation distinction at the morphological level; the distinction between lexical and functional word classes at the word level; or the distinctions between the major lexical classes of nouns, verbs, and adjectives), I investigate lexicality across multiple levels of linguistic structure, showing new parallels and connections between them. That being said, I limit my focus to *sub-phrasal* linguistic units (morphemes and words), leaving phrases, semantic frames, and more schematic constructions to future work.

Cross-linguistic investigation This thesis focuses on the *cross-linguistic* consistency of lexicality distinctions.² A large body of work in linguistic typology has argued that language-specific categories do not map onto some clean set of universal grammatical categories (Haspelmath, 2007; Croft, 2001; Dixon, 1977). Instead, typologists have argued for the importance of cross-linguistically valid *comparative concepts*—which need not necessarily map onto the structure of individual language’s grammar (Haspelmath, 2010; Croft, 2016). Studies that focus on the distinction between inflection and derivation or between lexical and functional word classes which consider only a single language risk conflating language-particular properties and categories with cross-linguistic generaliza-

²While in Chapter 6, I do conduct a language-particular analysis of Japanese word classes, the motivation for this analysis is to investigate whether cross-linguistically validated measures of lexicality can explain unusual language-particular patterns.

tions. While finding a consistent distinction between inflection and derivation, or between lexical and functional word classes in an individual language is interesting, it does *not* a-priori tell us whether such a distinction has cross-linguistic descriptive value. Thus, this thesis aims to cover a large and diverse sample of languages wherever possible.

Computational and quantitative methods To study the cross-linguistic consistency of lexicality distinctions, I take inspiration from the successes of empirical grounding in certain areas of typology (like vowel, color, and kinship systems; ?; ?; ?) and from recent advances in deep learning models of language. These models have been shown to learn rich representations of semantics and the world, without requiring direct instruction on this structure, but rather learning it implicitly from learning to predict words in a (linguistic and/or visual) context. This capacity makes these tools ideal for operationalizing semantic dimensions of lexicality distinctions. Further, in the second part of the thesis I leverage *multimodal* models which ground languages in images. This image grounding provides a language- and form-neutral representation of semantics, which enables a new method for separating out contextual contentfulness from formal linguistic predictability. The computational approach also aids in our goal of cross-linguistic investigation—while psychological and neurological evidence for lexicality distinctions exists, scaling it to typological study is challenging. Computational methods require only corpus data, which enables the simultaneous study of many languages, though it biases the study towards languages with sufficient digital resources. Further, the quantitative focus of this thesis enables the study and quantification of consistency, in contrast with many previous studies that focus primarily on problematic cases.

1.3 Structure of the Thesis

I investigate three key research questions in this thesis:

RQ1: What is the interplay between *form* and *semantics* across the lexicality spectrum? (Chapters 4, 5, 6)

RQ2: How can we operationalize semantic contentfulness in a cross-linguistically applicable way? (Chapters 3, 5)

RQ3: How *cross-linguistically consistent* are lexicality-related divisions like the division between the lexical and functional word classes, or the inflection–derivation distinction? (Chapters 4, 5)

Chapter 2: Background In this chapter, I expand on the theoretical framework for the thesis. I introduce in more detail the problems of cross-linguistic category comparison, and the method of comparative concepts for resolving these issues in typological research. I review the ways in which semantic function has been handled in typological comparative concepts, highlighting a role for deep learning models of language in the creation of functional³ comparative concepts. I then provide an overview history of the ways comparative concepts have been (at times, implicitly) employed, handled, and studied in computational typology. Through this background, I highlight how the creation of empirically grounded comparative concepts through the development of technologies and measures outside of typology (like perceptual theories of vowels and colors) has enabled major advances in typological research in the domains where it has been possible. This serves as further motivation for the computational approach to comparative concepts taken in this thesis. Finally, I provide a broad-scale overview of the lexicality spectrum, identifying different manifestations of a correlation between semantic contentfulness and formal linguistic structure,

³In this sense, semantic

providing the connective tissue for the studies that follow. More detailed background on specific lexicality-related distinctions is provided in the relevant chapters.

Part I: Inflection and Derivation

Part I of this thesis consists of Chapters 3–4, which focus on the inflection–derivation distinction drawn in morphology. These chapters are based primarily on the following journal article:

Haley, C., Ponti, E. M., and Goldwater, S. (2024). Corpus-based measures discriminate inflection and derivation cross-linguistically. *Journal of Language Modelling*, 12(2):477–529

Chapter 3: Corpus-based Measures for Inflection and Derivation In this chapter, we introduce a computational framework for the inflection–derivation distinction. Inspired by Spencer (2013)’s description of the distinction, we introduce a set of four quantitative measures of morphological constructions, including measures of both the magnitude and the variability of the changes to *form* and *distribution* introduced by each construction. Crucially, these measures can be computed directly from a linguistic corpus, allowing us to consistently operationalise them across many languages and morphological constructions. In contrast to prior computational studies that focus on a single language, we investigate 26 languages using the UniMorph 4.0 corpus (Batsuren et al., 2022). Using these measures, we find differences between inflection and derivation for all four measures, but substantial overlap for each individual measure. We demonstrate that the measures are not explained by simple frequency effects, and that the distributional measures capture a limited amount of syntactic information in addition to semantic information.

Chapter 4: Predicting Inflection and Derivation Using the measures from Chapter 3, we train classifier models to predict whether a construction is labeled as inflection or derivation in UniMorph. We find that language-agnostic classifier models over our measures are able to predict inflectional–derivational status with high accuracy (90%). We investigate linguistic categories of inflection, finding inflectional transpositions like participles are *not* more likely to be misclassified as derivational, in line with ?’s argument that these are best considered inflectional. Overall, our results are in line with a *consistent, yet gradient* view of the inflection–derivation distinction. Our results suggest that distributional and formal *variability* are the most important dimensions for the distinction, but the magnitude of distributional and formal change also play a role. While there is substantial overlap between the two categories on each individual dimension, the combination of all four dimensions provides a robust signal for the distinction in our sample.

Tricky knot to tie—magnitude seems less important, which is kind of at odds with the intro.

Part II: Word Classes

Part II of this thesis consists of Chapters 5–6, which focus on lexicality among (functional *and* lexical) word classes.

Chapter 5: Groundedness and the Lexical-Functional Distinction In this chapter, I introduce *groundedness*, a new semantic-contentfulness measure based on multimodal models. Focusing on the domain of image captions, I am able to treat an image as a proxy for a caption’s meaning. Using a language model and an image captioning model, I am able to estimate the pointwise mutual information between a token and the image as a surprisal difference under the two models. In this chapter, I focus on the **lexical-functional distinction** in

parts of speech.

Using image captioning data in 30 languages from 10 language families, I find this groundedness measure largely rediscovers the distinction between lexical and functional word classes across 30 languages. Further, though it correlates only weakly with norms like imageability and concreteness in English, it provides a ranking suggested by cognitive linguists between nouns, verbs, and adjectives (noun > adjectives > verbs) across languages but contradicts the view of adpositions as a “semi-lexical” class. However, our results suggest grammatical word classes still carry semantic content. These results suggest the utility of this measure as a general tool for studying contentfulness in linguistics, and of taking a visually grounded approach to typological problems. This chapter is based on a conference paper at which appeared at NAACL 2025:

Haley, C., Goldwater, S., and Ponti, E. M. (2025). A Grounded Typology of Word Classes. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 10380–10399, Albuquerque, New Mexico. Association for Computational Linguistics.

Chapter 6: Splitting and Lumping In this chapter, I investigate the relationship between visual groundedness and cross-linguistic variation in **lexical parts of speech**. While there has been substantial work in linguistic typology investigating cross-linguistic variation in the expression of major lexical categories like nouns, verbs, and adjectives, this work has previously been largely disconnected from work on semantic contentfulness, the lexical–functional distinction, and grammaticalisation. Building on the visual groundedness measure introduced in Chapter 5, I connect existing continuum and prototype theories of lexical categories and meanings with groundedness. I argue that the role of semantic contentfulness in lexical categories can help explain cross-linguistic variation in lexical category organisation. In particular, I focus on languages which have been argued to “split” or “lump” major lexical categories.

To establish this, I first investigate Japanese. In Japanese, words denoting “properties” have the unusual property of constituting two formally very distinct word classes, rather than a single “adjective” class. Building on the insight that one of these classes is more formally “nominal” (*na*-adjectives) and one more “verbal” (*i*-adjectives), I hypothesise that we should see analogous trends in function: one class serving more prototypically nominal functions and one more prototypically verbal. In terms of visual groundedness, this corresponds to higher values for the nominal class. I investigate two manually captioned datasets and one machine translated dataset, finding significantly higher groundedness for *na*-adjectives in the manually captioned datasets, in line with the theoretical predictions. This stands in contrast to previous studies, which have indicated little synchronic functional difference between the two classes.

To investigate lumping phenomena, I turned to the Tensedness Correlation, which correlates the formal similarities of adjectives to verbs in languages with a lack of obligatory tense marking on verbs. In languages with obligatory tense marking, the expression of adjectives is more similar to nouns. I investigate whether this correlation is reflected in groundedness, drawing on previous hypotheses for the cause of the Correlation. I find no significant relationship, which I argue is due to issues with directly comparing groundedness scores across languages, suggesting the need for careful study design for groundedness-based research, and the difficulty of grounding verbs in particular.

Chapter 7: Conclusion In this chapter, I summarise the contributions of this thesis, discuss limitations, and outline directions for future work.

1.4 Contributions

Is this needed?

- groundedness
- consistency of lexicality-related distinctions
- connections across levels of linguistic structure
- new computational methods and approaches to typological research