Doctoral thesis in computational linguistics

Semantic change in interaction

Studies on the dynamics of lexical meaning

Bill Noble

April, 2023

UNIVERSITY OF GOTHENBURG

Semantic change in interaction:
Studies on the dynamics of lexical meaning
Supervisors: Staffan Larsson and Asad Sayeed

The author is supported by grant 2014-39 from the Swedish Research Council (VR) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg

Cover by Noah Mease
Printed in Sweden by Stema Specialtryck AB
Publisher: University of Gothenburg (Dissertations)

Distribution:
Department of Philosophy, Linguistics and Theory of Science,
University of Gothenburg
Box 100, SE-405 30 Gothenburg

ISBN:
978-91-8069-205-2 (print)
978-91-8069-206-9 (PDF)

Part I is available online at
http://hdl.handle.net/2077/74969.

for my parents

Abstract

This compilation thesis investigates how word meanings change. In particular, it's concerned semantic change at the levels of *interaction* and the *speech community*. To this end, the compiled studies employ methods from both formal and computational semantics.

The first study presents a model for, and companion annotation study of, *word meaning negotiation*, a conversational routine in which the meaning of a word becomes an explicit topic of conversation. The next two studies introduce and apply *classification systems*, a model of communal conceptual resources for ordering and talking about a particular domain. We use a formalization thereof to model how *genus-differentia definitions* can be used in interaction to update lexical knowledge of perceptual categories. The next study considers a related phenomenon, *perceptual category description*, but this time from a computational perspective. By modeling a short interaction between two neural networks, we investigate how different ways of representing perceptual categories affect linguistic grounding. Following that, we turn to the dynamics of social meaning, particularly the meaning of implicit conversational assumptions called *topoi*, with a focus on situations of involving uncertainty about the speaker's social identity. The final two studies of the thesis shift the focus from particular interactions to the level of the community. First, we investigate linguistic variation using *community conditioned language models* to learn vector representations for a collection of online communities. These language-based representations are found to correlate with community representations based on community membership alone. Finally, we use diachronic distributional word vectors to study *short-term semantic shift* in online communities. We find that semantic change has a significant yet nuanced relationship with the social structure of the community.

Altogether, the compilation offers two main insights. First, semantic plasticity is directly related to the complexity of the lexical semantic system. Words exhibit both perceptual and inferential meaning potential, each of which play a role in conveying and learning new meanings. Monolithic representations of word meaning belie a structured flexibility that guides how words can be used, while providing opportunities for innovation. It is this flexibility that is often the site of new conventionalized meanings. Second, semantic change is rooted in the interactive practices of the community. Communities sustain the communicative norms that govern how linguistic interaction takes place. These norms also provide a framework for negotiating meaning, and comprise the social and semiotic context that supports semantic innovation and change.

Sammanfattning

Denna sammanläggningsavhandling undersöker hur ordbetydelser förändras. Mer specifikt handlar den om semantisk förändring på *interaktionsnivå* och på *språkgemenskapsnivå*. I detta syfte använder de i avhandlingen ingående studierna metoder från formell och komputationell semantik.

Den första studien presenterar en modell för, och en tillhörande annoteringsstudie av, *ordbetydelseförhandling*, ett samtalsmönster där ett ords betydelse blir det explicita samtalsämnet. De följande två studierna introducerar och tillämpar *klassificeringssystem*, en modell av gemensamma begreppsliga resurser som används för att organisera och tala om en viss domän. Vi använder en formalisering av klassificeringssystem för att modellera hur *genus-differentiae-definitioner* kan användas i interaktion för att uppdatera lexikal kunskap om perceptuella kategorier. Nästa studie behandlar ett besläktat fenomen, *perceptuella kategoribeskrivningar*, men denna gång från ett komputationellt perspektiv. Genom att modellera en kort interaktion mellan två neurala nätverk undersöker vi hur olika sätt att representera perceptuella kategorier påverkar språkligt delande av information. Därefter vänder vi oss till den sociala betydelsekomponentens dynamik och då särskilt med avseende på betydelsen hos underförstådda antaganden, så kallade *topoi*, och med fokus på situationer där det finns en osäkerhet om talarens sociala identitet. De två sista studierna i denna avhandling skiftar fokus från specifika interaktioner till språkgemenenskapsnivån. Först undersöker vi språklig variation med hjälp av *gemenskapsvillkorade språkmodeller* som lär sig vektorrepresentationer för grupper av onlinegemenskaper. Dessa språkbaserade representationer visar sig korrelera med gemenskapsrepresentationer som enbart grundas i gemenskapstillhörighet. Slutligen använder vi oss av diakroniska distributionella ordvektorer för att studera *kortsiktig semantisk förändring* i onlinegemenskaper. Vi finner att semantisk förändring har signifikanta men nyanserade samband med gemenskapens sociala struktur.

Sammantaget ger sammanställningsavhandlingen två huvudsakliga insikter. För det första är semantisk plasticitet direkt kopplad till det lexikala semantiska systemets komplexitet. Ord har både perceptuella och inferentiella betydelsepotentialer, och båda dessa aspekter spelar en roll i överförandet och inlärandet av nya betydelser. Monolitiska representationer av ordbetydelse bortser från en strukturerad flexibilitet som vägleder hur ord kan användas samtidigt som de erbjuder möjligheter till språklig innovation och förändring. Det är denna flexibilitet som ofta ligger till grund för uppkomsten av nya konventionaliserade betydelser. För det andra är semantisk förändring rotad i språkgemenskapens interaktiva praktiker. Språkliga gemenskaper upprätthåller de kommunikativa normer som styr den språklig interaktionen. Normerna erbjuder också ett ramverk för förhandladet av betydelser, och utgör den sociala och semiotiska kontext som möjliggör semantisk innovation och förändring.

Contents

Acknowledgements

It has been a long journey. My language and my world have changed a lot along the way. And there are so many people responsible for making that journey and this thesis possible.

First, I would like to acknowledge the substantial contributions of my PhD advisors, Staffan Larsson and Asad Sayeed. I spent so many hours talking through the ideas of this thesis with Staffan. How many times did I obstinately take another route, only to come back to his initial suggestion? Certainly more than once. Staffan also diligently helped me to push the thesis to completion and deserves much of the credit for making sure that the defense will, in fact, happen. Asad has had a major influence, not only on the thesis, but also on how I have come to view academic life and the project of research. Staffan and Asad were instrumental in helping to figure out how to frame the compilation and gave very important comments on multiple drafts of the kappa.

In addition to my advisors, I was fortunate to have many other co-authors on papers in the compilation. Each and every one of them have been a joy to work with and the compilation is at least as much theirs, collectively, as it is mine. Robin Cooper has, been a generous and careful mentor. I greatly admire the way he deftly draws out and nurtures the ideas of others. It was Raquel Fernández who first introduced me to computational linguistics research, and it was such a thrill to collaborate with her again, along with Staffan and Asad. I have had countless conversations with Ellen Breitholtz about social and pragmatic meaning in interaction. Her influence on the direction of the thesis goes well beyond our collaboration on Chapter 11. Jean-Philippe Bernardy spent many hours talking through algebraic and probabilistic modeling with me. Working with him has had a substantial impact on how I think about computation and formalization in linguistics. Kate Viloria's contribution to Chapter 7 was made as part of an independent study course for the Master's of Language Technology, in which she was an excellent student and collaborator. A special thanks goes to Nikolai Ilinykh for working with me to prepare a new and improved version of Chapter 10 in time for inclusion in the thesis, much to the chagrin of the poor people at Victor's Cafe, who were forced to, yet again, listen to us go on and on about birds.

I would also like to thank my other research collaborators at GU and beyond: Vladislav Maraev, Adam Ek, Julain Grove, Ben Clarke, Fahima Ayub Khan, Eleni Gregoromichaelaki, Christine Howes, Chiara Mazzocconi, Simon Dobnik, and Vidya Somashekarappa. I have been fortunate to work with so many people on a wide range of topics in computational linguistics. These collaborations have been hugely influential in shaping what would ultimately become the thesis, even when our work was outside the narrow

scope of the compilation.

Andy Lücking was the "green reader" for the thesis. His insightful comments were instrumental in shaping the final version of the kappa and will undoubtedly influence any future work I do on this topic. Nikolai Ilinykh and Noah Mease also gave comments on drafts of the kappa that helped a lot to improve its readability and cohesiveness.

I want to thank the members of CLASP, especially Sharid Loáiciga, Stergios Chatzikyriakidis, and Shalom Lappin for fostering a rich, nurturing research environment. I'd also like to acknowledge dialogue reading group, which has served as my academic "home base" during my PhD. Many of the ideas developed in the thesis were first incubated there.

The administration at the Department of Philosophy, Linguistics and Theory of Science (FLoV) supports our research in ways that I only begin to understand. I would especially like to thank Susanna Myyry, Hannna Edblom, and Iines Turunen. Christopher Kullenberg and Johan Söderberg also helped me a great deal with administrative questions leading up to the defense.

My fellow PhD students in FLoV have made our department a welcoming place that I was excited to go to each day. Thank you for your companionship.

Thank you to the Wannerskog family Maria, Lasse, Anna-Sofia, Lars, Miranda, Svante, Olaf, Kalle, Åsa, Axel, and Johanna. From the day I first landed in Göteborg you have made me feel like I belong. I can't begin to express my gratitude. Tack så jätte mycket.

Thank you to my friends and family back home for always being there for me. For countless hours on the phone, for board games over the internet, for visiting me from all the way across the Atlantic Ocean, and for keeping a place for me to come home to. Mom and Dad, Ryan, Brendan, Jack, Claire, Todd, Khanh-Anh, Dustin, and Ray, thank you.

And, of course, to Noah, thank you. How could I be so lucky?

Preface

This is a compilation thesis, meaning that the main scientific contributions come mostly in the form of studies that have previously been published as conference papers.[1] All these papers are, in one way or another, computational studies of variation and change in natural language. Those papers are reproduced in Part II of the print version of this thesis. Summaries of the studies, with links to the archival version of the papers can be found in Section 6.1.

Part I of the thesis, is what is colloquially referred to as the *kappa*, (English: *coat* or *cover* neé *hat*). Chapters 1 to 4 set up the theoretical framework that underpins the work in Part II of the thesis and Chapter 5 discusses the methodologies that are used. Finally, Chapter 6 (in addition to the summaries) offers some concluding remarks.

Naturally, some background and methodological exposition can be found in the introductory sections of the individual papers, but it is brought together a way that motivates the overall research outlook of the thesis. I hope, too, that the kappa makes the work available to a broader audience. Conference papers tend to be written with attendees (and especially reviewers) of the specific conference mind. Given the space constraints of a typical conference paper, this often means that a lot of theoretical background is assumed or introduced in a more cursory way than it would be for a more general audienceeven perhaps an audience familiar with the field of computational linguistics more broadly.

Computational linguistics is a highly collaborative field. All of the work in this compilation was all carried out in close collaboration with my PhD supervisors and other members of the computational linguistics community in Gothenburg, particularly at the Centre for Linguistics and Studies in Probability (CLASP) where I have been fortunate to be employed as a PhD student. I use the word *we* a lot in the thesis. Often times that's because the work being described was very much a joint effort. In other places, I'm just hoping to include you, the reader, in this adventure we're about to embark on.

Part II is not included in the PDF version of the thesis. However, the papers that comprise the chapters of Part II are all freely available online in their original format. Throughout Part I, the studies included in the compilation are referred to by their chapter numbers in Part II. Both the original citation for each of the chapters and links to the online versions can be found below.

[1] With the exception of Chapter 10, which has been published as a pre-print on ArXiv.

Chapter 7 Noble, B., Viloria, K., Larsson, S., & Sayeed, A. (2021). What do you mean by negotiation? Annotating social media discussions about word meaning. *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*

http://semdial.org/anthology/papers/Z/Z21/Z21-3016/

Chapter 8 Noble, B., Larsson, S., & Cooper, R. (2022a). Classification Systems: Combining taxonomical and perceptual lexical meaning. *Proceedings of the 3rd Natural Logic Meets Machine Learning Workshop (NALOMA III)*, 11–16

https://aclanthology.org/2022.naloma-1.2

Chapter 9 Noble, B., Larsson, S., & Cooper, R. (2022b). Coordinating taxonomical and observational meaning: The case of genus-differentia definitions. *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*

http://semdial.org/anthology/papers/Z/Z22/Z22-3020/

Chapter 10 Noble, B., & Ilinykh, N. (2023). Describe me an Aucklet: Generating Grounded Perceptual Category Descriptions. https://doi.org/10.48550/arXiv.2303.04053

https://arxiv.org/abs/2303.04053

Chapter 11 Noble, B., Breitholtz, E., & Cooper, R. (2020). Personae under uncertainty: The case of topoi. *Proceedings of the Probability and Meaning Conference (PaM 2020)*, 8–16

https://aclanthology.org/2020.pam-1.2/

Chapter 12 Noble, B., & Bernardy, J.-P. (2022). Conditional Language Models for Community-Level Linguistic Variation. *Proceedings of the 5th Workshop on NLP+CSS at EMNLP 2022*, 59–78

https://aclanthology.org/2022.nlpcss-1.9/

Chapter 13 Noble, B., Sayeed, A., Fernández, R., & Larsson, S. (2021). Semantic shift in social networks. *Proceedings of *SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics*, 26–37. https://doi.org/10.18653/v1/2021.starsem-1.3

https://aclanthology.org/2021.starsem-1.3

Part I.

Kappa

1. Introduction

> [...] the whole theory of language can be reduced to one question: what is the relationship between prevailing usage and the speech of an individual? How is the speech of an individual determined by prevailing usage in the community, and how in turn does the individual's speech affect prevailing usage?
>
> Herman Paul (1886)
> trans. Peter Auer (2015)

We know that words change. In the early 20th century the word *gay* meant *happy* or *joyous* in English. Now it almost always refers to sexuality. *Awesome* used to mean something awe-inspiring, frightening, even. Now it can also be used to mean *really very good*. Historical changes like these are well-documented, occurring in every language and at every time in history. We also know that speakers can coordinate a special vocabulary of word-meaning parings for collaborating on a project (Brennan & Clark, 1996), or even as a way of building and expressing intimacy (Hopper et al., 1981). But what do these two kinds of semantic plasticity have to do with one another? Is there a connection between these changes that take place on a historic time-scale and across whole languages and the *ad hoc* conventions that we develop for a particular communicative context? It would seem that there must be. After all, what makes a language if not the all of its individual occasions of use? And yet it is difficult to observe the transition from one to the other. **How does semantic coordination at the level of interaction relate to lexical change on the community level?**

On the other hand, when we consider variation across communities, we have something of a paradox. One view of language is that it is a channel for communication — a *code* in which one person can transmit information about the world to another person. Such a code functions best, one might assume, if its symbols and their meanings are perfectly aligned between the two speakers. To maximize efficiency the code should be stable be shared among as many speakers as possible. And yet this is not the situation in which we find ourselves. We have *many* different languages — not only among what are sometimes called the *macrolanguages* of the world. We also see a great deal of variation in the way that different communities communicate *within* these traditionally defined languages and dialects. There is a special *lingo*, *slang*, *jargon*, etc. associated with just about every vaguely communal activity you can think of. **Why does word meaning change across time and context?**

Unfortunately, this thesis will not answer either of these grand questions directly.

But they will nevertheless serve as two guiding stars as we set forth. The questions we do address in this thesis are small steps towards a better understanding of the dynamics of lexical meaning.

We start on the level of interaction, specifically interactions involving explicit talk about meaning to investigate the following questions:

1. What interactive resources are drawn upon when word meaning becomes an explicit topic of conversation? (Chapter 7)

2. When someone defines a word for us, how do we incorporate that meaning into our existing conceptual structure? (Chapters 8 and 9)

3. How does the way that perceptual categories are represented affect descriptions of those categories? (Chapter 10)

Continuing on theme of the dynamics of interaction, we consider the plasticity of certain *social signals*:

4. How does the meaning of a social signal change depending on what we learn about a speaker's social identity or ideology? (Chapter 11)

Finally, we shift our focus to the community level, addressing questions about how the character of communities relates to linguistic variation and change:

5. How do the linguistic particularities of a community correlate with its social makeup? (Chapter 12)

6. How does the social structure of a community affect the rate at which its word meanings change? (Chapter 13)

Along the way, we will come face to face with some of the most challenging questions in lexical semantics. Why is word meaning so flexible? How can it be that words have multiple senses? How do we synthesize seemingly incompatible aspects of word meaning? Does it matter to linguistics how words are associated with their meanings on a cognitive level? In Chapter 2, we set the stage to deal with these issues as they arrive. Chapter 3 situates lexical meaning in its natural habitat — in communities and in interaction. Finally, Chapter 4 gives some background on semantic variation and change, which is necessary to contextualize the contributions of the thesis.

The process of *lexicalization* — when new meanings become conventional — is difficult to observe directly. Because of this, the studies in Part II employ a variety of methods, from formal semantics to machine learning to investigate (1) interactions that *might* result in a speaker updating their understanding of the language of the community and (2) analysis of variation and change in small communities and across short

time periods. It's important to keep in mind how all these methods can be used to answer questions — what their limitations are and how insights gleaned using a certain methodology should be synthesized with the rest of the work. Chapter 5 introduces the methods used in the compilation.

Finally, Chapter 6 summarizes the studies presented in Part II and offers a synthesis of the conclusions that emphasizes the importance of lexical complexity and community-level interactive practice in understanding semantic change.

2. Lexical meaning

Why is a raven like a writing-desk?

Alice's Adventures in Wonderland
Lewis Carroll

In order to understand how words change in meaning, it's important to have a bit of background on what the field of linguistics understands meaning to *be*. In contemporary linguistics, *semantics* (the study of meaning) encompasses two fairly distinct sub-fields: **lexical semantics**, which studies the meaning of words (or, as we will discuss shortly, *lexical items*), and **compositional semantics**, which studies the meaning of larger linguistic units formed according to the language's *syntax*. Although we can never stray far from questions of compositional meaning, this thesis is primarily concerned with the dynamics of lexical meaning.

Lexical meaning implies a *lexicon*, i.e., a *book of words*, which suggests a certain model of natural language semantics in which the compositional and lexical aspects of a language are distinct modules. That model is reflected in the disciplinary division mentioned above, and is implicit in the **principle of compositionality**, which says that the meaning of a natural language expression is a *function* of *the meaning of its parts* (which come from the lexicon) and *how they are combined* (which comes from the syntax). This means that two expressions that use the same words can have different meanings:

(1) a. The man points at a dog.

b. The dog points at a man.

In English, this particular grammatical construction assigns different semantic roles (usually called *agent* and *patient*) to the two syntactic positions (*subject* and *object*). The meaning of (1a) differs from that of (1b) as a result of the compositional semantics, not because of the meaning of the words involved. But, of course, the meaning of the words does also matter:

(2) The man points at an apple.

To the extent that (2) differs from (1a) the principle of compositionality says that this difference must be explained by differences in the meaning of the words *dog* and *apple*.

The division of semantic labor implied by the principle of compositionality points to a *dictionary and grammar book* view of linguistic competency (Taylor, 2012), which says that with these two (more or less distinct) sources of knowledge, we have everything that is needed to understand and produce meaningful expressions in a given language. When studying variation and change of word meaning, it's helpful to consider how this "dictionary", which linguists call the **lexicon**, is structured. As a first attempt, we might consider the structure of a literal dictionary. Consider this entry for the word *point*:[1]

Point

noun (*plural*: **points**)

1. A discrete division of something.
    a) An individual element of a larger whole; a particular detail, thought, or quality. *The Congress debated the finer **points** of the bill.*
    b) A particular moment in an event or occurrence; a juncture. *There comes a **point** in a marathon when some people give up.*
    c) A focus of conversation or consideration; the main idea. *The **point** is that we should stay together, whatever happens.*
    d) A purpose or objective, which makes something meaningful. *Since the decision has already been made, I see little **point** in further discussion.*
2. A sharp extremity.
    a) The sharp tip of an object. *Cut the skin with the **point** of the knife.*
    b) An object which has a sharp or tapering tip. *His cowboy belt was studded with **points**.*
    c) A peninsula or promontory.
    d) [*falconry*] The perpendicular rising of a hawk over the place where its prey has gone into cover.

verb (*third-person singular simple present*: **points**:, *present participle*: **pointing**, *simple past and past participle*: **pointed**)

3. (*intransitive*) To extend the index finger in the direction of something in order to show where it is or draw attention to it. *It's rude to **point** at other people.*
4. (*intransitive*) To draw attention to something or indicate a direction *The arrow of a compass **points** north. The skis were **pointing** uphill.*
5. (*transitive, sometimes figuratively*) To direct towards an object; to aim *to **point** a gun at a wolf, or a cannon at a fort*
6. [*nautical*] (*intransitive*) To sail close to the wind *Bear off a little, we're **pointing**.*

The remainder of this chapter will use the dictionary model of the lexicon as a starting point for introducing lexical semantics topics that relate to variation and change. In Section 2.1, organizational structure of the dictionary which is, at its heart, a list of words where each item has its own distinct entry. Section 2.2 will talk about the

[1] This example is abridged and edited from the English-language Wiktionary entry for *point*, https://en.wiktionary.org/w/index.php?title=point, accessed December 1, 2022.

structure of the entry, which is itself a list of *senses*. Since words by their very nature exhibit variation in their contexts of use, how do we decide when, if at all, to make a distinction between different senses of the same word? When we do distinguish between senses, how do they relate to each other? This brings us to the question of semantic *generality*, which is discussed in Section 2.3. A word like *point* can apply to broad range of different real-world situations. How do we know what is included in its extension? Finally, Section 2.4 discusses two different *kinds of meaning* that arise from expectations we have of competent speakers.

2.1. Lexicality

Linguists disagree about the degree to which lexical and grammatical information can be considered distinct. But there does seem to be something to the idea of a flexible store of discrete lexical information that can slot in to a relatively stable compositional semantics.

To borrow an example from Larsson (2021), suppose you've never heard of a *wax jambu* (perhaps you haven't). And suppose I say to you *a wax jambu is pear-shaped fruit with a texture similar to that of an apple.* Your knowledge of wax jambus is still pretty incomplete, but you probably already have a pretty good idea of how to use the word in a sentence. Your knowledge of English morphology lets you construct and recognize the (admittedly awkward) plural *wax jambus*. You can even understand sentences like this one:

(3) The man points at a wax jambu.

You may not be able to call up a perfectly vivid image of a situation described by (3), but you can be assured that the difference in meaning between (3) and (2) is a function of the difference between apples and wax jambus, not because of how the sentence is formed or because of something about the meaning of the word *point*.

Lexicality is the dual of the principle of compositionality. Together, lexicality and the principle of compositionality give us a situation where lexical knowledge of specific words can change over time or vary across communities while the rest of the language remains stable. That the lexicon is so mutable means that we can study lexical change over relatively short periods of time and variations across relatively small communities of speakers, whereas grammatical changes are slower to manifest.

Example (3) brings up another issue which we should address now, and which is central to the notion of lexicality We've said that the topic of this thesis is how *word meanings* differ across communities and change over time. But this isn't quite true, depending what is meant by *word*. In deciding what to focus on as the unit of analysis, it is useful to look ahead to the phenomena we're interested in. Namely, we want to investigate differences in the *conventional association* between linguistic symbols and

their semantic meaning. These conventional associations are what is captured by the lexicon.

Orthographically *wax jambu* is two words (there is a space in between). Grammatically, *wax jambu* is an English adjective-noun construction. But the meaning of this expression cannot (at least not completely) be understood as a function of the meanings of *wax* and *jambu* and the grammar of adjective-noun constructions — otherwise we might think that a wax jambu is a jambu made out of wax (it isn't). Since the meaning can't be derived compositionally, *wax jambu* must have its own *lexical entry*. What we are really interested in in this thesis is not words in the grammatical sense, but the meaning of **lexical items** — any expression that has meaning which can't entirely be understood as composed of smaller parts. Informally, though, lexical items will still be referred to as *words* where it does not cause confusion.

Expressions like *laser printer*, *paperback book*, and *cell phone* might also be considered lexical items. "Phrasal verbs" like *let on*, *look up*, or *break down* might also be considered lexical items. There are also idioms like *cut corners* or *easy as pie* that encode some conventional meaning that can't be derived from the compositional semantics.[2] We might also like to consider expressions that you wouldn't find in a traditional dictionary. Think of sounds like *uh-huh*, which are often used as *backchannels* during another speakers turn to indicate understanding or agreement. Are such sounds words? They certainly play a meaningful role in conversation. The same can be said of laughter, including different qualities of laughter, which can have a variety of different communicative functions (Mazzocconi et al., 2022). Different morphological inflections of the same stem might be considered different words — for example, *jump*, *jumps* and *jumping*. But they derive their meaning from a single lexical item, which interacts with English morphology in a predictable way.

Non-compositional meaning can also accrue to longer expressions, such as idioms. *Kick the bucket* has a meaning that can't be discovered through compositional analysis. Should idioms be considered lexical items? By the criteria we have established they should, though including idioms in the lexicon poses a threat to our intuition that lexical items are word-like. Some linguists go so far as to take the view that *constructions*, which includes words as well as longer phrases and even syntactic patterns, are the fundamental building blocks of meaning (Croft, 2001). A related project is the *Generative Lexicon* (Pustejovsky, 1995), which redistributes the work of compositional meaning to the lexical level.

The question of what to count as a lexical item becomes operationally important when we want to conduct a corpus-based study of semantic variation or change. Of particular relevance are the pre-processing steps of **tokenization** (breaking up text into

[2] As with *wax jambu* it may be possible to get a rough idea of what some of these expressions mean by understanding the meaning of the words that make them up and the rules of English noun phrase composition. (Moon, 2015) argues that there is continuum in the degree to which such phrases can be considered together as multi-word expressions (see also Bücking (2010)). The important thing as far as we are concerned for the moment is that there is *some* aspect of the meaning which is conventional, not derivable from the composition of smaller units or general principles of communication (i.e., pragmatics; see Section 3.4.1).

discrete units of analysis) and lemmatization (normalizing different morphological inflections of the same lexical item). Section 5.2 discusses these issues in more detail, but for the most Part II proceeds with the assumption that lexical meaning *mostly* resides at the word level.

2.2. Polysemy

A word that has multiple *senses* is said to be **polysemous**. In the dictionary entry for *point*, each of the listed items represents a different sense. Polysemy is ubiquitous in natural language — it is hard to think of a word in English that *isn't* polysemous, at least to some degree. Different senses differ in meaning, and can also have different syntactic types. *Point*, for example, has both noun and verb senses.

Polysemy is important background because one of the main ways that words change in meaning is to gain or lose senses. Likewise, when a word is used differently in some community, it is often because it has an additional sense with a meaning specific to that community. In our example, senses 2d and 6 are special senses of *point* specific to falconry and nautical settings. Chapter 4 will discuss the relationship between polysemy and semantic variation and change in more detail, but for now we maintain a synchronic perspective.

When the meaning of an expression is undetermined with respect to two or more alternative interpretations, it is said to be **ambiguous**. Polysemy, then, is *lexical ambiguity*. When a polysemous word is used in a context that does not make clear which sense is meant, it can result in an ambiguous compositional expression, as in this example:

(4) When asked about his favorite 19th century American author, Jack pointed to the works of Louise May Alcott.

Since *works* itself is ambiguous in this context, the speaker could mean that Jack literally *pointed*$_3$ to a physical collection of books or it could mean that he figuratively *pointed*$_4$ to the abstract collection of literature that is the works of Louise May Alcott.[3]

Polysemy is sometimes distinguished from **homonymy** in that polysemous senses are *semantically related*, whereas homonymous senses (sometimes considered to be separate lexical items) are not. *Bank*, for example, has the *financial institution* sense and the *river bank* sense. The relation between these two senses is usually considered to be one of homonymy, since they are less semantically related than, for example, senses 1 and 2 of *point*. It is worth pointing out, however, that it is difficult to make a hard distinction between polysemy and homonymy (Murphy, 2003).

[3] In context, the gesture described by the first interpretation has a pragmatic meaning similar to the second interpretation since gestural pointing is used to draw attention to something, and a physical book can metonymously stand in for its contents. Insofar as the pragmatic meaning is what is important, the ambiguity of the sentence may not need to be resolved.

But *how* are polysemous senses related? The ways in which senses can be related can be split into two kinds: **regular polysemy**, which follows certain regular patters that can be found across many lexical items and **idiosyncratic polysemy**, where the semantic relation between senses does not follow an established route of connection.[4]

As we will discuss in Section 3.4, the semantic relations that hold between senses that are related by regular polysemy (certain kinds of metaphor, for example) often make it possible to use words in ways that are not lexicalized as senses.

One proposed test is to see whether two "senses" can be joined with a coordinating conjunction. If they can, they may simply be different contextual meanings of the same sense (Deane, 1988). Consider:

(5) The captain and the ship were both pointing.

Does this sentence admit an interpretation where the captain is pointing with his finger (sense 1a) and the ship is sailing close to the wind (sense 2)? If not, this might be evidence that they really are two distinct senses of the word.

It is not always so clear cut, however, whether a sense distinction is present. It would be more difficult to find a similar example that distinguishes between $point_{1c}$ and $point_{1d}$. In the process of writing a dictionary, lexicographers have to make decisions about when to *lump* different uses of a word into one sense or *split* them into multiple senses. This has lead some linguists to prefer a *monosemous* approach in which words are, in general, assumed to have a single meaning (Ruhl, 1989).

2.3. Generality and vagueness

Polysemy is one source of lexical flexibility, but it is far from the only one. Suppose we modify (4) as follows:

(6) When asked about his favorite 19th century American author, Jack pointed *dramatically* to the works of Louise May Alcott *sitting on his desk.*

This sentence no langer ambiguous in the way that (4) was since the context makes it clear that the literal sense of pointing is meant. But many of the details of the situation are still **underspecified**. In truth-theoretic terms, there are aspects of the situation that could change without affecting whether (6) is true. For example, we don't know, the color of the desk. We don't know how old Jack is. It's not specified whether he's pointing to a single *collected works* anthology, or if it's a pile of books. Particularly relevant to the word *point*, we still don't know the exact realization of the gesture Jack made. This is because of the **generality** of the meaning of $point_3$ with respect to some

[4] Deane (1988) calls idiosyncratic polysemy *lexical polysemy*, but here we follow the terminology of A. Blank (2003), who points out that this is somewhat confusing since both kinds of polysemy can be considered lexical.

aspects the gesture. Perhaps *point* implies an extended arm and index finger, but there certainly isn't any conventional (i.e., lexicalized) specification regarding whether it's with the left or the right (or how for the arm is extended or any number of aspects of the motion that gestures may depend on).

That lexical items create this kind of uncertainty may seem like a weakness of language, but in another way of thinking, underspecification *is the meaning* of the word. $Point_3$ picks out situations in which *pointing* is happening. That it doesn't on its own discriminate *between* those situations is exactly what brings them together and gives the word its meaning. Relatedly, lexical underspecification contributes to the flexibility of natural language. Broad lexical interpretations allow a finite vocabulary of words to apply to a wide variety of situations.[5]

Vagueness is a special kind of lexical underspecification in which the borders of the category are not precisely specified by the interpretation. A classic example is gradable adjectives like the word *tall* — there is no conventional height cutoff for when someone (or something) is considered tall. Following the implications of vagueness to their logical conclusions can lead to a paradox known as the Sorites paradox. Since *tall* has vague boarders, it isn't sensitive to very small differences in height — someone who is imperceptibly (say one millimeter) shorter than someone who is tall is still tall. This is what is known as the *tolerance principle* (Wright, 1975). But then we could imagine a long line of people, each one millimeter shorter than the last and by the previous reasoning, we would be committed to saying that someone who is clearly not tall is in fact tall.

There are various ways of dealing with this paradox, many of which involve rejecting or weakening one of the premises (Cobreros et al., 2012). Another is to say that the interpretation of vague terms is probabilistic, in which case the meaning can be represented with a probability distribution that captures uncertainty in where the boundary lies or represents the likelihood that the speaker would use that term in a given situation (Fernandez & Larsson, 2014; Lassiter & Goodman, 2017; Sutton, 2015).[6]

Another aspect of vagueness is that its interpretation is sensitive to what is sometimes called the *comparison class*. What is tall *for a person* may be different from what is tall *for a basketball player*. What is tall for a basketball player is certainly shorter than what is tall *for a skyscraper*. We will discuss context sensitivity and its relationship with semantic variation and change more thoroughly in Section 3.4, but it is necessary to mention here because it turns out that it is a general property of vague terms that they also exhibit this kind of context sensitivity, suggesting that a proper treatment of vague predicate boundaries must somehow take into account the comparison class.

[5] Words can also be used *outside* their conventional interpretations in what is known as *semantic innovation*. This is discussed in more detail in Section 3.4.

[6] See Sutton (2018) for an overview of such approaches.

2.4. Perceptual meaning

Formal semantics is mainly concerned with compositional meaning — given a string of words (perhaps enriched with a certain syntactic structure) and given the meanings of those words, how is the meaning of the string computed? The framing of this question takes lexical meanings for granted. Formal semantic theories often assume that the meaning of predicate-denoting words like *yellow*, *point*, *square* and *democratic* can be modelled as a certain kind of mathematical object. Once the *kind* of mathematical object is decided the semanticist's job is to decide how those meanings interact with each other in compositional expressions, *given their content*. For example, if the formal semantic theory says that predicates denote sets of entities, the semanticist might decide that the meaning of an expression like *yellow square* is the intersection of the meaning of yellow and the meaning of square. It doesn't matter exactly what sets *yellow* and *square* denote, the semanticist only cares that the meaning of adjective-noun phrases of this sort are computed by set intersection.

The fact remains, though, that the meaning of at least certain words *is* closely related to perception.

(7) a. A: Please pick up the yellow square.

b. B: Okay.

If we want to give an account of how this interaction can be successful — how *the yellow square* can successfully refer — we need to explain how B's interpretation of (7a) relates to their perception. In other words, the denotations of *yellow* and *square* must be **grounded** in perception.

Insofar as formal semantics is interested in inference, perceptual meaning cannot be ignored. Certain relations between words can be encoded by **meaning postulates**. We can imagine for example, encoding that *all ravens are birds* by restricting the denotation of *raven* to be a subset of the denotation of *bird*. However, Marconi (1997) argues that a speaker who only had access to lexical meaning encoded in this way could not ever be considered fully *competent* in the meaning of those words. This is especially apparent in situations where *referential competence* is required, as in (7), but it also extends to *inferential competence* in certain cases.

Computational models of meaning have a similar problem, as they often rely on the **distributional hypothesis** (Firth, 1957; Harris, 1954), which says that the meaning of a word can be approximated by the distribution of linguistic contexts in which it appears. A model based on the distributional hypothesis may be able to recognize that *yellow* and *orange* are similar in certain ways (they both appear in proximity to words like *paint*, *pigment*, perhaps even *sunrise* or *flower*), but also have some differences (perhaps *yellow* appears with *canary* and *orange* does not). This sort of model has been criticized for not explaining how symbols in the language are **grounded** in the actual world (Bender et al., 2021; Harnad, 1990; Lücking et al., 2019); such a model

cannot learn meaning representations like the ones humans have since they only relate text to other text, not to perception. A strong version of this argument might claim that even words like *democracy* that don't obviously relate to perception can't, in principle, be truly understood by such a model since meaning in the language system is interconnected and *democracy* is grounded in relation to the rest of the system.

One way of grounding perceptual meaning is to say that the meaning of a predicate-denoting perceptual word like *yellow* is at least in part determined by a **perceptual classifier** — something that computes a function that takes perceptual data as input and produces a category judgment. This approach can be used to ground the meaning of words in computational models, using machine learning classifiers (Schlangen et al., 2016; Silberer et al., 2017). Furthermore, classifier-based word-level meaning representations are subject to compositional analysis, at least in the case of referring expressions (Kennington & Schlangen, 2015).

Such an approach can also be made compatible with formal semantics and information state update models of dialogue (Larsson, 2013, 2020), which we will discuss further in Section 5.1. In brief, the classifier takes the place of a set of entities in the more traditional version of predicate denotation. This better tracks intuitions about how predicate denotations work for actual speakers — it's not that we carry around a list of all the yellow things in the world, but rather than we have the ability to determine if something is yellow, should the need arise. Furthermore, this classifiers as a basis for predicate denotation opens up possibilities for semantic learning based on linguistic and perceptual feedback (Larsson & Bernardy, 2021; Larsson & Cooper, 2021), something particularly important if we are interested in modeling semantic change.

2.5. Cognitive approaches

Wittgenstein (2009) points out that the meaning of a word can almost be described with a complete set of necessary and sufficient conditions. The word *game* is an example. It's very hard to come up with a definition that would cover everything we call a game. We can think of some examples that are *typical* games, but insofar as anything else is a game, it is through a sort of *family resemblance* to the other things we call games. **Prototype theory** holds that cognitive categories themselves are defined not by a set of features but rather in reference to certain ideal *prototypes*. Membership in the class, then, is judged in reference to the prototype (Rosch, 1975).

Some linguists, in turn, have argued that most of what is thought of as polysemy can be explained without making sense distinctions — that for the most part, there are just more and less prototypical realizations of the categories that words refer to (Ruhl, 1989). However, there are reasons to think that sense distinctions are real.

But then how do we explain that a situation described by $point_3$ seems more prototypical of *pointing* than one described by $point_4$? It may be that these two senses are not actually distinct (though the ambiguity of (4) suggests that they are), but others

have suggested that prototypically effects can obtain on two levels — on the *conceptual level*, between instances, as well as on the *semantic level*, between senses (Kamp & Partee, 1995; Tyler & Evans, 2001). Regardless, it seems difficult to make a hard distinction between when two different meanings come from different senses of a word versus when they result from different interpretations of the same sense, draw out by different contexts.[7] This difficulty is reflected in the nested list format often adopted by lexicographers when enumerating senses, which offers readers different options for granularity at which sense distinctions might be made.

Related to prototype theory is **exemplar theory** (e.g., Medin & Schaffer, 1978; Nosofsky, 1984), which is, in some way, an even stronger version of the same idea. In exemplar theory, a concept is still defined in relation to an ideal, but an exemplar is not an abstract idealization, but rather an ideal *member* of the very category. Put another way, exemplars of *of the same kind* as the members of the category, whereas prototypes need not be. Category membership is determined, then, in relation to one or more exemplars of the category. There is some experimental evidence to suggest that both exemplar and prototype-based strategies are involved in classification (H. Blank & Bayer, 2022; Malt, 1989).

While most of this work is not explicitly linguistic, it is of interest to the study of language since there is presumably some connection (via lexical meaning) between the processes linguistic production and interpretation and the representation of conceptual categories. In Chapter 10 we investigate this question using a neural language generation model provided with exemplar and prototype theory-inspired representations of visual categories.

[7] Section Section 3.4 further discusses the use of context for situated meaning making.

3. Sources of meaning

> We die. That may be the meaning of life. But we do language. That may be the measure of our lives.
>
> Toni Morrison
> 1993 Nobel Prize ceremony

The idea of *the lexicon* suggest a big book of words arranged in a list. In the previous chapter, we suggested that the structure of lexical knowledge might be a little more complex than what can be reasonably represented by a list, and that interrelationships between lexical items might have implications for how semantic change happens. In this chapter we'll question the idea that the "book" itself is a monolithic thing. Instead, we have many different sources of lexical meaning, which we draw on in different interactive contexts. Not only that, but context can allow us to draw in sources of meaning from outside the lexicon, or extend the meaning of words beyond what they would normally reach. All of this is important for change because change happens in a particular communicative context, and the meanings that are available in that context are also the ones that have the potential to become part of some lexicon.

We can't point out a language in the material world, and no more can we put our hands on its lexicon. Yet we talk about them as if they are individual entities that have properties, that can come into and go out of existence, and so on. *Ken and Helen both know French*. *Modern English appeared in the 15th century. Latin is a dead language.* But this way of talking about languages belies much of the complexity at the heart of this thesis. Do Helen and Ken have *exactly* the same knowledge of the French language? Of course not! Is the English of the 15th century the same as what is spoken around the world today? No! — it was very different, as are the many varieties of English that are spoken contemporaneously. The reason we call Latin a *dead* language is not because it doesn't have any speakers (indeed, some people do still learn a version of Latin in school), but because it doesn't have a *community of speakers* who use it, who breath life into it, whose communicative needs it serves and with whom it changes.[1]

The perspective on language adopted throughout Part II is that it exists *through people* and *in communities*, available as a resource for interaction. In dialogue, speakers

[1] Latin *is* spoken for ceremonial purposes and is even used in official documents in the Vatican but isn't generally used in *interactive* communicative contexts which, as we will see, are particularly important for engendering linguistic change.

generally assume that the language they are speaking is *common knowledge* among the interlocutors — that the meanings of words, how to construct and interpret utterances, etc. are shared by everyone, that everyone *knows* that they are shared by everyone, that everyone *knows that everyone knows* they are shared, and so on. This *and so on* makes things tricky. Practically, how do we get to a point where we don't need to go through an infinite regress of social deduction just to be assured that communication is possible? **Common ground** (Lewis, 1969; Stalnaker, 2002) is a model of common knowledge that solves this problem. We say that something is common ground for a group of people if there is a **shared basis** that indicates that it is true. A basis b is shared for a group G when:

1. everyone in G has information that b holds, and

2. b indicates to everyone in G that (1) is the case.

Clark (1996) identifies two kinds of common ground. **Communal common ground** is shared based on joint membership in a community. At a geology conference, the fact that *limestone is a sedimentary rock* may be considered common ground, based on joint membership in a community of geologists. Similarly the meaning of those and other geological terms may be taken to be common ground: the meaning of *limestone* is part of the lexicon for the community of geologists because everyone in the community knows what *limestone* is and because being part of the geology community is generally understood to entail knowing the meaning of *limestone*.

Personal common ground is grounded on a *perceptual* or *actional* basis.[2] Such a basis is shared because of their joint attention on some event or *situation*. Imagine we are at a baseball game. We're both intently watching a crucial at-bat. The batter hits the ball. It's a home run! Now it's common ground among us that the batter has hit a home run. This is the case on the basis of the situation just described, since you and I both have perceptual access to the situation (we were both paying attention) and since our joint attention is itself evident in that very situation we both have access to.

It is important to emphasize that common ground is a subjective notion — it depends on what an individual *takes to be* common ground.[3] It is not uncommon, for example, for dialogue participants to find that their construal of what has been said in a conversation is misaligned and in need of repair. What someone takes to be common ground in an interaction depends on the requirements of the interaction. For example, speakers may at times be rather loose with assuming that certain lexical items are common ground, trusting that misunderstandings will be identified and repaired.

[2] The main difference between actional and perceptual bases is that actional bases are events that are brought about by the joint action of the participants, usually by way of exploiting pre-existing common ground. For Clark (1996), actional bases are key to explaining how dialogue works

[3] In some cases, we may wish to consider what *all speakers take to be* common ground as a corollary to the objective notion. In other cases it makes sense to take a more explicitly agent-centric notion.

As we will discuss in the following two sections, lexical meaning can be grounded in both both communal and personal common ground, a fact that is particularly relevant to lexical semantic change.

3.1. Communal lexicons

Any community can serve as a basis for communal common ground, which would suggest that any community of language users could have its own lexicon. Indeed, this is the idea behind the notion of what Gumperz (1972) terms the **speech-community**, which is

> any human aggregate characterized by regular and frequent interaction by means of a shared body of verbal signs and set off from similar aggregates by significant differences in language usage. (Giglioli, 1972, p. 219)

A *language*, then, can be defined as the accumulation of the linguistic norms and practices grounded in a particular speech community.

Of course, this is quite a different notion of language from the one that is in common usage. We don't usually think of geologists and fire fighters as speaking *different languages*. But in a sense they do, especially when speaking with one another about topics of special importance to their respective communities. But this also reveals that there is a hierarchical relationship between speech communities. While geologists may have special terminology in the domain of geology, they default to what we will call a *macro-language* (English, for example) where the norms of the geologist community have no special bearing. There may of course be intermediary communities as well. Perhaps there are conventions among natural scientists, a designation which includes geologists.

Even this picture is a bit simplistic. French geologists mainly speak French. The ways in which their speech differs from the macro-language may in some ways be similar to how English geologists' speech differs from English (perhaps based in joint membership in an international community of geologists) and may in some ways be particular to the community of French geologists.

We usually think of macro-languages as at the top of this hierarchy, but some macro-languages are at least partially mutually intelligible. In these cases, it's not necessarily that there is a community that encompasses both, but rather that the norms of the two respective communities are close enough that certain linguistic conventions can be considered common ground for the purposes of conversation.

3.2. Interpersonal lexicons

Communal lexicons are probably what we usually think of when we think of lexical knowledge, but lexical meaning can also come from personal common ground. These interpersonal lexicons generally don't constitute a whole language, but rather, as with in a specialized community lexicon, supplement another more general linguistic resource that the participants also share.

Special idioms, nicknames, expressions of affection and more are often shared between families, close friends and romantic partners (Hopper et al., 1981). Not only do interpersonal lexicons facilitate communication (including possibly covert communication), they serve to express solidarity and closeness among intimates (Bell & Healey, 1992).

Interpersonal lexical resources are not limited to novel words, though. When someone uses a word in a particular way, their dialogue partner might take note and expect that sort of usage in the future. This is particularly true if the intended meaning wasn't obvious at first and required extra reasoning or especially repair (G. J. Mills & Healey, 2006). In general, when speakers have to coordinate on the meaning of a lexical item, the coordinated meanings may carry over to future dialogues, i.e., as part of an interpersonal lexicon.

3.3. Semantic coordination

So where do these interpersonal lexical resources come from? How do we go from not sharing any partner-specific meanings with someone to having them? The answer is semantic coordination. To understand how that works, we first need to discuss how personal common ground is built up over the course of an interaction.

The Collaborative Model (Clark & Schaefer, 1986; Clark & Wilkes-Gibbs, 1986) is a theory of conversation that explains, from a psycholinguistic point of view, how speakers collaborate through **communicative grounding**. The model describes a hierarchy of grounding levels that dialogue participants must move through in order to reach (and maintain) mutual understanding. To coordinate effectively, participants must tailor their actions to what has been grounded so far while also providing evidence (positive and negative) of grounding to facilitate their interlocutors doing the same. When evidence of understanding is demonstrated, participants can consider what was said to be common ground.

Communicative grounding is subject to *opportunistic closure*, meaning that grounding at higher levels are taken as evidence of grounding at lower levels. Closure can also work compositionally — if B gives evidence they understood A's utterance, that can be taken as evidence that they understood the words it was composed of in the way they were meant.

Such an understanding can be achieved even when the speaker uses a word outside of

	Level	Speaker (A) and addressee (B) actions
1	contact	A and B pay attention to each other
2	perception	B perceives the signal produced by A
3	understanding	B understands what A intends to convey
4	uptake	B accepts/reacts to A's proposal

Table 3.1.: Levels of communicative grounding (Fernandez, 2014)

its normal semantic range or in a way that the addressee is not familiar with. We discuss the ways in which extra-linguistic context interacts with word meaning in more detail in Section 3.4. In these situations, the addressee may shift their understanding of the word for the purposes of the conversation, what we call **implicit semantic coordination**.

Implicit coordination has been studied in experimental settings where participants develop **lexical pacts** (Brennan & Clark, 1996). These temporary, flexible conventions emerge as a consequence of successful interaction and persist as a resource as the interaction continues (or even in future interactions). Such conventions are not limited to isolated lexical items. G. Mills and Healey (2008) observed that participants asked to perform a collaborative maze-solving task would create a **conceptual pact** — a unified semantic model of how to refer to locations in the maze.

On the other hand, if they cannot figure out the intended meaning or wish to raise a meta-linguistic objection to that use of the word, they may initiate a **word meaning negotiations** (WMN) (Myrendal, 2015). Here, *negotiation* is meant in the sense that a group of friends might negotiate paying for a restaurant bill — WMNs are collaborative on the level of interaction, with the implicit goal of reaching a mutually agreed upon result. WMNs, *can* of course be adversarial in terms of the outcome, but they need not be and, as in any dialogue, some level of cooperation is needed to coordinate the interaction.

It's difficult to characterize exactly how common WMNs are in every-day conversation, since they take many surface-level forms, making them difficult to search for exhaustively in a corpus. Myrendal (2015) studied word meaning negotiation in Swedish discussion forums, collecting a corpus of exchanges by searching for the phrases like *Vad menar du med // what do you mean by*. In Chapter 7 we used variations of the same phrase in English to find WMNs.

Corrective feedback is another form of **explicit semantic coordination**. Consider these examples of adult-child speech (Larsson & Cooper, 2009):

(8) a. A: That's a nice bear.

b. B: Yes it's a nice panda.

(9) a. A: Mommy, where my plate?

b. B: You mean your saucer

Corrective feedback can be seen as a special case of WMN where there is an implied epistemic inequality between the participants.

Both implicit and explicit coordination have the potential to affect affect lexical resources beyond the current dialogue. After an interaction speakers may remember an unusual way they used a word or remember a word meaning that was negotiated. These newly coordinated meanings can be made available for use in future dialogues based on interpersonal common ground. Under the right circumstances, a speaker may also take the dialogue as evidence that new lexical information holds for a particular community, resulting in community-level change (for the speaker).

3.4. Meaning in context

Chapter 2 introduced the idea that words have a certain amount of semantic flexibility. A single word often has multiple related senses (polysemy) and meanings can describe a variety of different situations (generality). Although there are different ideas about where these sources of flexibility reside and how they interact with each other, it is clear that in-context meaning is dramatically *less* underspecified than lexical meaning in the abstract.

Abstract or lexical meaning is sometimes called **meaning potential**, which is considered to be something of an entirely different kind from in-context meaning. Norén and Linell (2007) describes meaning potentials as semantic *affordances*. They are something that "afford language users with semantic potentialities to be exploited in situated use." In Gibson (1966)'s theory of perception, an affordance is what the environment provides as an interactive possibility. Similar to the way that a cup with a handle may offer *grasping* as an affordance.

Lexical meanings (or meaning potentials) combine with linguistic and extralinguistic context to produce **situated meanings**. One way this happens is that context can narrow the generality of interpretation of a word. Consider a word like *eat*. Eating a soup and eating a sandwich are quite different activities, yet the sentential context can make clear what is meant. Such narrowings can persist even outside of a disambiguating sentential context by relying on discourse context.

(10) a. A: I'm eating soup.

b. ...

c. A: Leave me alone, I'm still eating!

To see how extralinguistic context can affect situated meaning, consider the interpretation of definite referring expressions, for example.

(11) a. *That red car over there* is mine.

b. *The red one* is mine.

c. *The head of department* will be at lunch tomorrow.

d. I'm going to *the library* later.

Expressions like these often require some additional perceptual common ground to successfully refer. We might imagine 11a being used in a situation where the interlocutors have joint attention on some visual scene. If the visual and communicative context makes it clear that the speaker is referring to a car, the 11c might do the trick Communal common ground could be required to interpret 11c, and 11d might draw on some shared relevance ordering on libraries.

Different modes of interaction are available depending on the *genre*. Situated meaning can also be affected by the **genre** of communication (Bakhtin, 1987). Genre is something similar to Wittgenstein (2009)'s *language games*. They are different modes of interaction that serve as communicative resources in different situations. Consider, for example, a waiter at a diner speaking to a colleague. The waiter may metanonomyously refer to a patron who ordered a ham sandwich as *the ham sandwich*. But this same referring expression wouldn't be available to another patron or to the waiter talking to someone else who isn't working there (A. Blank, 2003).

3.4.1. Pragmatics

Like semantics, *pragmatics* refers to both a subfield of linguistics and the collection of linguistic phenomena that the field studies. A classic way to make the distinction between semantics and pragmatics is to say that pragmatics has to do with **speaker meaning**, which is different from meaning in the abstract.

Of course, this distinction rests on the potentially precarious assumption that there *is* something like meaning in the abstract. Nevertheless, it is the case that there are a number of kinds of communicative situations where the interpretation of the speaker meaning is somehow based on a prior interpretation of the "literal" semantic meaning. Consider this classic example:

(12) a. A: Can you pass the salt?

b. B: Sure thing. *[passes the salt]*

We could imagine situations where (12a) is uttered as a genuine question, but typically one would interpret it as a request for the salt. Searle (1975) refers to this kind of utterance as an indirect speech act because the primary intention (requesting) is performed by performing an action with a different *literal* interpretation. B confirms the indirect interpretation, with their reply (12b), which grounds A's utterance as a request by responding to it as such.

So how does B know that (12a) is a request, given that it is literally a question? One story goes like this: B first processes the utterance as a question, using their "normal" faculty of semantic interpretation. From this interpretation, B reasons about why A might have asked this question, given that there is no reason for uncertainty about B's salt-passing ability. B realizes that (1) their ability to pass the salt is a prerequisite for actually passing it, (2) it would be considered impolite for A to make their request as an imperative (i.e., *Pass the salt.*), and perhaps (3) they are in a situation where it would be normal for A to want B to pass the salt. From here, B concludes that A must have uttered (12a) as an indirect way of requesting that they pass the salt.

This follows the classical *Gricean* account of *conversational implicature*, wherein the pragmatic meaning is derivative of the **cooperative principle** of communication, which says that in general, cooperative speakers try to be informative, truthful, relevant, and clear (Grice, 1975). When it would seem that a speaker is in violation of or *flouting* one of these maxims, a cooperative listener will go searching for some alternative interpretation under which their interlocutor is adhering to them.

Now, this is a rather long and involved story to tell about what would seem to be a rather simple and (crucially) routine interaction in (12). Indeed, Grice (1975) might instead explain this example in terms of *conventional implicature*. On this account, the meaning may have at one point been calculated as previously described, but it has since become *conventionalized*, obviating the need to perform the pragmatic inference.

Herein we see a clear connection between pragmatics and semantic change. If the phrasing *can you pass...* (or more generally *can you...*) becomes conventionally associated with the act of requesting, that would mean by definition that some lexicalization of the requesting function has occurred i.e., that the meaning has changed. It could be the case, as Morgan (1978) argues, that speakers can make a distinction between literal and indirect uses of a linguistic unit, even when the indirect function is conventionalized. This further fuzzies the border between semantics and pragmatics. If studying semantic change means studying changes in what thing speakers can *use words to mean*, we really can't avoid pragmatics.

Pragmatic inference often seems to rely on some shared assumptions about what follows from what.[4] In argumentation theory, an *enthymeme* is an argument in which one or more of the premises are not explicitly stated. A **topos** is that which supplies the missing premise (J.-C. Anscombre, 1995). Put another way, it is a function from enthymemes to full arguments. J. C. Anscombre and Ducrot (1983) observes that even language that is not *prima facie* argumentative still has a structure that can be analyzed as argumentation. Often this means that there are implicit (enthymematic) steps in a discourse. Breitholtz (2020) develops a theory in which topoi are a resource that can be drawn on in in linguistic interaction — part of the common ground in the same way that lexical items are. Indeed, we discuss topoi as having something like lexical

[4] Although we do not discuss it in the thesis, relevance theory is a theory of pragmatics that centers ethememetic reasoning (Sperber & Wilson, 2001).

meaning in Chapter 11, but in that work we focus not so much on their role in pragmatic inference, but rather their role in implicitly communicating social information about the speaker — their *social meaning*.

3.4.2. Social meaning

Social meaning is similar to pragmatic meaning in that it goes beyond the literal interpretation of an utterance. It is also similar in that recovering the social meaning usually involves considering the communicative context. In the case of social meaning, however, the relevant communicative context is usually not so much on the level of interaction, but rather the *social context* in which the interaction takes place. Very often the social meaning of an utterance communicates something about how the speaker themself relates to the social context — for example, by revealing something about their social position or ideology.

Eckert (2019) classifies the progression of sociolinguistics as a field in three *waves*, each of which have a different take on how social meaning functions. Early sociolinguistic work largely sought to describe regional linguistic variation and variation among macro-social categories of speaker (i.e., based on age, gender, class, etc.). First-wave sociolinguistics already acknowledged the potential for variables to carry social meaning. Labov (1963), for example, acknowledged that phonetic changes in the speech of certain non-native residents of Martha's Vineyard may have something to do with a desire to be associated with the working-class resident population as opposed to the upper class summer visitors. However, he also noted that the diphthong centralization he observed was not consciously salient to speakers he interviewed, suggesting that the changes were largely a matter of subconscious identification with a particular ideology. The idea that speakers use sociolinguistic variables as a way to (more or less) intentionally *construct* a social identity is the hallmark of second-wave sociolinguistics. However, it is in third-wave sociolinguistics that the variable moves from being seen as a theoretical tool to something with a "social and cognitive reality" (Campbell-Kibler, 2010).

This social and cognitive reality can be seen in the concept of the **persona**, which is a sort of stereotypical *kind of person*, which is a common ground resource (in the sense of Clark (1996)) that speakers can draw on to construct a social identity. Personae are not real people, but they do have a kind of social reality, give their common ground status as social reference points. People can draw on personae to construct an identity by way of social signals that are indexically associated with the personae in what Eckert (2008) calls the *indexical field*.

Importantly, social signals are not exclusively linguistic. They can include all kinds of things, including dress, body language, and so on. Furthermore, linguistic social signals are not limited to *how* something is said (although this has tended to be the focus of sociolinguistics), but also *what* is said. In Chapter 11 for example, where we develop

a probabilistic model of social signalling based on Eckert (2008)'s indexical field, we use topoi as the case study. A topos is evoked by ethememetic speech, something which is more related to content than style.

4. Semantic variation and change

> Gretchen, stop trying to make fetch happen. It's not going to happen!
>
> Regina, *Mean Girls* (2004)

Linguistic variation is an important concept in sociolinguistics, which is mainly concerned with variation across social groups.[1] On a structural level, change is just variation over time. Many of the same corpus-based methods used to study change can also be used to study synchronic variation (see Section 5.2.2). But it is also worth considering variation and change separately since each have a role to play in explaining the other. Variation leads to change as one community of speakers adopts ways of speaking from another community.

Change leads to variation as communities diverge in their language. Variation and change also have different social implications and different relationships to lexical meaning. The next two sections will draw out some distinctions that are made in *types of variation* and *types of change* we might observe. Some of these categories apply to both variation and change, while others are specific to one or the other.

4.1. Types of variation

One of these distinctions is related to the different factors driving variation. Sometimes, variation stems from the fact that speakers draw on different bases for linguistic common ground in different situations. An expert in some field may speak differently when talking to peers in her community than she would talking to non-experts who nevertheless speak the same macro-language. A teenager writing a message in a video game forum can relay a story about something that happened in the game differently from how they would tell the same story to their parents at the dinner table. This kind of variation is an example of **code-switching**, which can also involve multiple macro-languages.

However, it is somewhat naive to think that sociolinguistic variation is always rooted in differences in common ground. A politician from Skåne (in southern Sweden)

[1] There are also, of course, differences in language use across individuals which do not present on any particular axis of social identity. These differences usually fall under of *linguistic style* and, while they have received some attention (e.g., Johnstone, 1996), are not a main focus of sociolinguistic inquiry.

may use different vowel articulation, speech patterns, and lexical items depending on whether they are speaking to rural constituents or meeting with business leaders in Malmö. This code-switching may have nothing to do with common ground *per se* — both registers would probably be just as well understood in both situations, but rather the choice of register is explained by how the politician wants to be perceived — what *persona* they want to *project* (see Section 3.4.2). Different ways of speaking carry different *social meaning*

That variation is not always a result of different common ground is also evident in the fact that the social categories along which sociolinguists study variation are not always speech communities. Indeed, classical sociolinguistics is much more often concerned with variation across macro-social categories like gender, race, class, and even sexual orientation (Labov, 1963; Podesva, 2007). Eckert (2008) attempts to introduce further nuance, arguing for an approach in which linguistic variants are not mere markers of social identity, but a collection of signs in a complex semiotic system through which individuals may project their social identity in relation to social archetypes or *personae*. The personae themselves stand in relation to local and macro-social categories, but are not wholly constituted by them.

These two sources of variation are, of course, not easily separable. Variants that are understood across communities and whose primary function is to mark community membership or project personae may, with time, evolve into something that requires a certain common ground to understand. Similarly, variants that are only understood within a certain community may come to be understood more broadly while retaining their status as a social signal.

Type 1 and type 2 variation Another distinction has to do with the perspective we take on variation as observers of the system. We have already spoken loosely of *variants* as the units along which variation is observed, but this needs to be made more precise. To identify something as a *variant* means that there is a difference with respect to some reference point that stays the same. There is the *variation* and there is the thing it is a *variation of*. So when we talk about linguistic variation, what is it that changes and what stays the same?

In classical linguistic theory, language is thought of as a hierarchical system of semiotic relations where signs on a lower layer signify meanings on a higher layer (Fig. 4.1). Phonemes are signified by particular speech sounds, morphemes by particular phonemes, words are made up of morphemes, syntactic structures are determined by the grammatical categories of a string of words, and the meaning of units of speech larger than words (sentences, for example) is determined in part by the meaning of words that make it up and the syntactic structure they produce.

Without getting too into the weeds about what the appropriate unit of analysis is at each of these levels (or indeed where the lines between levels should be drawn, if at all), the classic assumption is that units at one level of analysis (sometimes alone

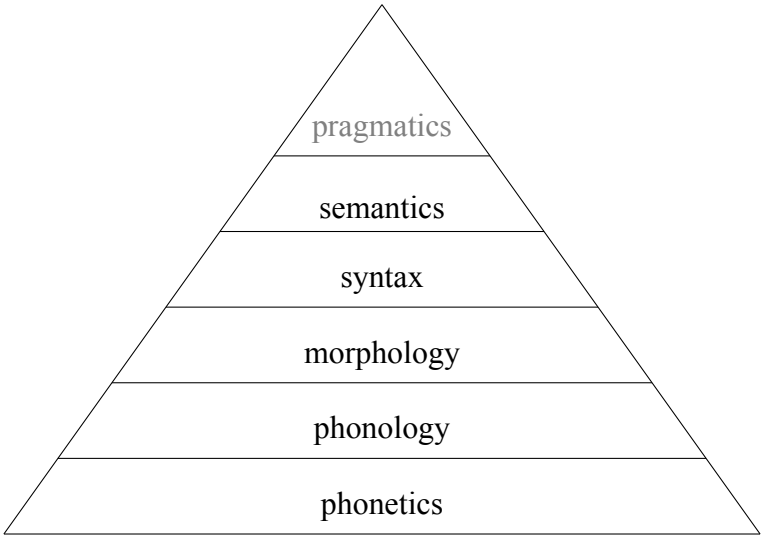


Figure 4.1.: The classical linguistic hierarchy. Forms at lower levels of the hierarchy combine to create meanings at higher levels. Pragmatics is somewhat different because pragmatic meaning depends on extra-linguistic context as well as semantic meaning, which might be why it is often left out of the classical linguistic hierarchy.

(a) Type 1 variation

(b) Type 2 variation

Figure 4.2.: Two types of linguistic variation. In type 1 variation, different forms variously signify the same meaning (for example, in communities A and B). In type 2 variation, the same form variously signifies different meanings.

and sometimes in combination with other units), which we will call *forms* stand in a signification relation with units at higher levels, which we will call *meanings*. Anttila (2004) points out that this leaves us with two kinds of variation in the relationship between forms and meanings (Fig. 4.2).[2] .

Sociolinguistics is almost always concerned with type 1 variation. Perhaps the most clear-cut example is sociophonetics, which studies variation in the relationship between speech sounds and phonemes. A *variant* in sociophonetics is a phoneme that can be signified by multiple different speech sounds. Think back (Section 3.4.2) to the centralized diphthongs that Labov (1963) observed in Martha's Vineyard. Those vowel sounds were phonetically different across groups, but interpreted phonologically

[2]This hierarchical model persists in spite of many arguments and counter-examples against it. In fact, there are interactions at many levels of the hierarchy, not only adjacent ones and not only in one direction (Cann et al., 2000). Indeed, pragmatics is not only about determining *why* something was said, but also *what* was said (Korta & Perry, 2008). If our linguistic theory computes meanings one level at a time from bottom to top, then it is going to run into problems.

to mean the same thing.

It can of course also be the case that the same speech sound is used to mean two different phonemes (in two different regional accents, for example). But this isn't really considered sociolinguistic variation. Why? It has to do with social meaning. When someone's way of saying M_1 is F_2 instead of F_1, that's a salient difference that we can ascribe meaning to. When someone's way of saying M_2 is F_1 (while someone else might use F_1 to mean M_1), that relationship is less salient because ambiguity is ubiquitous in language — just because the speaker used F_1 to mean M_2 doesn't mean they don't *also* use it to mean M_1.

This puts us in an awkward position if we want to investigate semantic change from a sociolinguistic perspective since semantics falls at the top of the linguistic hierarchy.[3] It's (relatively) easy to search for a particular word and see how its contexts of use vary, whether across time or some other factor. It's much more difficult to search for *contexts in which someone might want to express a particular meaning* and see what words they used. In fact, this is exactly what Hasan's (2009) work on semantic variation does. For example, in a corpus of child-directed speech, she investigates the different ways mothers have of issuing a command to their children (Hasan, 1989).

In lexicography, there is a parallel distinction between semasiological approaches (organized around form) and onomasiological approaches (organized around meaning). Historical linguists have adopted these terms: type 1 lexical variation over time is called **onomasiological** change, and type 2 lexical variation over time is called **semasiological** change. Historical linguistics mostly considers semasiological change for much the same reason (think of the classic examples like *awesome* and *gay* that we introduced in Chapter 1).

This thesis, likewise, mostly considers type 2 variation. Certainly in Chapter 13, when we use computational methods to quantify how much particular words change across corpora, we are working with type 2 variation. But type 1 variation is present in questions of semantic change as well. In Chapter 7 we study explicit conversations about word meaning (a phenomenon which we will introduce in Section 3.3). The word in question is very often a word that everyone already knows *some* meaning for, but the meaning in the particular context is unfamiliar (or even disagreeable) to someone.

4.2. Types of change

There are many different ways of categorizing lexical change, and different ideas about what should go into those taxonomies. Most change typologies take a semasiological perspective, but there are also onomasiological categories. Perhaps the most important of which is *lexical replacement*, which is when a word is replace by another word in a particular situation. We'll talk about two semasiological typologies here. The first

[3] See Hasan (1989) for discussion of how pragmatics fits (or doesn't) into this picture.

Change type	Meaning
Novel word	A new word form (and associated sense, possibly also novel)
Novel sense	A new sense for an existing word
Word death	A word form that goes out of use
Sense death	A sense of a word that goes out of use
Sense split	What was one sense of a word is now considered two senses
Sense join	Two senses are no longer distinguished

Table 4.1.: Sense-structure classification of lexical semantic change. Table adapted from Tahmasebi et al. (2021), which also includes an analysis of how these types are construed in the field of computational change detection.

is to classify according changes in the lexical **structure** — that is, with respect to the inventory of words, senses, and relations between the two (Table 4.1). Of course these distinctions assume a list-of-senses model of lexical meaning, which as we discussed in Section 2.2, is not without its problems.

Another way of classifying lexical change is in terms of the **explanations** for why the change happened or what might have made it possible (Table 4.2). Again, these explanations assume sense distinctions since they often are described in terms of a relationship between an old sense and a new sense. For example, the word *mouse* has a new (20th century) sense, which refers to a computer mouse. The new sense has a metaphorical relationship to the animal sense.

At the risk of just-so story, we can easily imagine that the old sense might have licensed innovative uses in contexts covered by the new sense before the change was lexicalized. For example, the change in the meaning of *mouse* (by adding a sense) came about by the conventionalization of metaphorical extensions of *mouse* to a new artifact (the computer input device) that played on visual similarity. These relations suggest a synergy between polysemy, lexical innovation, and change, although it's important to point out that there is no one-to-one correspondence between relations between senses and types of change (A. Blank, 2003).

Change type	Meaning
Metonymy	A new sense is related by metaphorical comparison
Metaphor	A new sense is related by metaphorical comparison
Co-hyponymous transfer	A new sense denotes something that shares a hypernym with the existing sense (the new sense and the old sense are co-hyponymous)
Semantic extension	The word now applies to more related situations (also called semantic *broadening*)
Semantic restriction	The word now applies to a more restricted set of situations (also called semantic *narrowing*)
Antiphrasis	A word gains a sense which is opposite to an existing sense (for example through innuendo or humorous innovation)

Table 4.2.: Explanatory classification of lexical semantic change. Abbreviated from A. Blank (2003), which includes a comparison with synchronic sense relations.

5. Methodology

> ... in that Empire, the Art of Cartography achieved such Perfection that an entire City was occupied with the Map of a Single Province, and a Province was required to display the Map of the whole Empire.
>
> In time, even these Vast Maps ceased to satisfy, so the Cartographers' Guilds unfurled a new Map of the Empire the Size of the Empire, each point overlaying exactly what it mapped.
>
> Later, Generations, less addicted to the Act of Mapping, understood that this Immodest Map was Useless — they irreverently surrendered it to the ravages of Sun and Snow. In the Western Deserts, tattered Ruins of the Map remain, home to animals and vagabonds; these are the Country's last vestiges of the Geographic Disciplines. (1658)
>
> ---
>
> *on Scientific Rigor*
> Jorge Luis Borges
> trans. Noah Mease

Part II uses a variety of different methods, including both formal and computational models of natural language semantics. The reasons for this methodological diversity are twofold. First, the nature of short-term semantic change places us in the liminal space between interaction, interpersonal relationships, and speech communities. We need to, on the one hand, investigate semantic plasticity in the context of concrete interactions and, on the other hand, investigate change over short time periods abstracted over communities of practice.

Second (and related to the first) is because of the scientific goals of the thesis. Some studies in the thesis test hypotheses about how change and variation takes place in communities (e.g., Chapters 12 and 13) or how metalinguistic communication can be implemented in neural models (e.g., Chapter 10). Other parts are more focused on developing theoretical models of semantic meaning and interaction that allow for change (e.g., Chapters 7 to 9 and 11).

Every study in the compilation uses some kind of *model*. But what *is* a model? Perhaps the most prototypical models are ones that mimic the workings of something they are *modeling*. Think of an environmental geologist who builds a physical model of a riverbed to asses the risk of flooding. They might try some experiment like building or removing a dam and extrapolate the effects observed in the model to what would happen if the analogous actions were carried out in on the real-life stream. A com-

putational model can play a similar role. Such a model would attempt to extract and quantify key aspects of the stream bed, many of which the physical model also captures (rates of flow, soil permeability, etc.). The model can then be used to make predictions as a function of those parameters.

But *models* in computational linguistics and NLP don't always have the same relationship to the *modeled*. In fact, it is not always very clear what, if anything, is being modeled. The mechanisms of language processing in a large language model like BERT are not at all clear, and there is no reason to think that, on a low level, it is doing anything at all analogous to what humans do when they process language. In this way, the relationship between such a model and human language processing is more like the relationship between a bird and a quadrocopter drone. Sure, they do *some* of the same things, but they do them by entirely different means. But that doesn't mean that models like BERT are useless for studying natural language. Just as one might use a drone to get closer to the habitat of birds, to measure the flow of wind over a cliff, or capture what a rabbit in a field looks like from high in the sky, machine learning models can, through careful analysis, be useful for investigating relationships in the linguistic environment in which human language use takes place.

Formal models, on the other hand, are usually *descriptive*, attempting to mirror real-world processes in a way that elucidates some theoretically important aspect of them. Computational models *can* be descriptive, but more often they are primarily designed to be *predictive*. If a computational model is good at predicting the outputs of a certain real-world process from its inputs, we might conclude that the real world process resembles the mechanism of the model in certain ways (for example, we might conclude something about the computational complexity of the real-world process or something more detailed by *probing* the model for relations between its internal mechanisms). Alternatively, we might just be interested in the practical applications of a computational model. If certain kind of model performs well enough over the long-term to be useful in applications, this might provide a different kind of evidence that it "gets something right" about how the real-world process works.

As the goals of the thesis would suggest, the models we use have a variety of different methodological roles to play. A useful question to ask is what *level of description* the model is targeting. The neural machine learning models described in Section 5.2.1 are *inspired by* biological neural networks, but no one would claim that neural language models process linguistic input in the same way humans do at the level of the neuron. If these models do mirror human language processing — and there is at least some evidence that they do, to some degree (Bhattasali & Resnik, 2021) — then the relationship between material parts of the brain and parametrized functions in the model is much more abstract than a one-to-one mapping between the two. Formal models mostly don't attempt to model psychological processes at all (although there are exceptions). Rather formal models tend to focus on structural phenomena that emerge from psychological processes and the causal factors (including mathematical, and log-

ical factors) that govern them.[1]

Formal modeling can have a symbiotic relationship with more empirical approaches. Formal models provide a language to pose hypotheses and can provide inspiration structuring computational and statistical models. Computational models, on the other hand, can provide a way to test hypotheses on large amounts of real-world data. This thesis has a small role to play in that relationship, but it is a great pleasure (and sometimes even scientifically productive) to play on both sides of it.

5.1. Formal methods

Understood broadly, formal methods are ways of making a theory precise, usually by the use of an abstract symbolic system borrowed from logic or mathematics. This is often what is meant to *formalize* a theory — the theory is translated into logic or math, which aids in precisely formulating (and often generating) hypotheses that can be tested empirically. Under this characterization, formal methods have long been employed in linguistics, though exactly what is meant by *formal* can vary by sub-discipline and across research traditions.

This section describes the formal methods adopted in the thesis. Not all of the work in this thesis uses formalization, but Chapters 7 to 9 and 11 all include at least some formalization in Type Theory with Records (TTR) and it's probabilistic counterpart ProbTTR.

To give context and motivate this choice of system, this chapter includes a broader introduction to the methodology of the formal semantics tradition sometimes termed *logical grammar*. Montague semantics, named after mathematician-turned-linguist Richard Montague. Montague's work on natural language semantics (1970, 1973), is certainly the most influential progenitor of the logical grammar approach. His paper *English as a formal language* 1970 defied the conventional wisdom that semantics was not a candidate for formalization. As Barbara Partee (1973) described the situation,

> Logicians seem to have felt that natural languages were too unsystematic, too full of vagueness and ambiguity, to be amenable to their rigorous methods, or if susceptible to formal treatment, only at great cost. Linguists on the other hand, emphasize their own concern for psychological reality, and the logicians' lack of it, in eschewing the logicians' approach. (p. 509 B. Partee, 1973)

These tensions are still present in semantics today, often manifesting as disagreement about what work counts as *formal*, what counts *natural* language, and what the goals of formalization in semantics should be. The overall trend, however, is to provide formal descriptions of more and more of those "unsystematic" features that motivated

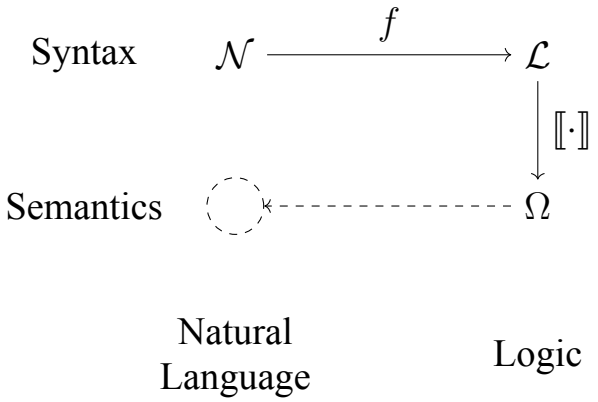
[1] See B. H. Partee (1979) and Teichman (n.d.) for further discussion.

the initial skepticism of logicians. This often means reaching for logical systems with more expressive power than what was used in Montague's initial work.

So what does it mean to treat a natural language *as a formal language*? And how does this result in a theory of meaning? A formal language is a mathematical object. It starts with a set of symbols, Σ. These symbols might be words or letters or even audio symbols like phonemes — anything at all really, as long as they can be reproduced, distinguished from one another, and put into sequence. A finite sequence of symbols from Σ (possibly including repeats) is called a Σ-**string**. A **formal language**, $\mathcal{L}$ *over* Σ is a particular subset of all the possible Σ-strings. A formal language is usually defined in terms of a rules that describe admissible strings — that is, strings that are part of $\mathcal{L}$.

Logics have a *syntax* and a *semantics*. The syntax is given by a formal language. The semantics lies in some other "realm" Ω, which is often *set theoretic* — made up of mathematical sets, including especially functions. A **logic** connects a syntax and semantics by means of an interpretation function, $[\![\cdot]\!] : \mathcal{L} \to \Omega$.

The final piece of the puzzle is a translation function, $f : \mathcal{N} \to \mathcal{L}$ which maps expressions in the natural language to strings of the formal language.[2] In practice, $\mathcal{N}$ is a **fragment** of some natural language, often defined in terms of a lexicon of words of various syntactic types and a set of rules for forming expressions. This means that elements of $\mathcal{N}$ are not just strings, but come along with a syntactic structure that can be used in the definition of f. Together, these three collections of objects and two functions that map between them give us the following (albeit simplistic) schema of a model theoretic logical grammar.



A potentially misleading aspect of this schema is that while there is a vertical divide between the syntactic and semantic of both logic and natural language, it is really the relationship between *all three* of these realms ($\mathcal{N}$, $\mathcal{L}$, and Ω) that comprises the semantic theory. By composing f and $[\![\cdot]\!]$, the theory proposes a mapping from the

[2] One complication to this picture worth noting already is that f is not always, strictly speaking, a function, since many theories account for *semantic ambiguities* — situations where a single syntactic form can have multiple semantic interpretations. This is the usual analysis given to sentences like *everyone pointed at someone*, where the two quantifiers create a *scope ambiguity*; is there one specific person who everyone pointed at or did everyone point at various (possibly different) people? Although the grammatical structure is the same, this sentence can be translated to to different predicate logic formulas which capture the two different readings.

meaning of natural language expressions to the semantic space of the logic. This means that f is just as much a part of the semantic theory as $[\![\cdot]\!]$; in practice, it is often the more consequential part.[3] Putting these mappings together, given a natural language expression $e \in \mathcal{N}$, there is an object $o \in \Omega$ such that $[\![f(e)]\!] = o$. The semanticist who proposes this theory asserts some modeling relationship (the dashed line) between the o and the *actual meaning* of e.

To make all of this a little more explicit let's consider a very simple example of a formal semantic theory for a fragment of English including the word *and*.

$\mathcal{N}$	$\mathcal{L}$
• *John is pointing* $\in \mathcal{N}$	• $p \in \mathcal{L}$
• *Kim is pointing* $\in \mathcal{N}$	• $q \in \mathcal{L}$
• $s_1, s_2 \in \mathcal{N} \Rightarrow [s_1 \textit{ and } s_2] \in \mathcal{N}$	• $\varphi, \psi \in \mathcal{L} \Rightarrow (\varphi \wedge \psi) \in \mathcal{L}$

f	$[\![\cdot]\!]$
• $f(\textit{John is pointing}) = p$	• $[\![p]\!] = 1$
• $f(\textit{Kim is pointing}) = p$	• $[\![q]\!] = 0$
• $f([s_1 \textit{ and } s_2]) = (f(s_1) \wedge f(s_2))$	• $[\![(\varphi \wedge \psi)]\!] = [\![\varphi]\!] \times [\![\psi]\!]$

In this theory, the objects in Ω (the semantic realm) are the boolean values 1 and 0, and functions from boolean values to boolean values. In particular, *and* is interpreted as boolean multiplication — a binary function that gives and 1 if both the arguments are 1 and 0 if one or both of them is 0.[4] Typically an analysis such as this one is presented as a **truth conditional** semantic theory, meaning that these boolean values are taken to correspond to the concepts of *truth* and *falsity*. In fact, the symbols T and F are often used for these values to emphasize that relationship, but here we use 1 and 0 to make the point that Ω is a realm of formal objects with only a meta-theoretical relation to the concepts of truth and falsity.

Notice that $\mathcal{N}$ is defined in such a way that expressions are endowed with syntactic structure, here conveyed with square brackets:

(13) a. [*John is pointing and Kim is pointing*]

b. [*Kim is pointing and John is pointing*]

c. [[*John is pointing and Kim is pointing*] *and John is pointing*]

[3] Indeed, some semantic theories see fit to do away with Ω all together, instead allowing a syntactically defined notion of logical consequence (i.e., a proof system) give content to the natural language expressions.

[4] In this presentation, *and* is interpreted *syncategorematically* since we don't give it a denotation directly, but rather provide a rule for interpreting expressions involving *and*. The categorematic approach would interpret the word *and* as the multiplication operation directly. In this case, the two approaches are essentially equivalent.

Each of the sentences of (Section 5.1) are part of $\mathcal{N}$. The brackets make it easy to define the translation function, f, which maps (Section 5.1) to $(p \wedge q)$, (Section 5.1) to $(q \wedge p)$, and (Section 5.1) to $((p \wedge q) \wedge p)$. Importantly, the semantics of the logic gives the same interpretation for all three of these sentences, due to the commutativity and associativity of boolean multiplication. This is an example of a prediction that makes the formal semantic theory falsifiable by empirical data like speaker judgments — if speakers don't think these three sentences have the same meaning, there might be a problem with the theory.[5]

Naturally this simplistic theory has quite a few shortcomings as a formal semantics. First of all, it fails to capture the compositional semantics internal to the expressions that are translated to p and q. What if we want to talk about *other* people pointing? Do we really need to enumerate every such sentence? Surely that's not analogous to how it works in the natural language — if we know who *Lisa* is, we can understand *Lisa is pointing* by analogy, even if we've never heard that exact sentence before.[6] Perhaps even more damning (insofar as this formalization is supposed to give a theory of the meaning of the word *and*), we can't seem to account for sentences like these:

(14) a. [[*John and Kim*] *are pointing*]

b. [*John is* [*pointing and laughing*]]

How does the word *and* behave when it is not joining propositions but expressions denoting entities or actions? Our theory has nothing to say here since there sentences aren't part of our fragment of English. We might expect a more complete theory to satisfy certain intuitions, such as that (Section 5.1) is given the same interpretation as (Section 5.1), or that (Section 5.1) entails *John is pointing*. Such an analysis would require a more expressive system than propositional logic, of which $\mathcal{L}$ and $[\![\cdot]\!]$ are a subset. More generally, a satisfactory theory ought to cover a more complete fragment of English.

Another problem is that this theory equates the meaning of an expression with its truth value. This means that *John is pointing* and [*Kim is pointing and John is pointing*] have the same meaning, 0 (or *False*). This is seems like an odd conclusion since intuitively each of these sentences have a different meaning, even if they both *happen* to be false. After all, isn't it possible for someone to *believe* the first sentence while *not believing* the second? If I utter one of these two sentences am I not conveying different *information*? This is what is known as the problem of **intensionality**. A good semantic theory should give interpretations that go beyond the contingent state of the world,

[5] There is some wiggle room for the theorist here if f and $f \circ [\![\cdot]\!]$ are taken to model two different kinds of semantic interpretation — note that each of the sentences in (Section 5.1) are translated to different propositional formulas by f even though the interpretation is the same.

[6] Implicit in this criticism is a particular modeling goal for the semantic theory — the meaning of a compositional expression should be computed in a way that is (at some level of description) analogous to the way that actual speakers compute that meaning.

since natural language functions in so-called *intensional contexts* such as when speaking hypothetically or talking about belief states. The success of Montague Semantics is due in part to the fact that its formal analysis extended to a very large fragment of English, including to many expressions that require dealing with intentionality.

Montague Semantics Montague's formal semantics (1970, 1973) uses a categorical grammar to capture a fragment of English and a combination of simply typed lambda calculus and higher order predicate logic for the semantics. Aside from giving a fully compositional treatment of a fragment of English, (*a/an*, *the*, *every*, *some*, etc.), Montague (1973) was principally interested in giving a compositional analysis of intensional contexts and quantified noun phrases (and especially quantified noun phrases *in* intensional contexts).[7]

In an **intensional context**, you cannot replace a constituent with a co-extensive expression without changing the meaning.

(15) a. Kim seeks the President of the United States

b. John is the President of the United States

c. ⁄∴Kim seeks John.

Here, we can tell that 15a includes an intensional context. Since substituting co-extensive terms (*John* for *the President of the United States* is not truth-preserving; it is possible for 15a to be true and 15c to be false if, f or example, Kim does not know that 15b is the case. Compare this, to the following:

(16) a. Kim points at the President of the United States

b. John is the President of the United States

c. ∴Kim points at John.

It would seem that this inference *does* go through. Note that the only difference between 15 and 16 is the verb. *Seeks* creates an intensional context where as *points* usually[8]does not. For this reason, Montague gives an analysis of intensionality where intensional contexts are created by particular lexical items.

Montague Semantics uses a context-free categorical grammar for the syntax of English and intensional logic (a combination of simply typed lambda calculus and model-theoretic higher-order predicate logic) for the semantics. The grammar defines the possible syntactic categories of expressions, CAT as follows:

[7]More detailed introductions to Montague Semantics can be found in Dowty et al. (1981) and Gamut (1991).

[8]There certainly are senses of point that do create an intensional context. Consider, for example, the sense where *point at* is used metaphorically to mean *to make an accusation*. Depending on the situation there might even be contexts where even literal pointing needs to be interpreted intensionally if, for example, the pointing is serving the purpose of making an accusation. This is just another example of how lexical meaning is deeply dependent on communicative context.

1. $CN, IV, S \in CAT$
2. If $A, B \in CAT$ then $A/B \in CAT$

The basic categories correspond to common nouns, intransitive verbs, and sentences. The slash categories can be thought of as something that gives you an A if you provide it with a B (this is specified more formally in the semantics). For example, *individual terms* like *John* are not a basic category, but are instead represented as S/IV — something that given an intransitive verb, will give you a sentence.[9]

Each syntactic category corresponds to a semantic type, which can be described as a formal language:

1. e is a type and t is a type
2. if σ and τ are types, then $(\sigma \to \tau)$ is type
3. if σ is a type, then $(s \to \sigma)$ is a type

The basic type e corresponds to entities and t corresponds to truth values. Combining two types forms a higher-order "function" type. The third "basic" type s (corresponding to possible worlds), is unlike the other two in that it cannot appear alone, but only as the antecedent to higher order types. Function types with s as the antecedent are what introduces intensionality into the system.

The correspondence between syntactic categories and semantic types is defined as follows:

1. $f(CN) = (e \to t)$
2. $f(IV) = (e \to t)$
3. $f(s) = t$
4. $f(A/B) = ((s \to f(B)) \to f(A)$

Finally, a lexicon assigns a grammatical category and an intensional logic formula (of the appropriate type) to each English lexical item. Intensional logic uses Kripke models (cite) for its semantics. Kripke models are set-theoretic constructions that include functions from possible worlds (corresponding to the semantic types of the form $(s \to \sigma)$, which allow the system to account for intensional phenomena.

It's not an exaggeration to say that contemporary formal semantics largely defined by variations and extensions of Montague Grammar. Some of these extensions use

[9] In fact, Montague's original presentation was a bit different, taking individual terms as a basic category and common nouns and intransitive verbs as derivative. Bennett (1976) came up with this version, which simplifies the grammar somewhat.

more powerful syntactic formalisms that can deal with phenomena like discontinuous constituents. Another line of work extends formal analysis to more semantic phenomena by using *rich type theories* in the semantics. We will briefly introduce this trend more generally before giving and overview of TTR (Section 5.1.1).

Type theoretic semantics Since Montague, the field of formal semantics has grown rapidly, extending formal analysis to fragments of to cover more and more semantic phenomena (including, importantly, phenomena that do not occur in English). A subfield of formal semantics has focused on applying new methods in type theory, a field which has independently seen a flourishing in recent decades, with applications in programming language theory and foundational mathematics as well as linguistics.[10] These type theories differ in character in a number of ways from the simply typed lambda calculus. For one, they are often **many-sorted**, meaning that the basic types are not a closed class as they are in Montague's intensional logic (limited to e, t and s), but are formally more akin to propositions in propositional logic or predicates in predicate logic in that the system could, in principle, include any number of them, depending on the lexicon in the fragment of natural language being modeled.

A many-sorted type theory might, for example, might have a type *Man* corresponding to the noun *man* and a word *Point* corresponding to the verb *to point*. We would write

$$j : \mathit{Point} \tag{5.1}$$

to mean that the object j (corresponding to *John*) is pointing — that is, John is the type of thing that is pointing. If the type judgment expressed by (Eq. (5.1)) holds, we would say that j is a **witness** for the type *Point*.

Another feature that makes rich type theories attractive for formal semantics is the **types as propositions** interpretation of types. Under this interpretation, a type stands for a proposition; namely the proposition that the type has a witness. In this interpretation *Point* would stand for the proposition that *someone is pointing* and $j : \mathit{Point}$ would constitute a proof of that proposition with j as the witness. For systems based on intensional logic, logically equivalent propositions are indistinguishable. But in **hyperintensional** contexts such as belief, it may be necessary for a formal semantics to distinguish between them.

Types as propositions has a convenient relationship with the Austinian notion of truth (Austin, 1950) in which propositions are not fundamentally true *simpliciter*, but rather truth *of* some part of the world. Barwise and Perry (1983) expand on this notion by developing a theory of *parts of the world*, which they call **situations** and types of situations, which correspond to propositions. Just as a proof may be a witness for a proposition, a situation may be a witness for a situation type. Cooper (2005) formalizes this relationship in Type Theory with Records (TTR), which we will present briefly in the next section.

[10] See Chatzikyriakidis and Cooper (2018) for an overview of type theory for natural language semantics.

5.1.1. Type Theory with Records

Type Theory with Records extends many-sorted dependent type theory with structured objects called **records**, defined as labeled sets of objects (including possibly records):[11]

$$r = \begin{bmatrix} \mathrm{k}_1 & = & \mathrm{a}_1 \\ \vdots & = & \vdots \\ \mathrm{k}_n & = & \mathrm{a}_n \end{bmatrix} \tag{5.2}$$

and corresponding structured types called **record types**, which are labeled sets of types (including possibly record types):

$$T = \begin{bmatrix} \mathrm{l}_1 & = & \mathrm{T}_1 \\ \vdots & = & \vdots \\ \mathrm{l}_m & = & \mathrm{T}_m \end{bmatrix} \tag{5.3}$$

Here, $\{k_1, ..., k_m\}$ and $\{l_1, ..., l_n\}$ are sets of *labels*, drawn from a special set of symbols reserved for labeling records and record types. We write $r.k$ to refer to the object r with corresponding to the label k. The record r is of type T (written $r : T$) just in case for each l_i, there is some k_j such that $r.k_j : T.l_i$.[12]

Type Theory with Records (TTR) is a logical system like the simply typed lambda calculus — on its own, it doesn't offer any theory of natural language semantics as such. However, formal semantic theories that use TTR do tend to have certain aspirations in common, which are supported by the expressive features of TTR. For one, these theories usually try to persevere much of the compositional analysis afforded by Montague semantics. This is made possible by the fact that TTR is a dependent type theory, meaning that it has all of the expressive power of the simply typed lambda calculus. TTR theories are often oriented towards going beyond sentence-based theories of meaning, focusing instead on interaction a starting point for natural language semantics. Modeling action (and the change that results from action) is core to these theories and type theory with records is particularly well-suited to model that kind of dynamics (Cooper, 2012).

This focus on action and interaction also gives TTR-based theories an agent-oriented outlook. Types and type judgments are often taken to be relative to a particular agent, whose *information state* is modeled with a record. Actions (for example an utterance of their own or by another agent) then result in updates to this information state, in what is called the **information state update** (ISU) approach to semantics (Larsson, 2002; Traum & Larsson, 2003). Related to this is the **dialogue game board**, which models the public component of dialogue participans' information state. Dialogue game board theories of dialogue seek to understand the structure of the common ground that is built

[11] A full formal description of TTR can be found in Cooper (2023). Cooper and Ginzburg (2015) also gives a brief introduction to TTR and describes a range of applications in semantics and dialogue.

[12] Technically the type judgment definition for record types also allows *re-labellings*, but we will ignore that detail in this presentation.

up during dialogue, and how speakers make use of it to facilitate communication. Insofar this thesis is interested in how interaction affects common-ground lexical semantic resources, is is important that we can connect the work in the thesis to a dialogue game board account (KoS is one such theory that uses TTR (Ginzburg, 2012)).

TTR-based semantics usually takes an Austinian notion of truth in which, a situation (or alternatively, an agent's *take on* a situation) is modeled by a record and propositions are modeled by a record type. Consider a type judgment like the following:[13]

$$\begin{bmatrix} \mathrm{x} & = & \mathrm{jack} \\ \mathrm{y} & = & \mathrm{helen} \\ \mathrm{c}_1 & = & \mathrm{s}_1 \\ \mathrm{c}_2 & = & \mathrm{s}_2 \end{bmatrix} : \begin{bmatrix} \mathrm{x} = \mathrm{jack} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{c}_1 & : & \mathrm{PointAt}(x, y) \end{bmatrix}$$

Here, the situation (on the left) is judged to be of the type of situation where Jack is pointing at someone. This implies that $\mathrm{s}_1 : \mathrm{PointAt}(\mathrm{jack}, \mathrm{a})$ for some individual a. The object s_1 can be thought of as a part or aspect of a situation. Note that if it's the case that $\mathrm{s}_2 : \mathrm{PointAt}(\mathrm{helen}, \mathrm{jack})$ then the record is *also* a situation of the type where Helen is pointing at Jack. The definition for record type judgments mirrors the intuition that situation types (infons in Barwise and Perry (1983)'s terminology) can involve some underspecification. This allows situation types to model underspecification in a word's lexical meaning.

The type of situation in which Jack points at Helen is a **subtype** of the type of situation where they are both pointing at each other, since something of the second type is always also of the first type:

$$\begin{bmatrix} \mathrm{x} = \mathrm{jack} & : & \mathit{Ind} \\ \mathrm{y} = \mathrm{helen} & : & \mathit{Ind} \\ \mathrm{c}_1 & : & \mathrm{PointAt}(x, y) \\ \mathrm{c}_2 & : & \mathrm{PointAt}(y, x) \end{bmatrix} \sqsubseteq \begin{bmatrix} \mathrm{x} = \mathrm{jack} & : & \mathit{Ind} \\ \mathrm{y} = \mathrm{helen} & : & \mathit{Ind} \\ \mathrm{c}_1 & : & \mathrm{PointAt}(x, y) \end{bmatrix}$$

This subtype relation can be verified by examining the structure of the two record types — for every label on the right-hand side there is a label on the left-hand side corresponding to a type that is either equal to– or a subtype of– (but in this case always equal to) the type on the right. This subtype relation also holds:

$$\begin{bmatrix} \mathrm{x} = \mathrm{jack} & : & \mathit{Ind} \\ \mathrm{y} = \mathrm{helen} & : & \mathit{Ind} \\ \mathrm{c}_1 & : & \mathrm{PointAt}(x, y) \end{bmatrix} \sqsubseteq \begin{bmatrix} \mathrm{x} = \mathrm{jack} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{c}_1 & : & \mathrm{PointAt}(x, y) \end{bmatrix}$$

To see that the type situation where Jack points at Helen is a subtype of the type of situation where Jack points at *someone*, we need to know that $T_{\mathrm{helen}} \sqsubseteq Ind$, which is true by the definition singleton types (see Footnote 13).

[13] The notation $\mathrm{x} = \mathrm{jack} : \mathit{Ind}$ is a *manifest field* (Coquand et al., 2003), which is shorthand for $x : \mathit{Ind}$ and $x : \mathit{Ind}_{\mathrm{jack}}$, where $\mathit{Ind}_{\mathrm{jack}}$ is a *singleton type*. In general, for any object $a : T$, $b : T_a$ if and only if $b : T$ and $b = a$.

5.1.2. Probabilistic Type Theory with Records

Probability theory provides a mathematical formalization of uncertainty. Insofar as natural language interpretation deals with uncertainty, probabilistic concepts can be useful in formal semantics. Probabilistic Type Theory with Records (Cooper et al., 2015) adapts TTR to the probabilistic setting by definition a probabilistic type judgment:

$$p(\mathrm{a} : T) = r,$$

there, r is a real number between 0 and 1. We can read this as saying that the probability that a is of type T is r. The value of p is given by a probability model.[14] Conditional type judgements can be expressed similarly:

$$p(a : T_2 \mid a : T_1) = r$$

In the probabilistic setting there are multiple candidate notions for subtype, but a minimal requirement for $T_1 \sqsubseteq T_2$ would be that whatever something is certainly of T_1 it is certainly of type T_2.[15] That is,

$$T_1 \sqsubseteq T_2 \Rightarrow p(a : T_2 \mid a : T_1) = 1.$$

5.1.3. Classifier-based meaning

In order to ground perceptual meaning in classification in TTR, we need to do two things: (1) we need to give an account of how, given a classification function, the semantics of the TTR types it is based on are determined, and (2), we need to encode the classifier in TTR in such a way that makes the classification function available. Another approach would be to forego (2) and instead use the classifier as a witness condition for some type corresponding to the meaning (in the place of a set theoretic model, for example). The problem with this approach is that if we want linguistic activity to serve as a basis for semantic learning, we need to make the parameters of the classifier, not just the classification function, available to our theory of interaction, which is stated in TTR (see Fernandez & Larsson, 2014; Larsson & Cooper, 2021).

[14] Cooper et al. (2015) first defines the model as a probability function over a set of possible worlds, following van Eijck and Lappin (2012). Doing so guarantees adherence to the standard Kolmogorov (1950) probability axioms, but at the cost of completeness and cognitive plausibility. They suggest that probabilities might alternatively be assigned to situation types, as this is analogous to the assumption that is commonly made in probabilistic AI in which the universe of *worlds* is not made up of maximally consistent sets of propositions, but rather a local set of alternative possible outcomes (Cooper & Ginzburg, 2015, §1.2). Another approach might be break from classical probability theory and use a model theory that assigns probability to type judgments directly in the style of de Finetti (see de Finetti, 1992). Indeed, this is essentially what we do in Chapter 8 when we use classifiers as witness conditions for certain types. More work is needed to ensure that this approach would yield a well-behaved probabilistic type system in the general case, however.

[15] A stricter requirement, for example might be that $p(a : T_1) \leq p(a : T_2)$ in every possible interpretation.

Larsson (2013) demonstrates how to encode a linear perception in TTR, using the example of providing grounded semantics for the terms *left* and *right*. In Chapter 8, we expand this treatment to multiclass classifiers. To do so, we define a categorical variable type $\mathbb{A}$, which ranges over a set of value types $\mathfrak{R}(\mathbb{A}) = (A_1, ..., A_n)$. Where each A_i is a record type.[16] A classifier $\kappa_{\mathbb{A}}$, for $\mathbb{A}$ is a function of the following type:

$$\Pi \rightarrow Sit_{\mathfrak{Y}} \rightarrow \{ \begin{bmatrix} \text{sit} & : & Sit_{\mathfrak{Y}} \\ \text{sit-type} & : & RecType_{A_i} \\ \text{prob} & : & [0, 1] \end{bmatrix} \mid A_i \in \mathfrak{R}(\mathbb{A}) \}.$$

Here, Π is the type of the parameters needed by by the classifier, $Sit_{\mathfrak{Y}}$ is the type of situations that yield perceptual input, and $RecType_{A_i}$ is the (singleton) type of records identical to A_i. We assume that for parameters $\pi : \Pi$ and input $x : Sit_{\mathfrak{Y}}$, we have $\sum_i \kappa_{\mathbb{A}}(\pi)(x)(A_i) = 1$.[17]

5.2. Computational methods

Formal methods seek to precisely state a theory of how meaning works in natural language. As we have seen in Section 5.1, the scope of such a theories has been extended (or perhaps shifted) in some cases to focus not only sentence meaning, but the meaning of a wide range of types of utterances situated in an interactive context. Computational semantics starts with a similar goal, to understand meaning in natural language. The methodology in computational semantics — its relationship it's relationship to *models*, *data*, and *hypotheses* tends to be much different, however.

Data and preprocessing Humans receive linguistic input as combination of audio and visual signals,[18] most typically perceived in the course of an interaction in which we ourselves take part. Most psycholinguists would agree that processing speech signals involves some degree of discretization by way of classifying the continuous input (sounds into phonemes, strings of phonemes into words, etc.). Nevertheless, the continuous signal remains available, for example, for use in communicative repair.

In computational linguistics, we rarely work with the raw speech signal,[19] applying some discretization before we even start modeling. For spoken data, this means working with a transcript, which is the result of a laborious human transcription process or a noisy speech-to-text system. Transcription comes with a lot of choices about what

[16] In Chapter 9 these correspond to the meaning of lexical item.

[17] In Chapter 9 this is ensured by the standard standard softmax function used in neural multiclass classification models.

[18] Why visual signals? In addition to the many signed languages of the world, gesture serves an important communicative function in-person spoken and signed dialogue. In the following *speech*, is used to refer to both verbal and gestural communication spoken and signed interaction.

[19] There are exceptions — computational phonology, for example. But this work generally stays on the level of phonology. It is rare for raw speech signal to be used as input to computational models that work further up the classical linguistic hierarchy.

aspects of the speech and how much of the interaction to capture. Speech-to-text processing is noisy and error-prone and captures an extremely impoverished record of the speech, especially in interactive settings.

For these reasons it is much more common to work with text data in computational linguistics. Digital text is already discretized into characters and hand-written text can be converted to digital text through automatic character recognition which, though not without errors, captures a more faithful representation of the original data than speech-to-text.

This is not where preprocessing ends, however. **Tokenization** is a key preprocessing step for most work in computational linguistics. Tokenization divides a sequence of characters into multi-character strings from a finite vocabulary. Often these tokens are meant to correspond to something like words or lexical items, though there is probably no way to do this perfectly in principle since, as we discussed in Section 2.1, there is no clean separation between lexical and compositional meaning. Sub-word tokenization strategies are also popular. They divide text into hopefully meaningful sub-word units, either by employing some morphological analysis, or with strategies that group together common sequences of characters. Once tokenized, text is represented as a sequence of tokens, each of which is drawn from a finite vocabulary of token *types*. One could, for example count up the tokens in a piece of text and compute a distribution of token types over the vocabulary. This sort of thing is the basis for modeling in computational linguistics, including for machine learning models.

Machine learning The *machine learning paradigm* is pervasive in computational semantics and computational linguistics more generally. In this paradigm, an abstract **task** is defined, which seeks to approximate some human-like competency involving language use. *Sentiment analysis* (judging if a piece of text expresses positive or negative sentiment), *natural language inference* (determining if a premise sentence entails a hypothesis, if they are contradictory, or if there is no relation), and *image captioning* are all examples of machine learning tasks that involve natural language semantics.

A **dataset** is the concrete manifestation of a task. A dataset consists of a set of pairs $D = \{\langle x, y, \rangle \mid x \in X, y \in Y\}$, where each x is some input (a sentence, a pair of sentences, or an image for example, respective to the above) and each y is a ground-truth *labels* (a sentiment score, entailment relation, or image caption), usually produced by a human annotator. A **model** is a function, $\varphi(\theta, x)$, that given some parameters, θ and an input, produces something of the same kind as the elements of Y.

Standard practice is to split into disjoint *train* and *test* sets.[20] A *loss function* is defined such that $\mathcal{L}(\hat{y}, y)$ measures the distance between a model prediction and ground-truth label. Then, a learning algorithm is used **train** the model — to find the parameters that minimize the loss over the training set; that is, to find:

[20] And often also a *validation* set, which is used to select hyperparameters, such as model size, and to check during training if the model is *overfit* to the train set.

$$\hat{\theta} = \arg\min_{\theta \in \Omega} \sum_{\langle x,y \rangle \in D_{\text{train}}} \mathcal{L}(\varphi(\theta, x), y), \tag{5.4}$$

where Ω is the space of all possible values of θ.

Generally speaking Ω can be extraordinarily high-dimensional. Together with a complex φ, this means that an analytic solution to Eq. (5.4) often doesn't exist or is completely intractable — you can't just *solve for* $\hat{\theta}$ as you would in algebra class. The core of the discipline of machine learning is to define a model, φ, loss function $\mathcal{L}$ and an algorithm for estimating $\hat{\theta}$ from D_{train} such that $\varphi(\hat{\theta}, x)$ has good *performance* on the test set according to one or more **performance metrics**, which measure how well the model approximates the ground truth output of the test set.[21] If the dataset is a faithful realization of the task, performance on the test set will indicate how well the model *generalizes* beyond the training data — that is, how good it is at performing the task *in general* without respect to the particular examples it "saw" during training.

Scientific knowledge is not always the goal of machine learning — a model that performs well on a particular task can have useful real-world applications. But it's worthwhile to consider how computational methods differ from formal methods when the aim is to discover something about natural language as an empirical phenomenon. What does it mean if a model performs well on a particular task? This depends somewhat on the dataset, but if the model exhibits non-trivial generalization to the test set, that means it has learned some patterns that connect the input and the output. If we take natural language inference (NLI) as an example, we could see a machine learning model that performs well to be a model of inferential meaning in natural language in the same way that formal models with the same goal are. Such a model could be taken as evidence that (1) the training data is sufficient to learn how inference works in general and (2) the model architecture has sufficient computational power to determine entailment relations, as well as to learn *how* to determine them from the training data. In practice, there are reasons to be skeptical of claims that NLI models, even ones that perform very well, really capture inferential meaning the way that human semantic interpretation does. For one thing, the NLI datasets available capture a certain only a certain *kind* of inference which is a bit different from what is assumed by formal theories and doesn't generalize to all contexts (Bernardy & Chatzikyriakidis, 2019). Furthermore, some models can perform almost as well when they are trained using the premise sentence alone, suggesting that the model is taking a "shortcut", using certain correlations between the premise sentences and the entailment relation without considering the hypothesis sentence at all (Gururangan et al., 2018).

[21] The loss function is often different from the performance metrics are used in testing. This is counter-intuitive; you might think that maximizing the same performance metrics during training would be the best way to maximize them during testing. However, the loss function must be chosen carefully in conjunction with the learning algorithm. For example, most strategies for training neural networks (Section 5.2.1) require that a gradient of the loss function can be computed with respect to the model's current parameters. Doing so requires a differentiable loss function, which is not generally the case for performance metrics.

Language modeling is an especially important task that has a complicated relationship with both practical applications and linguistic theory. In the strict sense, a **language model** is a function that estimates a probability distribution over strings of tokens. This is usually done by training the model to perform *next token prediction*, since a model that computes the probability of a token given its preceding context can be used to compute the probability of the string:

$$P(w_1, ...w_m) = \prod_{i=1}^{m} P(w_i \mid w_1, ..., w_{i-1})$$

Of course, it is also impossible to compute the conditional probability on the right, since in principle the context string, $w_1, ..., w_{i-1}$, could be anything. A language model must estimate this probability, for example by substituting it for $P(w_i \mid w_{i-n}, ..., w_{i-1})$ for a small value of n, as in a *n-gram* model.

One especially salient feature of language modeling as a task is the fact that it doesn't require any annotation. Since the training objective is to predict the next word in a sequence, tokenized text can serve as its own labelled data. This makes language modeling a **self-supervised** learning paradigm.

Since language models estimate a probability distribution over possible strings, they can be seen as the statistical corollary to a formal language (Section 5.1) It has been argued (Lau et al., 2017) that this means they can be interpreted as capturing grammatical competence in a particular language. But they are also enormously useful in downstream tasks. For example, in machine translation, a language model can serve as a prior distribution of strings in the target language, meaning that the translation model need only estimate a probability of source language strings *given* a string in the target language. This is what is known as a *noisy channel* translation model. The intermediary representations learned by a language model can also be useful. This is especially relevant for neural language models, as we will discuss in the next section.

5.2.1. Neural network models

A neural network is a kind of machine learning model inspired by the way synaptic signals pass through biological brains. Most neural network models are arranged in layers, with the output of the previous layer supplying the input for the next layer. A network with multiple layers is called **deep**. Intermediary layers are called **hidden layers** and their outputs are **hidden states**.

Each layer consists of a collection of "neurons" (the gray circles in Fig. 5.1), each of which compute a real-number value based on the outputs of the neurons of the previous layer and some trainable parameters. An *activation function* is usually applied as a final step in computing the neuron's output value. Activation functions are often non-linear, amplifying the output value if it reaches a certain (soft) threshold. The activation function is supposed to mimic the behaviour of biological neurons whose

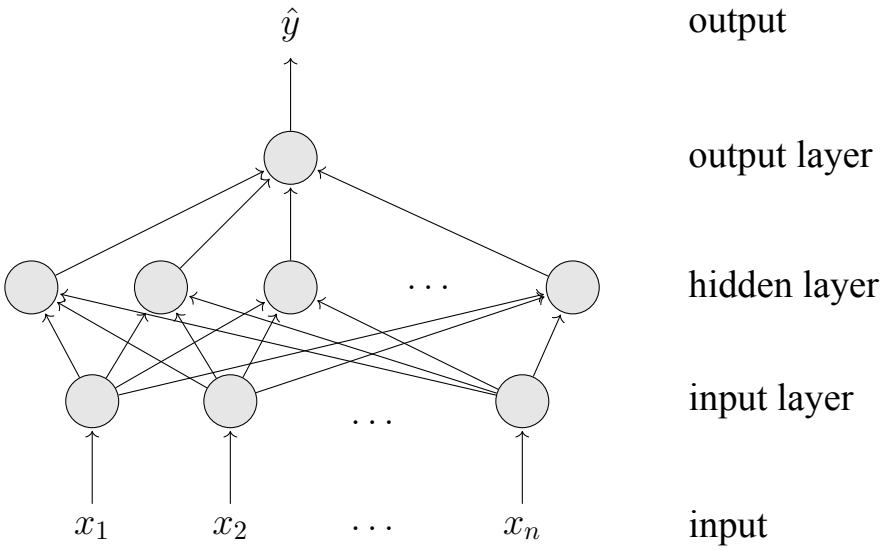


Figure 5.1.: Simple neural network with 3 fully-connected layers and a one-dimensional output.

synapses *fire* given a certain amount of stimulus. In practice, the non-linearity of the activation function is what lets deep neural networks learn functions where the desired output can not be computed as a linear combination of the input.

Figure 5.1 may look fancy, but in fact it is just a series of matrix multiplications with a bias term added:

$$\begin{aligned} \hat{y} &= \sigma_2(\mathbf{W_2} \cdot \mathbf{h_2} + \mathbf{b_2}), && \text{where} \\ \mathbf{h_2} &= \sigma_1(\mathbf{W_1} \cdot \mathbf{h_1} + \mathbf{b_1}) && \text{and} \\ \mathbf{h_1} &= \sigma_0(\mathbf{W_0} \cdot \mathbf{x} + \mathbf{b_0}). \end{aligned} \tag{5.5}$$

Here, for example, the ith neuron of the hidden layer is parametrized by the ith row of $\mathbf{W_1}$ and the ith element of $\mathbf{b_1}$, and σ_1 is the activation function for that layer.

Different neural network architectures have different patterns of connections between neurons. Convolutional layers, for example, compute outputs based on a sliding context window over the input. Recurrent layers are made up of a sequence of *recurrent units*, where each item in the sequential output is a function of the previous item and a (possibly) sequential input.

Modern neural networks are commonly trained with the aid of *back-propagation*, an algorithm that estimates the gradient of the model's parameters with respect to the loss (Rumelhart et al., 1986) This gradient is used by optimization functions like *gradient descent* to incrementally adjust the parameters to minimize the loss by moving the parameters in the direction indicated by the gradient. The *learning rate* controls how big of a step in the direction of the gradient the optimizer takes. Some optimizers also include a momentum hyperparameter, which biases the change in parameters to

continue on in the direction it went in previous steps. The Adam optimizer (Kingma et al., 2015), which is used to train the neural networks used in Chapters 10 and 12, is such an optimizer.

Neural networks dealing with text data usually make use of an **embedding** layer, which converts tokens into vectors of a particular dimensionality or *size*, n. It's called an embedding because, given the right learning objective, it *embeds* the vocabulary in the $\mathbb{R}^n$ vector space. More precisely, given a vocabulary of tokens V, and a embedding size n, a embedding layer is a function: $Emb : \mathbf{W} \rightarrow (V \rightarrow \mathbb{R}^n)$, with parameter matrix $\mathbf{W} \in |V| \times n$ and where

$$Emb(v_i) = \mathbf{W}_i.$$

Converting discrete tokens into continuous-valued vectors makes them available as signals for later layers of the network. Like the parameters of other layers, $\mathbf{W}$ is learned in training. When the tokens are roughly word-level units of text, the rows of $\mathbf{W}$ are called **word vectors**. But embeddings are suitable for representing any kind of discrete input, not just text tokens — especially if the "vocabulary" items have relationships between each other that can be learned during training. In Chapter 12 we use an embedding to represent a discrete set of communities. Each community was represented by a vector of size $n = 64$ and the model learned to represent similar communities with similar vectors.

Neural networks are highly *modular*, meaning that it is relatively easy to swap out layers or extend a model with additional layers. This is the principle behind **pre-training** in which a model is trained on one task (language modeling is particularly popular, since it is self-supervised) and one or more of the initial layers along with their learned parameters are joined with additional untrained layers to perform a different *target task*. The new combined model is then trained on on data for the target task. The parameters of the new layers are trained while the parameters of the pre-trained layers are either kept *frozen* (meaning the layers act as a constant function of the input) or *fine-tuned* — also trained, but often with a slower learning rate.

Until recently, the typical way to make use of pre-training in NLP was to use pre-trained word embeddings. A common practice, for example, was to train word embeddings using a model like skipgram (Mikolov et al., 2013). And then use those word embeddings as input for recurrent neural model like an LSTM (Hochreiter & Schmidhuber, 1997) to perform sequence-level tasks. The skipgram model learns word embeddings by trying to predict context words drawn from a context window of a certain size around the input word.

Now it has become much more common to use multiple pre-trained layers of a deep neural network. These models produce representations that are a function of a whole sequence of tokens, meaning that they can, in principle, capture not just lexical but also some degree of compositional meaning. BERT (Devlin et al., 2019) is one such model. It is trained by *masked token prediction*, where it tries to guess the identity of

one or more tokens that have been masked out, as in a cloze task.[22]

5.2.2. Semantic change detection

There is a still relatively young but growing subfield of computational semantics that uses distributional methods to study semantic change.[23] Following the machine learning paradigm, this has been construed as a task, **lexical change detection** (LCD), where the objective is to automatically determine which words in a diachronic corpus have changed in meaning (or at least change in *usage*) over a certain time period.[24]

There are wide range of approaches to mode design for LCD. There are also a variety of ways the task can be realized, and any realization of the task needs to make certain assumptions about the nature of semantic change. In some variations the task involves detecting the *kind of change* that takes place for a certain word over the period of interest. The DURel corpus (Schlechtweg et al., 2018), for example, distinguishes between *innovative meaning change* and *reductive meaning change*, a distinction that the authors justify by demonstrating good inter-annotator agreement in the dataset.

The same can be said of different *methods* of semantic change detection. For example, one methodology involves **word sense induction** (WSI) or **word sense disambiguation** (WSD). WSD is the task of assigning a sense from a pre-determined sense inventory to each instance of a certain set of vocabulary items in a corpus. WSI is the same, but no pre-determined sense inventory is provided. On the assumption that semantic change typically involves adding or removing senses from a word's sense inventory, a WSI or WSD model can then be used for semantic change detection by measuring how sense distributions change over a certain time period (see, for example Mitra et al., 2015; Tahmasebi et al., 2013).

Another very popular methodology, is to use **diachronic word vectors**. This method involves training separate sets of word vectors across multiple time periods and ensuring the vocabularies are embedded in a shared vector space across time.[25] In Chapter 13 we used a *diachronic skip-gram* model (Kim et al., 2014) with noise-rectified change scores (Dubossarsky et al., 2017) to compare semantic change across a collection of online communities. Diachronic word vectors were preferable to a methodology using WSD/WSI for this study because it took relatively little training data and because we

[22] Both skipgram and BERT are sometimes referred to as language models, but it's important to point out that the are not language models in the sense introduced previously since they don't perform next-token prediction, and so are not trained to estimate a probability distribution over strings. They are similar, however in that the training objective allows for self-supervised learning.

[23] See Tahmasebi et al. (2021) and Kutuzov et al. (2018) for recent surveys. The former is a comprehensive survey including a diversity of methodologies and the later focuses on diachronic word vectors.

[24] A recent example of an LCD is the 2020 SemEval shared task in unsupervised lexical semantic change detection (Schlechtweg et al., 2020). A *shared task* is an event where different teams concurrently design and train models to perform a certain task using the same dataset.

[25] This is achieved by either by sharing some model parameters across time periods or by post-hoc alignment (see Hamilton et al., 2016, for detalis).

could measure change for a given word as a scalar value, comparable to other items in the vocabulary and across communities.

Diachronic word vectors, as most LSC methodologies, rely heavily on the distributional hypothesis. It is important to interpret the results of any study involving LCD with this in mind. As we've discussed, meaning representations based on the distributional hypothesis really measure change in *usage*, which may or may not correspond to a change in lexicalized meaning *potential* with respect to a certain community. This can have surprising implications for the aspects of meaning change that are captured. In Chapter 13 we measured meaning change in a collection of online communities. We found that one word consistently appeared among the words that changed most in each community: *2016*. As it happens, the two time time periods in our corpus were from 2015 and 2017, so *2016* went from referring to a year in the *future* to referring to a year in the *past*. One could argue that it didn't really change in *meaning* since the denotation was the same, but the contexts in which the word appeared nevertheless did change.

5.3. Statistical modeling

Statistical techniques can be applied to almost any kind of data. In general, they are useful when there is some underlying stochasticity, perhaps due to features that aren't observed as variables in the dataset. In this situation, we might want to know if some relationship between variables is "real", or if it can be explained by chance alone. With this in mind, statistical modeling can have essentially there different goals: (1) to **test** a hypothesis about a relationship between variables, (2) to **explore** relationships in the data, and (3) to **predict** some unknown or not-yet-realized values based on known values. While all of these are perfectly valid goals conflating them can lead to the appearance of statistical significance where there is none. For example, if one does some exploratory analysis to find relationships between variables, confidence metrics like the p-value are no longer a good indicator of whether the relationship is statistically significant since the p-value assumes the researcher is testing an *a priori* hypothesis.[26] Chapter 13, wherein we do perform exploratory analysis of the relationship between community structure and semantic change, is the only study in the thesis with sophisticated statistical modeling, but there are instances of statistical testing elsewhere in the thesis.

Agreement statistics In Chapter 7 we developed a new annotation scheme for word meaning negotiation. We wanted to test how much our annotators agreed on their annotations, since high agreement means that the results of the annotation can be considered **reliable** — suitable to use as the basis for further analysis and modeling.

[26] See McGill (2013) for further discussion.

High agreement can also seen be seen as validation that the annotation schema captures categories that correspond to "real" categories, although as we discuss in Chapter 7 this can be controversial since some real-world phenomena are inherently subjective.

Agreement statistics test whether annotators agree more than one would expect by random chance. Suppose we can always frame an annotation task as a collection of items, I, where annotators select one of a set of labels, L for each item. For annotator A, let $L_A : I \to L$ represent their annotations — i.e., let $L_A(i) \in L$ be the label that annotator A assigns to item $i \in I$. The naive agreement statistic, A_0, measures what proportion of items the two annotators agree on. Assuming we have two annotators A and B,[27]

$$A_0 = \frac{|\{i \in I \mid L_A(i) = L_B(i)\}|}{|I|} \tag{5.6}$$

The problem with A_0 as a metric is that it doesn't account for the possibility that some annotations will agree by chance, and that this is more likely to happen for labels that are more common. This makes A_0 incomparable between label sets and difficult to interpret in general. Agreement statistics thus try to adjust the agreement score based on the prior distribution the labels. This is done by computing the expected chance-level agreement, A_e, then computing a ratio which tells us what proportion of agreement beyond chance-level was actually observed:

$$\frac{A_0 - A_e}{1 - A_e} \tag{5.7}$$

The difficulty is, since we don't have access to an objective ground truth, we don't know what the prior distribution is and therefore have no objective way of computing A_e. In Chapter 7 we use two different statistics, which make different assumptions about how A_e should be estimated. For **Scott's pi**, the chance-level agreement, A_π is estimated from the data by assuming that each label has a different prior, which does not depend on annotator. **Cohen's kappa**, on the other hand, estimates A_κ by assuming that labels have annotator-specific prior distributions.[28] The agreement statistics π and κ are computed by plugging A_π and A_κ respectively in to Eq. (5.7). One reason to compute both statistics is that getting very different results for π and κ would indicate that annotators have different priors for the labels.

Confidence in embeddings There are two occasions in the thesis where we use statistics to measure the significance of measurements taken on *embeddings* (as described in Section 5.2.1). In Chapter 12 we have two embeddings S and L, which each represent the same set of n communities, but are computed in different ways. We

[27] All of these metrics can be generalized to n annotators. For the annotation study in Chapter 7 we had four annotators total, but each item was annotated by two people.

[28] See Artstein and Poesio (2008) for precise definitions of the agreement statistics and an extensive analysis of their use in computational linguistics.

want to test if there is a correlation between them. To do this, we align the coordinate systems, performing *orthogonal Procrustes* by *singular value decomposition*.[29] As a correlation metric, we compute

$$d(L, S) = n - Tr(\Sigma), \tag{5.8}$$

where Σ is the square matrix computed by singular value decomposition and Tr is the sum of its diagonal entries (that is, the sum of the *singular values*). As explained in Chapter 12, if $Tr(\Sigma)$ is equal to n, this would correspond to a perfect correlation between the two matrices, so Eq. (5.8) provides a normalized correlation metric between embeddings.

The problem is that singular value decomposition will *always* find some correlation between embeddings, even if they are completely random. To establish the significance of the correlations we measured in the paper, we wanted to compare the measured correlation to what one would expect to measure by chance. Since this expectation is difficult to compute analytically, we took 10 random embeddings L'_i and measured $d(S, L'_i)$ for each of the random embeddings. This gave us a mean, $\bar{x}_d$, and Bessel's corrected standard deviation,[30] s_d. From there we computed

$$\frac{d(S, L) - \bar{x}_d}{s_d}, \tag{5.9}$$

the *number of standard deviations* between the similarity computed for the real embedding and the mean of the similarities computed for each of the random embeddings. This was observed to be between 60 and 70 for each of the versions of L we tested, meaning that we could be very confident in concluding that the observed correlations were not by chance.

In Chapter 13 the situation was a bit different. We again had two aligned embeddings, but this time they were diachronic word embeddings (see Section 5.2.2), and we wanted to measure word-level change. In general, we can measure change for a word as the cosine distance between its two vector representations:

$$\Delta^{\cos}(\vec{w}_0, \vec{w}_1) = \frac{\cos^{-1}(\cos\,\text{sim}(\vec{w}_0, \vec{w}_1))}{\pi} \tag{5.10}$$

where

$$\cos\,\text{sim}(\vec{w}_0, \vec{w}_1) = \frac{\vec{w}_0 \cdot \vec{w}_1}{\|\vec{w}_0\|\|\vec{w}_1\|}. \tag{5.11}$$

[29] In general, two coordinate systems can encode the same information in different ways. Consider, for example CMYK color coding versus RGB, or a faucet that control water temperature and pressure with two knobs versus one. All the same results are achievable in both cases, but coordinate (or control) systems represent them differently. Orthogonal Procrustes is the problem of finding a transformation that aligns two vector spaces. Singular value decomposition is a kind of matrix factorization that can be used to solve the orthogonal Procrustes problem (Schönemann, 1966).

[30] Bessel's correction is a way of estimating a true prior standard deviation from the standard deviation of a sample.

The problem is that over a very large vocabulary, some amount of change is bound to observed by chance. This is related to the fact that word vectors are meaning representations based on the distributional hypothesis. Even if a word has not changed in meaning at all, there might, just by chance, be statistical regularities in the differences in contexts that it appears in across time periods. This is *especially* true for words that appear in highly variable contexts (for example, because of polysemy or lexical flexibility).

We corrected for this problem using a method described by Dubossarsky et al. (2017). First, we constructed 10 pseudo-diachronic corpora by shuffling the data from the two time periods and splitting them in half again. This resulted in 10 corpora with the same structure as the genuinely diachronic corpus, but where the actual dates of the texts were evenly distributed across time periods. Then, we trained the two embeddings again on 10 pseudo-diachronic corpora, resulting in a pseudo-diachronic embedding pair $\langle w'_{i,0}, w'_{i,1} \rangle$ for each word w and random trial i. In theory we would expect to measure $\Delta^{\cos}(\vec{w}'_{i,0}, \vec{w}'_{i,1}) = 0$ everywhere since the dates are roughly uniformly distributed over the corpora, so no genuine change can be measured. Of course, due to the reasons stated above, these values will be positive and tend to be larger for words with higher contextual variability. Similar to what was done in Chapter 12, we measured the mean, $\bar{x}_w$, and Bessel's-corrected standard deviation, s_w of $\Delta^{\cos}(\vec{w}'_{i,0}, \vec{w}'_{i,1})$ for each word across the pseudo-diachronic embeddings. We then defined the *rectified change score* as the t-statistic:

$$\Delta^*(\vec{w}_0, \vec{w}_1) = \frac{\Delta^{\cos}(\vec{w}_0, \vec{w}_1) - \bar{x}_w}{s_w \sqrt{1 + 1/10}} \tag{5.12}$$

Note that this is very similar to the metric computed in Eq. (5.9), but the t-statistic is slightly more interpretable — on the assumption that the $\Delta^{\cos}(\vec{w}'_{i,0}, \vec{w}'_{i,1})$ scores are normally distributed on a word level (we checked that they roughly are), the t-statistic can be used to compute confidence intervals. For example, if we observe $\Delta^*(\vec{w}_0, \vec{w}_1) = 4.74$, we can be sure with 95% confidence that continuing to sample $\bar{x}_w$ in the long-run will still show change for w above what can be explained by random noise.

Generalized mixed-effects modeling A generalized linear model is a statistical model that attempts to predict a response variable, based some number of *fixed effect* predictors. The model is called *generalized* because the response variable is not assumed to be normally distributed. A generalized linear *mixed effects* model (GLMM) also includes some number of *random effects* as predictors, which split the data points into groups. The response variable is modeled as sampled from an exponential distribution, which is parametrized by a linear combination of all the predictors — both fixed and random effects. In statistical modeling, these parameters are called a *design matrix*. To *fit* the model is to find the design matrix that explains the maximal amount of variance in the response variable, similar to how a machine learning model is trained

to find parameters that optimize the loss function with respect to the data. Statistical software like the lme4 package for R (Bates et al., 2015) includes various algorithm for fitting the model to the data.

Random effects are used to capture effects that correlate with a predictor, but aren't necessarily a function of its value. In Chapter 13, for example, we use GLMMs to model word-level semantic change (i.e., with change as the response variable) in 45 different online communities. There, we used community ID as a random effect, since we assumed there would likely be idiosyncratic community-level factors affecting the rate semantic change that wouldn't be captured by the other community-level fixed effects (like community size) that we included.

GLMMs, like all linear models, fit the model as a linear combination of the feature variables. But some of the variance in the response variable may also be explained by non-linear combinations of the features. To account for this, it is common to include **interaction features**, which are typically computed as products of two or more of the fixed effect features.

Fitting the model results in coefficients and standard errors (from the design matrix) for each of the included features. Unlike with neural network parameters, these coefficients are nicely interpretable, since they define the linear combination that explains the maximal variance in the response variable. For example, a positive coefficient means that there is a positive correlation between the corresponding feature and the response variable. The standard error can be used to compute a p-value, which helps to asses whether the relationship is statistically significant.

However, in order to ensure that the results are interpretable as described above, it can be necessary to test for **multi-collinearity** among the fixed effects. If one of the predictors can itself be reliably predicted as a linear combination of the other features, then it would be dubious to use the model coefficients to infer effects among the predictors, since they may be acting as proxies for each other in ways that can't be easily identified. For that reason, it is good practice to do some multi-collinearity detection before fitting a GLMM. This can be done by calculating the *variance inflation factor* (VIF) on a simple linear regression model (Fox & Monette, 1992). The VIF is used to find a set of predictors where the overall multi-collinearity of the model is low enough that the results of the GLMM will be reliably interpretable.

In Chapter 13, after eliminating predictors to reduce multi-collinearity, we performed our exploratory analysis by backwards model selection. We started with six fixed effects and all interactions between the three community-level and the three word-level features. Then, we removed features one-by-one and compared the overall predictive power of the model with and without those features. This allowed us to asses which features had a significant effect on the response variable, per-community word-level semantic change.

5.4. Social network modeling

In social network theory, social networks are modeled by graphs. Social networks analysis is a collection of methodologies used in various social sciences, especially sociology, political science, and economics, but it has also been used in linguistics, especially in sociolongusitics. In general, social network modeling attempts to capture the structure of communities and answer questions about how social structure affects the flow of information and ideas, material resources, and even contagious diseases. Graph models are a very good way of capturing social structure,[31] A graph is a set theoretic object consisting of two components,

$$G = \langle V, E \rangle, \tag{5.13}$$

in which V is a set of *nodes* (also called *vertices*) and $E \subseteq V \times V$ is a set of *edges* that connect the vertices. The nodes (usually) represent individuals and an edge between two nodes $\langle v_1, v_2 \rangle \in E$ represents a (directed) relationship between v_1 and v_2. A graph can also be represented as an *adjacency matrix*, $M : \{0, 1\}^{|V| \times |V|}$ where

$$M_{i,j} = \begin{cases} 1 & \text{if } \langle v_i, v_j \rangle \in E \\ 0 & \text{otherwise} \end{cases} \tag{5.14}$$

Various extensions of the graph-based model are possible. One might like to define multiple sets of edges for different types of relations, for example. Some relations might be directed (like *boss of*) while others are symmetric (like *coworker of*). Edges can also be *weighted*; that is, where $E : (V \times V) \to \mathbb{R}$.

Social network models like the one we use in Chapter 13 have not seen very extensive use in sociolinguistics because it is, in general, difficult to get complete information on all the relationships in a given community.[32] Certain types of social media data make this a possibility, however. While one can never be sure that there aren't interactions going on between members of the community in a different venue, a forum-style social network like Reddit allows the researcher to compile all the interactions that take place on a given forum.

Once one has a graph model of a social network, there are various node- and network-level metrics that can be computed on the graph. *Centrality metrics*, for example, are ways of measuring the "importance" or centrality of a given node in the community. For example, *betweenness centrality* is the proportion of all of the shortest paths between pairs of nodes that go through a given node. *Eigenvector centrality* uses measures how connected a node is to other highly central nodes.[33]

[31] See Jackson (2010) for an introduction to graph-based social network modeling and its applications.

[32] Sharma and Dodsworth (2020) gives survey of social network theory in sociolinguistics, including a detailed explanation of the different types of models.

[33] Eigenvector centrality is the basis for Google's PageRank algorithm.

In Chapter 13 we are interested in comparing different social networks to each other. In particular, we want to measure the effect of network cohesion on the pace of lexical change. As a measure of cohesion we use the **clustering coefficient**, which is defined as follows: First, for a given node v_i, let the *neighborhood* of v_i be the set of nodes connected to i:

$$N(v_i) = \{v_j \in V \mid \langle v_i, v_j \rangle \in E\}. \tag{5.15}$$

The clustering coefficient for a node v_i is defined as the proportion of a nodes neighbors that are also connected to each other

$$C(v_i) = \frac{|\{\langle v_j, v_k \rangle \in E \mid j, k \in N(i)\}|}{|N(i)| \cdot (|N(i)| - 1)}. \tag{5.16}$$

We use this metric to define the community-level metric as the average clustering coefficient across all its nodes.

6. Exposition

you are a participant in the future of language

Ocean Vuong
from *On Being with Krista Tippett*

With both theoretical and methodological background out of the way, we can now turn to the contributions of the thesis. Broadly speaking, the studies can be thought of in two categories. Chapters 7 to 11 are geared towards interaction. With the exception of Chapter 10, which uses neural language models, all of theses studies employ some formal interaction modeling. Chapters 12 and 13 investigate community-level variation and change using machine learning models trained on social media corpora. In contrast to Chapter 10, the neural networks in the final two chapters do not act as models *of agents*, but rather as models of the community-level linguistic norms, aggregating over the data in the corpora.

6.1. Part II summaries

This section contains summaries of each of the studies included in the thesis, with an eye towards how they fit together to tell a cohesive story. In the final part of this chapter we make some concluding remarks that draw insights from across the studies in Part II.

Chapter 7: What do you mean by negotiation?

> Noble, B., Viloria, K., Larsson, S., & Sayeed, A. (2021). What do you mean by negotiation? Annotating social media discussions about word meaning. *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*

Word meaning negotiation (WMN) is a conversational routine in which speakers explicitly discuss the meaning of a word or phrase — the so-called *trigger word* (because it triggered the discussion). This study has two parts. In the first part, we develop a model of WMN as a formal interaction game. In the second part, we use that model to

develop an annotation protocol and report on the results of an annotation study of 150 WMNs collected from Twitter.

The goal of the WMN interaction game model is to describe the structure that these interactions take, what moves are possible at different game states, and effect of different moves on the dialogue state, especially as it pertains to word meaning. The model we describe is built on previous work on WMNs, especially by Myrendal (2015) and Larsson and Myrendal (2017). Additional background on WMN can be found in Section 3.3 of the thesis.

Our model starts with the observation that WMNs involve setting up certain other reference points, which we refer to as semantic **anchors**. These reference points may be introduced with another lexical item, a description of a type of situation, or even a particular individual or situation which is either in the environment or commonly known to the participants. Participants then use these anchors to triangulate the meaning of the trigger word by relating the anchor to the trigger word with semantic **relations** (*X is an example of Y* or *X is a partial definition for Y*), which can then be grounded or rejected by other participants. As the WMN progresses, participants may even draw relations between non-trigger anchors in an attempt to find common ground.

In summary, the game state is represented by a graph structure that includes a set of anchors and a set of relations between those anchors. Relations are decorated with labels that indicate which speakers have committed to the relation (or its negation). The game state defines what future actions are possible (e.g., it is possible to introduce a new relation between anchors that have already been introduced; grounding existing relation is possible if the relation has been proposed). We can also read off **semantic updates** from the game state. The update is computed recursively on the sub-graph of relations that all speakers have committed to. The update works by minimally accommodating the grounded relations — e.g., if for two anchors A and B, it is grounded that B is an example of A, then A is updated such that its interpretation includes B. It should be noted that this constitutes a very conservative update. In Chapter 9 we explore semantic update from a certain type of definition in more detail.

In the second part, we report on the results of an annotation study of 150 WMN interactions from Twitter. The annotation protocol, which was developed by carrying out a series of pilot studies, suggests a sequence of steps for annotating the WMN: (1) identify the trigger word, (2) find text spans that introduce or refer to anchors and determine the relation they describe, (3) connect co-referring anchors, and (4) find explicit statements of commitment or grounding. The annotation protocol results in annotations that can be used to recover game states, as described in the formal model. We also annotated whether the interaction overall was one originating in non-understanding or disagreement. We found good agreement on token-level relation type (example or definition) and polarity (positive or negative), but poor agreement on statements of grounding. We found only moderate agreement on non-understanding or disagreement. Our error analysis found that most annotator disagreements about text spans indicating relations between anchors was disagreement about the extent of the span, or whether it refers to

two anchors or one. We also noticed that a number of disagreements result from different interpretation by the annotators due to different background knowledge about the topic of the Twitter interaction. This highlights the fact that WMNs and the meanings they negotiate can be highly specific to the context of a particular speech community.

Author contributions I developed the initial interaction model in close consultation with Staffan Larsson and Asad Sayeed. Kate Viloria and I conducted the pilot annotations and developed the annotation guide in consultation with Staffan Larsson, which also resulted in adjustments to the interaction model. All the authors performed annotations for the annotation study. I performed the agreement analysis of the results and Staffan Larsson, Asad Sayeed and I conducted the error analysis together. All authors read and approved the final manuscript.

Chapter 8: Classification systems

Noble, B., Larsson, S., & Cooper, R. (2022a). Classification Systems: Combining taxonomical and perceptual lexical meaning. *Proceedings of the 3rd Natural Logic Meets Machine Learning Workshop (NALOMA III)*, 11–16

As we discussed in Section 2.4 of the thesis, lexical meaning seems to have both referential and inferential aspects, though there is no clear separation between the two. We consider the domain of **classification systems** as a case study for unifying these two aspects of meaning. A classification system, as we conceive of it, is a common ground resource for a particular community of practice, which sets out a conceptual structure and methods for classifying entities within that structure for a particular domain. Having these classification systems as common ground facilitates teaching and learning the how to identify new classes within the community. Some examples of classification systems might include the way that a community of mushroom foragers identifies mushrooms, how a group of birders distinguish between local bird species, and the system by which professional astronomers categorize celestial objects. The goal of this paper is to develop a model of lexical meaning that synthesizes referential and inferential aspects of meaning in the context of classification systems. Ideally this model should be compatible with a Montague-style account of compositional semantics.

To do this, we use ProbTTR, which is introduced in Section 5.1.1 of the thesis. Our account starts with two components, a **folk taxonomy**, which represents the structural relations between concepts, and a set of multiclass **perceptual classifiers**, which give content to the concepts. A folk taxonomy is represented by a particular kind of tree structure, where each node can support multiple sets of branches. It can equivalently be represented as a set of **distinctions**, which, consist of a pair including a base concept

and a set of sub-concepts that partition the base concept. With these two ingredients, we define a ProbTTR type system with types representing concepts in the folk taxonomy. We use perceptual classifiers as **witness conditions** for auxiliary types, which are then combined with structural witness conditions that ensure the types in the classification system respect the inferential relationships specified by the taxonomy.

In the end, we have a type system in which probabilistic type judgements can be used to classify where objects belong in the taxonomy. In a small experiment using simulated data, we compare a classification system defined in this way two other methods for classifying in a hierarchical label set and find that, using the same underlying classifier architecture, the classification system outperforms those methods in both precision and recall.

Author contributions I originated the idea of combining perceptual classifiers and taxonomies in classification systems. The type theoretic model was developed in close collaboration by all the authors. I was responsible for the empirical comparison. All authors read and approved the final manuscript.

Chapter 9: Genus-differentia definitions

> Noble, B., Larsson, S., & Cooper, R. (2022b). Coordinating taxonomical and observational meaning: The case of genus-differentia definitions. *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*

Classically, a genus-differentia definition has two parts: First, it gives a **genus**, a super-concept of which the definiendum is a part. Second, it gives a method for differentiating the definiendum from other species of the same genus, the **differentia**. Although genus-differentia definitions are known from their role in the Aristotelian philosophical tradition, many real-world examples from dialogue can be analyzed as genus-differentia definitions, including utterances occurring as moves in a WMN, and corrective feedback, as in in child-directed speech. This chapter builds directly on Chapter 8 to formalize the semantic update incurred by grounding a genus-differentia definition of previously unknown concept in the context of a classification system. Throughout the paper, we use the utterance, *a raven is a large black corvid*, as our canonical example.

The goal of our account is to, given an existing classification system, define a new type, *Raven*, that (1) is a subtype of *Corvid* and, (2) is such that the properties described by the differentia (i.e., being *Large* and *Black*) are *taken as evidence* of that something is of type *Raven*, given that it is of type *Corvid*.

To do this, we first must define record types that correspond to multi-class classifiers. In contrast to Chapter 8 where classifiers were used as witness conditions for

basic types, we need to represent classifiers in the type system so that the type of the definiendum can be defined. We assume that, in addition to the **distinction classifiers** that we previously postulated as part of a classification system, there may be certain **feature classifiers** corresponding to features like *Large* and *Black* that don't directly define types in the taxonomy, but which may be used in conjunction with each other to define those types (for example in a naive Bayes classifier). Some features like *large* may require a comparison class for their interpretation. These we represent as dependent types which, given a certain context type, result in a classifier type (Fernandez & Larsson, 2014).

We first attempt a constructive definition which simply defines *Raven* as something that is a corvid, and large-for-a-corvid, and black (i.e., $Raven = Corvid \wedge (Large(Corvid) \wedge Black)$), but this definition results in the subtype relations $Raven \sqsubseteq (Large(Corvid))$ and $Raven \sqsubseteq Black$), which is undesirable since the subtype relation is intensional and we can at least *imagine* contexts where an individual of the type of the definiendum is nevertheless not a witness of one of the differentia types (an albino raven or a raven chick, for example). Instead, we argue that the definiendum should be represented as an **underspecified type** — a type with no explicit witness conditions, but where certain relationships with other types are specified as constraints on the type system as a whole. We show that an underspecified type can meet our previously stated desiderata under the following conditions: (1) all ravens are corvids ($Raven \sqsubseteq Corvid$), and (2) all else equal, something that is a raven is assumed to be black and large-for-a-corvid ($p(Large(Corvid) \wedge Black \| Raven) = 1$).

Author contributions I conceived of the general approach and made some initial attempts at formalization. Staffan Larsson was responsible for defining multiclass classifiers as a ProbTTR type. The remaining parts of the paper were developed in close collaboration with all the authors. All authors read and approved the final manuscript.

Chapter 10: Describe me an Aucklet

> Noble, B., & Ilinykh, N. (2023). Describe me an Aucklet: Generating Grounded Perceptual Category Descriptions. https://doi.org/10.48550/arXiv.2303.04053

There are many language and vision tasks in machine learning that require some degree of perceptual grounding. Image captioning and visual question answering are two examples of such tasks. But both of these setups put forth a particular image as the focus of each trial in the task (i.e., the image that is being captioned or that the questions are about). When humans use language, though, we can talk about perceptual experience at a level of concepts. Moreover, we argue that the *grounding* can't be

abstracted from a particular *communicative context*. How can you tell if language use is grounded if you don't know what was supposed to be communicated, or what the norms are under which the communication is taking place? The best contexts in which to investigate perceptual grounding in machine learning models are contexts that center communication.

We propose a task that we call *perceptual category description* for this purpose. The scenario is very much like the one described in Chapter 9. A teacher model, which has knowledge of a large set of perceptual classes must describe one or more the classes to a student model. The student model then uses those descriptions of classes they didn't previously know about to classify among all the new and previously known classes. The role of the student model is to perform *zero-shot classification*, which is not in itself a novel task. What we hope to contribute with this is the idea of using the classification performance of the student model as a way of measuring communicative success and obliquely evaluate the generation model.

In this study, we investigate how well different cognitively-inspired neural network architectures perform in the task of perceptual category description. In particular, we investigate generation models that use prototype-based representations, models that use exemplar representations, and hybrid models that use both. Both the generation and interpretation models have two modules that are trained jointly: a classifier module and a grounded language module. The interpretation model is trained to take text descriptions of categories and produce a vector representation close to the representation learned by the classifier. The generation model is trained to take class representations and use them to generate descriptions of the corresponding class.

For the **prototype** models, we simply use the representation learned by the classifier as the class representation. For **exemplar** models, we let the model use its classifier to select the highest-scoring training image for each class and used that as the class representation. A third model used **both** of these representations by concatenating them together.

The results showed that our models were able to achieve modest communicative success, but that the interpretation model still performed better when using the ground-truth descriptions of the unknown classes (written by human annotators). In general, exemplar models achieved the highest communicative success, which suggests that the other models aren't learning to abstract visual information to the class level. In essence the exemplar model converts the task back into one that can be solved by referring to a particular image. Finally, we found that certain generation strategies resulted in poor communicative success despite generating descriptions that were more statically discriminative among the classes. This could have to do with the way those descriptions expressed the information. Perceptual grounding is not only about packing perceptual information into text, but also doing so in a way that will be understood (for example in a particular speech community or by a model trained on a particular dataset).

Author contributions Nikolai Ilinykh trained and evaluated the generation models. I trained and evaluated the interpretation models. The task of perceptual category description was developed in close collaboration by both authors. Both authors read and approved the final manuscript.

Chapter 11: Personae under uncertainty

> Noble, B., Breitholtz, E., & Cooper, R. (2020). Personae under uncertainty: The case of topoi. *Proceedings of the Probability and Meaning Conference (PaM 2020)*, 8–16

A **topos** is an unstated assumption, which is necessary for interpreting certain *enthymematic* arguments or utterances in dialogue (J.-C. Anscombre (1995); also see Section 3.4.1 of this thesis). When someone makes an utterance that requires a certain topos to be interpreted, we say that the shared topos is **evoked**, since, although it is unstated, the listener must use the topos to bridge a certain chain of reasoning that is required to understand the meaning of the utterance. Topoi operate as background assumptions and, as such, may be associated with (or even constitutive of) certain ideologies.

A **persona** is a commonly recognized archetypical *kind of person*, which, in third-wave sociolinguistics, has an important interpretation as a source of social meaning (see Section 3.4.2). When someone speaks (or dresses, or acts, etc.) in such a way that indicates their idealogical alignment with a certain persona, we say that they are **projecting** that persona. In third-wave sociolinguistics, the **indexical field** of a social signal is the "constellation of ideologically related meanings" (Eckert, 2008) that arises in virtue of the variable's relationship with one or more personae. When someone projects a persona, it is understood that this is not the *only* persona they associate with. Instead, people construct a multifarious social identity as a **bricolage** of aspects of different personae with different idealogical associations.

Topoi are particularly interesting as social signals because, whereas many sociolinguistic analyses make a clean distinction between *what is said* and *how it is said*, topoi are, first of all, not *said* at all, but rather evoked by omission. Furthermore, the social meaning is not cleanly separable from the inferential meaning, since there are situations where the evoked topos is ambiguous and the listener must rely on what they know about the speaker's social identity to infer which topos they meant to evoke.

Chapter 11 has two main goals: (1) to develop a probabilistic model of social meaning based on the indexical field, and (2) to account for the social meaning of topoi in terms of updates to the speaker's perceived bricolage of personae. We proceed by introducing two probabilistic models. Both models consider a situation where a listener, Self, updates their representation of the social identity of a speaker, Other. Both

models also associate each persona, π, with a prior distribution, φ_π, over topoi, which captures the idealogical associations of the persona.

The *first-order* model represents social identity as a categorical distribution, θ, over personae. When a speaker evokes a topos, this distribution is updated by Bayesian update based on the likelihood (computed from φ_π, with θ as a prior) that a certain persona would project that topos. This model is nice because of it simplicity, but it doesn't achieve all of our modeling goals. We can interpret θ as either Self's uncertainty about Other's (singular) persona, or a representation of Other's personae bricolage (without uncertainty), but it can't represent both without conflating the two.

In the *second-order model* we seek to address this limitation by representing Self's understanding of Other's social identity as a Dirichlet distribution, α, *over* categorical distributions of personae. Given an utterance that evokes a topos, τ, we compute the *projected persona* as the persona that maximizes the probability of τ, give the prior α and the likelihood of each π_i resulting from φ_{π_i}. We again update α by Bayesian update, this time relying on the fact that the Dirichlet distribution is a conjugate prior for the categorical distribution.

This model of interpreting of social signals in the presence of social uncertainty about the speaker can be characterized as a kind of **category adjustment effect**, something which has been observed in the interpretation and recollection of perceptual stimuli. Essentially, the effect results in stimuli being biased towards the mean of the perceptual category in which they fall. Something similar goes on in our second-order model — the social meaning we assign to a certain topos is biased based on our priors about the speaker and the persona (or personae) it is associated with.

Finally, we show how to incorporate the parameters of the second-order social meaning model in a *dialogue gameboard* (see Section 5.1.1), with the aim of modeling social meaning as resulting in incremental updates in an idealogical context. To this end, we define an information state update (based on the Bayesian update defined in the second-order model), which is licensed by the evoked topos and the projected persona. The information state update is implemented as an asymmetric merge of record types, resulting in a new dialogue game board.

Author contributions I developed the probabilistic models of social signalling and conducted the signaling game simulations. Robin Cooper and Ellen Breitholtz created the dialogue game board interpretation representation of the model and defined the information state update function. All authors read and approved the final manuscript.

Chapter 12: Conditional language models for community-level linugistic variation

Noble, B., & Bernardy, J.-P. (2022). Conditional Language Models for Community-Level Linguistic Variation. *Proceedings of the 5th Workshop on NLP+CSS at EMNLP 2022*, 59–78

Language models make use of left-to-right text context to predict the next word in a sequence. But they can make use of additional extra-linguistic context as well (consider, for example an image captioning model, which is is trained to generate text conditioned on an image). In this study we introduce **community-conditioned language models** (CCLMs) as a technique for investigating community-level linguistic variation. We experiment on a dataset of social media posts from 510 different Reddit communities.

Experiments were carried out on LSTM and Transformer language models with a word embedding layer and three stacked sequence-to-sequence layers before the final prediction layer. The CCLMs also include a community embedding layer, which is concatenated to the hidden state of the language model at 4 different layer depths (directly to the word embedding and between each of the sequence-to-sequence layers). We compare the CCLMs to vanilla language model without community information. The models are assessed according to their **perplexity**, which measures performance on the language modeling task, and **information gain**, which measures the reduction in entropy of the CCLM over its un-conditioned counterpart.

We found that almost all models benefit from community-level information, but the distribution of average information gain for messages across different communities was highly skewed right. That is, the model benefits a little from community information for the majority of communities, but a lot for a small minority of communities.

Since the conditioned language models are trained with a community embedding, we also "incidentally" learn a vector representation of communities, similar to how a neural language model with word embeddings learns a vector representation of words as a consequence of optimizing for the next-word prediction task. Since the community embeddings are optimized for the same goal, we refer to them as *linguistic embeddings*. We compare these embeddings to another embedding which is trained based on user-community co-occurrence, with no linguistic information whatsoever. We refer to this embedding as the *social embedding*.

As an initial analysis, we examine pairs of communities that are similar (with respect to the cosine similarity of their vector representations) in the linguistic embeddings *and* in the social embeddings, and pairs of communities that are similar in one but not the other. We find that we can identify pairs of communities in all three conditions: socially and linguistically similar, socially similar but linguistically different,

and linguistically similar but socially different.[1] This suggests that the while the two types of embedding *do* capture something different about the communities, what they capture is nevertheless highly correlated.

Although this initial analysis is encouraging, we are limited to comparing *pairs of communities* across embeddings, since the two vector spaces represent the embeddings differently in their axes. To solve this problem, we use orthogonal Procrustes by singular value decomposition to align the axes. We find that all of the linguistic embeddings are correlated with the social embedding to a high degree of confidence (see Section 5.3 for details).

The main results of this study are that (1) information abut which community a message came from is useful for the next-word prediction task in almost all communities and language model architectures we tested, and (2) socially similar communities are also linguistically similar, which provides further evidence for the *homophilic* hypothesis from sociolinguistics. We also make a number of qualitative observations, perhaps the most evident of which is that our models make the most use of community information for messages from communities with highly routinized patterns of interaction (such as communities centered around organizing trades of different kinds). This provides support for the idea (discussed in sec Section 3.1) that the *community of practice* is the site of linguistic convention.

Author contributions I was responsible for training the models. Jean-Philippe Bernardy developed the method for testing correlations between embeddings. Both authors were responsible for the analysis and the remaining aspects of the research. Both authors read and approved the final manuscript.

Chapter 13: Semantic shift in social networks

> Noble, B., Sayeed, A., Fernández, R., & Larsson, S. (2021). Semantic shift in social networks. *Proceedings of *SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics*, 26–37. https://doi.org/10.18653/v1/2021.starsem-1.3

Most work on language change, both in historical linguistics and in computational linguistics, has focused on change at the level of the macro-language and on a time scale of decades or even centuries. In this study, we turn our focus to short-term lexical change in relatively small online communities. As in the previous study, we use a corpus of Reddit comments. This time we limit our focus to 45 randomly selected sub-forums and use a diachronic corpus split into two time periods (2015 and 2017 with a one-year gap in between).

[1] Of course the vast majority of pairs of communities are dissimilar in both types of embedding.

To measure semantic change, we use a diachronic skip-gram model (Kim et al. (2014); also see Section 5.2.1 of this volume) and compute **rectified change** scores to account for the possibility that words appearing in more variable contexts will have inflated cosine change scores (Dubossarsky et al. (2017); also see Section 5.3 of this volume). We observed that naive (un-rectified) cosine change assigned high scores to discourse connectives and other words with a distinctly rhetorical function like *possibly*, *however*, and *;* (semicolon). This is consistent with the hypothesis that this metric over-estimates change for words that appear in highly variable contexts. The words recording the highest rectified change scores were much more varied across community and tended more towards nouns and verbs. We also observed that while there is an (apparently) strong (albeit non-linear) relationship between naive change and log word frequency, that relationship is not present for rectified change.

In addition to the community-level change scores, we also measured semantic change on a larger collection of Reddit comments (not restricted to any forum) over the same time period. This **generic change** score is intended to help distinguish between change that originates at the community level and change that is happening on a broader scale but reflected in the community.

We also considered word **frequency** and **change in frequency** as factors that might predict lexical change.

Next, we induced a social network graph on each of the communities in the dataset by drawing edges between users that interacted at least once in 2015. We then computed the mean **clustering coefficient** for each community (see Section 5.4 for details). We also defined a number of other community-level metrics that we thought might be correlated with semantic change, including community **size**, **stability** (overlap in active members between 2015 and 2017), and **mean posts** per member.

Finally, we performed an exploratory analysis by backward model selection on generalized linear mixed effects models to investigate the relationship between rectified change (as the response variable) and the community- and word-level features as predictors. We found a significant positive effect between change in frequency and community-level change and also between generic change and community-level change. Word frequency had a small but also significant negative effect. Among the community-level features, we found that there is a significant three-way interaction between community size, stability and clustering coefficient. In particular, in loosely-connected communities (those with low clustering), more stability among the members is correlated with more semantic change. For more densely connected communities (with average or high clustering), the positive relationship between stability and change only holds in smaller communities. For large and dense communities, the relationship between stability and change actually trends negative.

Author contributions I was responsible for training the models and computing the community-level metrics. Asad Sayeed was responsible for the GLMM exploratory

analysis. The research questions as well as the qualitative analysis and conclusions were developed in close collaboration with all the authors. All authors read and approved the final manuscript.

6.2. Conclusions

Using a variety of methodologies can make it difficult to draw direct connections from one project to the next, but it does afford us the benefit of multiple perspectives from which to make sweeping conclusions. To distinct patterns of insights emerge from the compilation.

Lexical complexity supports semantic plasticity. Words rarely, if ever, have a monolithic meaning. Whether or not it is correct to make sharp sense distinctions, it is clear that all words have a range of situations in which they can be used and that a word can carry a different meaning depending on the situation. This non-uniformity of meaning creates opportunities for lexical innovation. Words also have a range of communicative affordances. They have inferential as well as referential potential. They can be associated with other words, situations, or feelings by connotation. They can carry social meaning. Innovative uses draw on these different affordances to extend a word's range. To understand how that happens and what it means when innovations are lexicalized, (or when someone is a explicitly taught a completely new sense of a word), we can't avoid getting into the messy details of lexical structure.

- In **Chapter 7**, we saw many examples of WMNs about common words where it was clear that both participants *knew* the word, but didn't understand how it was being used in the current situation or as part of a particular construction. This could be either because the use was an innovative or because it was conventional in some community that the WMN initiator wasn't familiar with. This means that (1) other senses of the word can be used as a resource to help negotiate the meaning and (2) if the new meaning is grounded, it may only apply in situations like the one that initiated the WMN. While these dynamics were evident in the annotation study, our interaction model would have to be extended to fully accommodate them.

- We used monolithic vector representations of words in **Chapter 13**. While this allowed us to easily quantify change from one time period to the next. It did mean we were limited in what we could understand about *how* a word was changing. Similar to the triggers in Chapter 7, many of the most-changed words by community were words already present in the community's vocabulary (although increase in frequency *was* highly correlated with change). Relatedly, the distributional approach can't tell us whether changes in word representation are

a result merely of changes in the *distribution of use*, or if those changes reflect (or engender) underlying changes in the word's *meaning potential*.

- In **Chapter 8**, we showed how referential and inferential aspects of meaning can be synthesized by broadening our perspective from considering lexical items one-by-one to considering a classification system as a lexical resource from which individual lexical meanings can be derived. In **Chapter 9**, we used that structure to give an account of how genus-differentia definitions can be interpreted to create a new lexical entry that carries both inferential and referential meaning.

- In **Chapter 10**, we showed that the cognitive structure of perceptual concepts matters for how they can be described. Our models performed best when they generated textual descriptions from exemplar *instances* of a perceptual class, rather than aggregated class representations. More work is needed to understand whether the same might be true of humans, or whether there are machine learning architectures that would better model the way people represent perceptual classes.

- It is not only words that carry meaning. In **Chapter 11**, we assigned a prior over topoi to each persona, imbuing the personae with idealogical content, but also giving social meaning to the topoi by Bayesian inference. Of course, there are all sorts of indexical relationships in the world that we wouldn't necessarily want to consider as part of the lexicon (smoke *means* fire, for example). But topoi point to the fact that it is not always easy to make a distinction between the lexical and the non-lexical. This suggests lexical change is related to the more general cognitive phenomena of inference and uncertainty that govern how indexical relationships are established.

Methodologically, we can get access to new ways of understanding the process of semantic change by starting with frameworks that acknowledge the complexity of lexical structure and its implications for both compositional meaning and interaction. On the formal side, systems like TTR make it possible to represent structured lexical information. There may also be benefits to adopting a construction grammar approach to meaning, since it seems that multi-word constructions are often the site of coordination and change. Finally, modeling lexical meaning at the level of cognition can give us a more fine-grained understanding of what happens when a word's meaning potential changes in the mind of a speaker.

Community-level change stems from the interactive practices of the community. All language use takes place in a communicative context. When that context includes a particular community, lexicalization is possible. Some interactive practices (WMN, for example) are explicitly oriented towards lexicalization. In other cases,

semantic coordination is more implicit, and whether or not the coordinated meaning "sticks" (or is propagated to the community level) may depend on multitude of factors, including the communicative utility of the innovation, and whether it is compatible with existing community norms.

- Our WMN interaction model from **Chapter 7** relied on on the concept of semantic *anchors*, which highlight the importance of existing common ground when negotiating new meanings. As an interaction game, WMN itself relies on community norms about how the game proceeds, what moves are possible at different times, and how different moves should be interpreted to maintain a shared understanding of the state of the joint activity.

- In **Chapter 9** we gave an account of how an agent might update their lexicon based on a genus-differentia definition. Importantly, this account relied on community-level norms about how to classify entities in a particular domain. Sharing a classification system is a way, not only to classify for oneself, but to make it possible to teach and learn new concepts among the community.

- Similarly in **Chapter 10**, our generation model was able to successfully describe novel perceptual categories to the interpretation model. This success depended not only on an existing set of shared perceptual categories, but also on norms (implicit in the training data) about how how a bird should be described to maximize class-level discriminativity of the description.

- In **Chapter 12**, the interactive practices of the community appeared to be related to how linguistically idiosyncratic the community was. Communities with highly formulaic patterns of interaction tended to be more informative to the language model, whereas communities where interactions tended towards general conversation were less informative. Although this is *prima facie* a synchronic observation about linguistic variation, it suggests that a certain task-orientedness can serve as motivation for innovation and conventionalization.

- In **Chapter 13**, we saw that in more loosely-connected communities, stability of membership was always correlated with more change, but the same was not true for densely connected communities, especially large, dense communities. Anecdotally, it seemed that densely connected communities tend to have more extended interactions involving multiple parties. It could be that in these more intensely interactive environments, changes are more easily propagated to the community level, relying less on the pairwise common ground that is preserved by a more stable membership.

Again, these insights suggest certain methodological recommendations. An approach that centers interaction can yield a lot of new insights about language change.

If we are interested in *why* change takes place, we must go to the site of change — the particular communicative context or interaction. This is where the rubber meets the road: where we try out new semantic innovations, accommodate unfamiliar language, and learn from each other. A flexible mutable language is what gives linguistic interaction its distinctly human character. And it is in interaction that we make our mark on the language.

Bibliography

Anscombre, J. C., & Ducrot, O. (1983). *L'argumentation dans la langue*. Editions Mardaga.

Anscombre, J.-C. (1995). La théorie des Topoï : sémantique ou rhétorique ? *Hermès*, (15), 185. https://doi.org/10.4267/2042/15167

Anttila, A. (2004). Variation and Phonological Theory. In *The Handbook of Language Variation and Change* (pp. 206–243). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756591.ch8

Artstein, R., & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, *34*(4), 555–596. https://doi.org/10.1162/coli.07-034-R2

Auer, P. (2015). Reflections on Hermann Paul As a Usage-Based Grammarian. In P. Auer & R. W. Murray (Eds.), *Hermann Paul's Principles of Language history revisited: Translations and reflections*. De Gruyter.

Austin, J. L. (1950). Truth. *Aristotelian Society Supp*, *24*(1), 111–29.

Bakhtin, M. M. (1987). *Speech Genres and Other Late Essays* (C. Emerson & M. Holquist, Eds.; V. W. McGee, Trans.; 2nd Edition). University of Texas Press.

Barwise, J., & Perry, J. (1983). *Situations and Attitudes*. MIT Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bell, R. A., & Healey, J. G. (1992). Idiomatic Communication and Interpersonal Solidarity in Friends' Relational Cultures. *Human Communication Research*, *18*(3), 307–335. https://doi.org/10.1111/j.1468-2958.1992.tb00555.x

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bennett, M. (1976). A Variation and Extension of a Montague Fragment of English. In *Montague Grammar* (pp. 119–163). Elsevier. https://doi.org/10.1016/B978-0-12-545850-4.50010-8

Bernardy, J.-P., & Chatzikyriakidis, S. (2019). What Kind of Natural Language Inference are NLP Systems Learning: Is this Enough?: *Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, 919–931. https://doi.org/10.5220/0007683509190931

Bhattasali, S., & Resnik, P. (2021). Using surprisal and fMRI to map the neural bases of broad and local contextual prediction during natural language comprehension. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 3786–3798. https://doi.org/10.18653/v1/2021.findings-acl.332

Blank, A. (2003). Polysemy in the lexicon and in discourse. In B. Nerlich, Todd, V. Herman, & C. David D. (Eds.), *Polysemy: Flexible Patterns of Meaning in Mind and Language* (pp. 267–293). Mouton de Gruyter.

Blank, H., & Bayer, J. (2022). Functional imaging analyses reveal prototype and exemplar representations in a perceptual single-category task. *Communications Biology*, *5*(1), 1–13. https://doi.org/10.1038/s42003-022-03858-z

Breitholtz, E. (2020). *Enthymemes and Topoi in Dialogue: The Use of Common Sense Reasoning in Conversation*. Brill.

Brennan, S. E., & Clark, H. H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Bücking, S. (2010). German Nominal Compounds as Underspecifiesd Names for Kinds. *Linguistische Berichte. Sonderheft*, (17), 253–281.

Campbell-Kibler, K. (2010). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change*, *22*(3), 423–441. https://doi.org/10.1017/S0954394510000177

Cann, R., Grover, C., & Miller, P. H. (Eds.). (2000). *Grammatical interfaces in HPSG*. CSLI Publications.

Chatzikyriakidis, S., & Cooper, R. (2018). Type Theory for Natural Language Semantics. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.329

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1986). Collaborating on contributions to conversations. *Language and Cognitive Processes*, *2*(1), 19–41.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2012). Tolerant, Classical, Strict. *Journal of Philosophical Logic*, *41*(2), 347–385. https://doi.org/10.1007/s10992-010-9165-z

Cooper, R. (2005). Austinian Truth, Attitudes and Type Theory. *Research on Language and Computation*, *3*(2), 333–362. https://doi.org/10.1007/s11168-006-0002-z

Cooper, R. (2012). Type Theory and Semantics in Flux. In R. Kempson, T. Fernando, & N. Asher (Eds.), *Philosophy of Linguistics* (pp. 271–323). North-Holland. https://doi.org/10.1016/B978-0-444-51747-0.50009-3

Cooper, R. (2023). *From Perception to Communication: A Theory of Types for Action and Meaning*. Oxford University Press.

Cooper, R., Dobnik, S., Lappin, S., & Larsson, S. (2015). Probabilistic Type Theory and Natural Language Semantics. *Linguistic Issues in Language Technology, Volume 10, 2015*.

Cooper, R., & Ginzburg, J. (2015). Type Theory with Records for Natural Language Semantics*. In *The Handbook of Contemporary Semantic Theory* (pp. 375–407). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118882139.ch12

Coquand, T., Pollack, R., & Takeyama, M. (2003). A Logical Framework with Dependently Typed Records. In M. Hofmann (Ed.), *Typed Lambda Calculi and Applications* (pp. 105–119). Springer. https://doi.org/10.1007/3-540-44904-3_8

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

de Finetti, B. (1992). Foresight: Its Logical Laws, Its Subjective Sources (H. E. Kyburg Jr., Trans.). In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics: Foundations and Basic Theory* (pp. 134–174). Springer. https://doi.org/10.1007/978-1-4612-0919-5_10

Deane, P. D. (1988). Polysemy and cognition. *Lingua*, *75*(4), 325–361. https://doi.org/10.1016/0024-3841(88)90009-5

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dowty, D. R., Wall, R. E., & Peters, S. (1981). *Introduction to Montague semantics*. D. Reidel Pub. Co. ; sold and distributed in the U.S.A. and Canada by Kluwer Boston Inc.

Dubossarsky, H., Weinshall, D., & Grossman, E. (2017). Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1136–1145. https://doi.org/10.18653/v1/D17-1118

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, *12*(4), 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

Eckert, P. (2019). The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language*, *95*(4), 751–776. https://doi.org/10.1353/lan.2019.0072

Fernandez, R. (2014). Dialogue. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics 2nd edition* (Second). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199573691.001.0001

Fernandez, R., & Larsson, S. (2014). Vagueness and Learning: A Type-Theoretic Approach. *Proceedings of the Third Joint Conference on Lexical and Computational Semantics (*SEM 2014)*, 151–159. https://doi.org/10.3115/v1/S14-1019

Firth, J. (1957). A synopsis of linguistic theory 1930–1955. *Studies in Linguistic Analysis (special volume of the Philological Society)*, *1952–59*, 1–32.

Fox, J., & Monette, G. (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, *87*(417), 178–183. https://doi.org/10.2307/2290467

Gamut, L. T. F. (1991). *Logic Language and Meaning, Volume 2: Intensional Logic and Logical Grammar*. University of Chicago Press.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Mifflin OCLC: 1222716492.

Giglioli, P. P. (1972). *Language and social context: Selected readings*. Harmondsworth : Penguin.

Ginzburg, J. (2012). *The Interactive Stance*. Oxford University Press.

Grice, H. P. (1975). Logic and Conversation. In P. Cole (Ed.), *Speech acts* (5. print, pp. 44–55). Acad. Pr.

Gumperz, J. (1972). The Speech Community. In P. P. Giglioli (Ed.), *Language and social context: Selected readings*. Harmondsworth : Penguin.

Gururangan, S., Swayamdipta, S., Levy, O., Schwartz, R., Bowman, S., & Smith, N. A. (2018). Annotation Artifacts in Natural Language Inference Data. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 107–112. https://doi.org/10.18653/v1/N18-2017

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1489–1501. https://doi.org/10.18653/v1/P16-1141

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Harris, Z. S. (1954). Distributional Structure. *Word*, *10*(2-3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hasan, R. (1989). Semantic variation and sociolinguistics. *Australian Journal of Linguistics*, *9*(2), 221–275. https://doi.org/10.1080/07268608908599422

Hasan, R. (2009). *Collected works of Ruqaiya Hasan. Vol. 2, Semantic variation: Meaning in society and in sociolinguistics*. Equinox.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hopper, R., Knapp, M. L., & Scott, L. (1981). Couples' Personal Idioms: Exploring Intimate Talk. *Journal of Communication*, *31*(1), 23–33. https://doi.org/10.1111/j.1460-2466.1981.tb01201.x

Jackson, M. O. (2010). *Social and Economic Networks* (Illustrated edition). Princeton University Press.

Johnstone, B. (1996). *The Linguistic Individual: Self-Expression in Language and Linguistics*. Oxford University Press.

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, *57*(2), 129–191. https://doi.org/10.1016/0010-0277(94)00659-9

Kennington, C., & Schlangen, D. (2015). Simple Learning and Compositional Application of Perceptually Grounded Word Meanings for Incremental Reference Resolution. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 292–301. https://doi.org/10.3115/v1/P15-1029

Kim, Y., Chiu, Y.-I., Hanaki, K., Hegde, D., & Petrov, S. (2014). Temporal Analysis of Language through Neural Language Models. *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 61–65. https://doi.org/10.3115/v1/W14-2517

Kingma, D. P., Ba, J., & Amsterdam Machine Learning lab (IVI, FNWI). (2015). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)*.

Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. New York: Chelsea Pub. Co.

Korta, K., & Perry, J. (2008). The pragmatic circle. *Synthese*, *165*(3), 347–357. https://doi.org/10.1007/s11229-007-9188-3

Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018). Diachronic word embeddings and semantic shifts: A survey. *Proceedings of the 27th International Conference on Computational Linguistics*, 1384–1397.

Labov, W. (1963). The Social Motivation of a Sound Change. *WORD*, *19*(3), 273–309. https://doi.org/10.1080/00437956.1963.11659799

Larsson, S. (2002). *Issue-based Dialogue Management* (Doctoral dissertation). University of Gothenburg. Gothenburg, Sweden.

Larsson, S. (2013). Formal semantics for perceptual classification. *Journal of Logic and Computation*, *25*(2), 335–369. https://doi.org/10.1093/logcom/ext059

Larsson, S. (2020). Discrete and Probabilistic Classifier-based Semantics. *Proceedings of the Probability and Meaning Conference (PaM 2020)*, 62–68.

Larsson, S. (2021). The role of definitions in coordinating on perceptual meanings. *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*.

Larsson, S., & Bernardy, J.-P. (2021). Semantic Classification and Learning Using a Linear Tranformation Model in a Probabilistic Type Theory with Records. *Proceedings of the Reasoning and Interaction Conference (ReInAct 2021)*, 14–22.

Larsson, S., & Cooper, R. (2009). Towards a formal view of corrective feedback. *Proceedings of the EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition - CACLA '09*, 1–9. https://doi.org/10.3115/1572461.1572464

Larsson, S., & Cooper, R. (2021). Bayesian Classification and Inference in a Probabilistic Type Theory with Records. *Proceedings of the 1st and 2nd Workshops on Natural Logic Meets Machine Learning (NALOMA)*, 51–59.

Larsson, S., & Myrendal, J. (2017). Dialogue Acts and Updates for Semantic Coordination. *SEMDIAL 2017 (SaarDial) Workshop on the Semantics and Pragmatics of Dialogue*, 52–59. https://doi.org/10.21437/SemDial.2017-6

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, *194*(10), 3801–3836. https://doi.org/10.1007/s11229-015-0786-1

Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, Acceptability, and Probability: A Probabilistic View of Linguistic Knowledge. *Cognitive Science*, *41*(5), 1202–1241. https://doi.org/10.1111/cogs.12414

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Wiley-Blackwell.

Lücking, A., Cooper, R., Larsson, S., & Ginzburg, J. (2019). Distribution is not enough: Going Firther. *Proceedings of the Sixth Workshop on Natural Language and Computer Science*, 1–10. https://doi.org/10.18653/v1/W19-1101

Malt, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 539–555. https://doi.org/10.1037/0278-7393.15.4.539

Marconi, D. (1997). *Lexical competence*. MIT Press.

Mazzocconi, C., Tian, Y., & Ginzburg, J. (2022). What's Your Laughter Doing There? A Taxonomy of the Pragmatic Functions of Laughter. *IEEE Transactions on Affective Computing*, *13*(3), 1302–1321. https://doi.org/10.1109/TAFFC.2020.2994533

McGill, B. (2013). In praise of exploratory statistics.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *NIPS Proceedings*, 9.

Mills, G., & Healey, P. (2008). Semantic negotiation in dialogue: The mechanisms of alignment. *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, 46–53.

Mills, G. J., & Healey, P. (2006). Clarifying spatial descriptions: Local and global effects on semantic co-ordination. *Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*.

Mitra, S., Mitra, R., Maity, S. K., Riedl, M., Biemann, C., Goyal, P., & Mukherjee, A. (2015). An automatic approach to identify word sense changes in text media across timescales. *Natural Language Engineering*, *21*(5), 773–798. https://doi.org/10.1017/S135132491500011X

Montague, R. (1970). English as a Formal Language. In B. Visentini (Ed.), *Linguaggi nella societa e nella tecnica* (pp. 188–221). Edizioni di Communita.

Montague, R. (1973). The Proper Treatment of Quantification in Ordinary English. In P. Suppes, J. Moravcsik, & J. Hintikka (Eds.), *Approaches to Natural Language* (pp. 221–242). Dordrecht.

Moon, R. (2015). Multi-word Items. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 120–140). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641604.013.031

Morgan, J. L. (1978). Two Types of Convention in Indirect Speech Acts. In P. Cole (Ed.), *Pragmatics* (pp. 261–280). BRILL. https://doi.org/10.1163/9789004368873_010

Murphy, M. L. (2003). *Semantic Relations and the Lexicon: Antonymy, Synonymy and other Paradigms*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486494

Myrendal, J. (2015). *Word Meaning Negotiation in Online Discussion Forum Communication* (Doctoral dissertation). University of Gothenburg. University of Gothenburg.

Noble, B., & Bernardy, J.-P. (2022). Conditional Language Models for Community-Level Linguistic Variation. *Proceedings of the 5th Workshop on NLP+CSS at EMNLP 2022*, 59–78.

Noble, B., Breitholtz, E., & Cooper, R. (2020). Personae under uncertainty: The case of topoi. *Proceedings of the Probability and Meaning Conference (PaM 2020)*, 8–16.

Noble, B., & Ilinykh, N. (2023). Describe me an Aucklet: Generating Grounded Perceptual Category Descriptions. https://doi.org/10.48550/arXiv.2303.04053

Noble, B., Larsson, S., & Cooper, R. (2022a). Classification Systems: Combining taxonomical and perceptual lexical meaning. *Proceedings of the 3rd Natural Logic Meets Machine Learning Workshop (NALOMA III)*, 11–16.

Noble, B., Larsson, S., & Cooper, R. (2022b). Coordinating taxonomical and observational meaning: The case of genus-differentia definitions. *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*.

Noble, B., Sayeed, A., Fernández, R., & Larsson, S. (2021). Semantic shift in social networks. *Proceedings of *SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics*, 26–37. https://doi.org/10.18653/v1/2021.starsem-1.3

Noble, B., Viloria, K., Larsson, S., & Sayeed, A. (2021). What do you mean by negotiation? Annotating social media discussions about word meaning. *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*.

Norén, K., & Linell, P. (2007). Meaning potentials and the interaction between lexis and contexts: An empirical substantiation. *Pragmatics*, *17*(3), 387–416. https://doi.org/10.1075/prag.17.3.03nor

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *10*(1), 104–114. https://doi.org/10.1037//0278-7393.10.1.104

Partee, B. (1973). Some transformational extensions of Montague grammar. *Journal of Philosophical Logic*, *2*(4), 509–534. https://doi.org/10.1007/BF00262953

Partee, B. H. (1979). Semantics–Mathematics or Psychology? In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics From Different Points of View* (pp. 1–14). Springer Verlag.

Paul, H. (1886). *Prinzipien der Sprachgeschichte*. Max Niemeyer.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona1. *Journal of Sociolinguistics*, *11*(4), 478–504. https://doi.org/10.1111/j.1467-9841.2007.00334.x

Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*(4), 532–547. https://doi.org/10.1016/0010-0285(75)90021-3

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics*. State University of New York Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Schlangen, D., Zarrieß, S., & Kennington, C. (2016). Resolving References to Objects in Photographs using the Words-As-Classifiers Model. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1213–1223. https://doi.org/10.18653/v1/P16-1115

Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. (2020). SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 1–23.

Schlechtweg, D., Schulte im Walde, S., & Eckmann, S. (2018). Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 169–174. https://doi.org/10.18653/v1/N18-2027

Schönemann, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika*, *31*(1), 1–10. https://doi.org/10.1007/BF02289451

Searle, J. R. (1975). *Indirect Speech Acts*. Brill. https://doi.org/10.1163/9789004368811_004

Sharma, D., & Dodsworth, R. (2020). Language Variation and Social Networks. *Annual Review of Linguistics*, *6*(1), 341–361. https://doi.org/10.1146/annurev-linguistics-011619-030524

Silberer, C., Ferrari, V., & Lapata, M. (2017). Visually Grounded Meaning Representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(11), 2284–2297. https://doi.org/10.1109/TPAMI.2016.2635138

Sperber, D., & Wilson, D. (2001). *Relevance: Communication and cognition* (2nd ed). Blackwell Publishers.

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, *25*(5-6), 701–721.

Sutton, P. R. (2015). Towards a Probabilistic Semantics for Vague Adjectives. In H. Zeevat & H.-C. Schmitz (Eds.), *Bayesian Natural Language Semantics and Pragmatics* (pp. 221–246). Springer International Publishing. https://doi.org/10.1007/978-3-319-17064-0_10

Sutton, P. R. (2018). Probabilistic Approaches to Vagueness and Semantic Competency. *Erkenntnis*, *83*(4), 711–740. https://doi.org/10.1007/s10670-017-9910-6

Tahmasebi, N., Borin, L., & Jatowt, A. (2021). Survey of computational approaches to lexical semantic change detection. Zenodo. https://doi.org/10.5281/ZENODO.5040302

Tahmasebi, N., Niklas, K., Zenz, G., & Risse, T. (2013). On the applicability of word sense discrimination on 201 years of modern english. *International Journal on Digital Libraries*, *13*(3), 135–153. https://doi.org/10.1007/s00799-013-0105-8

Taylor, J. R. (2012). The dictionary and the grammar book: The generative model of linguistic knowledge. In J. R. Taylor (Ed.), *The Mental Corpus: How language is represented in the mind* (pp. 19–43). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199290802.003.0002

Teichman, M. (n.d.). Greg Kobele discusses mathematical linguistics https://elucidations.vercel.app/posts/transcript-episode-111/.

Traum, D. R., & Larsson, S. (2003). The Information State Approach to Dialogue Management. In J. van Kuppevelt & R. W. Smith (Eds.), *Current and New Directions in Discourse and Dialogue* (pp. 325–353). Springer Netherlands. https://doi.org/10.1007/978-94-010-0019-2_15

Tyler, A., & Evans, V. (2001). Reconsidering Prepositional Polysemy Networks: The Case of Over. *Language*, *77*(4), 724–765. https://doi.org/10.1353/lan.2001.0250

van Eijck, J., & Lappin, S. (2012). Probabilistic Semantics for Natural Language. *Logic and Interactive Rationality (LIRA)*, *2*, 11–35.

Wittgenstein, L. (2009). *Philosophische Untersuchungen =: Philosophical investigations* (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans.; Rev. 4th ed). Wiley-Blackwell.

Wright, C. (1975). On the Coherence of Vague Predicates. *Synthese*, *30*(3/4), 325–365.