All together now... This time with meaning: A hierarchical lexicon for semantic coordination

MSc Thesis *(Afstudeerscriptie)*

written by

Bill Noble
(born October 11, 1989 in Takoma Park, Maryland, USA)

under the supervision of **Dr. Raquel Fernández Rovira**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:
June 29th 2015

Members of the Thesis Committee:
Dr. Maria Aloni
Dr. Raquel Fernández Rovira
Dr. Robert van Rooij
Dr. Henk Zeevat

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

Classically, semantic theories have assumed that words are endowed with universal, immutable meanings. This assumption is not tenable when modeling natural language dialogue; far from treating word meanings as fixed entities, linguistic agents are constantly coordinating useful semantic conventions. They disambiguate polysemous words, construct *ad hoc* interpretations particular to their communicatory goals, and ultimately learn new ways to use old words based on these collaborations. This thesis proposes a lexical model for dialogue semantics that supports semantic coordination.

The model that is proposed employs a hierarchical lexicon to distinguish between the different kinds of shared semantic information. It is used to build a theory of semantic coordination that drives lexicosemantic learning and tends to introduce polysemy in the lexicon. The lexical model and theory of semantic coordination are formalized using Type Theory with Records, which uses feature structure-like objects to store semantic information. Finally, the thesis presents some empirical work: A corpus study on contrastive focus reduplication—one strategy for semantic coordination—confirms predictions made by the pragmatic and semantic theory presented earlier, and a signaling games simulation supports the connection between semantic coordination and polysemy that is predicted by the hierarchical lexical model.

Acknowledgments

There are so many people who have had a hand in getting me here. I'll try to be brief though.

Starting with the obvious, my incredibly generous advisor, Raquel Fernández. I feel so lucky to have had the chance to work with Raquel. This experience, writing this thesis and the other work we've done together, has opened my eyes to a whole fascinating world of research that I didn't even know existed. I have so many new things to try, questions to ask, threads to pull. I couldn't have asked for a better introduction to Computational Linguistics.

I'd also like to thank my committee for their careful reading of my thesis. As a Master's student, it's a real privilege to have four genuine experts read my work and talk with me about it. It was a great experience.

I owe a special thanks to Thomas Brochhagen for introducing me to Signaling Games and for our valuable conversations about polysemy. I was glad to be able to add some simulation results to the thesis and I'm looking forward to doing more work in that area.

I'd like to thank my friends in the MoL and from Sarphatistraat for making the last two years fun, interesting, productive, a time I will remember well. I could not have made it through without all of you.

And finally, my incredible family and friends back home who've put up with this crazy stuff I've been doing. I'm so lucky to have your love and support. Thank you.

Contents

Chapter 1

Introduction

1.1. Motivations

Semantic Coordination The main motivation of this thesis is to develop a lexical framework that supports a theory of semantic coordination. Semantic coordination is a joint action that linguistic agents perform as part of a discourse, often with the goal of improving semantic alignment.

In dialogue, people produce and interpret utterances according some semantic schema that represents the meaning of relevant expressions. *Semantic alignment* depends on how much agents agree on the meaning of an expression as it is used in context. But simply agreeing on a meaning is not always enough; even perfectly aligned semantic schemas may be better or worse at facilitating communication depending on the communicatory goals of the agents involved. Therefore it is also important that the agreed upon representations can express the right concepts. The *local expressivity* of a semantic schema depends on how well it meets an agent's immediate communicatory goals in this regard. Thus semantic coordination has two main goals:

1. improve semantic alignment and

2. improve the local expressivity of the aligned semantic schema.

There are various strategies by which agents improve semantic alignment and local expressivity. Not all of these strategies require coordination; an agent may independently realize that his representation of some expression's meaning differs from that of his interlocutor and unilaterally align his representation to her's in response. Improvements to expressivity do require semantic coordination, however. Agents finding their semantic schema wanting for expressivity will seek to develop new representations that facilitate better communication. In order to preserve alignment, new meanings must be arrived at jointly:

(1) A: A docksider.
B: A what?.
A: Um.
B: Is that a kind of dog?
A: No, it's a kind of um leather shoe, kinda preppy pennyloafer.
B: Okay okay got it.

In this dialogue (Brennan and Clark, 1996), A has an adequately expressive expression (*docksider*), but finds that it is not aligned with B's lexical scheme. The speakers then coordinate an expression to refer to the shoe in question (or perhaps to the kind of shoe that it is). This coordination affects their shared lexical schema—afterwords they refer to it as *the pennyloafer*.

Semantic coordination has long been observed by psycholinguists (e.g. Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987; Brennan and Clark, 1996) and has more recently been recognized as an important phenomenon in dialogue modeling (e.g. Pickering and Garrod, 2004; Ginzburg, 2012). In contrast to existing models, this thesis seeks to develop a framework in which semantic coordination is the driver of long-term lexicosemantic change. When people engage in semantic coordination to meet an immediate communicatory goal, they make use of their existing semantically aligned lexical resources to do so. Rather than coining a new expression to take on the coordinated meaning, agents typically attach it to an existing expression (as is done with *pennyloafer* in example 12) The "conventional" meaning is then ignored in favor of the coordinated *ad hoc* meaning. If an expression is used on an *ad hoc* basis to communicate the same meaning with enough frequency, it may, under the right circumstances, be lexicalized as one of the expression's conventional meanings. Since *ad hoc* usages of expressions are typically semantically related to their conventional meanings, this naturally leads to lexical polysemy. Thus any lexical model that is dynamically dependent on semantic coordination must also deal with polysemy. Semantic coordination, in turn, must itself deal with the fact that the expressions from which meanings are coordinated may themselves be polysemous.

Lexical Semantics as Formal Semantics Semantic theories have classically assumed that words are endowed with universal, immutable meanings. This assumption is not tenable when linguistic agents are constantly coordinating new semantic conventions. We must use the lexicon to interpret expressions rather than reading their meanings off a fixed interpretation function. Furthermore, we must account for the fact that the very utterances whose meaning we seek to determine may be used to change the lexicon itself (i.e., through semantic coordination). Thus another motivation of this thesis is to develop a lexical model that can be incorporated into a formal semantic theory in a way that it both provides input in the form of semantic interpretations and accepts changes in the form of semantic coordination.

A lot of contemporary formal semantics is motivated by the attempt to address an ever-broader range of natural linguistic phenomena. Partee (2011) calls this movement the "naturalization of formal semantics". The naturalization of formal semantics includes such developments as situation semantics (Barwise and Perry, 1981), discourse representation theory (Kamp, 1981), and dynamic predicate logic (Groenendijk and Stohof, 1991). Semanticists in this tradition seek to address more of the cognitive and interactive aspects of language use while maintaining a connection to the early success of formal semantics. The goal of bringing lexical semantics—particularly dynamic lexical semantics—into the fold of formal semantics makes this thesis part of that tradition.

Applied Formal Semantics A final motivation for this thesis is a consideration for its possible applications both in dialogue systems and in data-driven semantics. The possible applications to dialogue systems are clear: Since semantic coordination is part of dialogue, artificial linguistic agents must be able to engage in it in order to naturally interact with humans. Furthermore, agents need to know when coordinated meanings are available for use and how they relate to an expression's conventional meaning.

Applications in data-driven semantics are somewhat less overt, but perhaps even more immediately promising. The distributional hypothesis (Harris, 1954)—the idea that words appearing in similar contexts tend to have similar meanings—has been extremely productive in statistical models of semantics (e.g., Lenci, 2008; Turney and Pantel, 2010). It does have well-known limitations, however. One of these limitations stems from the fact that the contexts in which *ad hoc* usages of a word appear may be distributionally distinct from those where it is used with its conventional meaning. Statistical models that take semantic coordination into account will do better both at using *ad hoc* contexts to compute conventional meanings and at predicting the *ad hoc* meanings a word may take on in other contexts.

1.2. Contributions

This thesis has four main contributions:

1. The Interactive Lexical Hierarchy (ILH), a lexical model for dialogue that allows for polysemy, semantic coordination, and linguistic change;
2. a formalization of the ILH based on Type Theory with Records (TTR);
3. a theory of semantic coordination based on the ILH; and
4. empirical results supporting the predictions made by 1 and 3.

In the ILH, an agent's semantic resources are represented by a hierarchy of lexicons where more general lexical resources are kept in higher level lexicons and more specific resources in lower level ones. In a given discourse, agents deploy a lexicon or not depending on what common ground is shared among the interlocutors. As the discourse progresses, agents jointly develop lexicosemantic resources at the lowest level of the lexical hierarchy. These jointly constructed resources then may or may not work their way into agents' higher level resources. The ILH is bound to break down in some places and simply fail to address other aspects of what it is modeling, but there is good coverage of the core phenomena that motivate this thesis. Furthermore, this thesis explores one possible formalization where meanings are represented using TTR. An important aspect of this formalization is that it demonstrates the compatibility of the ILH with traditional formal semantics.

The main goal of the ILH is to give an account of semantic coordination in which coordination plays a central role in driving semantic change—both on a personal level (language learning) and on a community level (language change). The theory of semantic coordination is given in two parts: A semantic theory that describes how different strategies of semantic coordination interface with the lexical hierarchy in general, and a pragmatic theory that gives accounts of particular coordination strategies using that framework.

In addition to the more theoretical contributions, the empirical research presented in chapter 6 is consequential in its own right. Section 6.1 uses a signaling games simulation to provide evidence for the claim that there is a connection between semantic coordination and polysemy. Section 6.2 is a corpus study of a linguistic phenomenon known as *contrastive focus reduplication* (Ghomeshi et al., 2004) that gives evidence for the semantic and pragmatic accounts given in chapter 4.

1.3. Overview

The thesis is structured as follows: Chapter 2 provides psycholinguistic background justifying the cognitive principles on which the ILH depends and introduces the type system on which its formalization is based. Chapter 3 sets up some semantic framework including how conceptual space is structured and the interpretation of expressions. Chapter 4 introduces the ILH and gives a theory of linguistic coordination. Chapter 5 formalizes the work presented in chapters 3 and 4 using TTR. Chapter 6 presents the results of a corpus study and a signaling games simulation inspired by issues raised in chapter 4. Finally, chapter 7 offers some concluding remarks.

Chapter 2

Background

The present chapter gives background on work that parts of this thesis immediately depend upon. Sections 2.1 and 2.2 introduce the psycholinguistic theory that justifies cognitive principles underlying the ILH and section 2.3 introduces the type system on which the its formalization is based.

2.1. Relevance Theory

According to Grice (1961), speakers produce utterances that automatically create expectations leading listeners to the intended meaning. Those expectations of listeners are the maxims of Quality (truthfulness), Quantity (informativeness), Relation (relevance), and Manner (clarity) (Grice, 1975). Relevance theory (Sperber and Wilson, 1986) questions the necessity of three of those maxims, suggesting that relevance is enough to guide listeners to the intended meaning. Furthermore, relevance theorists claim that relevance is not just a communicative convention, but rather a basic feature of human cognition that communication may exploit. The *cognitive principle of relevance* states that human cognition tends to be geared towards the maximization of relevance (Wilson and Sperber, 2008). To communicate effectively, people must be relevant. This usually, but not always entails satisfying Grice's other maxims. Brennan and Clark (1996), for example, find that speakers are often more informative than they need to be when deciding how to refer to objects.

Relevance is a property that can hold of both external stimuli and internal representations—anything providing input to cognitive processes. Those inputs that positive cognitive effects—that is, they produce a "worthwhile difference in the individual's representation of the world" (Wilson and Sperber, 2008)—are considered relevant. For example, cognitive effects such as making contextual implications and modifying assumptions are positive if they lead to true conclusions. Relevance also depends on cognitive processing effort. Inputs that require more cognitive processing effort to produce a positive effect are less relevant.

The cognitive principle of relevance affects the way that people communicate—both in the way that they produce and interpret utterances. Communication typically follows the form of *ostensive inferential communication* which includes:

1. the intention to inform an audience of something and

2. the intention to inform the audience of one's informative intention.

That is, communication where it is intentionally made apparent to the audience that the communication is directed at them. Compare ostensive inferential communication to the properties of a shared basis (§2.2.1). Information that is conveyed by ostensive inferential communication is eligible to be part of the common ground between speaker and audience.

The *communicative principle of relevance* states that every ostensive stimulus conveys a presumption of its own relevance. The argument for it is as follows: Suppose that the speaker makes an ostensive stimulus. By definition, she intends to inform an audience of something. The cognitive principle of relevance tells the speaker that the audience tends to interpret stimuli in a way that maximizes relevance. Since the cognitive principle of relevance is common knowledge, and since the stimulus produces by the speaker is ostensive, the audience knows that the speaker will expect them to interpret her stimulus in this way. Thus the audience is justified in presuming that the speaker, intending to inform the audience of something, will craft the stimulus in such a way that takes the cognitive principle of relevance into account.

2.2. The Collaborative Model

The collaborative model (Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989; Clark, 1996) is a theory of communication where agents coordinate primarily by showing positive evidence of understanding (both implicitly and explicitly), and where that positive evidence is the basis on which coordinated meanings are grounded. Thus establishing common ground is both the object of and a prerequisite for language use.

The joint action model is a dialogue model based on the framework of the collaborative model. Dialogue models of language take communication through dialogue to be the fundamental activity of language.

Section 2.2.1 defines common ground as it is understood in the collaborative model. Section 2.2.2 gives an overview of the joint action model, a model of how the common ground is both used and developed by language use.

2.2.1. Common Ground

Stalnaker (1978) proposes common ground as a kind of common knowledge that has real-world applications. The common ground for a group of agents (a community) is defined by way of a *shared basis*. A basis is a *shared basis* for a community C if and only if:

1. every member of C has information that b holds, and

2. b indicates to every member of C that (1) is the case.

The *principle of justification* says that agents take a proposition p to be common ground only when they believe there is a shared basis indicating that p.[1]

While common ground has been defined from an objective viewpoint, it is important to note that only an omniscient being could truly know whether a basis is shared. What we are really interested in is agents' *beliefs* about the common ground. In general agents will take a proposition to be common ground if they think there is sufficient evidence that a shared basis for that proposition exists. What constitutes sufficient evidence—the *grounding criterion*—will depend on the purposes at hand.

Each agent maintains her own version of the common ground for each community. There may be discrepancies in what agents A and B find to be common ground for a community C (say, the participants in a discourse). Often these discrepancies will be small and go unnoticed, but sometimes they must be rectified. In dialogue, those discrepancies are typically handled with clarification requests.

Clark (1996) distinguishes between two kinds of common ground: Communal common ground, which is established on the basis of belonging to some cultural community, and personal common ground which, has a perceptual basis. The basis for a communal common ground can be very broad. For example, we take the laws of nature and certain aspects of folk psychology as a shared basis for the common ground between most humans. Personal common ground holds between a group of individuals based on shared experience. It therefore has a perceptual basis.

There are two main kinds of perceptual bases that establish personal common ground: A joint perceptual basis is the mutual awareness of an event that indicates mutual awareness of itself. Suppose, for example, that two people hear a loud scream from the next room. This forms a perceptual basis since both persons heard the scream and the scream, loud as it was, indicated to each that the other was aware of it.

The other kind of basis for personal common ground, an *actional basis*, is the kind of basis most relevant to a theory of language use (Clark, 1996, p. 114). One might say, for example, that if A utters σ to B, then under the right conditions, the action of A's utterance is taken as a shared basis for A and B to ground the fact that A asserted p, where p is the proposition (jointly taken to be) expressed by σ. Crucially, an action basis requires some preexisting common ground in order to serve as a shared basis. In the example above, A and B need a shared basis for the interpretation of σ. Exactly what conditions are required to establish such a basis is fundamental to Clark's analysis of language use.

The notion that all bases require some preexisting common ground leads to an apparent problem of regress:

> *If every new piece of common ground is built on an old one, where does it start? Is there a first piece of common ground, and if so, what is it based on?*

[1] Exactly what kind of thing propositions and bases are, or what it means for a basis to *indicate* proposition are left open. For now, it suffices that a proposition is the kind of thing that can be *believed* or that an agent can *be aware of*, and that a basis is the kind of thing that may *cause an agent to believe* or to be *made aware of* a proposition. A basis may itself be a proposition or collection of propositions, or it may have non-propositional aspects such as sense data.

> *The paradox is more apparent than real. Each of us has built up information about others from infancy. Originally, we may have taken much of this information as common ground—as children often do—with out a proper basis. Children first appear to thing that their interlocutors are omniscient, and it is only with age that they set higher standards.* *(Clark 1996, p. 120)*

Thus common ground should not be treated as an idealized form of reasoning, but as cognitive theory of how agents model others' beliefs. Agents are subject to mistakes in what they take as common ground, and grounding requirements depend on the utility of taking something as common ground as well as the consequences of being wrong.

2.2.2. Joint Action Model

The joint action model (Clark and Schaefer, 1989) is a dialogue model that views communication as the process of establishing common ground. Coordinating any kind of joint action requires that agents involved share some common ground.

There are four level of common ground that are maintained as part of a joint activity (table 2.1). To achieve some level of grounding, positive evidence—not just the lack of negative evidence—must be given that can serve as the basis for adding to the common ground at that level (Clark, 1996, p. 225). Communication that gives evidence for grounding is called feedback. Feedback is meta-communicative; it does not necessarily assert anything to do with the topic of conversation, but rather serves to establish a basis on which contentful utterances can be grounded.

Taking assertion as an example of a linguistic action requiring grounding, suppose that A makes some assertion p. To ground that assertion at level 3, B must give feedback indicating that he has *understood* the assertion so that "A asserted that p" may be added to the common ground. To ground the assertion at level 4, his feedback must indicate that he *accepts* the assertion. Then p may be added to the common ground.[2]

	Level	**Grounding Action**
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 2.1.: Grounding levels and feedback for dialogue (Fernández, 2014)

Each successive level of grounding requires that all of the lower levels are established. It is possible to achieve multiple levels of grounding with the same grounding action, though. By the *principle of opportunistic closure*, agents assume that lower level ground is established from evidence for higher level grounding, so feedback indicating

[2]This thesis is mostly concerned with grounding (and failure to ground) at level 3; however, it is sometimes unclear where a grounding failure occurred.

understanding carries with it evidence of contact and perception. Likewise, feedback indicating uptake is evidence for understanding.

2.2.3. Coordination Problems

A *coordination problem* (Schelling, 1960) is a situation where two or more agents must choose between some alternative actions and where the outcome of those actions (positive or negative) is determined jointly by all of the actions taken. Thus agents will seek to coordinate their actions to receive a positive outcome. A simple example is two people who are trying to meet up with each other. Perhaps it doesn't particularly matter where the meeting takes place, but in order for a positive outcome to occur (i.e., they actually meet), the two people must coordinate their actions and go the same place. To make things somewhat more complicated, some positive outcomes may be better than others. Suppose that they each prefer the park over a café. Meeting in either location is a positive outcome as long as they both show up, but supposing it is common ground that the park is preferred, they should be able to infer to go to the park even without communicating explicitly.

Achieving semantic alignment is a similar coordination problem. The actual content of linguistic conventions (such as word meaning) matters little in the sense that communication would be just as successful if, for example, *cat* meant dog and *vice versa.* It does matter that agents coordinate their language so that what I mean by *cat* is the same as what you think it means, and so on. Furthermore, good expressivity is preferred as it facilitates better communication. If *cat* meant dog and *dog* meant dog, the convention would still be in semantic alignment, but it may be somehow suboptimal if one of the agents needs to talk about cats.

In the first example, the agents were able to implicitly coordinate their meeting location based on some prior common ground (their mutual preference for parks over cafés). A similar thing happens with semantic alignment. Recall that by the communicative principle of relevance, every ostensive stimulus conveys a presumption of its own relevance. Thus agents will assume that expressions take on their most relevant interpretation in a given context. Clark (1996) calls this the *principle of joint salience.*

2.3. Type Theory with Records

A formal theory of semantic coordination needs to represent semantic units in a way that allows the meaning of expressions to be updated compositionally. When two agents coordinate the meaning of an expression, they start with their own personal cognitive representations of what it means. The meaning that they negotiate must be a function of those representations as well as the semantic or pragmatic circumstances under which the coordination takes place. This function models the agents' cognitive processes in constructing the new representation; it must treat semantic units not as atomic objects, but as entities with components that can be compared and manipulated. Type theory with records (TTR) (Cooper, 2005a) exhibits special *record types* which meet these requirements. Additionally, TTR provides a compositional framework that promises compatibility with traditional Montagovian formal semantics.

This section gives an overview of the aspects of TTR that are especially relevant for the semantic framework on which the interactive lexical hierarchy is built. The formal details of the type system as it is applied in this thesis can be found in appendix A. A more exhaustive introduction can be found in Cooper (2012). Cooper (2005b, 2008) discuss TTR as a unifier of various semantic theories (discourse representation theory, situation theory, head-driven phrase structure grammar, unification grammar, and others). Cooper and Larsson (2009) specifically uses TTR for interactive lexical dynamics (though in this regard it is used somewhat differently here—see §4.3). Finally, TTR is applied in the KoS dialogue framework (Ginzburg, 2012). Cooper and Ginzburg (2015) gives a detailed look at how TTR unifies KoS's analysis of dialogue structure with some of the more traditional concerns of formal semantics.

Records and record types are feature structure-like entities that encode structural information in a type theoretic framework. A record is an object that consists of a finite set of labels with associated objects. Likewise, a record type is a finite mapping from labels to types. Records and record types are typically displayed in tabular form.

$$\texttt{librarian} = \begin{bmatrix} x & : & \mathit{ind} \\ y & : & \begin{bmatrix} z & : & \mathit{ind} \\ c_2 & : & \mathit{building}(z) \\ c_3 & : & \mathit{lends\_books}(z) \end{bmatrix} \\ c_1 & : & \mathit{work\_at}(x, y) \end{bmatrix}$$

This example demonstrates some of the key aspects of record types.[3] First, note that record types may be nested. Here `librarian` is defined as an individual who works at a library, which itself is a record type defined as a building that lends books.[4]

TTR is a Martin Löf-style intuitionistic type theory. Martin Löf type theory is typically more closely associated with constructive foundational mathematics than with natural language semantics. In such uses, types typically represent something like propositions and objects are something like proofs. Thus the central relation in type theory—the type judgment relation—relates propositions to their proofs. The judgment $a : t$ (a is of type t) is typically interpreted as meaning that a is a proof for of t.

In cognitive models of semantics, types and objects may represent cognitive aspects of the semantic agents in question. In this thesis, record types represent concepts (see section 3.1). Objects are agents' cognitive representation of aspects of the world (individuals, events, etc.). Type judgments represent an agent deciding wether a given object is an instance of some concept. Given a record type p and a record r, r is of type p ($r : p$) if for each label l in p, l is a label in r and the object corresponding with l in r is of the type of the type corresponding with l in p (definition A.0.5).

[3] Terminal leaves of record types contain basic types or predicate types. In this case *ind* is a basic type for individuals.

[4] Of course this example is simplified for expository purposes, but the utter insensibility of this as a definition for *librarian* may cause the reader to doubt that it is possible to come up with a TTR record that represents the concept at all. This difficulty is further discussed in section 3.1.1.

Consider the following record as a possible representation of the individual Jack:

$$\texttt{jack} = \begin{bmatrix} x & = & jack \\ y & = & lib\_a \\ c_1 & = & obs_{123} \\ age & = & 31 \end{bmatrix}$$

The agent judges that Jack is a librarian on the basis of the contents of labels x, y, and c_1; that is that he is an individual ($jack : ind$), that there is a library ($lib\_a : \texttt{library}$),[5] and that Jack works at the library ($obs_{123} : work\_at(jack, lib\_a)$). Note that the record may contain additional information about Jack (for example, that he is 31 years old) that do not factor into this particular judgment.

Exactly what kind of thing objects ($jack$, obs_{123}, etc.) are depends on the cognitive features of the agent in question. One suggestion (Cooper, 2005a) is that they are something like infrons from situation theory (Barwise and Perry, 1981). For artificial agents, they may be sensor reading (or enriched representations there of) (Cooper and Larsson, 2009). Whatever cognitive entities objects represent, the important thing is that agents make judgments about basic types (i.e., that says $jack$ is an individual or not) and predicate types (i.e., that say obs_{123} confirms that $jack$ works at lib_a). Thus agents must have some model of the world that relates objects to basic and predicate types. This requirement gives TTR a model theoretic flavor.[6]

This section has given a fairly informal overview of Type Theory with Records as it is used in this thesis. A more formal specification of the type system can be found in appendix A.

[5]Note that lib_a must itself be a record.

[6]Although technically it is not model theoretic since the "model" also effects which types can be formed and therefore what the well-formed terms of the language are (see Cooper (2012) for details).

Chapter 3

Semantic Framework

Before presenting the ILH, some background is required. Section 3.1 describes concepts, the objects of lexical interpretation, while section 3.2 tells how expressions relate to these objects through semantic interpretation. This chapter and the next discuss important choices that are made in the semantic model and attempt to justify these choices linguistically. The exact formalization is not immediately important in these discussions—though it is of course important that a suitable one exists. Chapter 5 will pick up the formal details of what is proposed in chapters 3 and 4, offering a possible formalization using Type Theory with Records. Where appropriate, these chapters make reference to theorems and definitions found in chapter 5.

3.1. The Conceptual Domain

Concepts are tricky philosophically. This thesis understands *concept* in a way that may not be the most common or philosophically popular. Concepts are used here as the basic constituents of semantic content.[1] They may be thought of as something like Fregean senses.[2] That concepts are Fregean senses is a position defended by some philosophers (e.g. Peacocke, 1992). The difficulty with adopting an unqualified version of that view is that Fregean senses are an abstract, platonic object, while concepts cannot be entirely abstract entities for our purposes. Given that this thesis is developing a cognitive model of lexical change, it is important that concepts meaningfully relate to agents' cognition. The more traditional view that concepts are just mental states or representations of states is problematic too since that may make concepts too fine grained to, for example, be compared across agents. To avoid having to develop a whole philosophical theory, concepts will just be taken to be some kind of abstractions *over* mental representations that record all of the semantically salient features of those representations. Of course as a definition this is a bit circular since concepts are used here as semantic atoms, but a more expository description is beyond the scope of this thesis.[3]

[1]In fact, the meaning of an expression is not exactly a concept, but rather meanings are multi-pointed sets of concepts (§3.2).

[2]Indeed, *sense* and *concept* refer to the same kind of object in this thesis, though *sense* is reserved for discussing concepts qua interpretation of expressions, while *concept* does not carry any necessarily linguistic connotations.

[3]Indeed, the divide between cognitive and abstract semantic content is arguably at the center of what separates traditions that study natural language as an empirical, essentially social phenomenon

The *conceptual domain* is part of the cognitive resources of linguistic agents. It consists of those concepts of which the agent can conceive, in addition to some structure that lets agents reason about their concepts. To be part of an agent's conceptual domain, it is not required that a concept is actively being conceived or even that it has ever been conceived by the agent at some point in the past, but merely that the agent have all of the cognitive resources necessary to conceive it. Since concepts provide provide semantic content, the conceptual domain is an aspect of agents' linguistic resources. Furthermore, language may influence an agent's conceptual domain; it is often through linguistic interaction that new concepts are learned or old ones are updated. With this in mind, it is important to note that the conceptual domain is not itself an essentially linguistic object. Concepts can just as well be formed by non-linguistic cognitive processes using, for example, sensory data.

The following section presents a model of the conceptual domain that represents concepts as record types from Type Theory with Records (see §2.3). TTR is suitable for these purposes because it is easy to define a subconcept/superconcept relation and a measure of conceptual similarity. It is also notable that many other advancements in the naturalization of formal semantics can be encoded in TTR. However, there may be other equally well suited ways of formalizing the concept domain. One possible alternative is Formal Concept Analysis (FCA) (Ganter and Wille, 1996). FCA may be better suited for computational methods since it gives an easy way of updating the concept domain given observational input. That said, FCA concepts do not have the nice ready-to-use logical properties of TTR record types.

3.1.1. The Concept Domain: A Lattice of Records

A TTR-based conceptual domain is defined over a set of basic types, and predicate types (with respective arities). Record and function types may be generated from there.

In examples, things like individuals (*ind*), real numbers ($\mathbb{R}$), times (*time*), events (*event*), and locations (*loc*) are generally taken as basic; however no especially strong claims about human cognition or its ontology are intended by doing so—analyses that use fewer basic types and make heavier use of predicates to distinguish these classes bay also be possible.

Recall that a record type is a set of ordered pairs—specifically, it is the graph of a function with a finite domain of labels, and whose range is a set of types (possibly including other record types).

$$\texttt{bear} = \begin{bmatrix} ref & : & ind \\ c_1 & : & size(ref,\ MuchBigerThanMe) \\ c_2 & : & shape(ref,\ BearShape) \end{bmatrix}$$

An object is said to be an *instance of* or *belong to* or *fall under* a concept if it is of the type of that concept (considered as a TTR record type). The set of objects belonging

from those that study it as an abstract formal system. For now we can only observe that there is much to be gained by ignoring that divide and using the tools developed by both traditions. Certainly many insightful contributions to the understanding of natural language are made by doing just that.

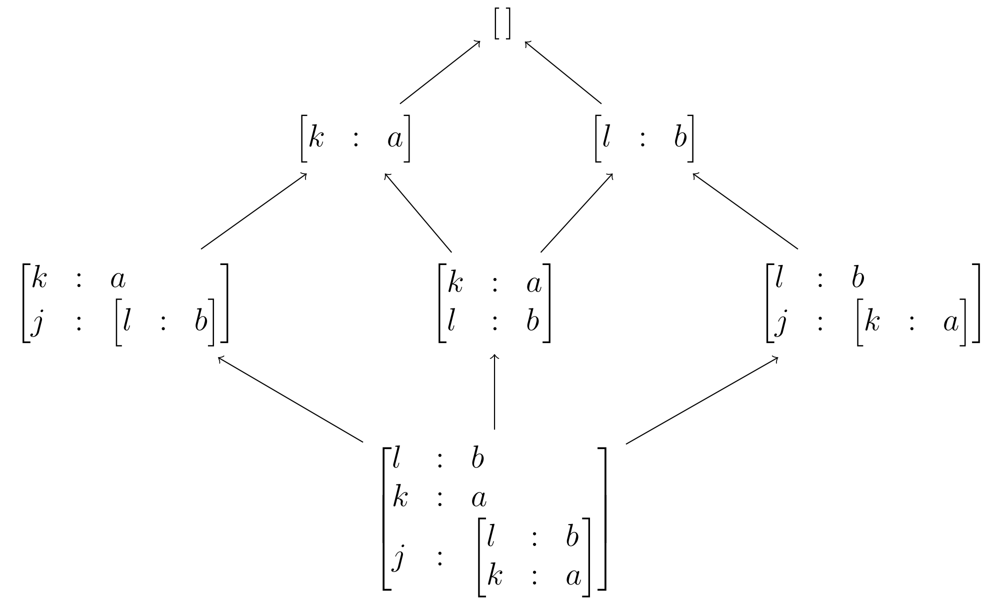


Figure 3.1.: Example concept lattice ordered by the subconcept relation

to a concept is the concept's *extension*. In the example above from Cooper and Larsson (2009), the concept `bear` has in its extension those individuals whose size and shape are of the corresponding predicate types.[4]

As discussed in section 2.3, records and record types are both examples of feature structures. A space of feature structures exhibits a natural lattice structure induced by *subsumption* (Shieber, 1986). Where record types are concerned, subsumption corresponds to the subconcept relation. A concept p is a subconcept of q ($p \sqsubseteq q$) if for ever label in $\text{dom}(p)$, there is an identical label in $\text{dom}(q)$ whose values are either equal or the p's value is a subconcept of q's value (definition 5.1.2). If $p \sqsubseteq q$, we also say that q is a *broader* concept than p or that p is *narrower* than q.

Note that there is something slightly counter-intuitive about the concept/subconcept relation since narrower concepts always have a domain at least as large as broader ones. The reason for this has to do with the way that record type judgments are made: It's easier for a record to belong to a concept with fewer constraints.[5] The most important property of concept/subconcept relation is that objects belonging to a concept necessarily belong to all of its superconcepts (proposition 5.1.1).

At this point it is worthwhile to address a difficulty that is both important and potentially elucidating. When comparing record types as concepts, predicates present a challenge. Consider the following records:

[4]We distinguish the concept name `bear` (metalanguage) from the word *bear* (object language). Recall that concepts are non-linguistic entities.

[5]By definition A.0.5, $a : P$ if and only if for each label $l \in \text{dom}(p)$ we have $l \in \text{dom}(a)$ and $a.l : p.l$; however the record may have labels outside of $\text{dom}(p)$ and those values are unrestricted by p.

$$\texttt{bird} = \begin{bmatrix} x & : & indv \\ c_1 & : & has\_fethers(x) \\ c_2 & : & bird\_shaped(x) \end{bmatrix} \texttt{eagle} = \begin{bmatrix} x & : & indv \\ c_1 & : & has\_feathers(x) \\ c_2 & : & eagle\_shaped(x) \\ c_3 & : & flies(x) \end{bmatrix}$$

Clearly it should hold that $\texttt{eagle} \sqsubseteq \texttt{bird}$; however that is not the case since there are incompatible predicates attached to the same label: $bird\_shape \neq eagle\_shape$. Perhaps we could define $eagle\_shape \sqsubseteq bird\_shape$ if and only if all eagle shaped things are bird shaped. Naturally this solution is that it brings up issues of intensionality: We wouldn't want one predicate to be a sub-predicate of another just because it happens to be extensionally included—there needs to be some explanatory force behind the subconcept relation. It is possible to extend TTR to a modal system (see Cooper, 2012, §3.3) and define an intensional sub-predicate relation; however that solution to this particular problem belies the utility of record types to analyze concepts. If all eagle shaped things are necessarily bird shaped things, then there should be a reason for that, and if TTR is going to be useful, it must be robust enough to capture those reasons. Indeed, the intuition that there is a relationship between *bird shaped* and *eagle shaped* suggests that they ultimately shouldn't be analyzed as "atomic" predicates, but rather concepts in themselves.

$$\texttt{eagle\_shaped} = \begin{bmatrix} x & : & indv \\ y & : & shape \\ c_1 & : & shape\_of(y,x) \\ c_2 & : & quality_1(y) \\ c_3 & : & quality_2(y) \\ c_4 & : & quality_3(y) \end{bmatrix} \texttt{bird\_shaped} = \begin{bmatrix} x & : & indv \\ y & : & shape \\ c_1 & : & shape\_of(y,x) \\ c_2 & : & quality_1(y) \\ c_3 & : & quality_2(y) \end{bmatrix}$$

Unsurprisingly, it becomes difficult to make elucidating examples at this level of granularity. Predicates like $quality_1$ signify some property of however things of type *shape* are represented by the agent. Exactly what those properties are depends heavily on the cognitive properties of the agent involved. For artificial agents, it may be some feature of a vector representation of the shape obtained with computer vision techniques.

Another important aspect of concepts is that they can be combined to form new concepts. Two such operations are defined formally: Given any set of concepts, the *generalization* or *join* of those concepts is the smallest concept that is a superconcept of every thing in the set (lemma 5.1.5).[6] The join of p and q is denoted $p \vee q$.

If two concepts share a subconcept, the largest such subconcept is called their *merge* or *meet* (proposition 5.1.8).[7] We give a definition of merges that can easily be adapted as an algorithm for finding it from the constituent concepts, but it is important to note that the merge of two concepts does not always exist. Where it does exist, the merge of p and q is denoted $p \wedge q$.

[6] Formally the join is defined as a binary operation, but since we are working with finite conceptual domains, it can also be used to find a least upper bound for any subset of the domain.

[7] Also see Cooper (2012, §3.2).

3.1.2. Vague Concepts

In terms of issues of imprecision, thesis is primarily concerned with ambiguity and broadness and does not deal explicitly with vagueness; however it is important to note that there are resources available to integrate many popular accounts of vagueness with a concepts-as-record-types model. One such tool is probabilistic TTR (Cooper et al., 2015). Rather than a precise threshold for concepts such as `tall`/ ¬`tall`, the probabilistic record types give a noisy threshold.

An earlier version of probabilistic TTR is used to give an account of learning vague concepts where perceptual knowledge is integrated a little at a time (Fernández and Larsson, 2014). In that account, judgments about vague concepts are represented as probabilistic linguistic knowledge (cf. Lassiter, 2011).

Probabilistic TTR of either the metaphysical or epistemic variety is subject to many of the same objections that fuzzy logic encounters when dealing with vagueness (cf. Kamp and Partee, 1995). More research is needed to see what frameworks that handle vagueness well also meet the requirements to formalize the work in this thesis.

3.1.3. Conceptual Similarity

One of the great successes of distributional models of semantic content is the ability to easily define similarity metrics on the space of meanings (e.g., Resnik, 1995). Distributional representations, however, generally do not project logical structure, which makes them difficult to work with compositionally. One strength of using feature structures to represent concepts is that they admit similarity metrics while being themselves logical structures. This fact is one of the oft-cited advantages of TTR (Larsson and Cooper, 2009; Cooper, 2012); however there do not appear to be any examples of such metrics in the existing literature. This section, gives one possible metric for quantifying the similarity of two records. Let p and q be two concepts belonging to the same conceptual domain. The similarity of p and q is defined as follows:

$$sim(p,q) = \frac{|\{l \mid p.l = q.l\}| + sim'(p,q)}{|\operatorname{dom}(p)| + |\operatorname{dom}(q)|}$$

where $sim'(p,q)$ sums the distances between p and q's values which are not identical, but nonetheless are record types:

$$sim'(p,q) = \sum\{sim(p.l, q.l) \mid p.l \neq q.l \text{ and } p.l, q.l \in \mathbf{RType}\}$$

The similarity of p and q is a number between 0 and 1, where 1 means that they are the same concept and 0 means that the have nothing in common (definition 5.1.5). The similarity measure essentially reports the proportion of the labels of p and q that have counterparts with the same value in the other concept (with the caveat that "same value" may actually be a measure of similarity if they are themselves both record types).

Since the meaning of a word is identified with a set of concepts (and not a single concept), the similarity measure given in this section is not on its own a measure of semantic similarity, but it will serve as the basis for one (see §3.2.3).

3.2. Interpretation of Expressions

Recall that the motivation for defining a concept space was to provide the basic atoms on which semantic units—meanings—are built. Expressions may express concepts, but the overall meaning of an expression has more structure. This section introduces that structure.

3.2.1. Polysemy: Lexical Interpretation Sets

So far, two two sources of semantic imprecision have been addressed: concept vagueness and concept breadth. Both of these phenomena are fundamentally non-linguistic since they have to do with concepts. The source of imprecision with which this thesis is primarily concerned is *lexical ambiguity*. A lexicalized expression is said to be ambiguous if it has more than one interpretation. Certain expressions with multiple related interpretations are said to be *polysemous*. Expressions with multiple unrelated expressions are homophonous. What makes some expressions polysemous and others homophonous is a difficult question. The notion of polysemy employed in this thesis will be introduced in the next section. For now we focus on lexical ambiguity in general.

This analysis claims that any completely disambiguated interpretation refers to a single concept. That concept may still be broad. It may also be vague, but it is a single concept. The meaning of a word consists of a set of such concepts—its *interpretations* or *senses*. Sometimes, putting all of these interpretations together in the same set seems quite strange.

Consider the word *fruit*. It seems that there are two major classes of interpretation for *fruit*: There are technical interpretations, wherein a fruit is part of a plant identified by the specific botanical tissues it comes from, or by its biological function (seed dissemination, etc.), and there are colloquial interpretations where fruits are identified by their texture and flavor, or by their culinary function.[8] Perhaps it is natural to assume that the lexical semantics of *fruit* should represent some of the structure suggested by these two classes of interpretation (technical/botanical versus colloquial/culinary). Moreover, although we have discussed the noun interpretations of *fruit*, it also has verb interpretations as in *the apple tree did not fruit this year* (meaning that it did not produce any apples). Surely the lexicon makes a structural distinction between verb uses and noun uses of the same word? Section 3.2.2 introduces some additional structure on the interpretation set, accounts for the intuition that there are distinct classes of the senses of words like *fruit*; however it is not always so clear when a sense belongs to one class or another—there may very well be interpretations of *fruit* that combine botanical and culinary criteria. For that reason the semantics of an expression are analyzed as a flat set that includes all of its interpretations.

The remainder of this section, gives two additional justifications for analyzing the semantics of an expression as a flat set. The first reason has to do with the requirements

[8] Of course the ambiguity does not end there, for even among the technical definitions there are multiple interpretations (exactly which botanical parts of the plant count as a fruit, etc). Likewise, there are ambiguities in the colloquial interpretations (is a tomato a fruit? an avocado?).

of action bases for establishing common ground. Recall (§2.2.1) that an action basis is a joint action that establishes some common ground. Suppose that A says to B:

(1) A: Elizabeth is going to bring fruit to the picnic.

Under the right circumstances, the proposition p that asserts that Elizabeth is bringing fruit will be added to the common ground. Let e be the event where A said 1 to B. The agents will consider p to be common ground if:

1. A and B are aware of e,
2. e indicates to both A and B that each was aware of e, and
3. e indicates to both A and B that A asserted p.

Propositions, unlike utterances, are not ambiguous, which makes this third requirement especially problematic. Consider for now only the ambiguity in 1 arising from the polysemy of the word *fruit*. In order for p to be added to the common ground, e must indicate that A's use of *fruit* refers to precisely one concept. Certainly e suffices to rule out a lot of possible interpretations of *fruit*. Pragmatic context dictates that Elizabeth is not bringing poisonous berries, for example (since one usually brings food to a picnic). Syntactic context dictates that *fruit* is to be interpreted as a mass noun (rather than given one of its singular or verbal interpretations). There may be additional facts in the common ground that rules out other interpretations. However, it is very unlikely that e indicates a *single* interpretation out of all of the possible interpretations of *fruit*. Thus the assertion that e indicates (if e is to indicate any assertion at all) must remain ambiguous. What gets added to the common ground, then, is a set of propositions $\mathbf{p}^i$, those propositions corresponding to concepts from the set of interpretations of *fruit* that are not ruled out by e. In support of this approach, Palmer et al. (2007) finds that human inter-annotator agreement on fine-grained senses is only around 73%, and that testing with sets of related senses instead reduces the occurrence of disagreements by more than a third. Inter-annotator agreement is a good indicator for what may be added to the common ground since it is in the nature of the common ground that interlocutors must agree on it.

The second justification for analyzing the semantics of an expression as a set of interpretations has a more cognitive basis. Reading time studies suggest that all of the meanings of an expression are primed even within a disambiguating context (William Onifer, 1981; Seidenberg et al., 1982). These studies support maintaining a single set of interpretations even for homophones that have little or no semantic relation to one another.[9]

[9]Note, however, that separate lexical entries are still assumed for heterophonic homographs (e.g., wĭnd/wīnd). Frost and Bentin (1992) find that the priming effects for of heterophonic homographs are weak compared to homophonic homographs—subordinate interpretations are only available after 250 ms as opposed to 100 ms in the later case.

3.2.2. Prototypicality and Polysemy: Multi-Pointed Interpretation Sets

Prototype theories of concepts have been developed to deal with so-called prototypicality effects wherein some members of a class are seen as more typical of the class than others. To take a classic example, robins are reported to be more typical of the class *bird* than are penguins (Rosch, 1973). In contrast with prototype theories that make use of fuzzy membership criterion (e.g., Osherson and Smith, 1981), Kamp and Partee (1995) offer a version of prototype theory that is compositional. This account likewise seeks to maintain compositionality while accommodating prototypicality effects used in the pragmatic analysis of certain semantic coordination strategies (§4.3.1 and §4.3.3); however the prototypes employed here are purely semantic—not conceptual.

Kamp and Partee (1995) discuss the difficulty separating semantic and conceptual prototypicality effects. In the system detailed in this thesis, concepts can be fuzzy or sharp and they can be broad or narrow, but prototypicality effects are only explained with respect to the denotation of an expression. In other words, if some birds seem to be more typical birds than others, it is not because the concept to which the word *bird* refers has more or less prototypical instances, but rather because the word itself has many interpretations, some of which are more *semantically* prototypical than others.

To achieve this, the meaning of an expression is taken to be a multi-pointed set, i.e., a pair consisting of the interpretation set (as discussed in the previous section) and a (possibly empty) set of distinguished elements from that set—the prototypes:

$$\mathbf{p} = \langle \mathbf{p}^i, \mathbf{p}^p \rangle$$

where $\mathbf{p}^i$ is called the interpretation set and $\mathbf{p}^p \subseteq \mathbf{p}^i$ is called the prototype set. In addition to distinguishing a special set of concepts as an expression's prototype interpretations, this set induces a graded prototypicality measure on the whole interpretation set by considering how similar an interpretation is to its closest prototype.

In some prototype theories (e.g., Osherson and Smith, 1981), membership in a concept's extension is graded according to its similarity to the prototype. Such approaches have been criticized on the grounds that even though some things may be more prototypical of a class than others, the actual membership criterion for the class may still be sharp (Armstrong et al., 1983): Even though a robin is more a prototypical bird than a penguin, it is still true without qualification that penguins are birds. The prototype theory presented in this thesis accounts nicely for that intuition: Penguins fall under a concept in the interpretation set of *bird*, but not in its prototype set and robins fall under a concept that is in both the prototype set and its interpretation set.

In addition to modeling prototypicality effects, the prototype set let us make a natural distinction between polysemous and homonymous interpretations. Each prototype interpretation of an expression gives rise to a different *sense class* of that expression. The sense class associated with a given prototype p is the subset of the interpretation set whose members are most conceptually similar to the prototype in question (definition 5.2.2). Note that this operation, induces a near-partition on $\mathbf{p}^i$.[10]

[10]It is only a near-partition because there may some interpretation q and prototypes p and p' such

Now an expression is polysemous to the extent that it has related senses; that is, multiple interpretations belonging to the same sense class. A classic example illustrating the difference between polysemy and homophony is the interpretations of *window* and *bank*. Of course the actual semantics for these two words may be significantly more complicated, but for illustration purposes, consider the following possible interpretations:

$\texttt{bank}_1$ - a financial institution that lends money, etc.

$\texttt{bank}_2$ - the building that houses such a financial institution

$\texttt{bank}_3$ - the side of a river

$\texttt{window}_1$ - a non-door opening in the facade of a building

$\texttt{window}_2$ - the glass, etc. that goes in such an opening

$\texttt{window}_3$ - $\texttt{window}_1 \vee \texttt{window}_2$

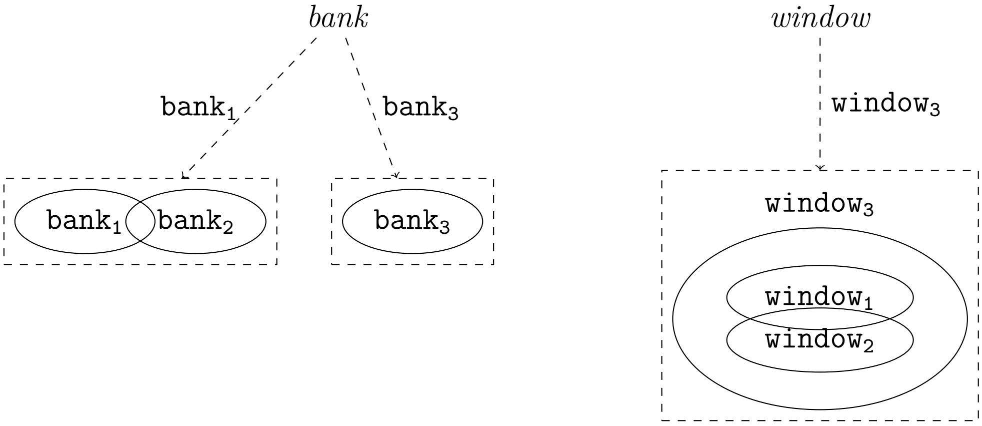


Figure 3.2.: Homonymy and polysemy in *bank* and *window*.

These interpretations may not align with everyone's intuitions—perhaps the most tenuous among them is $\texttt{window}_3$, but consider a sentence like:

(2) Frederick looked out the window, but it was dirty.

In such a sentence, *window* refers to both the glass that covers the hole in the wall and the hole in the wall itself, conceptualizing them as one object. This interpretation is taken to be more prototypical than either of the narrower interpretations.

In figure 3.2, dashed lines indicate prototype interpretations. The dashed boxes at the end of those lines encompass interpretations that belong to that prototype's sense class. Note that in addition to the semantic relationships between interpretations there are also conceptual relationships that are not explicitly encoded in the interpretation set. For example, $\texttt{bank}_1$ and $\texttt{bank}_2$ have some intensional overlap, and $\texttt{window}_3$ is a superconcept of the other two interpretations. *Bank* exhibits homonymy because it

that $sim(p, q) = sim(p', q)$. In that case, we let q belong to both sets. In practice, most concepts are fine-grained enough that this should rarely happen.

has two prototypes, `bank`$_1$ and `bank`$_3$. *Window* is polysemous because it has just one prototype, `window`$_3$, with two additional interpretations in its sense class.

3.2.3. Semantic Similarity

Having a measure of semantic similarity is important because it allows agents to compare the meanings of two different words and estimate how much their interpretation of a word differs from that of their dialogue partner's.

Since meaning is defined as a set of concepts, semantic similarity measures the similarity of two sets of concepts.[11] Since a similarity measure is already defined on concepts, defining semantic similarity very similar to the classic problem of defining a measure on subsets of a metric space. The only difference is that the conceptual domain defines a measure of similarity, not distance.

One popular solution to this problem is to use Hausdorff distance, which can be computed in linear time (Achermann and Bunke, 2000). A modified version of Hausdorff distance—Hausdorff similarity—is used to define semantic similarity. The semantic similarity between $\mathbf{p} = \langle \mathbf{p}^i, \mathbf{p}^p \rangle$ and $\mathbf{q} = \langle \mathbf{q}^i, \mathbf{q}^p \rangle$ is defined as follows:

$$Sim(\mathbf{p}, \mathbf{q}) = \max(Sim'(\mathbf{p}, \mathbf{q}), Sim'(\mathbf{q}, \mathbf{p}))$$

where Sim' is the following asymmetric Hausdorff similarity measure:

$$Sim'(\mathbf{p}, \mathbf{q}) = \min_{p \in \mathbf{p}^i} \max_{q \in \mathbf{q}^i} sim(p, q).$$

and where *sim* is the measure of conceptual similarity defined in section 3.1.3.

Semantic similarity has three primary uses:

1. Measuring how close the meanings of the same word in two different lexicons are.
2. Measuring how close the meanings of two different words in the same lexicon are.
3. Measuring how related two sense classes are.

[11] Only the interpretation set is involved in this semantic similarity measure, though more complicated measures making use of prototypicality are also possible.

Chapter 4

Discourse Semantics

The following chapter sketches a model for discourse semantics that gives primacy of place to semantic coordination. The Interactive Lexical Hierarchy (ILH) structures the semantic information available to linguistic agents during a discourse. The ILH posits a three-tiered hierarchy of lexical resources in which changes to lower level (more specific) resources may percolate up to change higher level (more general) resources. Semantic alignment results in changes to the lowest level of the hierarchy. A theory of semantic change has two parts: a theory of how change percolates through the lexical hierarchy, and a theory of semantic coordination.

The structure of the chapter is as follows: Section 4.1 gives a general description of how lexicons are to be thought of in this thesis, and how they encode the *meaning* of expressions (as defined in the previous chapter). Section 4.2 describes each of the three levels of lexicon involved in a discourse and how they relate to each other. Section 4.3 develops a semantic theory of explicit semantic coordination and goes on to give pragmatic accounts of several explicit coordination strategies. Implicit coordination is also briefly discussed. Finally, section 4.4 further develops the thesis that all semantic change originates from alignment and discusses how language learning and linguistic change look in this framework.

4.1. Lexicons

A lexicon is a resource for linguistic agents that contains information about the meaning of words and lexicalized multi-word expressions so that they may be combined to construct more complicated utterances. In general a lexicon must contain grammatical information too (e.g., the syntactic contexts in which an expression may appear). Since this thesis is primarily concerned with the semantic aspects of the lexicon, there is little mention of grammar in the lexicon, though it is difficult to ignore entirely. For example, lexical grammar is needed to explain how two senses of a word can be disambiguated depending on the syntactic context in which they appear. Some work has been done that demonstrates the suitability of record types in representing syntactic roles (e.g., Cooper, 2008), so there is good reason to think that grammatical information could easily be incorporated in the lexical model presented here.[1]

[1]Earlier accounts of lexical semantics that make use of feature structures also make include syntactic information. E.g., head-driven phrase structure grammar (Pollard, 1994), the Generative Lexicon

Ignoring these concerns, a lexicon is defined as a mapping from expressions to multi-pointed interpretation sets (as in §3.2).[2] Given a lexicon L and an expression σ, the interpretation of σ under L is written

$$\mathsf{L}(\sigma) = [\![\sigma]\!]_{\mathsf{L}} = \langle |\sigma|_{\mathsf{L}}, \wr\sigma\int_{\mathsf{L}} \rangle,$$

where $|\sigma|_{\mathsf{L}}$ and $\wr\sigma\int_{\mathsf{L}}$ denote the interpretation set and the prototype set respectively.[3]

There are two senses in which a lexicon is a linguistic resource: it is a *cognitive* linguistic resource i.e., a relevant cognitive feature of linguistic agents, and it is a *joint* linguistic resource i.e., common ground information that is drawn upon to coordinate a joint linguistic activity (e.g., a discourse) among a given class of agents. As with any common ground information, joint linguistic resources properly reside in the heads of individual agents; in fact, the joint linguistic resources of an agent are just those cognitive resources that stand in a particular relation to the group in question—namely, they are grounded. An agent may ground different lexical information for different groups, meaning agents have many lexicons. How a given lexicon is used depends on the group with respect to which it is grounded and the type of basis that supports it.

4.2. The Interactive Lexical Hierarchy

In natural language there is a fundamental conflict between expressivity and ambiguity. A language is expressive if it can convey many different concepts. It is ambiguous when the concept being conveyed is not fully determined by the lexicon. In general, languages are more useful for communication the less ambiguous and the more expressive they are. Given a lexicon of fixed size, it would seem that reducing ambiguity and increasing expressiveness are in opposition since reducing ambiguity means removing concepts from the interpretation set of an expression (which reduces its expressiveness) and increasing expressiveness means adding to the interpretation set (which increases ambiguity). One way that natural language sidesteps this fundamental conflict is through compositionality. The expressivity of a language is not simply a function of the size of the lexicon and ambiguity of its expressions since lexicalized expressions are not ultimately what speakers use to express themselves. Instead, the grammatical rules of the language provide ways of combining lexicalized expressions so that the resulting expression has an interpretation set that may be at once less ambiguous and possibly even disjoint from the concepts expressed by any of its parts.

(Pustejovsky, 1995) and Sign-Based Construction Grammar (Boas and Sag, 2012).

[2]Formally lexicons are not functions (see footnote 4), but for the present purposes they are best thought of as such.

[3]The Generative Lexicon (Pustejovsky, 1995) seeks to address many of the some phenomena as the ILH—most notably the notion that an expressions can be used to mean something that is based on, but not manifest in their (permanent) lexical entry. The Generative Lexicon eschews sense enumeration in favor of a qualia structure that generates interpretations on the fly. The ILH allows the generation of *ad hoc* interpretations through semantic coordination while maintaining a sense enumeration structure which is easier to work with computationally.

Another way that natural language relieves the tension caused by the expressivity-ambiguity conflict is by storing lexical information hierarchically. Expressivity may be obtained at a higher level while low level domain-specific lexicons (constructed based on the more general resources) reduce ambiguity and retain what expressivity is required by the communicatory task at hand.

Hierarchy implies some kind of order. The hierarchy that the ILH tries to model is one in which more general resources are *higher* than more specific ones. This notion of generality and specificity is captured by *lexical inheritance*. When a lexicon L inherits from a more general lexicon K, that means that L defers to K on the meaning of expressions for which it has no lexical entry. For any σ such that $\sigma \notin \mathrm{dom}(\mathsf{L})$ but $\sigma \in \mathrm{dom}(\mathsf{K})$, we let $\mathsf{L}(\sigma) := \mathsf{K}(\sigma)$.[4]

The lexical information relevant to a discourse is a sequential hierarchy of three lexicons: the community lexicon, CL, the shared lexicon SL, and the discourse lexicon, DL. Since each agent has her own representation of the community, shared, and discourse lexicons, it is sometimes necessary to index the lexicons by agent (e.g., SL^A). Since lexicons may change as the discourse progresses, it is also useful to index the lexicon to indicate position in the discourse, usually by utterance (e.g., $\mathsf{DL}_0^A, \mathsf{DL}_1^A, \dots$).

Lexicon		With respect to	Bases	Inherits from
community lexicon	CL	a community	communal	$\varnothing$
shared lexicon	SL	a set of agents	actional (past discourses)	CL
discourse lexicon	DL	a discourse	actional (current discourse) perceptual (context)	SL

Table 4.1.: Lexical hierarchy

As part of her general lexical resources, an agent may have many different community, shared, and (at different times) discourse lexicons, but this model assumes that, for a given agent, one of each is in play in any given discourse. Each lexicon in the hierarchy inherits from the previous one and is supported by a different kind of common ground basis. The following section describes each of the three kinds of lexicon.

4.2.1. The Community Lexicon

Recall that Clark (1996) makes a distinction between communal and personal common ground. Personal common ground is built up between particular groups of individuals based on shared experiences. Communal common ground holds between individuals based solely on assumptions about the communities to which they belong. A community lexicon is the common ground lexical information that is supported by a particular

[4]This is a good intuition for how lexical inheritance works, but formally somewhat problematic. To deal with lexical inheritance formally, a lexicon must be defined not as a mapping but as an object that gives rise to a mapping. For details see section 5.2.

communal basis. In any given discourse, there is a (hopefully non-empty!) set of communities to which all of the interlocutors belong. Which of these communities determines the community lexicon depends on further context of the discourse (e.g., when, where and why it is taking place).

Community lexicons roughly correspond to the lexicons of *languages* in the conventional sense (also including sub-languages, dialects, etc). A language in the conventional sense is tied to a community of speakers. Likewise, a community lexicon consists of the expressions and interpretations that are commonly used among members of that community i.e., those interpretations that are supported by the community's common ground. The main difference between a language in the conventional sense and a community lexicon (apart from the fact that community lexicons as defined here only contain semantic information) is that community lexicons are relative to a particular agent—they a community lexicon is a given agent's representation of the lexical common ground of a particular community.

As the top resource in the three-part lexical hierarchy, the community lexicon need not inherit from any other lexicon. Nonetheless, inheritance is common among community lexicons. Domain-specific languages and some dialects inherit from more general community languages. Firefighters, Twitter users, academics, linguists, mathematicians, and investment bankers all have their own set of expressions that they know they can use with each other and be understood. That jargon refines and expands upon a more general lexical resource (say, English). Some expressions in the more general resource may also be off-limits for use in the sub-language. Determining the inheritance relations between community languages is not an easy task. Languages may inherit from a more general resource in some areas, but behave independently in others. Some languages may appear to inherit from multiple larger languages. Furthermore, inheritance between community lexicons may be categorically different from inheritance between lexicons in the ILH. It could be, for example, that a community lexicon inherits from some more general community lexicon in some situations or subject matters and a different one in others. A full investigation of inheritance between community lexicons is beyond the scope of this thesis.[5]

At the start of a discourse between people with no personal common ground (i.e., people who don't know each other), the community lexicon is the only shared lexical resource available—only those interpretations shared by a language community to which all interlocutors belong (and for which that belonging is common ground) can possibly be grounded. This gives communal lexicons a special role as the starting point from which a shared lexicon is built. Interpretations in the community lexicon are grounded based on mutual community membership. Those interpretations may be overruled by circumstances which are recorded in the shared and discourse lexicons.

[5]Nevertheless, the topology on the membership of lexicons' respective communities may give a good first approximation. E.g., the community lexicon of set theorists inherits from that of mathematicians since all set theorists are mathematicians. Of course this heuristic is not perfect—certainly Welsh does not inherit from English, though it is probably the case that all Welsh speakers also speak English.

4.2.2. Shared Lexicon

A specific group of individuals may have more common ground than just what is based on their shared community membership. In particular, they may also have personal common ground. Recall (§2.2.1) that there are two kinds of personal basis, actional bases and perceptual bases. An actional basis consists of some joint action. In this case, the joint action in question is dialogue.

The shared lexicon of a group of agents consists of that persistent lexical common ground information which is not encoded by the community lexicon. It is the personal lexical conventions that those agents share. The shared lexicon is based on semantic coordination that has occurred in discourses involving (at least) those agents. These persistent conventions may be established based on a specific instance of semantic coordination in a particular dialogue, or by the continued use of some convention or family of conventions over many dialogues.

When an agent enters into a discourse, her shared lexicon for that discourse consists of those conventions she believes count as common ground among the discourse participants, based on some prior personal common ground. Of course if that is all she had to go on, the lexicon may be very limited; for most sets of interlocutors, the shared lexicon is fairly small, consisting mostly of proper names. The shared lexicons tends to be larger among smaller groups of intimate conversants. Romantic partners, for example, commonly have idioms that they only use with each other (Hopper et al., 1981), but even they are few in number. For this reason it is important that the shared lexicon inherits from some community lexicon to fill in the gaps.

There are two ways to think of lexicons in the ILH: As the lexical information it adds to the common ground or as the entirety of the lexical information it admits to the common ground (i.e., including that which it inherits).[6] Unless otherwise indicated, future mentions of the shared lexicon refer to the later conceptualization (i.e., including the community lexicon).

The shared lexicon accounts for most of the lexical information in the common ground of a discourse—it provides the raw resources with which a more specialized lexicon may be crafted. In fact, at the start of a discourse, the shared lexicon encompasses *all* of the participants' joint lexical information. But an important point about the shared lexicon is that it does not take context into account. The discourse lexicon is what accounts for dialogue-specific lexical common ground.

4.2.3. Discourse Lexicon

The discourse lexicon is what is actually used to interpret agents' utterances in a discourse. The shared lexicon (and by extension community lexicon) may be thought of as a resource which is used to build (usually more precise) interpretations for the purposes of the current discourse. The discourse lexicon inherits from the shared lexicon. To demonstrate the difference between the discourse lexicon and the shared lexicon, consider the following example:

[6] This distinction is formalized in definition 5.2.1.

(1) Peter: Did John pay back the money he owed you?
Mary: No. He forgot to go to the bank.

Clearly, what Mary means by *bank* is a financial institution ($\texttt{bank}_1$) and not the edge of a river ($\texttt{bank}_2$). Since the topic of conversation is the money that John owes Mary, the discourse lexicon picks out $\texttt{bank}_1$ over $\texttt{bank}_2$.

Suppose that the shared lexicon gives only those two possible interpretations for *bank*. In addition to narrowing the denotation of *bank* to one or the other of those concepts, the discourse lexicon may contain *ad hoc* interpretations that are based on those in the shared lexicon. For example, the exchange above may result in an interpretation for *bank*, $\texttt{bank}_1^*$ that entails the kind of bank where money can be withdrawn (i.e., a consumer financial institution rather than, say, the World Bank). It can also be that *ad hoc* interpretations are broader than the shared interpretations they are based on. Perhaps the discourse lexicon in example 1 contains an interpretation, $\texttt{bank}_1^{**}$, that refers to any place where John might withdraw money (so including ATM's).

The discourse lexicon serves both to disambiguate the lexical common ground (e.g., deciding between $\texttt{bank}_1$ and $\texttt{bank}_2$) and to enrich it by developing *ad hoc* interpretations that do not belong to the more general shared lexicon (e.g., $\texttt{bank}_1^*$ and $\texttt{bank}_1^{**}$). Herman Paul (1891) would call an expression's meaning in the discourse lexicon its "occasional signification".

> *The occasional signification is very commonly richer in content than the usual one, and narrower in extent. In the first place the word in its 'occasional' sense may denote something concrete, while in its 'usual' sense it denotes only something abstract, i.e. some general conception under which different concrete ones may be ranged.* *(Paul 1891, page 66)*

The discourse lexicon deviates from the shared lexicon in a way that makes it at once less ambiguous (narrower) and more expressive (richer). These deviations are supported by both perceptual and actional bases. The perceptual basis for the discourse lexicon, which is broadly referred to as *discourse context* mostly supports disambiguation of the shared lexicon, whereas the actional basis, which consists of semantic coordination supports both disambiguation and the addition of *ad hoc* interpretations.

First we consider the effect of context on the discourse lexicon. During a joint activity, agents must share in some aspects of their experience of the world. For example, in a face-to-face conversation, the state of the participants' immediate surroundings is (barring unusual circumstances) common ground. This common ground may be used to support lexical information. In context, *the big tree* may be taken to refer to a particular large tree in view of the conversants.

The *question under discussion* (cf. Ginzburg, 2012), or the topic of a conversation more generally may also contribute to discourse context. If Andrea and Franz are discussing what to have for dinner, and Franz suggests they *pick up some fish from the store*, it is clear that he does not want to purchase a goldfish from the pet store. In this example, context (in this case, the topic of conversation: dinner) narrows the readings of *fish* and *store* to exclude goldfish and pet stores.

Finally, semantic and syntactic context may affect the discourse lexicon. Consider the interpretation of the word *ran* in the following:

(2) a. Khalid ran the company.
    b. Khalid ran for president.

Clearly the sense of *ran* is disambiguated by the meanings of the other words with which it appears.[7]

There is experimental evidence (Barsalou, 1987, 1992) that the results of stereotypical narrowings are affected by discourse context too. This suggests that, in addition to cutting down on the interpretation set, context may also have an effect on which of those interpretations are in the prototype set.

Agents are justified in their presumption that context serves as lexical common ground on the basis of the communicative principle of relevance. Recall that communicative principle of relevance (Wilson and Sperber, 2008) states that every ostensive stimulus (e.g., an utterance) presumes its own optimal relevance: since it is commonly understood that human cognition is geared towards maximal relevance, the most relevant interpretation of an utterance must be the one that was meant.[8]

Despite the disambiguation effect of the communicative principle of relevance, a key feature of the theory presented here is that ambiguity is commonplace in natural language communication. Not only may meanings be ambiguous, but that that very ambiguity is part of the common ground—part of the discourse lexicon. To be more precise, it is possible for A to utter an expression σ such that $|\sigma|_{\mathsf{DL}_A} = |\sigma|_{\mathsf{DL}_B} = \mathbf{p}$ and $|\mathbf{p}| > 1$. Not only is the ambiguity of σ present in the discourse lexicon of all the discourse participants, but the ambiguity itself is (correctly) assumed to be part of the common ground.

The argument that this can happen closely resembles the argument for the communicative principle of relevance itself (§2.1): Communication in natural language is undertaken for some purpose. Context dictates how much ambiguity is acceptable for the purpose at hand. This context constitutes a perceptual basis that grounds the acceptable level of ambiguity. According to the relevance-theoretic comprehension procedure (Wilson and Sperber, 2008), agents follow a path of least resistance in computing the cognitive effects of a stimulus, stopping when expectations are satisfied. For the same reason that agents stop when expectations of relevance are satisfied, agents also stop

[7]Word sense disambiguation is a well-studied task in the field of Natural Language Processing. Some of the most successful strategies for automatic sense disambiguation rely on the assumption that different senses of a word will have different co-occurrence distributions (Yarowsky, 1995, e.g.,).

[8]Experimental results supporting the so-called *post-decision* hypothesis (cf. Swinney, 1979)—where agents use context disambiguate between senses only after the word has been uttered—are problematic with respect to cognitive justification for the disambiguating role of the discourse lexicon. That being said, one of the conclusions of Swinney (1979) is that all of the information associated with a lexical item is available in sentence comprehension (regardless of context). Indeed this *is* something that the ILH model predicts since agents have access to the shared lexicon even when the discourse lexicon has been disambiguated.

when expectations of (dis)ambiguity are satisfied.[9] This is an extension of what Clark (1996) calls the *principle of least collaborative effort*. Since this principle is common knowledge, speakers take it into account when forming their utterances just as speakers take into account the cognitive principle of relevance. Knowing that listeners will give ambiguous interpretations to her utterance, a speaker would not be justified in adding an unambiguous proposition p to the common ground herself. Anticipating that ambiguity will result, speakers must *intend* their utterances to be ambiguous. Then both speaker and listener are justified in adding an ambiguous proposition $\mathbf{p}$ to the common ground.

This section has given a sketch of the perceptual basis (discourse context) for the discourse lexicon. As we have seen, discourse context serves primarily to disambiguate polysemy in the discourse lexicon. This disambiguation is a form of passive semantic alignment; however, it is important that new interpretations can be added to the discourse lexicon too. For this we need a theory of semantic coordination.

4.3. Semantic Coordination

Now we can define semantic coordination more precisely as the joint process by which agents expand and improve how their discourse lexicons align. Agents adopt various *strategies* to effectively coordinate new meanings. These strategies may be encoded as different dialogue acts that, when they are grounded, change the discourse participants' lexicons in some way. According to Larsson and Cooper (2009), an adequate theory of semantic coordination must have two parts:

1. a semantic part that accounts for how meanings (and concepts) can be updated; and
2. a pragmatic part that relates these updates to dialogue and particular dialogue moves.

This account of semantic coordination follows that framework with one caveat: explicit and implicit semantic coordination are given slightly different semantic treatments.

In explicit coordination, the coordination strategy is signaled by an expression or syntactic marker or by some other meta-communicative utterance. In implicit coordination, a speaker simply uses an expression in a discourse-innovative way without any special meta-communicative indication that he is doing so. The listeners are expected to use context to figure out which coordination strategy is being employed. Note that even though implicit coordination has no explicit discourse markers, it is still achieved by ostensive inferential communication; implicit coordination strategies are initiated with the intention to inform the audience—otherwise it wouldn't be coordination.

This thesis is primarily concerned with explicit semantic coordination, so that is what we focus on here, especially in the pragmatics part. Nevertheless, implicit semantic coordination is ubiquitous in natural language, so it is essential that the proposed model can accommodate both.

[9]Strictly speaking, they stop only when *both* are satisfied.

To illustrate the difference between implicit and explicit coordination, consider the following two examples.

(3) A: I'm looking for a bank.
B: The Federal Reserve Bank is just down the street.
A: Not that kind of bank. I need a BANK-bank—the kind I can cash a check at.

(4) A: Look I picked all these blueberries.
B: You need a bank to put them in so you can pick more.
A: Maybe that tree stump over there would make a good bank.

Example 3 is a case of explicit coordination. An *ad hoc* meaning for *bank* (something like $\mathtt{bank}_1^*$ from the previous section) is being negotiated. A special syntactic form (the reduplication of the word *bank*) indicates the coordination strategy by which the new meaning is calculated. This particular form and the associated coordination strategy are called *contrastive focus repduplication* (Ghomeshi et al. 2004; also see §4.3.1 for its pragmatics).

In example 4 there is no such explicit indication of the coordination strategy. *Bank* is being used under a particular conceptual schema where only its features as "a safe place to store things" are relevant. *B* proposes this schema implicitly by using *bank* in a way that is not compatible with any of its usual (shared lexicon) interpretations. The proposal is then grounded by *A*. Although the coordination is not explicit, *B*'s proposal still constitutes ostensive-inferential communication: Not only does *B* intend to convey the new meaning for *bank*, but he does so in a way that makes it clear that that is his intention.

In explicit coordination there are more distinct signals that have semantic roles to play. There are three kinds of expression involved:

1. The *topical expression*, σ, whose meaning brought about the need for explicit semantic coordination (either because it was ambiguous in DL, or because of a perceived significant misalignment of $\mathsf{DL}(\sigma)$).
2. The *coordinating expression*, τ, which is the expression that characterizes the coordination act. It is used to explicitly indicate the coordination strategy being employed.[10]
3. *Auxiliary expressions*, $\rho_1, \rho_2, \ldots$, which are used to aid in the interpretation of the coordinating expression, but whose interpretations are not subject to change as a result of the coordination.

The first stage of explicit coordination is the creation of the *coordinated meaning*. For a given coordination strategy, this meaning is a determined by the topical expression, the coordinating expression, any auxiliary expressions. The function that determines the coordinated meaning is called the *coordination function*. Given a coordination strategy c, the coordination function for c is denoted δ^c,

$$\delta^c(\sigma, \tau, \rho_1, \rho_2, \ldots) = \mathbf{p} = \langle \mathbf{p}^i, \mathbf{p}^p \rangle$$

[10]Note that the identity of the coordinated meaning depends on the pragmatic strategy involved; it is not necessarily $\mathsf{DL}(\tau)$.

where **p** is a multi-pointed interpretation set.[11] Specifying δ^c for a given coordination strategy is the pragmatic part of semantic coordination. Some examples of coordination functions are given later in the pragmatics part of this section.

The second stage is assigning the coordinated meaning to some expression in the discourse lexicon. Intuitively, it may seem that the goal of semantic coordination is to modify the semantics of the topical expression, but that is not always the case. As we will see, the coordinated meaning may also end up being assigned to the coordinating expression. Some coordination strategies express a preference for one or the other outcomes, while others appear to be compatible with either. The discourse lexicon is updated so that σ', the expression whose meaning is changed, looses its old meaning and now has the following coordinated meaning:

$$\mathsf{DL}_1 := \Big(\mathsf{DL}_0 \setminus \{\langle \sigma', \mathsf{DL}_1(\sigma')\rangle\}\Big) \cup \{\langle \sigma', \mathbf{p}\rangle\}$$

where σ' is either σ or τ. Thus $\mathsf{DL}_1(\sigma') = \mathbf{p} = \gamma^c(\sigma, \tau, \rho_1, \rho_2, \ldots)$.

Now we briefly consider the somewhat simpler semantics of implicit coordination. Implicit coordination occurs in situations where there is some concept p that A wants to express, but there is no σ such that $p \in |\sigma|_{\mathsf{DL}}$—or perhaps more generously, there is no such σ that is suitably unambiguous. The semantics are simpler than that of explicit coordination—since there are no explicit discourse markers to initiate implicit coordination, there are no coordinating or auxiliary expressions. An implicit coordination strategy is characterized by the function, γ^c that creates a new meaning for σ based on the old one (and some other aspects of the discourse context particular to the strategy):

$$\gamma^c(\sigma) := \mathbf{p} = \langle \mathbf{p}^i, \mathbf{p}^p\rangle$$

The new meaning is then applied to σ:

$$\mathsf{DL}_1 = \Big(\mathsf{DL}_0 \setminus \{\langle \sigma, \mathsf{DL}_0(\sigma)\rangle\}\Big) \cup \{\langle \sigma, \mathbf{p}\rangle\}$$

When A and her interlocutors are sufficiently aligned in their present representation of the discourse, they will arrive at the same new meaning $\gamma^c(\sigma)$. But how do the interlocutors know which coordination strategy A is using? Indeed how do they know she is using an implicit coordination strategy at all? Unlike explicit coordination, implicit coordination doesn't necessarily come with any structural or phonological markers to indicate its initiation—that is exactly what makes it implicit. One possible clue would be if $\mathsf{DL}(\sigma)$ isn't *prima face* coherent in the context where it appears, but $\gamma^c(\sigma)$ is. Another consideration is that implicit strategies are sometimes *made* explicit by the use of phrases such as *"metaphorically speaking…"* or by combination with an explicit coordination strategy such as hedging (see §4.3.4).

What follows is a detailed pragmatic analysis of three explicit coordination strategies in the framework developed here. Several other explicit and implicit coordination strategies are also discussed in less detail.

[11] The shared and discourse lexicons are implicit arguments to the coordination function.

4.3.1. Contrastive Focus Reduplication

Contrastive focus reduplication (CR) is a phenomenon observed in colloquial English[12] where an expression is reduplicated to induce a more restrictive reading than the one given by the un-reduplicated expression. Many different kinds of expressions are subject to CR, including nouns, verbs, adjectives, pronouns and lexicalized (and certain unlexicalized) multi-word expressions.

(5) A: Would you bring some salad to the barbecue?
B: Sure, should it be a SALAD-salad or is something like tuna salad ok?
A: Salad-salad if possible.
. . .
B: Did you bring a salad?
A: Yes there is.

A key aspect of this exchange is that both dialogue participants are familiar with the broad interpretation of *salad* (which includes tuna salad) and the narrow interpretation (which does not). The word *salad* may be used in the discourse without any prejudice towards one interpretation or another up until the moment a pragmatic ambiguity arises. The clarification question of *B* and subsequent response of *A* (line 2) prompts a discourse-specific convention that *salad* shall take the narrower interpretation. When *salad* is used again later in the discourse, there is no need to disambiguate between interpretations.[13] Example 5 is formalized in appendix B.

Note that this exchange may have just as well resulted in a different convention where *salad* now refers to the broader interpretation while *salad-salad* continues to be used for the narrower one. In either case, *salad* looses some of its ambiguity. The difference is just in which interpretations remain attached to the original expression, and which get attached to new expressions (or dropped all together). A full pragmatic account of reduplication as a coordination strategy should have something to say about what semantic conventions result from its use, though there may be complicated factors such as social status and expediency.

Most accounts of CR agree on a general characterization that the reduplicated expression reduces ambiguity by restricting the semantics of the expression around some focus set. There is some discussion, though, about how that focus set is determined. Horn (1993) analyzes the focus as the prototype of the expression; thus SALAD-*salad* refers to garden salad, but not tuna salad since garden salad falls under the prototype denotation of *salad* (presumably because it involves leafy green vegetables). Later, Horn identifies four possible focuses denoted by CR:

a) prototype meaning
b) literal meaning

[12] CR is also observed in other languages including Italian, Spanish, Russian, and Persian (Ghomeshi et al., 2004).

[13] Of course, it is possible for some ambiguity in the interpretation of *salad* to persist even after this clarification. It may not be clear, for example, whether *salad* could be used to refer to a julienne salad, which is perhaps more salad-typical than tuna salad, but less so than, say, a garden salad.

c) intensified meaning
d) value-added meaning

Whitton (2006) likewise argues that the prototype account is insufficient to explain the wide variety of construals CR seems to have English. In various contexts DRINK-*drink* may refer to an alcoholic beverage, a *strong* alcoholic beverage, a non-alcoholic beverage, or a soft drink (as opposed to water). Whitton proposes that CR indicates a move along some scalar dimension that is determined by a contrast set (which may happen to be the non-prototypical readings of the original expression). Others have objected to this characterization too, however. Song and Lee (2011) argues for a revised prototype account, noting that there is wide agreement among English speakers on the meaning of most CR constructions even when without context (where a contrast set is presumably absent). The account of Song and Lee (2011) makes use of dynamic prototypes, which our model is well-equipped to accommodate. Ghomeshi et al. (2004) speculate that CR may itself be polysemous or that it may be able to pick out contextually salient readings in addition to objectively prototypical ones. It is possible that "contextually salient readings" may just be prototype meanings in the discourse lexicon. Regardless, we shall follow Ghomeshi et al. (2004) in restricting attention to CR as a phenomenon that focuses meaning around a prototypical interpretation of the expression.

Semantic Role	Name	Pragmatic Role	Example 5
Topical Expr.	σ	standard form	*salad*
Coordinating Expr.	σ^{CR}	reduplicated form	SALAD-*salad*
Auxiliary Exprs.	ρ_1	indicative	*garden salad*
	ρ_1	contrastive	*tuna salad*

Table 4.2.: Pragmatics of Contrastive Focus Reduplication

The coordination function for CR produces a meaning that is focused around the prototype of the expression in standard form:[14]

$$\delta^{\text{CR}}(\sigma, \sigma^{\text{CR}}, \rho_1, \rho_2) = \mathbf{p} = \langle \mathbf{p}^i, \wr \sigma \int_{\mathsf{DL}_0} \rangle$$

where:

1. $\mathbf{p}^i$ is a restriction of $|\sigma|_{\mathsf{DL}_0}$ around $\wr \sigma \int_{\mathsf{DL}_0}$,
2. if ρ_1 is present, there is a $p \in \mathbf{p}^i$ and $q \in |\rho_1|_{DL}$ such that $q \sqsubseteq p$, and
3. if ρ_2 is present, it is not the case that there is a $p \in \mathbf{p}^i$ and $q \in |\rho_2|_{DL}$ such that $q \sqsubseteq p$.

Thus the auxiliary expressions serve to dictate how much the interpretation set of the reduplicated expression should be restricted.

The assignment of the coordinated meaning $\mathbf{p}$ is left open by CR, but in general there are two possibilities: 1) the coordinated meaning is assigned to the standard form ($\mathsf{DL}_1(\sigma) = \mathbf{p}$), or 2) the reduplicated form takes on the coordinated meaning as a

[14] Note that δ^{CR} assumes that $| \wr \sigma \int_{DL_0} | = 1$. See section 6.2 for further discussion.

lexicalized expression ($\mathsf{DL}_1(\sigma^{\text{CR}}) = \mathbf{p}$) and the standard form is disambiguated so that it excludes the coordinated (more prototypical) meaning:

$$|\sigma|_{\mathsf{DL}_1} = |\sigma|_{\mathsf{DL}_0} \setminus \mathbf{p}^i$$

Note that in the second option, since the reduplicated form, σ^{CR} is now a lexicalized expression, so further uses of σ^{CR} do not invoke CR, and so do not depend on the new prototype $\wr\sigma\int_{\mathsf{DL}_1}$.

4.3.2. Corrective Feedback

The following are two examples are corrective feedback from adult-child dialogue (Larsson and Cooper, 2009):

(6) A: That's a nice bear.
B: Yes it's a nice **panda**.

(7) A: Mommy, where *my* **plate**?
B: You mean *your* SAUCER?.

Corrective feedback occurs in situations where one agent, A, uses an expression τ in such a way that her interlocutor, B, becomes aware of a significant[15] discrepancy in their respective common ground interpretations of another different expression, σ:

$$\mathsf{DL}^A(\sigma) \neq \mathsf{DL}^B(\sigma)$$

Crucially, corrective feedback requires that B infers (or believes he is able to infer) what A actually meant by τ. From there, B may take several courses of action. He might, for example, choose to silently align DL^B to resemble what he now believes A's looks like. Or he might refuse to fully ground A's utterance and instead try to correct A's interpretation to look more like his. This later option is what constitutes corrective feedback.

Corrective feedback, since it implies at least a partial grounding failure, interrupts the discourse to some degree. It also appears to violate the principle of least effort (cf. Clark, 1996, p. 224): we assume that B understands $\mathsf{DL}^A(\tau)$ (or believes he does), so the easiest thing to keep the dialogue moving would be to silently update his lexicon to match that interpretation of τ. Thus B must have a good reason for choosing corrective feedback over the more passive option mentioned above. Perhaps the most familiar such reason is that A's interpretation conflicts with the interpretation of some community language that B wants her to learn.

Agent B may choose corrective feedback for other reasons too. Perhaps, for example, the expression in question is being used in an *ad hoc* way and B believes his interpretation better facilitates the task at hand. In both cases the intent of corrective feedback is to update DL^A to look more like DL^B. Depending on the relationship of the interlocutors, the corrective feedback may have a side effect of updating SL^B or CL^B as well. Corrective feedback is unusual in that this "side effect" is probably the primary motivation for employing the coordination strategy.

Larsson and Cooper (2009) identify four forms that corrective feedback may take: in repair (example 6), clarification request (example 7), explicit replace, and bare. For each form, Larsson and Cooper identify expressions filling three pragmatic roles: the initial utterance (the utterance that prompts the correction), an offered form (the word that the corrector suggests instead), and an initial form (the word that the corrector proposes replacing with the offered form). Corrective feedback assumes that it is common ground which expressions fill these three pragmatic roles.[16] In each case, the effect is to modify A's (the correctee's) lexicon so that replacing the offered form for the initial form in her utterance would result in an utterance that enhances what A originally meant.

In what follows, we describe how the pragmatic roles identified by Larsson and Cooper (2009) fit into our semantic theory of coordination.

Semantic Role	Name	Pragmatic Role	Example 6
Topical Expr.	σ	offered form	*panda*
Coordinating Expr.	τ	initial form	*bear*
Auxiliary Expr.	ρ	initial utterance	*That's a nice bear*

Table 4.3.: Pragmatics of Corrective Feedback

Note another unusual feature of corrective feedback: the topical expression, σ, is not the expression that was used in the utterance immediately preceding the corrective feedback (table 4.3). In example 6, A's use of *bear* may not indicate any significant misalignment on $\mathsf{DL}_0(bear)$. Rather, B infers a misalignment on $\mathsf{DL}_0(panda)$ since it would be a more appropriate word to use in the present context. The coordination function for corrective feedback takes three expressions:

$$\delta^{\text{CF}}(\sigma, \tau, \rho) = \mathbf{p}.$$

What remains is to specify the coordination function and update assignment for corrective feedback; that is, how $\mathbf{p}$ is determined by δ^{CF} and for which expression it becomes the new meaning. Let a be the infron such that $a : \mathsf{DL}_0^A(\rho)$.[17] We assume that B has some (perhaps imperfect) access to a. B's corrective feedback conveys to A that σ would be more appropriate to use that τ in ρ. The discourse lexicon of A is thus updated so that the semantics for the initial form matches the meaning computed by

[16] This assumption is perhaps most tenuous with the initial form. How is the correctee supposed to know which part of her utterance is being corrected? If the feedback takes the form of in-repair, the correctee may use clues based on syntactic parallelism between the correction and the initial utterance (Cooper and Larsson, 2009). Otherwise the task is somewhat more nuanced. Often the initial form is the semantically rich expression in the utterance (i.e., the one with the most possible interpretations and prototypes), but if the correctee has any representation for the offered form that may supply clues too (since the initial form is probably the one that is the most semantically similar). A more in depth investigation of how the proposed form is inferred by the correctee is required.

[17] Here we assume that ρ is a positive statement (as in example 6); however, it may also be a question or some other dialogue act (as in example 7). In that case, a is a different kind of object. We do not give a semantics for questions in TTR, though it should be possible to do so in principle.

the coordination function for corrective feedback:

$$\mathsf{DL}_1^A(\sigma) := \mathbf{p}.$$

Thus $\mathbf{p}$ is the interpretation set closest to $\mathsf{DL}_0^A(\sigma)$ such that $a : \mathsf{DL}_1^A(\rho[\tau/\sigma])$, where $\rho[\tau/\sigma]$ is the utterance ρ with τ replaced with σ. Depending on context, this update may induce several side effects. First, and most importantly, corrective feedback usually induces an update to the community or shared lexicon that matches the one made to the discourse lexicon. Second, the semantics of τ may change as well: depending on B's tone and other contextual clues, A may infer that her semantics for τ should exclude a from $\mathsf{DL}_1(\rho)$.

4.3.3. Slack Regulation

We distinguish between words whose intensity can be modified with expressions like *very* (degree words) and those that cannot. This distinction appears even between words that are otherwise very semantically similar.

(8) a. The ball is very round.
b. *The ball is very spherical.

The discrepancy can be explained by analyzing the semantics of *round* and *spherical* differently. While *round*, applies different degrees (on some scale) to different objects, *spherical* merely denotes a set of objects (the spherical ones). The semantics of the qualifier *very* make direct reference to that scale, which is why it can be used with words like *round* but not words like *spherical.* For example, the cutoff for *very round* may be higher on the roundness scale than that of *round.* Since there is no spherical-ness scale, *very* cannot cannot modify *spherical* (Lasersohn, 1999).

There is a puzzle, however, since certain other intensifying schema *may* apply to degree and non-degree words equally.

(9) a. This ball is perfectly round.
b. This ball is perfectly spherical.

Lasersohn (1999) solves this problem by analyzing intensifying schema of the second kind as a separate pragmatic phenomenon. He posits that although it may not be *literally* true of any object (outside of a mathematical setting) that it is spherical, it is not unusual to talk about objects as being spherical even when their shape deviates somewhat from that of a sphere in the strictest sense. How much of a deviation is allowed—the *pragmatic slack*—depends on the pragmatic context. Words like *perfectly* operate on that pragmatic slack, reducing how much one is allowed to deviate from the literal truth.

The question of how much slack is allowed is certainly a pragmatic one; however this analysis allows statements like *the ball is spherical* to be interpreted as genuinely true (given that context allows for it) even when the ball deviates from the strictest interpretation of *spherical.* In the ILH framework, the discrepancy between the "literal" and "pragmatic" truth conditions of spherical corresponds to differences in its interpretation in the shared and discourse lexicon. While it is my intuition that intensifiers like *perfectly* do something to change the truth conditions of sentences in which they appear—and not just their conversational allowability—a more important consequence of taking such a view is that it allows intensifiers to act as coordinating expressions, potentially modifying the semantics of the word they appear with. Consider the following (fictional) exchange:

(10) A: How many spherical fruit are in the first basket?
B: Let's see. . . There's two oranges, an apple, a lemon—
A: Is the lemon perfectly spherical?
B: No, not so much.
A: OK. Just three then?
B: Yeah, three.
. . .
A: How many spherical fruit in the second basket?
B: Five.

A's clarification request, *perfectly spherical?*, on (line 3) is a consequence of her inference that *B* may be allowing more pragmatic slack in line 2 than she intended (since lemons are typically rather oblong). The question allows *B* to pick up on *A*'s pragmatic requirements, updating the interpretation of the (unqualified) *spherical* to exclude lemons. In subsequent exchanges, *spherical* is understood to carry the interpretation negotiated in lines 2–6.

Semantic Role	Name	Pragmatic Role	Example 10
Topical Expr.	σ	slack term	*spherical*
Coordinating Expr.	τ	slack regulator	*perfectly spherical*
Auxiliary Exprs.	ρ_1	slack indicator (indicative)	*orange*
	ρ_2	slack indicator (indicative)	*apple*
	ρ_3	slack indicator (contrastive)	*lemon*

Table 4.4.: Pragmatics of Slack Regulation

The coordinated meaning is determined by the coordination function as follows:

$$\delta^{\text{SR}}(\sigma, \tau, \rho_1, ...) = \mathbf{p} = \langle \mathbf{p^i}, \mathbf{p^p} \rangle$$

where $\mathbf{p}^p = \wr \sigma \int_{\mathsf{SL}}$ and $\mathbf{p_i}$ is subject to the following constraints:

- $\mathbf{p}^i \subsetneq |\sigma|_{\mathsf{DL}_0}$ (the negotiated interpretation set is strictly smaller than the slack term);
- $|\sigma|_{\mathsf{SL}} \subseteq \mathbf{p}^i$ (the "official" interpretation of σ is still inside the negotiated interpretation set);
- $|\rho^+|_{\mathsf{DL}_0} \subseteq \mathbf{p}^i$ (indicative slack indicators fall inside the negotiated interpretation set) ; and

- $|\rho^-|_{\mathsf{DL}_0} \not\subseteq \mathbf{p}^i$ (contrastive slack indicators fall outside the negotiated interpretation set).

Within these constraints, how tightly $\mathbf{p}^i$ is drawn around $|\sigma|_{\mathsf{SL}}$ may depend on other contextual features such as if there are particularly salient interpretations one way or another in the boundary. If no further negotiation is made, then whatever leeway remains remains in the negotiated interpretation set must be small enough that agents are confident in grounding their respective negotiated meanings.

As in the previous example, note that while the slack regulator may serve to update the interpretation of the word it appears with, it could also be that the unadorned word retains its original semantics, while the combination of the slack regulator and the word continues to convey the new meaning in subsequent utterances. There is no uniform way to predict whether σ or τ gets assigned the coordinated meaning. One can imagine a situation, for example, where *spherical* continues to be interpreted to include lemons and *perfectly spherical* is used whenever they should be excluded. Again, which of these possibilities is realized depends on expediency (as dictated by the pragmatic situation) and the relative social status of the interlocutors, among other things.

4.3.4. Other Explicit Coordination Strategies

The above treatments of contrastive focus reduplication, corrective feedback, and slack regulation give a more or less complete picture of how the coordinated meaning is pragmatically determined from the coordinating and topical expressions, and from relevant auxiliary expressions. This section merely lists some other explicit coordination strategies that can be incorporated semantically in the ILH without going into detail on their pragmatics.

Hedging There are two different uses for hedging. The first use is exactly the dual of slack regulating: to increase the pragmatic slack allowed in the discourse lexicon:[18]

(11) It's sort of spherical.

The other use is to make explicit an implicit coordination strategy such as semantic approximation. Recall the *pennyloafer* example (Brennan and Clark, 1996):

(12) A: A docksider.
B: A what?
A: Um.
B: Is that a kind of dog?
A: No, it's a kind of um leather shoe, kinda preppy pennyloafer.
B: Okay okay got it.

[18]Lasersohn gives a different analysis for hedges vs. slack regulators while our others have given them a uniform analysis (e.g., Lakoff, 1973).

Convolution Convolution strategies attempt to combine features of the concepts in the interpretation sets of various other expressions. These combinations may be carried out with meet and join operations.

(13) A unicorn is like a horse but with a horn.

Analogy Analogy is roughly an explicit version of metaphor (see §4.3.5). It induces a specific function from features of p' to the target concept, p based on the relationship between concepts q and q' (where the concepts p', q and q' are expressed by aligned-upon expressions). The canonical form of an analogy is as follows: σ *is to* σ' *as* τ *is to* τ', but in colloquial English, it most often appears in the form σ *is the* τ *of* σ' where τ' is implied by an already-established relation between the denotations of σ and σ':

(14) Montreal is the Paris of Canada.

Explicit Definition In explicit definition, agents use the interpretation of a compositional (non-lexicalized multi-word) expression as the coordinated meaning.

(15) A chair is a man-made object for sitting that has an upright back.

Relativization The meaning of some expressions depends on a *comparison class* against which it is evaluated. The comparison class is usually assumed to be fixed by context, but it can also be negotiated to affect the meaning of the relevant word:

(16) a: She is very short.
b: Shorter than you?
a: No. Actually she's pretty tall. I just meant she's short for a basketball player.

Navigating Lexical Structure Since the shared lexicon is common ground, agents may sometimes make explicit mention of its structure.

(17) A: How is our friend Hubert?
B: I think he might be mad.
A: What is he mad about?
B: I mean the *other* kind of mad.

B is trying to negotiate a meaning that is (very roughly) $\mathsf{SL}(mad)$ minus $\mathsf{DL}_0^A(mad)$. It is taken as common ground that there are two sense classes for *mad*; that is, there are two prototypes in $\wr mad \int_{\mathsf{SL}}$, which is why B is able to use the moniker "the *other* kind of mad" (since if B's intention was not in the sense class that A assumed, there can only be one other one).

This example is formalized in appendix B.

4.3.5. Implicit Semantic Coordination Strategies

This section looks at some implicit coordination strategies. Since we are mostly concerned with explicit coordination, this section will only sketch how the pragmatic functions for various implicit coordination strategies may look.

Semantic Approximation

Semantic approximations use a word to mean something that is semantically similar to its conventional meaning. Duvignau et al. (2007) observe this phenomenon in both children, who have underdeveloped lexicons, and in adults with dementia of the Alzheimer type, who use semantic paraphrases to "fill in" when they cannot remember the right words.

Semantic approximation assumes that there is some set of relevant concepts R that an agent A is expected to be expressed. In the experiments of Duvignau et al. (2007), participants are asked to describe the actions of a woman on a video screen (for example peeling an orange or squashing a tomato). Thus the set of relevant concepts are naturally the ones that appear in the short video. Then there is some concept $p \in R$ that A *wants* to express in that location. To approximate p, A chooses an expression σ with a prototype interpretation $q \in \wr\sigma\int_{\mathsf{DL}_0}$ close to p (and with no prototype interpretation closer to some other $p' \in R$). By implicit coordination, σ is granted the following interpretation:

$$\mathsf{DL}_1(\sigma) = \gamma^{\text{SA}}(\sigma) = \arg\min_{p \in R} \min_{q \in \wr\sigma\int_{\mathsf{SL}}} sim(p, q)$$

Recall that in order to qualify as coordination, its initiation must be an ostensive inferential stimulus. Here that stimulus is achieved by a comparative difference in relevance between $\gamma^{\text{SA}}(\sigma)$ and what σ expresses in the shared lexicon.

Conceptual Schemas

Conceptual schemas are similar to semantic approximation, except that they take what Brennan and Clark (1996) call *historical factors* into account.[19] Agents choose their referring expressions with future uses in mind and may refine their initial expressions to be more efficient as the discourse progresses. Furthermore, under a conceptual schema, semantic approximations take into account which conceptual features were particularly salient in past approximations. Table 4.5 tracks the referring expressions used in where agents are referring to parts of a grid-aligned maze (Mills and Healey, 2008):

Each semantic approximation highlights certain features to be used for referring in subsequent expressions. The first approximation highlights "figural" features of the maze. Later approximations only consider those features that were relevant in previous

[19] Conceptual schemas are not to be confused with conceptual pacts (Brennan and Clark, 1996), which specifically coordinate how a particular referent is to be conceptualized and referred to. Conceptual pacts may be negotiated using various different coordination strategies, including by a conceptual schema.

0 mins:	The piece of the maze sticking out
2 mins:	The left hand corner of the maze
5 mins:	The northernmost box
10 mins:	Leftmost square of the row on top
15 mins:	3rd column middle square
20 mins:	3rd column 1st square
25 mins:	6th row longest column
30 mins:	6th row 1st column
40 mins:	6 r, 1 c
45 mins:	6,1

Table 4.5.: Conceptual schema development in a maze task dialogue

approximations.[20] As referring expressions become shorter and more efficient, the highlighted set of conceptual features may shift or shrink. In the example above, highlighted features shift from what Garrod and Doherty (1994) call "figural" and "path" to "line" and "matrix" features.

Another aspect of conceptual schemas is that the negotiated set of relevant features can be used to refer to other objects of the same kind. For example, once the convention *6,1* has been established to refer to the maze cell in the sixth row, first column, other expressions such as *2,3* may be used likewise. This phenomenon is related to the idea of a *conceptual cover* (Aloni, 2001), but whether or not such an identification scheme exactly corresponds to conceptual covers requires further investigation.

Metaphor

Metaphor is similar to a conceptual schema in that it picks out a set of relevant conceptual features that should be considered in subsequent semantic approximation. It is different in that the relevant features are not evaluated as is, but rather subject to some transformation function. There are many accounts of metaphor that attempt an analysis along these lines (e.g., Carston, 2002).

4.4. Semantic Learning & Linguistic Change

When agents engage in semantic coordination, they are disambiguating the discourse lexicon and producing new *ad hoc* interpretations. These changes sometimes persist beyond the present discourse and become available for use in future discourses. Such changes are regarded as semantic learning. Formally, semantic learning occurs when changes to some lexicon are adopted by the lexicon from which it inherits (definition 5.3.3).

The question remains: When does semantic learning occur? To answer this question, we think back to the grounding criterion of the three lexicons in the ILH. A shared lexicon

[20] As record types, these "features" correspond to a set of labels. The approximation is carried out on the concept with domain restricted to those labels.

consists of the discourse-persistent shared lexical information of a group of agents. Thus an agent will promote a given change in the discourse lexicon to the level of the shared lexicon if she believes that it will also be remembered (and taken to be common ground) by those agents in future discourses. Importantly, it may take many uses of the same *ad hoc* interpretation or narrowing (possibly over many discourses) for this to happen. This implies that agents must have some persistent representation of the history of changes to the discourse lexicon that are not necessarily encoded by any higher level lexicon. Exactly what that representation looks like is not discussed here.

Changes to the community lexicon likewise happen when an agent comes to believe that a particular change in the discourse lexicon is actually in the common ground of a given community. Perhaps, for example, an agent observes that some semantic approximation is used with such regularity and by so many different speakers from a given community that she comes to believe that the speakers do not intend it as a semantic approximation at all, but as a proper (community) lexical denotation of the expression. In that case she will add the interpretation in question to her own representation of the community lexicon.

Changes in individual agent's lexicons are ultimately what drives community-level lexicosemantic change; that is, linguistic change in the conventional sense. As Herman Paul notes, changes in community-level languages also come about by repeated semantic coordination:

> *[D]epartures of the occasional meanings from the usual meaning are starting-points for true change of signification. As soon as these departures repeat themselves with a certain regularity, what was individual and momentary becomes gradually generic and usual. [...] In each individual case the beginning of the transition from an occasional to a usual meaning is made as soon as, on the employment or apprehension of the former, the recollection of an earlier employment or apprehension comes into play: the full transition may be deemed accomplished as soon as such recollection only is effective, and when employment and apprehension alike follow, without any reference to the usual signification of the word.* (Paul 1891, page 70)

Quite simply, a community's language changes when its speakers' representations of the community language change. This change has a cascading quality to it: More and more agents will come to believe that a given change is lexicalized at the community level as they interact with those who have already adopted the change. Of course this implies that, at least at the beginning, some people must make mistakes in their representation of the community lexicon. This is why errors such as the one in the example above are important.

Chapter 5

Formalization

This chapter formalizes the model presented in chapters 3 and 4.

5.1. Conceptual Domain

Formally, an agent's *conceptual domain* is a join-semilattice $(\mathcal{C}, \sqsubseteq)$ where $\sqsubseteq$ is the subconcept relation. This section defines the subconcept relation and proves that it induces a join-semilattice on a set of concepts. Concept meets are also defined, but do not always exist, so the concept domain is not necessarily a complete lattice.

The notion of record depth is useful in several of the proofs in this section. It is also important because it is necessary to require that concepts have finite depth (in addition to the the requirement that restricts records and record types to finite domains).

Definition 5.1.1 (Depth)**.** If a record type p has no record type values then $depth(p) = 0$. Otherwise,

$$depth(p) := 1 + \max\{depth(v) \mid q \in \text{ran}(p) \text{ and } q \in \textbf{RType}\}.$$

Elements of a conceptual domain—the concepts—are the record types of finite depth in a TTR system as defined in appendix A. Although non-record types of the type system are not considered concepts, they do play a role in the conceptual domain since they help to define the record types. The letters $p, q, r, \ldots$ will refer to concepts because of their similarity to propositions in the classical sense (see §3.1).

The intention of the subconcept relation is that $p \sqsubseteq q$ if and only if the extension of p (that is, $\{a \in \text{dom}(A) \mid a : p\}$) is analytically a subset of the extension of p. What is meant by analytically is that the extension of p is a subset of the extension of q regardless of A and F, the "model" components of the type system. It is important to give the subconcept relation a computationally friendly definition (i.e., one that doesn't require reasoning about every possible model) since agents routinely use subconcept judgments in their reasoning. The following definition of subconcept is well-defined on record types as long as they have finite depth.

Definition 5.1.2 (Subconcept)**.** Let p and q be concepts. We define $p \sqsubseteq q$ (p is a subconcept of q) if and only if

1. $\text{dom}(q) \subseteq \text{dom}(p)$ and
2. for all $l \in \text{dom}(q)$, either $p.l = q.l$ or $p.l \sqsubseteq q.l$.

Note that every concept is trivially a subconcept of the empty concept, denoted [ ].

Proposition 5.1.1 (Subconcept Correctness)**.** If $p \sqsubseteq q$, then the extension of p is a subset of the extension of q. In other words, if $a : q$ and $p \sqsubseteq q$, it follows that $a : p$.

Proof. Proof is by induction on the depth of p. □

Lemma 5.1.2 (Concept domains are partially ordered)**.** *Given a set of concepts $\mathcal{C}$, $(\mathcal{C}, \sqsubseteq)$ is a partial order.*

Proof. Let p, q, and R be concepts.

(Reflexivity): For every label $l \in \text{dom}(p)$, $p.l = p.l$.

(Anti-symmetry): Suppose that $p \sqsubseteq q$ and $q \sqsubseteq p$. By definition, $\text{dom}(p) = \text{dom}(q)$. For any label l such that $p.l \neq q.l$, it follows that $p.l \sqsubseteq q.l$ and $q.l \sqsubseteq p.l$. By induction, $p.l = q.l$.

(Transitivity): Suppose that $p \sqsubseteq q \sqsubseteq r$. Let $l \in \text{dom}(r)$ be given. Then $l \in q$ and either $r.l = q.l$ or $q.l \sqsubseteq r.l$. First, suppose that $r.l = q.l$. By hypothesis, either $p.l = q.l$ or $p.l \sqsubseteq q.l$ so either $p.l = r.l$ or $p.l \sqsubseteq r.l$. Now suppose that $q.l \sqsubseteq r.l$. Again, if $p.l = q.l$, then we are done. If $r.l \sqsubseteq q.l$ then we have $p.l \sqsubseteq q.l \sqsubseteq r.l$ (and proof proceeds by induction on the depth of p).

□

Definition 5.1.3 (Concept Generalization)**.** The generalization of concepts p and q is defined as the smallest concept of the form $\{\langle l, T\rangle \mid l \in \text{dom}(p) \cap \text{dom}(q)\}$ such that:

1. If $p.l = q.l$ then $\langle l, p.l\rangle \in p \vee q$, and
2. if $p.l \neq q.l$ and $p.l, q.l \in$ **RType** then $\langle l, p.l \vee q.l\rangle \in p \vee q$.

Lemma 5.1.3. *Generalization is an upper bound in $(\mathcal{C}, \sqsubseteq)$.*

Proof. We need to show that $p \sqsubseteq p \vee q$ for any q (since $\vee$ is a symmetric operation, this suffices). proof is by induction on the depth of p.

($depth(p) = 0$): Suppose p has a depth of 0. Since p has no record type values, it follows by definition that $p \vee q = \{\langle l, p.l\rangle \mid l \in \text{dom}(p) \cap \text{dom}(q) \text{ and } p.l = q.l\}$. Thus $p \vee q.L = p.L$ for every $l \in \text{dom}(p \vee q)$.

($depth(p) = k + 1$): Suppose that for any q' and any p' with $depth(p') \leq k$, we have $p' \sqsubseteq p' \vee q'$. Let $l \in \text{dom}(p \vee q)$ be given. If $p.l = q.l$ then by the same reasoning as in the base case, $p \vee q.l = p.l$. Otherwise $p.l$ and $q.l$ must be record types. By definition, $depth(p.l) \leq k$, so by induction hypothesis, $p.l \sqsubseteq p.l \vee q.l$. Thus $p \sqsubseteq p \vee q$. □

Corollary 5.1.4 (Concept generalizations are extentionally inclusive)**.** For concepts p and q, if $a : p$ or $a : q$ then $a : p \vee q$.

Proof. Suppose without loss of generality that $a : p$. By lemma 5.1.3, $p \sqsubseteq p \vee q$. Thus by proposition 5.1.1, $a : p \vee q$. □

Lemma 5.1.5. *Generalization is a join operation (least upper bound) for $(\mathcal{C}, \sqsubseteq)$.*

Proof. lemma 5.1.3 states that generalization is an upper bound in $(\mathcal{C}, \sqsubseteq)$, so we need only show that it is a *least* upper bound. Let p, q, and r be given such that $p \sqsubseteq r$ and $q \sqsubseteq r$. proof that $p \vee q \sqsubseteq r$ is by induction on the depth of r.

($depth(r) = 0$): Suppose that $depth(r) = 0$. By definition 5.1.2, $\mathrm{dom}(r) \subseteq \mathrm{dom}(p)$ and $\mathrm{dom}(r) \subseteq \mathrm{dom}(q)$ thus $\mathrm{dom}(r) \subseteq \mathrm{dom}(p) \cap \mathrm{dom}(q)$. By definition 5.1.3, $\mathrm{dom}(p) \cap \mathrm{dom}(q) \subseteq \mathrm{dom}(p \vee q)$, so $\mathrm{dom}(r) \subseteq \mathrm{dom}(p \vee q)$.

Now let $l \in \mathrm{dom}(r)$ be given. By definition 5.1.1 $r.l$ is not a concept, so by hypothesis (and definition 5.1.2) it follows that $p.l = r.l$ and $q.l = r.l$. Thus by definition 5.1.3, $p.l = q.l = p \vee q.l$ so $p \vee q.l = r.l$.

By definition 5.1.2 $p \vee q \sqsubseteq r$.

($depth(r) = k + 1$): By the same reasoning as in the base case, $\mathrm{dom}(r) \subseteq \mathrm{dom}(p \vee q)$. Let $l \in \mathrm{dom}(r)$ be given. By hypothesis (and definition 5.1.2), we have the following

1. a) $p.l = r.l$ or b) $p.l \sqsubseteq r.l$; and
2. a) $q.l = r.l$ or b) $q.l \sqsubseteq r.l$

Which gives rise to four cases. In each case we want to show that either $r.l = p \vee q.l$ or $p \vee q.l \sqsubseteq r.l$.

(1a and 2a): $p.l = r.l = q.l$, so $q \vee p.l = r.l$.
(1a and 2b): $p.l \sqsubseteq q.l$ so by lemma 5.1.5 $p \vee q.L = q.l$ and thus $p \vee q.l = r.l$.
(1b and 2a): $p \vee q.l \sqsubseteq r.l$ by a symmetrical argument.
(1b and 2b): $p \vee q.l \sqsubseteq r.l$ by induction hypothesis since $depth(r.l) = depth(r) - 1$. ☐

Theorem 5.1.6. *Given a set of concepts $\mathcal{C}$, the closure of $\mathcal{C}$ under $\vee$ is a join semilattice; moreover, if $\mathcal{C}$ is finite, its closure is too.*

Proof. Since the generalization (join) of two concepts always exists, by lemma 5.1.5 it suffices to show that the closure is finite. We define the closure in steps. A partial closure of $\mathcal{C}$ consists of adding the joins of all of the pairs in $\mathcal{C}$:

$$\mathcal{C}' = \mathcal{C} \cup \{p \vee q \mid p, q \in \mathcal{C}\}$$

Note that $\mathcal{C}$ is closed under join if and only if $\mathcal{C} = \mathcal{C}'$.

Since $\mathcal{C}$ is finite, there are finitely many pairs of concepts in $\mathcal{C}$, so its partial closure is finite too. It suffices to show that closure occurs after finitely many partial closures. This is the case since the number and maximum depth of new concepts at each closure are both monotone decreasing. ☐

Definition 5.1.4 (Concept Merge[1]). For concepts p and q, the merge of p and q ($p \vee q$) is defined as the smallest concept such that:

1. If $l \in \mathrm{dom}(p) \setminus \mathrm{dom}(q)$ then $\langle l, p.l\rangle \in p \wedge q$.
2. If $l \in \mathrm{dom}(q) \setminus \mathrm{dom}(p)$ then $\langle l, q.l\rangle \in p \wedge q$.
3. If $l \in \mathrm{dom}(p) \cap \mathrm{dom}(q)$ then:
   a) If $p.l = q.l$ then $\langle l.p.l\rangle \in p \wedge q$.
   b) Else if $p.l, q.l \in$ **RType** and $p.l \wedge q.l$ exists, then $\langle l, p.l \wedge q.l\rangle \in p \wedge q$.
   c) Otherwise $p \wedge q$ is undefined.

[1]Concept merge is the same as *unification* in Unification-Based grammars that make use of feature structures (Shieber, 1986).

In general, it is not always the case that the merge of two concepts (as defined above) exists. Viewing the concepts as types in a TTR system, it is possible to ensure that the merge always exists by adding *merge types* to the system such that the merge of T_1 and T_2 is a triple $\langle \wedge, T_1, T_2 \rangle$. Type evaluation for merge types is then defined to be necessarily extensionally intersective (i.e., $a : (\wedge, T_1, T_2)$ if and only if $a : T_1$ and $a : T_2$). The record type r can be found using a method that follows definition 5.1.4 with the caveat that where $\wedge$ is undefined, one would instead add the meet type to the new concept:

c) Otherwise $\langle l, \langle \wedge, p.l, q.l \rangle \rangle \in p \wedge q$

The consequence is that it is possible to define merges in TTR in such a way that the merge of two record types is always a record type, and otherwise it is a merge type. The details of this approach can be found in Cooper (2012, §3.2).

Proposition 5.1.7 (Concept merge is extensionally intersective)**.** For concepts p and q such that $p \wedge q$ exists, $a : p \wedge q$ if and only if $a : p$ and $a : q$.

Proof. Let object a be given and concepts p and q be given such that $p \wedge q$ exists.

($\Rightarrow$): Suppose that $a : p \wedge q$. Without loss of generality, consider p. Let $l \in \mathrm{dom}(p)$ be given.

If $l \in \mathrm{dom}(q)$ then since $p \wedge q$ exists, either $\langle l, p.l \rangle \in p \wedge q$ (in which case $a.l : p.l$ by hypothesis) or $p.l$ and $q.l$ are record types for which $p.l \wedge q.l$ is defined and $a.l : p.l \wedge q.l$. By induction, $a.l : p.l$ and $a.l : q.l$.

If $l \notin \mathrm{dom}(q)$ then by definition $\langle l, p.l \rangle \in p \wedge q$ so by hypothesis $a.l : p.l$.

($\Leftarrow$): Suppose $a : p$ and $a : q$ and let $l \in \mathrm{dom}(p \wedge q)$ be given. If $l \in \mathrm{dom}(p) \setminus \mathrm{dom}(q)$ then $p \wedge q.l = p.l$ and by hypothesis $a : p.l$ and if $l \in \mathrm{dom}(q) \setminus \mathrm{dom}(p)$ then likewise $p \wedge q.l = q.l$.

So suppose $l \in \mathrm{dom}(p) \cap \mathrm{dom}(q)$. If $p.l = q.l$ then $p \wedge q.l = p.l$ and again $a : p.l$. If $p.l \neq q.l$ then both are record types and $p \wedge q.l = p.l \wedge q.l$. By induction $a.l : p.l \wedge q.l$. ☐

Lemma 5.1.8 (Merge is a meet operation (greatest lower bound) for $(\mathcal{C}, \sqsubseteq)$)**.** *Given concepts p and q, $p \wedge q$ is the greatest lower bound of p and q under $\sqsubseteq$.*

Proof. By induction on the depth of p. ☐

When all meets exist, it follows immediately from lemma 5.1.2 and proposition 5.1.8 that $(\mathcal{C}, \sqsubseteq)$ is a meet semi-lattice; however it is important to note that in practice meets rarely, if ever, all exist without adding meet types. If meets are added, there is no guarantee that the meet closure of a finite concept domain is finite.

Corollary 5.1.9. The following are equivalent:

(a) $p \sqsubseteq q$

(b) $p = p \wedge q$

(c) $q = p \vee q$

Proof. Standard result for meet and join. ☐

Definition 5.1.5 (Concept Similarity)**.** Given a concept space $\mathcal{C}$ and concepts $p, q \in \mathcal{C}$, the similarity between p and q is defined by

$$sim(p,q) := \frac{|\{l \mid p.l = q.l\}| + sim'(p,q)}{|\operatorname{dom}(p)| + |\operatorname{dom}(q)|}$$

where $sim'(p,q)$ sums the distances between p and q's values which are not identical, but nonetheless are record types:

$$sim'(p,q) = \sum\{sim(p.l, q.l) \mid p.l \neq q.l \text{ and } p.l, q.l \in \mathbf{RType}\}$$

Again, note that this definition assumes finite depth.

The following facts about concept similarity are left unproven for brevity, but in all cases the proof is straight-forward. The first two facts (a and b) show that concept similarity is indeed a similarity measure (in the sense analogous to what it means to be a similarity relation). The later two facts (c and d) establish some degree of compatibility between concept similarity and the subconcept relation.

Proposition 5.1.10 (Facts about Concept Similarity)**.**

(a) $sim(p,q) = sim(q,p)$ (symmetry)

(b) $sim(p,q) = 1 \iff p = q$ (reflexivity)

(c) $p \sqsubseteq q \sqsubseteq r \implies sim(p,q) \geq sim(p,r)$

(d) If $p \sqsubseteq q$ then $sim(p,q) > 0$.

5.2. SEMANTICS

Definition 5.2.1 (Lexicon)**.** Given a set of expressions Σ and a concept domain $\mathcal{C}$, a lexicon is defined as a tuple

$$\mathsf{L} = \langle \mathsf{K}, \mathsf{L}^{i+}, \mathsf{L}^{i-}, \mathsf{L}^{p+}, \mathsf{L}^{p-} \rangle$$

where:

- K is a (possibly empty) lexicon[2];
- $\mathsf{L}^{i+}, \mathsf{L}^{i-}, \mathsf{L}^{p+}, \mathsf{L}^{p-}$ are all partial functions from Σ to $\mathcal{P}(\mathcal{C})$
- $\mathsf{L}^{i+}(\sigma) \cap \mathsf{L}^{i-}(\sigma) = \varnothing$ and $\mathsf{L}^{p+}(\sigma) \cap \mathsf{L}^{p-}(\sigma) = \varnothing$; and
- the resulting mapping (see below) respects the condition that all prototypes are in the interpretation set.

We say that L *inherits from* K. Along with the lexicon it inherits from, the other components of L give rise to the following partial mapping $\mathsf{L} : \Sigma \nrightarrow \mathcal{P}(\mathcal{C}) \times \mathcal{P}(\mathcal{C})$:

$$\mathsf{L}(\sigma) := \langle |\sigma|_{\mathsf{L}}, \wr\sigma\int_{\mathsf{L}} \rangle := \langle (|\sigma|_{\mathsf{K}} \smallsetminus \mathsf{L}^{i-}(\sigma)) \cup \mathsf{L}^{i+}(\sigma), (\wr\sigma\int_{\mathsf{K}} \smallsetminus \mathsf{L}^{p-}(\sigma)) \cup \mathsf{L}^{p+}(\sigma) \rangle$$

The interpretation set of σ under L is $|\sigma|_{\mathsf{L}}$ and its prototype set is $\wr\sigma\int_{\mathsf{L}}$. All lexicons must adhere to the condition that $\wr\sigma\int_{\mathsf{L}} \subseteq |\sigma|_{\mathsf{L}}$ for all $\sigma \in \Sigma$.

[2]Where the empty lexicon is defined as $\langle \varnothing, \varnothing, \varnothing, \varnothing, \varnothing \rangle$.

Definition 5.2.2 (Sense Class)**.** An expression σ has a *sense class* for each prototype in $p \in \wr\sigma\int_{\mathsf{L}}$. The sense class of p is the subset of $[\![\sigma]\!]_{\mathsf{L}}$ for whom p is the closest prototype:

$$|\sigma|^p_{\mathsf{L}} := \left\{ q \in [\![\sigma]\!]_{\mathsf{L}} \mid p \in \left(\arg\max_{r \in \wr\sigma\int_{\mathsf{L}}} sim(q,r) \right) \right\}$$

Definition 5.2.3 (Semantic Similarity)**.** Let σ and τ be expressions and L and L' be languages. The semantic similarity between $\mathsf{L}(\sigma)$ and $\mathsf{L}'(\tau)$ is defined as follows:

$$Sim(\mathsf{L}(\sigma), \mathsf{L}'(\tau)) := \min(Sim'(\mathsf{L}(\sigma), \mathsf{L}'(\tau)), Sim'(\mathsf{L}'(\tau), \mathsf{L}(\sigma))$$

where Sim' is the asymmetric Hausdorff similarity measure:

$$Sim'_{\mathsf{L}}(\sigma, \tau) := \min_{p \in [\![\sigma]\!]_{\mathsf{L}}} \max_{q \in [\![\tau]\!]_{\mathsf{L}'}} sim(p,q).$$

In addition to semantic similarity, we also define the semantic coherence of an expression. Semantic coherence measures how tightly clustered a word's senses are. It can be used to estimate how easily the senses of a word can be disambiguated by context.

Definition 5.2.4 (Semantic Coherence)**.** The semantic coherence of an expression σ is defined as the average similarity between each pair of its prototypes:

$$coh_{\mathsf{L}}(\sigma) := \frac{\sum_{p,q \in \wr\sigma\int_{\mathsf{L}}} Sim(|\sigma|^p_{\mathsf{L}}, |\sigma|^p_{\mathsf{L}})}{\binom{|\wr\sigma\int_{\mathsf{L}}|}{2}}$$

Definition 5.2.5 (Interactive Lexical Hierarchy)**.** In a given discourse, each agent $A \in \mathcal{A}$ maintains a hierarchy of three common ground lexicons $\mathsf{CL}^A, \mathsf{SL}^A$, and DL^A such that SL^A inherits from CL^A and DL^A inherits from SL^A.

5.3. Coordination & Learning

At the start of a discourse, without any dialogue having taken place, the discourse participants alone give rise to a shared lexicon, SL. Since no coordination has taken place, there are initially no lexical entries unique to the discourse lexicon:

$$\mathsf{DL}_0 = \langle \mathsf{SL}, \varnothing, \varnothing, \varnothing, \varnothing \rangle.$$

Definition 5.3.1 (Explicit Semantic Coordination)**.** If $\textsc{c}$ is an explicit coordination strategy and u_i (the ith utterance of the discourse) is an utterance using that strategy, then

$$\mathsf{DL}_i := \left(\mathsf{DL}_{i-1} \smallsetminus \{\langle \sigma', \mathsf{DL}_{i-1}(\sigma') \rangle\} \right) \cup \{\langle \sigma', \delta^{\textsc{c}}(\sigma, \tau, \rho_1, \rho_2, \ldots) \rangle\}$$

where

- $\delta^{\textsc{c}}$ is the coordination function associated with $\textsc{c}$;
- σ is the topical expression of $\textsc{c}$ in u_i;
- τ is the coordinating expression of $\textsc{c}$ in u_i;

- $\rho_1, \rho_2, \ldots$ are the auxiliary expressions of C in u_i; and
- σ' is either σ or τ.

Definition 5.3.2 (Implicit Semantic Coordination)**.** If C is an implicit coordination strategy and u_i (the ith utterance of the discourse) is an utterance using that strategy, then

$$\mathsf{DL}_i := \left(\mathsf{DL}_{i-1} \setminus \{\langle\sigma, \mathsf{DL}_{i-1}(\sigma)\rangle\}\right) \cup \{\langle\sigma, \gamma^{\mathrm{C}}(\sigma)\rangle\}$$

where

- γ^{C} is the coordination function associated with C and
- σ is the topical expression of C in u_i.

Definition 5.3.3 (Semantic Learning)**.** Let L be a lexicon that inherits from K. *Learning* occurs when some "change" in L moves to K; that is when

$$\mathsf{K}_1^e = \mathsf{K}_0^e \cup \{p\}$$

and

$$\mathsf{L}_1^e = \mathsf{L}_0^e \setminus \{p\}$$

for $p \in \mathsf{L}_0^e$ and $e \in \{i+, i-, p+, p-\}$.

Chapter 6

Empirical Analysis

This chapter presents some empirical work based on the theoretical framework developed in the previous chapters.

6.1. A Signaling Games Simulation

In this section, a signaling games simulation is used to investigate the relationship between conceptual similarity and lexical ambiguity. The purpose of this investigation is to find evidence for two claims made by the theoretical framework presented in chapter 4.

The first claim to investigate is that lexical ambiguity is, in general, beneficial for communication when it is commonly understood among linguistic agents that such ambiguity may arise. The expressivity of a lexicon is directly proportional to its ambiguity. Expressive lexicons are beneficial for communication because they allow agents to communicate about a greater range of concepts; however this expressivity is useless if the associated ambiguity causes too many failures of communication. The hypothesis is that linguistic agents who are aware of this situation can take measures to mitigate communication failures while still reaping the benefits of greater expressivity.

The second claim involves how polysemous expressions work their way up through the lexical hierarchy. Recall that *ad hoc* conventions (i.e., interpretations that are only in the discourse lexicon) commonly use an expression to mean a concept that is similar to those concepts that the expression refers to in higher-level lexicons (e.g., semantic approximation section 4.3.5). The hypothesis is that as these conventions are found to be successful at the level of discourse, they may become lexicalized in agents' more permanent resources.

To investigate these claims, we use a signaling game where agents have a limited number of expressions which may be used to refer to many different concepts, some of which are related (§6.1.2). Then sections 6.1.3 and 6.1.4 present the results of simulating that game and discusses how it relates to the hypotheses above. First, some background on signaling games is required.

6.1.1. Signaling Games

Lewis (1969) conceived of signaling games to model coordination problems (see section 2.2.3):

A signaling game consists of a set of types, T, a set of messages, M, a set of actions, A, and a utility function $U : T \times M \times A \to \mathbb{R}$. Typically there are two agents, a *sender* and a *receiver*. In that case, the flow of the game is as follows:

1. A type $t \in T$ is selected at random. (There may be a prior distribution on T, or it may be selected uniformly at random. If there is a prior distribution, the agents may or may not know it.)

2. Based on t, the sender selects a message $m \in M$ to send to the receiver.

3. Based on the message that was sent, the receiver selects an action $a \in A$.

4. The utility function assess the outcome of the exchange with $U(t, m, a)$.

5. The agents update their selection criterion based on the outcome of the exchange and the choices they made.

A (possibly probabilistic) mapping from types to messages is called a *sender strategy*:[1] A sender strategy is characterized by the probability that the sender will send message m given that she has type t.

$$P_s(m \mid t) \in [0, 1].$$

Likewise, such a probability function over actions that depends on messages is called *receiver strategy*:

$$P_r(a \mid m) \in [0, 1].$$

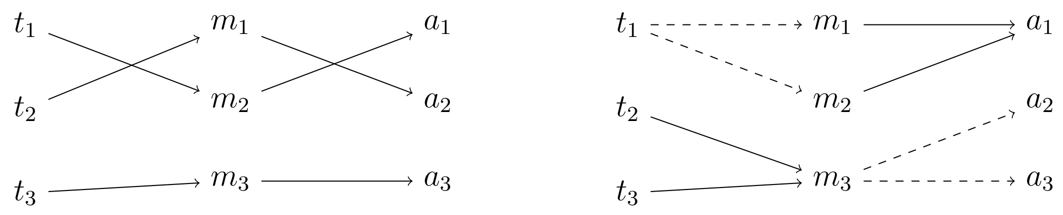
Thus the *expected utility* of a given game may be calculated based on the sender and receiver strategies as follows:

$$\mathbb{E}[U] = \sum_{t \in T} \sum_{m \in M} \sum_{a \in A} P(t) \cdot P_s(m \mid t) \cdot P_r(a \mid m) \cdot U(t, m, a)$$

A sender-receiver strategy pair that maximizes expected utility is in *optimal equilibrium*; however there may be other equilibria (so-called *pooling equilibria*) that produce sub-optimal expected utility. In figure 6.1a, agents have reached an optimal equilibrium: the sender and receiver strategies are coordinated such that any type t will result in successful communication. Figure 6.1b is in a pooling equilibrium: neither the sender nor the receiver can independently change strategies without lowering the expected utility.

After playing a round, the sender (and the receiver) must decide how to update their strategies based on the type (or message) she received, the message (or action) she chose, and the utility produced by the round. A function that updates play strategies is called a *learning strategy*. The general class of learning strategies we will investigate are *reenforcement strategies*. These are strategies where successful interactions make

[1]Following the usual restriction that for all $m \in M$, $\sum_{t \in T} P(t \mid m) = 1$ and likewise for P_r.



(a) An optimial equilibrium: $\mathbb{E}[U] = 1$. (b) A sub-optimal equilibrium: $\mathbb{E}[U] = \frac{1}{2}$.

Figure 6.1.: A signaling game where $U(t_i, m_j, a_k) = 1$ if $i = k$ and 0 otherwise. Lines from t_i to m_j represent the sender strategy, while lines from m_j to a_k represent the receiver strategy. Solid lines represent probability 1 (e.g., $P(a_k|m_j) = 1$) and dashed lines represent probability $\frac{1}{2}$ (Catteeuw and Manderick, 2014).

the agents more likely to play the same way in the future: If $U(t_i, m_j, a_k)$ is high, then $P_s(m_j|t_i)$ and $P_r(a_k|m_j)$ both increase. Some learning strategies are more prone to get stuck in pooling equilibria than others. One way to mitigate the risk of getting stuck in a pooling equilibria is to use *discounting* (Roth and Erev, 1995; Catteeuw and Manderick, 2014). When discounting is used, $P_s(m_{j'}|t_i))$ and $P_r(a_{k'}|m_j)$ decrease (for all $j' \neq j$ and $k' \neq k$) by some discounting factor λ.

6.1.2. Setup

The purpose of this experiment is to use a signaling game to investigate how polysemy arises from a need for greater expressivity. Thus the signaling game described in this section models a situation where agents need to communicate more concepts than they have expressions available. In this game types and actions both correspond to concepts and messages correspond to expressions. For consistency (and to avoid confusion with *type* as it is used elsewhere in this thesis), we refer to game objects as the things they represent in the ILH: types and actions as concepts, and messages as expressions. The set of concepts is denoted C.

The outline of the game (following the general outline of signaling games above) is as follows:

1. A concept $p_i \in C$ is selected uniformly at random.
2. Based on her lexicon L_s and the concept domain, the sender selects an expression $m_j \in M$ to send to the receiver.
3. Based on the message received and his lexicon L_r, the receiver selects a concept $p_k \in C$ which interprets the expression.
4. The utility function $U(p_i, m_j, p_k)$ scores the exchange based on how close p_i is to p_k in the concept domain.
5. Agents update their lexicons based on the utility received and the choices they made.

The concept domain and agent lexicons are described below.

The Concept Domain

The concept domain plays two roles in this signaling game: First, it determines the utility received after a round of play by defining a similarity measure on concepts. And second, it is available to both agents in their play and learning strategies. As elsewhere, we assume that both agents have perfect knowledge of the concept domain and that it is common ground.

Recall that concepts are related by a reflexive, symmetric similarity measure. For the purposes of this simulation, we are not interested in the subconcept relation, but only conceptual similarity. The conceptual domain is modeled by random weighted graph $G = \langle N, V \rangle$ where N is the set of concepts and $V : N \times N \to [0, 1]$ models similarity as follows:

$$sim(p_i, p_j) = \begin{cases} 1 & \text{if } i = j \\ V(p_i, p_j) & \text{if } (p_i, p_j) \in \text{dom}(V) \\ 0 & \text{otherwise} \end{cases}$$

The unweighted graph is created with the Holme-Kim growing random graph algorithm (Holme and Kim, 2002) and assigned a weight between 0 and 1 uniformly at random.

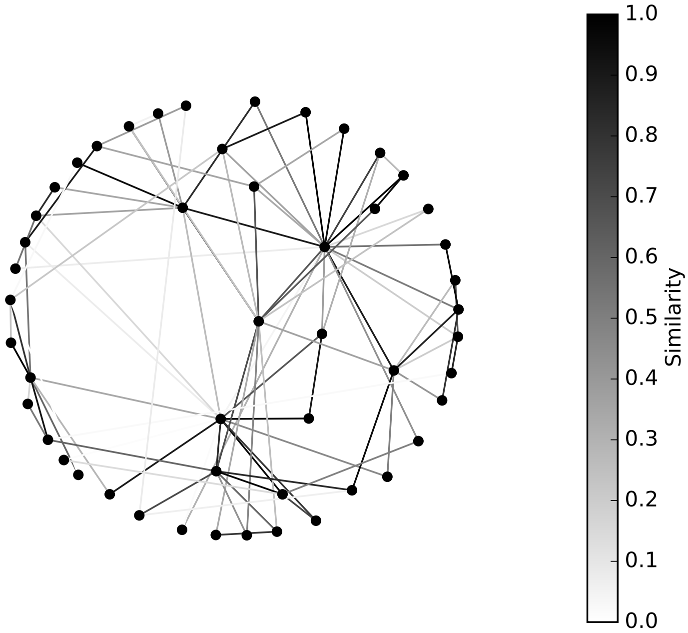


Figure 6.2.: A sample concept domain model of 50 concepts.

The Holme-Kim algorithm has two important properties that make it suitable for our purposes: First, it encourages clustering. Because of the way concept similarity works, if concepts p and q are similar and concepts q and r are similar, it is more likely than average that p and r are similar (see §3.1.3 for details). This means we should expect the conceptual domain we are trying to model to exhibit some clustering.

Second, the Holme-Kim algorithm scales well over different numbers of nodes. One factor of interest in these simulations is the proportion of concepts to messages. Thus an important feature of the Holme-Kim random graph algorithm is that it produces graphs with similar distributional properties regardless of the number of nodes.[2]

The Agent Lexicon

Both sender and receiver keep a lexicon (L_s and L_r respectively). To avoid the difficulty of having to keep track of a prior distribution on expressions, the lexicon is defined from the perspective of production: as a probabilistic mapping from concepts to expressions where $L(m_i|p_j)$ represents the degree to which m_i means p_j.

Agents do not just use their lexicons for producing expressions, but also for interpretation. Although L is defined to give the probability of an expression given a concept, Bayes theorem lets us use the lexicon in the other direction too:

$$L(p|m) = \frac{L(m|p) \cdot P(p)}{P(m)} = \frac{L(m|p) \cdot 1/|C|}{1/|C| \cdot \sum_{p'} L(m|p')} = \frac{L(m|p)}{\sum_{p'} L(m|p')}$$

A word is more ambiguous the more uncertain it is what it means. Thus we define the ambiguity of a lexicon L as the conditional entropy of p given m:

$$\begin{aligned}
H(p|m) &= \sum_i P(m_i) \cdot H(p|m = m_i) \\
&= -\sum_i P(m_i) \cdot \sum_j \Big[P(p_j|m_i) \cdot \log P(p_j|m_i)\Big] \\
&= -\sum_{i,j} P(m_i, p_j) \cdot \log P(p_j|m_i) \\
&= -\frac{1}{|C|} \sum_{i,j} L(m_i|p_j) \cdot \log \left[\frac{L(m_i|p_j)}{\sum_{j'} L(m_i|p_{j'})}\right]
\end{aligned}$$

It is important to note that while the lexicons are defined as probability distributions, they do not constitute play strategies on their own. In principle, play strategies may take into considerations other than pure semantics. The three strategies we use in simulations are described below:

Baseline

The baseline strategy (BL) against which we evaluate the other two is the one where the sender/receiver deterministically selects the most likely option based on the lexicon alone. At first brush, this may seem like a pretty good strategy (why choose anything other than what is most likely?), but it denies the agents the opportunity to explore

[2]In particular, the Holme-Kim algorithm produces so-called *scale free* networks—networks for which the fraction of nodes with k neighbors is proportional to $k^{-\gamma}$ for $2 < \gamma < 3$ (Barabási and Bonabeau, 2003).

the space of possible conventions, and is therefore very likely to end up in a pooling equilibrium.

$$P_s^{\text{BL}}(m_i|p_j) = \begin{cases} 1 & \text{if } i = \arg\max L_s(m_i|p_j) \\ 0 & \text{otherwise} \end{cases}$$

$$P_r^{\text{BL}}(p_j|m_i) = \begin{cases} 1 & \text{if } j = \arg\max L_r(p_j|m_i) \\ 0 & \text{otherwise} \end{cases}$$

Most Accurate

The "most accurate" strategy (MA) also only considers the lexicon, but retains its probabilistic content. The agent makes a choice at random based on how accurately the expression conveys the concept (sender) or how accurately the concept is described by the message (receiver).

$$P_s^{\text{MA}}(m_i|p_j) = L_s(m_i|p_j) \qquad\qquad P_r^{\text{MA}}(p_j|m_i) = L_r(p_j|m_i)$$

Best Bet

In the "best bet" strategy (BB), agents take into account not only the lexicon, but also similarity relations between concepts. Best bet should model situations where an agent wants to communicate concept p_i, but there are no expressions that refer to that concept especially well, so she uses the expression m_j because $L(m_j|p_k)$ is high, and the original concept p_i is very similar to p_k. This situation models semantic approximation: the agent creates an *ad hoc* meaning for m_j (namely p_j) based on her existing lexical resources (namely that m_j refers well to p_k).

Using BB, the signaling agent considers not just one, but *all* of the concepts a message might refer to, and how close they are to concept she intends to convey:

$$P_s^{\text{BB}}(m_i|p_j) \propto \sum_k L_s(m_i|p_k) \cdot sim(p_k, p_j)$$

We assume that it is commonly understood among agents adopting this strategy that signaling agents pick their expressions in this way. Thus the receiver strategy is again defined analogously:

$$P_r^{\text{BB}}(p_j|m_i) \propto \sum_k L_s(p_k|m_i) \cdot sim(p_k, p_j)$$

To illustrate that best bet captures the phenomenon mentioned above, consider the following example: Let the sender lexicon and concept domain be defined as follows:

$$L_s = \begin{array}{c} \begin{matrix} p_1 & p_2 & p_3 \end{matrix} \\ \begin{bmatrix} 0 & 0.3 & 0.9 \\ 1 & 0.7 & 0.1 \end{bmatrix} \end{array} \begin{matrix} m_1 \\ m_2 \end{matrix} \qquad\qquad sim = \begin{array}{c} \begin{matrix} p_1 & p_2 & p_3 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0.2 \\ 0 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{bmatrix} \end{array} \begin{matrix} p_1 \\ p_2 \\ p_3 \end{matrix}$$

Now suppose that the sender wants to express p_2. Using the MA strategy, her production frequency is dictated by L_s alone: $P_s^{\text{MA}}(m_i|p_2) = L_s(m_i|p_2) = [0.3, 0.7]$; however, using the BA strategy, the production frequency is calculated as follows:

$$\begin{aligned}
P_s^{\text{BB}}(m_1|p_2) &\propto L_s(m_1|p_1)sim(p_1,p_2) + L_s(m_1|p_2)sim(p_2,p_2) + L_s(m_1|p_3)sim(p_3,p_2) \\
&= 0 \cdot 0 + 0.3 \cdot 1 + 0.9 \cdot 0.8 \\
&= 1.02 \\
P_s^{\text{BB}}(m_2|p_2) &\propto L_s(m_2|p_1)sim(p_1,p_2) + L_s(m_2|p_2)sim(p_2,p_2) + L_s(m_2|p_3)sim(p_3,p_2) \\
&= 1 \cdot 0 + 0.7 \cdot 1 + 0.1 \cdot 0.8 \\
&= 0.78
\end{aligned}$$

Normalizing, $P_s^{\text{BB}}(m_i|p_2) = [0.57, 0.43]$. Even though m_2 better expresses p_2 in the lexicon, the agent is more likely to choose m_1. The biggest reason for this is that p_3 is very well expressed by m_1 (that is, $L_s(p_3|m_1)$ is very high), and p_3 is closely related to p_2. On the other hand, m_2 is most likely to express p_1 since $L_s(p_i|m_2) = [0.55, .39, .06] \propto [1, 0.7, 0.1]$, and p_1 is not at all related to p_2.

Learning Strategies

All learning in the signaling game is manifest as lexical reinforcement learning. Given a game where p_i, m_j, p_k are played, the lexicons are updated as follows:

$$L'_s(m_j|p_{i'}) \propto \begin{cases} \lambda L_s(m_j|p_{i'}) + sim(p_i,p_k) & \text{if } i' = i \\ \lambda L_s(m_j|p_{i'}) & \text{otherwise} \end{cases}$$

and

$$L'_r(p_i|m_{j'}) \propto \begin{cases} \lambda L_r(p_i|m_{j'}) + sim(p_i,p_k) & \text{if } j' = j \\ \lambda L_r(p_j|m_{j'}) & \text{otherwise} \end{cases}$$

Where λ is some discounting factor $\lambda \leq 1$ (if $\lambda = 1$, then there is no discounting). Note that alternative choices for the given state are discounted. So in the case of the sender, discounting only effects the weight that other messages give the concept she received. In the case of the receiver, discounting only effects the weight of other concepts given the message sent by the sender.

Hypotheses

To test the claims made at the beginning of this section, we make the following hypotheses:

- H1 Agents using BB will be more successful (i.e., they will form signaling conventions with higher expected utility) than those using strictly lexical strategies (BL, MA), especially on larger domains.
- H2 The lexicons of agents using BB will be more ambiguous than those of agents using strictly lexical strategies.

6.1.3. Results

The simulations were structured as follows: Concept domains ranged from 10 to 100 concepts (in intervals of 10); 250 random concept domains of each size were generated. Agents were afforded 10 messages regardless of domain size. For each strategy and on each domain, 5000 rounds of the game were played. For MA and BB, discounted and un-discounted learning strategies were tested. The discounted learning strategies were given a discounting factor of 0.99.

All strategies except for BB without discounting performed better than the baseline on all domain sizes ($p < 0.05$; $p < 0.001$ after 10 concepts).[3] While MA performed better than BB on smaller domains, BB performed better on larger domains (figures 6.3 and 6.4). This confirms H1. Discounted learning helped MA on smaller domains, but its impact diminished on larger domains. Meanwhile, the effect of discounting on BB was largest for small and medium-sized domains and actually became negative on the largest domains (table 6.1).

All strategies except for BB with discounting leveled off fairly quickly on all domains. Only BB with discounting continued to show much improvement after the initial 1000 rounds (figure 6.4).

In support of H2, lexical ambiguity was significantly higher for both BB strategies that either of the MA strategies (figure 6.5). Among BB strategies, ambiguity was higher for BB without discouning on medium-sized domains. Interestingly, these are the same domains where discounting had the strongest positive effect.

6.1.4. Discussion

In some regards, it is not at all surprising that BB outperforms the purely lexical strategies since it is in a sense tuned to the utility function. The sender picks an expression to maximize the expected utility of the present round (assuming that the receiver lexicon is aligned with her own) and the receiver picks a concept based on the knowledge that that is how the sender produces expressions. These simulations give good evidence that such pragmatic considerations of semantic approximation play a role in natural language.

There are other pragmatic considerations in natural language that are not considered in this simulation, but that could be modeled in future research. One could, for example, investigate the effects of relevance by assigning a prior distribution on concepts, a “context”. Agents would then be able to take the salience of various interpretations into account in their interpretation and production by considering those priors.

Another possibility that is not considered is the coining of new expressions. In this signaling game, agents’ lexicons are a fixed size. If we allowed agents to add new expressions to the lexicon, they would be able to alter the ratio of expressions to concepts. In real life it is relatively uncommon for natural language speakers to coin completely new expressions. Perhaps this is because attaching a meaning to an expression in a shared or community lexicon takes a lot of time. It is almost always better to use an existing expression in some way to jump-start the coordination process.

[3] All p-values indicate Welch’s t-tests.

	Number of concepts									
strategy	10	20	30	40	50	60	70	80	90	100
MA	0.57	0.20	-0.31	0.19	-0.16	0.04	0.12	0.00	0.01	0.03
BB	0.96	2.27	0.23	0.11	2.06	1.28	1.76	1.85	-0.05	-0.24

Table 6.1.: Effect of discounting on final expected utility (Cohen's d).

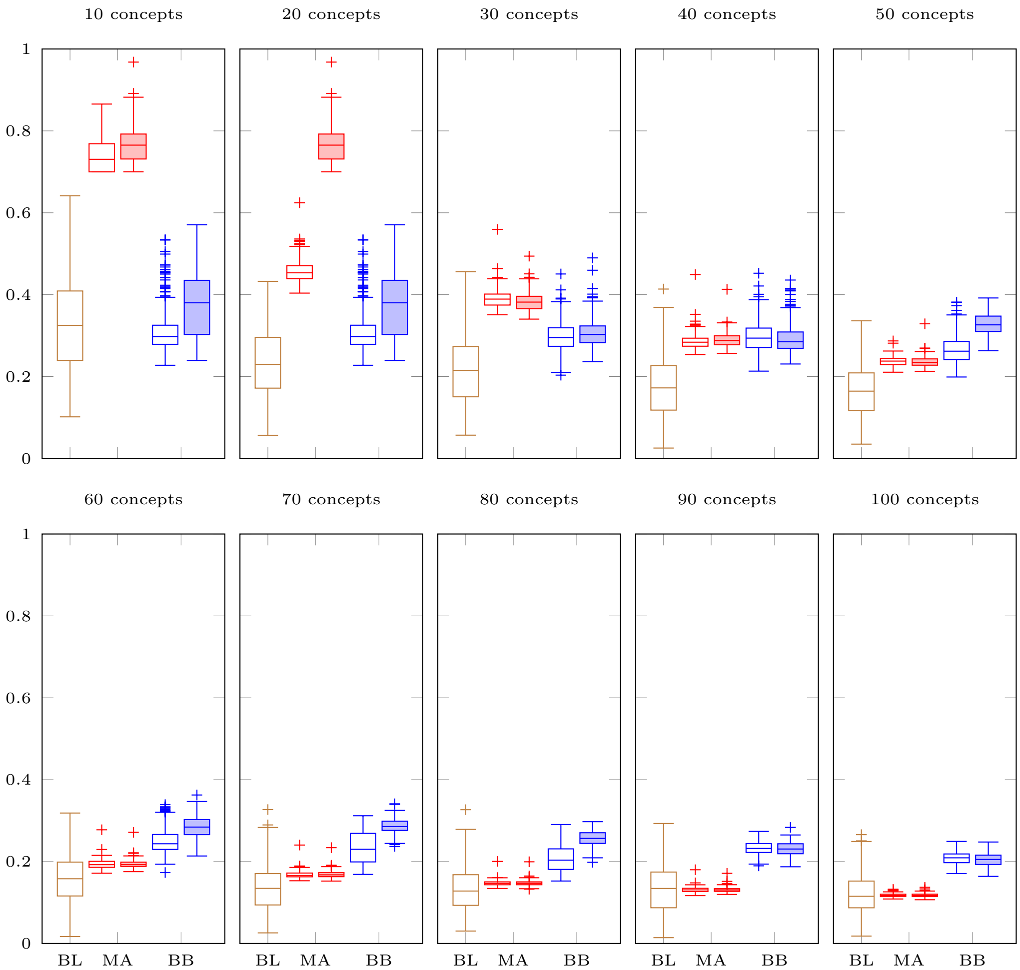


Figure 6.3.: Final expected utility (after 5000 iterations) over 250 trials. (Filled boxes indicate discounted learning)

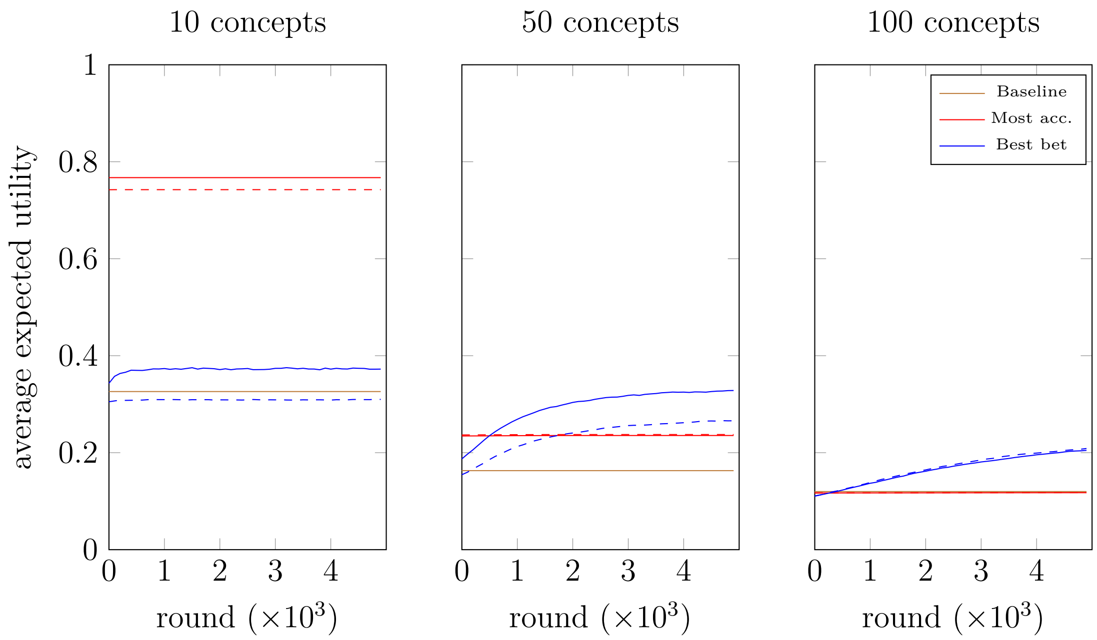


Figure 6.4.: Average expected utility over 250 trials. (Dashed lines indicate strategies without discounted learning.)

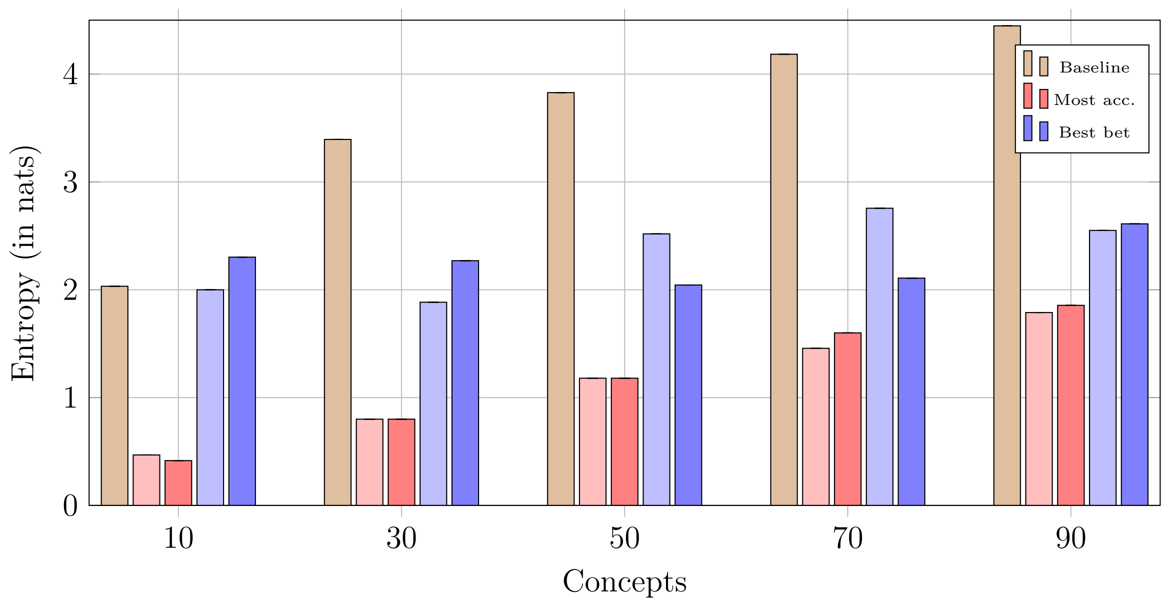


Figure 6.5.: Average lexical ambiguity by strategy (after 5000 iterations). (The darker red and blue bars use discounting.)

The fact that both agents give weight to *ad hoc* interpretations brings up issues of grounding since adding (or adjusting) an interpretation in the discourse lexicon is a joint action that must be grounded. How can it be grounded that a speaker means p by m when p is not in its interpretation set in the shared lexicon? There are two considerations to be made in this regard. The first is that relevance may play a role: perhaps given the current discourse context, p is simply more salient than q. That would give the speaker (and hopefully the listener) some grounding evidence for that interpretation. The other consideration is that the strictness of the grounding requirements are graded according to the consequences of a grounding failure (see §2.2.1). Since a grounding failure in this case is likely to result in the listener interpreting m as q (or from the listeners perspective, is likely to mean that the listener actually meant q by m), and since q and p are similar, the requirements for grounding m as p are greatly reduced. When the agents' play strategies are commonly known, there is even *positive* evidence for *ad hoc* uses of expressions where the meaning is conceptually related to a lexicalized meaning of that expression. In this game, that is modeled by the fact that both sender and receiver add weight to such interpretations in their production and interpretation strategies.

Another property of BB as a signaling game strategy is that it encourages "exploration" of the space of possible signaling conventions. To see why that is, consider again the simplified situation where the sender uses m to mean p even though it better expresses q. If the communication is successful—even if it is only partially successful because, for example, the receiver interprets m as q—the lexicalized interpretation $L_s(m|p)$ is reinforced. Successful *ad hoc* uses of expressions result in their being added to the lexicon (at least to some degree). This leads to a sort of "bleeding" effect where probability the mass of $L_s(q|m)$ dissipates to other concepts related to q. This is exactly analogous to the mechanism whereby *ad hoc* uses of of an expression lead to lexical polysemy. The following three conditions are sufficient to encourage polysemous lexicons:

1. There are fewer expressions available than concepts to express.[4]

2. Imperfect communication (i.e., the existence of partially successful communication by conceptual approximation) is acceptable.[5]

3. Agents follow a version of the "best bet" strategy of communication, using words to mean concepts that are similar to their (higher level) lexicalized interpretations.

The tendency of BB towards exploration also explains why it takes longer for expected utility to level off and why discounting has a bigger effect on this strategy (at least on large domains). While other strategies reach a convention fairly quickly and rarely

[4]Of course, ambiguity is not the only way to increase the expressivity of a lexicon of fixed size. Famously, the finite lexicons of natural language can be used to express arbitrarily many propositions (and therefore concepts) by using compositionality. For work on compositionality in signaling games, see (Smith et al., 2013; Franke, 2014; Brochhagen, 2015).

[5]In games where the utility function always produces either 0 or 1 the optimal equilibrium is always deterministic, even when there fewer messages than types/actions.

stray from it, any (lexical) convention reached in BB also encourages interpretations that are *outised* of that convention. Those interpretations then have a chance to be lexicalized themselves. Thus improvement of the convention continues for longer. When discounting is used, successful *ad hoc* interpretations can be forgotten. This is important because it may have been exactly the properties of the lexicon at the time (where the *ad hoc* interpretation had little or no weight) that made the exchange successful (or made other exchanges successful). This models a balance that must be maintained between adding useful *ad hoc* interpretations to the shared lexicon and keeping the shared lexicon sufficiently unambiguous to be useful for future discourses.

6.2. A Corpus Study in Contrastive Focus Reduplication

Ghomeshi et al. (2004) observe that contrastive focus reduplication is usually[6] used to "restrict the meaning of an item to its central or prototypical meaning". As evidence of this claim, they demonstrate that expressions with no prototypical meaning (such as function words, which lack semantic variation) are not valid objects of reduplication.

(1) a. #Are you sick, or ARE-are you sick?
b. #I didn't just read the book, I read THE-the book!

In the analysis in this thesis, contrastive focus reduplication narrows the discourse lexicon interpretation of an expression σ around the prototype; that is, the contrastive focus interpretation set is smaller, and every interpretation that was dropped is at least as far from the prototype as those that remain: If $[\![\sigma]\!]_{\mathsf{DL}} = \langle \mathbf{p}, \{p\} \rangle$ then $[\![\sigma^{cr}]\!]_{\mathsf{DL}} = \langle \mathbf{p}', \{p\} \rangle$ where $\mathbf{p}' \subsetneq \mathbf{p}$ and $\forall q \in \mathbf{p} \setminus \mathbf{p}'$ and $q' \in \mathbf{p}'$ we have $sim(p, q) \leq sim(p, q')$. The contrastive focus operation, as described above, assumes that $\mathsf{DL}(\sigma)$ has a *single* prototype—it is undefined otherwise:

(2) A: I am going to the bank.
B: Do you mean to the side of the river or to the place with money?
A: #I mean the BANK-bank.

Here the pragmatic intent of BANK-*bank* is unclear since $\wr bank \int_{\mathsf{DL}} = \{\texttt{bank}_1, \texttt{bank}_2\}$ (as indicated by B's either-or clarification request). Thus our model predicts that in order for contrastive focus reduplication with σ to be felicitous, it must be the case that $| \wr \sigma \int_{\mathsf{DL}} | = 1$

Recall that DL may have be disambiguated to a single prototype (even if SL or CL has more than one). This disambiguation is more likely to have happened in the discourse lexicon if the prototypes are far apart semantically:

[6] Other uses include shifting to a literal meaning, intensified meaning, or indicating a "value-added" meaning (Ghomeshi et al., 2004, p. 314).

(3) A: I need to withdraw some money.
B: There's an ATM over there.
A: No, I need to go to the BANK-bank.

In this example, although $\wr bank \int_{\mathsf{SL}}$ may indicate two sense classes for *bank* (as the DL does in example 2), $\wr bank \int_{\mathsf{DL}} = \{\mathtt{bank}_2\}$ since the other prototype is not relevant to the current context.

Assuming that contrastive focus reduplication can be successful like example 3, one would expect that reduplicated expressions tend to have either one sense class, or multiple sense classes that are easily disambiguated by context. There isn't an obvious answer to the question of what it means for sense classes to be easily disambiguated by context. But it's a reasonable assumption that those sense classes whose prototypes are conceptually dissimilar are more likely to be disambiguated by context: Given a situation, there is a range of concepts that might be considered relevant. It is more likely that two concepts will both fall within that range if they are similar to one another. Thus we assume that expressions with lower coherence (definition 5.2.4) in CL are more likely to have a single prototype in DL. The first hypothesis we would like to test, then, is that words with lower coherence are more likely to be subject to contrastive focus reduplication.

Now recall that contrastive focus reduplication also take auxiliary expressions. The primary function of these auxiliary expressions is to give positive and negative constraints on how tightly the focus set is drawn around the prototype. The other function they have is to disambiguate between prototypes.

(4) A: I need a BANK-bank, not an ATM.

Now even if DL hadn't disambiguated between $\mathtt{bank}_1$ and $\mathtt{bank}_2$ in the prototype set, example 4 picks out $\mathtt{bank}_2$ since the auxiliary expression providing negative a negative constraint on the focus set only makes sense for that prototype; that is, $\mathtt{bank}_2$ is the only prototype that has interpretations in its sense class under which ATM could possibly fall.

Since auxiliary expressions have this dual purpose, the second hypothesis is that the coherence of the topical expression tends to be higher in CR's that use auxiliary expressions.

6.2.1. SETUP

To test this hypothesis, we employ the contrastive focus reduplication corpus cited in Ghomeshi et al. (2004).[7] WordNet is used to approximate the CL coherence of expressions found in the corpus, and to provide baseline coherence for the sake of comparison.(Pedersen et al., 2004). To make use of WordNet, we must make the following assumptions:

[7] `home.cc.umanitoba.ca/~krussll/redup-corpus.html`

1. WordNet synsets correspond to sense classes in the community lexicon for English.[8]
2. The path similarity between WordNet senses is proportional to the similarity of their corresponding prototypes in the conceptual domain.
3. Senses whose part of speech differ are always disambiguated in the DL.

We use WordNet to approximate the coherence of English exrpessions as follows:

$$coh_{wn}(\sigma) = \frac{\sum_{p,q \in synset(\sigma)} path\_sim(p,q)}{\binom{|synset(\sigma)|}{2}}$$

Note that since coherence is being used to estimate the likelihood that more than one sense remain in the interpretation set, we define $coh_{wn}(\sigma) = 0$ when $|synset(\sigma)| = 1$.

Under these assumptions, the revised hypothesis is that reduplicated words tend to have a lower coherence (as defined above) .

An additional difficulty in testing this hypothesis is that to find out whether the reduplicated words have lower coherence, one needs some measure of baseline coherence. Ideally, we would test against the coherence of the other words in SL from which the dialogue is drawn; however, the examples in the small reduplication corpus do not come with a lot of context, so a good estimation of SL cannot be achieved. Thus we again use the assumption that SL, is well approximated by WordNet: For a baseline, we test against the coherence of all words (of the same part of speech) on WordNet.

6.2.2. Results

We find that for each part of speech tested, expressions appearing in reduplication constructions (in the CR corpus) have lower average coherence than the category as a whole. This confirms the prediction made by our account of contrastive focus reduplication.

	Reduplications	Baseline (All Words)
nouns	0.01633 (0.05214)	0.09097 (0.06757)
verbs	0.07352 (0.09136)	0.19293 (0.04237)
adjectives	0.00019 (0.01365)	0.03175 (0.17673)

Table 6.2.: Mean (and standard deviation) for reduplicated expressions and all expressions (in WordNet) of the same part of speech. Independence t-tests are $p < 0.001$ in all categories.

[8] See `https://wordnet.princeton.edu/wordnet/documentation/`.

Chapter 7

Conclusion

7.1. Overview of Main Contributions

The central contribution of this thesis is the Interactive Lexical Hierarchy (ILH), a model of lexical semantics for dialogue. This model is based on notions of common ground developed by Clark (1996)—different lexicons in the hierarchy are supported by different kinds of common ground. The lexical hierarchy is interactive because it allows for changes at lower levels in the hierarchy to be "learned" by lexicons at higher levels. In this way, semantic learning in the traditional sense (change in higher level lexicons) is propagated by semantic coordination (joint changes to the lowest level lexicon).

The ILH supports a novel theory of semantic coordination, which was the main motivation of this thesis. This theory of coordination is divided into a semantic part (how the discourse lexicon is updated) and a pragmatic part (how the coordinated meaning is determined). The pragmatic part of the theory identifies several implicit and explicit "coordination strategies" that are used in colloquial English and gives a detailed account of three of those explicit strategies.

Finally, this thesis contains some empirical work to support the theoretical choices that were made. A signaling games simulation is used to demonstrate the plausibility of the claim that semantic coordination leads to lexical learning (and therefore to lexicalized polysemy), and a corpus study confirms predictions made by the LIH and by the pragmatics of contrastive focus reduplication as a coordination strategy.

7.2. Directions for Future Research

This thesis is rich in potential future work both in improving model developed here and in its applications. Three promising directions are considered here.

Probabilistic Lexicon One avenue for extending the ILH is to move to a probabilistic lexicon where concepts in the interpretation set of an expression are given a likelihood probability according to the degree to which it expresses them (similar to how lexicons are defined in section 6.1). Discourse context would then provide prior probabilities over the concept space. From the likelihood of a given interpretation and the prior distribution over concepts, Bayesian production and interpretation would be carried out by agents. In order to fully integrate with the work in this thesis, some additional challenges would have to be broached. For example, introducing a probabilistic lexicon raises the question of how interpretation likelihood should interact with prototypicality. Must more prototypical interpretations also be more likely or are prototypicality and interpretation likelihood completely independent? Alternatively, some may wish to argue that all so-called prototypicality effects can be explained by differences in interpretation likelihood and there is therefore no need for prototypes.

Enhanced Lexical Inheritance Another avenue for future research is to further develop the notion of lexical inheritance to model more complicated relationships between different joint lexicons. The simple three-tier hierarchy presented in this thesis gives a concrete model of lexical relationships in dialogue. This is already an improvement over traditional lexical models that struggle to adequately account for semantic coordination; however there are certainly lexical relationships that are not captured. One example—complex hierarchical relations between community lexicons—is mentioned in section 4.2.1. Another phenomenon this thesis fails to account for is *translanguaging*; that is, when multilingual agents employ more than one community language resource in the same discourse. Molina (2011) speaks of a "common communicative arena" among speakers which may include resources from multiple languages. In a framework like the ILH, this would manifest as the shared lexicon inheriting from multiple community lexicons. Multiple inheritance is not defined in this model. Such an extension would have to account for shared lexicons where the community lexicon that expressions are drawn from depends on such things as subject matter and social context.

Distributional Semantics Finally, perhaps the most promising area for future research is applications in distributional semantics (cf. Lenci, 2008; Turney and Pantel, 2010). Recall that one of the motivations of this work is the observation that data-driven methods in semantics have a lot to gain from innovation that draws inspiration from work in the naturalization of formal semantics. The ILH is presented in this thesis as a cognitive model of lexical semantics. For that reason, each agent has her own representation of each joint lexicon in the hierarchy. Ultimately, the ILH (or something like it) should be taken into account in distributional models of semantics. To do so, it would probably be most productive to assume a single lexicon for each discourse, group of agents, and linguistic community. Each lexical level presents its own challenges for computing meaning representations. In the discourse lexicon, the problem is sparsity. Even given a very large corpus, it is unlikely that discourses go on long enough to provide sufficient data to build, for example, a vector representation of word meanings. Suppose that we had such representations though. How would the community and shared lexicons be computed from there? One possibility would be to take averages (for example component-wise averages of vector representations) over the discourses involving the agents in the corresponding group or community. This seems like a reasonable approach, but there isn't any *a priori* reason to believe that discourse lexicons will distribute word meanings that are centered around their meaning in the shared or community lexicon from which they are coordinated. One possible direction for addressing both these concerns is to incorporate what we know about semantic coordination into how the distributional representations are computed. For example, one might compute vector representations for the whole corpus (representing the community lexicon) and then modify those representation per-discourse based on observed semantic coordination. Shared lexicons could then be inferred from the discourse lexicons and even be used in turn to generate an improved estimate of the community lexicon. A natural follow-on question is whether iterating such a process would converge or not. If it does converge, the hope would be that accounting for semantic coordination will lead to more careful handling of contexts where expressions are used in an *ad hoc* manner and ultimately to better meaning representations at the highest lexical level.

Appendix A

Type Theory with Records

What follows is a brief, but formal definition of the system known as Type Theory with Records (TTR) (Cooper, 2005b, 2012). While the emphasis of presentation differs slightly, we closely follow the formalism and notation of (Cooper, 2012). An earlier version of the system can be found in Cooper (2005b).

Definition A.0.1 (Basic Type System)**.** A system of basic types is a pair

$$\mathbf{TYPE}_B = \langle \mathbf{Type}, A \rangle$$

where:

- **Type** is a non-empty set and
- A is a function such that $\text{dom}(A) = \mathbf{Type}$ and $\text{ran}(A)$ contains sets disjoint from **Type**.

Definition A.0.2 (Basic Type Judgments)**.** Let $\mathbf{TYPE}_B = \langle \mathbf{Type}, A \rangle$ be a basic system of types. For any type $T \in \mathbf{Type}$, a is of type T if and only if $a \in A(T)$:

$$a :_{\mathbf{TYPE}_B} T \overset{def}{\iff} a \in A(T).$$

Before moving on to complex type systems, we will say what it means for a type system to have records and record types. Since the existence of record types is agnostic to other features of the type system, we will consider an arbitrary type system **TYPE** with types **Type**.

Definition A.0.3 (Record Types)**.** **TYPE** is said to have record types based on a label set L if:

- there is a type $Rec \in \mathbf{Type}$; and
- for any sequence distinct labels $l_1, \ldots, l_n \in L$ and types $T_1, \ldots, T_n \in \mathbf{Type}$, we have a type $\{\langle l_1, T_1\rangle, \ldots \langle l_n, T_n\rangle\} \in \mathbf{Type}$.

For visual clarity, record types are typically displayed in tabular form:

$$\begin{bmatrix} l_1 & : & T_1 \\ \ldots & & \\ l_n & : & T_n \end{bmatrix}$$

Now we turn to record types.

Definition A.0.4 (Record)**.** Let **TYPE** be a system of types with records based on L. A record is a finite set of pairs $\{\langle l_1, a_1\rangle, \ldots \langle l_n, a_n\rangle\}$ such that each $l_i \in L$ and each $a_i :_{\mathbf{TYPE}} T$ for some $T \in \mathbf{Type}$. Records are likewise displayed as tables:

$$\begin{bmatrix} l_1 & = & a_1 \\ \ldots & & \\ l_n & = & a_n \end{bmatrix}$$

Note that records and record types are functional graphs (sets of pairs). We sometimes refer to the set of labels as the record's domain, and the set of values as its range. For example, given a record r, $\mathrm{dom}(r)$ refers to the labels of r and $\mathrm{ran}(r)$ to its values.

Furthermore, since records and record types may be nested, we use path notation to refer to specific values. Given

$$r = \begin{bmatrix} l_1 & = & a_1 \\ l_2 & = & a_2 \\ l_3 & = & \begin{bmatrix} k_1 & = & b_1 \\ k_2 & = & b_2 \end{bmatrix} \end{bmatrix}$$

We write $r.l_1 = a_1$ and $r.l_3.k_1 = b_1$.

Definition A.0.5 (Record Type Judgments)**.** Let $r = \{\langle l_1, a_1\rangle \ldots \langle l_n, a_n\rangle\}$ be a record in type system **TYPE** with record types based on L. Then:

- $r :_{\mathbf{TYPE}} Rec$; and
- $r :_{\mathbf{TYPE}} \{\langle k_1, T_1\rangle \ldots \langle k_m, T_m\rangle\} \overset{def}{\Longleftrightarrow}$ for each k_i there is an l_j such that $k_i = l_k$ and $a_i :_{\mathbf{TYPE}} T_j$.

In the following, we write $\mathrm{seq}(X)$ to denotes the set of finite sequences of elements of X.

Definition A.0.6 (Complex Type System)**.** A system of complex types is a quadruple

$$\mathbf{TYPE}_C = \langle \mathbf{Type}, \mathbf{BType}, \langle \mathbf{PType}, \mathbf{Pred}\rangle, \langle A, F\rangle\rangle$$

where:

- $\langle \mathbf{BType}, A\rangle$ is a system of basic types,
- **Pred** is a function whose domain is a set of predicates and

  $$\mathrm{ran}(\mathbf{Pred}) \subseteq \mathcal{P}(\mathrm{seq}(\mathbf{Type})).$$

  If P is a predicate, then $\mathbf{Pred}(P)$ is called P's *arity.*
- $P(a_1, ..., a_n) \in \mathbf{PType}$ if and only if $\mathbf{Pred}(P) = \langle T_1, \ldots, T_n\rangle$ and we have $a :_{\mathbf{TYPE}_C} T_1, \ldots, a_n :_{\mathbf{TYPE}_C} T_n$,
- F is a function with $\mathrm{dom}(F) = \mathbf{PType}$ and whose range is sets disjoint from **Type**, and
- $\mathbf{Type} \supseteq \mathbf{BType} \cup \mathbf{PType}$.

Note that predicates may have polymorphic arities.

In the following, Let $\mathbf{TYPE}_C = \langle \mathbf{Type}, \mathbf{BType}, \langle \mathbf{PType}, \mathbf{Pred}\rangle, \langle A, F\rangle\rangle$ be an arbitrary system of complex types.

Definition A.0.7 (Complex Type Judgments)**.**

- If $T \in \mathbf{BType}$ then $a :_{\mathbf{TYPE}_C} T \overset{def}{\iff} a :_{\langle \mathbf{TYPE}_B, A\rangle} T$.
- If $T \in \mathbf{PType}$ then $a :_{\mathbf{TYPE}_C} T \overset{def}{\iff} a \in F(T)$.

Note that in definition A.0.6, **Type** (and therefore $:_{\mathbf{Type}_C}$ in definition A.0.7) is underdetermined. In general, we assume **Type** to be the *smallest* set meeting some given criteria. We may expand those notions to meet the additional criteria described below.

Definition A.0.8 (Function Types)**.** $\mathbf{TYPE}_C$ is said to have function types if for any T_1, $T_2 \in \mathbf{Type}$, there is a type $(T_1 \to T_2) \in \mathbf{Type}$.

Definition A.0.9 (Function Type Judgments)**.** If $\mathbf{TYPE}_C$ has function types, then $f :_{\mathbf{TYPE}_C} (T_1 \to T_2)$ if and only if f is function with

$$\text{dom}(f) = \{a \mid a :_{\mathbf{TYPE}_C} T_1\} \text{ and}$$
$$\text{ran}(f) \subseteq \{a \mid a :_{\mathbf{TYPE}_C} T_2\}.$$

We want to be able to represent propositional attitudes as intensional relations between individuals and propositions. Suppose we have basic type *Ind* and a predicate *believe*. What is the arity of *believe*? It should take an individual and a proposition. Unfortunately, is no basic type for propositions—a proposition may be of *any* type. Thus we need a type *Type* such that for any type T we have

$$T :_{\mathbf{TYPE}_C} Type.$$

Then we let $\mathbf{Pred}(believe) = \langle Ind, Type\rangle$ as desired.

Stipulating a universal type puts us on dangerous territory though, since it allows us to construct Russell's paradox. The solution to this problem, although somewhat cumbersome, is to use stratification. Since it allows us to treat types as objects, we call it an *intensional* type system.

Definition A.0.10 (Intensional System of Complex Types)**.** An intensional system of complex types is a family of systems of complex types

$$\mathbf{TYPE}_{IC} = \langle \mathbf{Type}^n, \mathbf{BType}, \langle \mathbf{PType}^n, \mathbf{Pred}\rangle, \langle A, F^n\rangle\rangle_{n\in\mathbb{N}}$$

where:

- for all $n \in \mathbb{N}$, $\mathbf{Type}^n \subseteq \mathbf{Type}^{n+1}$ and $\mathbf{PType}^n \subseteq \mathbf{PType}^{n+1}$;
- for all $n \in \mathbb{N}$ and $T \in \mathbf{PType}^n$, $F^n(T) \subseteq F^{n+1}(T)$; and
- for all $n > 0$, there is a type $Type^n \in \mathbf{Type}^n$.

We will sometimes need to refer to the nth system of complex types in an intensional system of complex types:

$$\mathbf{TYPE}_{IC^n} = \langle \mathbf{Type}^n, \mathbf{BType}, \langle \mathbf{PType}^n, \mathbf{Pred}\rangle, \langle A, F^n\rangle\rangle.$$

In what follows, let $\mathbf{TYPE}_{IC} = \langle \mathbf{Type}^n, \mathbf{BType}, \langle \mathbf{PType}^n, \mathbf{Pred}\rangle, \langle A, F^n\rangle\rangle_{n\in\mathbb{N}}$ be an arbitrary intensional system of complex types.

Definition A.0.11 (Intensional Type Judgments)**.** The type judgments of the component systems with index > 0 are expanded to accommodate intensional judgments as follows: For any $n \in \mathbb{N}$ and $T \in \mathbf{Type}^n$ we have

$$T :_{\mathbf{TYPE}_{IC^{n+1}}} Type^{n+1} \overset{def}{\Longleftrightarrow} T \in \mathbf{Type}^n.$$

The intensional types system collects the judgments of all of its components:

$$a :_{\mathbf{TYPE}_{IC}} T \overset{def}{\Longleftrightarrow} a :_{\mathbf{TYPE}_{IC^n}} T \text{ for some } n \in \mathbb{N}.$$

Now we may use the property of predicate polymorphism so that *believe* can take an individual and any proposition:

$$\mathbf{Pred}(believe) = \{\langle Ind, Type^n \rangle \mid n \in \mathbb{N}\}$$

We don't usually want to worry about the details of stratification so we may write, for example, $\mathbf{Pred}(believe) = \langle Ind, Type \rangle$, but the use of the unindexed $Type$ indicates that formally, *believe* has a polymorphic arity as above. Likewise, we will sometimes refer to a set **Type** which is just the union of each $\mathbf{Type}^n$.

Definition A.0.12 (Dependent Function Types)**.** $\mathbf{TYPE}_{IC}$ is said to have dependent function types if

- each $\mathbf{TYPE}_{IC^n}$ has function types; and
- for any $n \in \mathbb{N}$, $T \in \mathbf{Type}^n$, and $\mathcal{F} :_{\mathbf{Type}_{IC^n}} (T \to \mathbf{Type}^n)$, there is a type $((a : T) \to \mathcal{F}(a)) \in \mathbf{Type}^n$.

Definition A.0.13 (Dependent Function Type Judgments)**.** If $\mathbf{TYPE}_{IC}$ has dependent function types, then $f :_{\mathbf{TYPE}_{IC}} ((a : T) \to \mathcal{F}(a))$ if and only if

- $\text{dom}(f) = \{a \mid a :_{\mathbf{TYPE}_{IC^n}} T\}$; and
- for any $a \in \text{dom}(f)$, we have $f(a) :_{\mathbf{TYPE}_{IC^n}} \mathcal{F}(a)$.

Definition A.0.14 (Dependent Record Types)**.** $\mathbf{TYPE}_{IC}$ is said to have dependent function types if it has record types and for each $n \in \mathbb{N}$: if $R \in \mathbf{RType}^n$, $R.\pi_1, ..., R.\pi_m$ are paths in R and $\mathcal{F} : ((a_1 : T_1) \to \ldots ((a_m : T_m) \to \mathbf{Type}^n) \ldots)$ is a function type, then $R \cup \{\langle l, \langle \mathcal{F}, \langle \pi_1, \ldots \pi_m \rangle \rangle \rangle\} \in \mathbf{RType}^n$ (where $l \notin \text{dom}(R)$).

Note that by definition, dependent record types are just a special kind of record type.

Definition A.0.15 (Dependent Record Type Judgments)**.** If $\mathbf{TYPE}_{IC}$ has dependent record types, then $r :_{\mathbf{TYPE}_{IC}} R \cup \{\langle l, \langle \mathcal{F}, \langle \pi_1 \ldots \pi_m \rangle \rangle \rangle\}$ if and only if $r : R$, $l \in \text{dom}(r)$, and $r.l : \mathcal{F}(r.\pi_1 \ldots r.\pi_m)$.

Appendix B

Coordination Examples Revisited

Example 5 (§4.3.1)

This example negotiates the meaning of the word *salad*. For ease of reference, we will refer to the first three utterances by their position in the dialogue:

u_1 = Would you bring some salad to the barbecue?
u_2 = Sure, should it be a SALAD-salad or is something like tuna salad ok?
u_3 = SALAD-salad if possible.

We assume that the discourse participants have two interpretations for *salad* in their shared lexicon:

$$|salad|_{\mathsf{SL}} = \{\texttt{salad}_1, \texttt{salad}_2\}$$

and that there is one prototype (and therefore one sense class):

$$\wr salad \int_{\mathsf{SL}} = \{\texttt{salad}_1\}$$

However u_1 (uttered by A) assumes a single sense for *salad*. This is because there is a discrepancy in the discourse lexicon—A thinks that *salad* has been disambiguated to exclude $\texttt{salad}_2$:

$$\mathsf{DL}_1^{i-(A)}(salad) = \{\texttt{salad}_2\}$$

While B does not:

$$\mathsf{DL}_1^{i-(B)}(salad) = \varnothing$$

Thus the interpretation of salad in their respective discourse lexicons are as follows:

$$|salad|_{\mathsf{DL}_1^A} = |salad|_{\mathsf{SL}} \setminus \mathsf{DL}_1^{i-(A)}(salad) = \{\texttt{salad}_1\}$$
$$|salad|_{\mathsf{DL}_1^B} = |salad|_{\mathsf{SL}} \setminus \mathsf{DL}_1^{i-(B)}(salad) = \{\texttt{salad}_1, \texttt{salad}_2\}$$

This prompts B to make a clarification request (u_2) that employs contrastive focus reduplication as a cooperation strategy.

$$\delta^{cr}(salad, \textsc{salad}\text{-}salad, tuna\text{-}salad^{(-)}) = \{\texttt{salad}_1\}$$

The coordinated meaning is focused around the prototype $\texttt{salad}_1$. Furthermore, $\texttt{salad}_2$ is guaranteed to be left out of the coordinated meaning since $p \subseteq \texttt{salad}_2$ for all $p \in \wr tuna\text{-}salad \int_{\mathsf{DL}_1}$.

Finally, u_3 confirms that the coordinated meaning is what A originally meant by *salad* and the discourse lexicon is coordinated appropriately:

$$\mathsf{DL}_3^{i-(A)} = \mathsf{DL}_3^{i-(B)} = \{\texttt{salad}_1\}$$

Thus the meaning of *salad* in the discourse lexicon is as follows:

$$|salad|_{\mathsf{DL}_3} = |salad|_{\mathsf{SL}} \smallsetminus \mathsf{DL}_3^{i-}(salad) = \{\texttt{salad}_1\}$$

Example 17 (§4.3.5)

Recall that example 17 uses a coordination strategy called *navigating lexical structure* where agents coordinate to disambiguate the discourse-meaning of an expression my making direct reference to its semantic structure in the shared or community lexicon. Although the pragmatics of this coordination strategy, δ^{NLS} are not discussed in section 4.3, it is nonetheless instructive because conceptual similarity comes into play.

Suppose for this example that there are three concepts in the shared interpretation set of *mad*; that is,

$$|mad|_{\mathsf{SL}} = \{\texttt{mad}_1, \texttt{mad}_2, \texttt{mad}_3\}.$$

The prototypes for being mad are being mad *at* someone or something ($\texttt{mad}_2$) or being mad, as insane ($\texttt{mad}_3$). Thus

$$\wr mad \int_{\mathsf{SL}} = \{\texttt{mad}_2, \texttt{mad}_3\},$$

with $\texttt{mad}_1$ as a subordinate meaning.

The concepts in question are identified with the following record types:

$$\texttt{mad}_1 = \begin{bmatrix} x & : & indv \\ y & : & mental\_state \\ c_1 & : & mental\_state\_of(x,y) \\ c_2 & : & angry1(y) \end{bmatrix} \texttt{mad}_2 = \begin{bmatrix} x & : & indv \\ y & : & mental\_state \\ z & : & indv \\ c_1 & : & mental\_state\_of(x,y) \\ c_2 & : & angry1(y) \\ c_3 & : & angry\_at(z,y) \end{bmatrix}$$

$$\texttt{mad}_3 = \begin{bmatrix} x & : & indv \\ y & : & mental\_state \\ c_1 & : & mental\_state\_of(x,y) \\ c_2 & : & insane(y) \end{bmatrix} \texttt{mstate} = \begin{bmatrix} x & : & indv \\ y & : & mental\_state(y) \\ c1 & : & mental\_state\_of(x,y) \end{bmatrix}$$

Note that $\texttt{mad}_1 \sqsubseteq \texttt{mad}_2$ and $\texttt{mad}_1, \texttt{mad}_2, \texttt{mad}_3 \sqsubseteq \texttt{mstate}$; even though there are two sense classes, all senses are related (by proposition 5.1.10 (d)) Thus it is not immediately obvious to which sense class $\texttt{mad}_1$ belongs. Since $sim(\texttt{mad}_1, \texttt{mad}_2) = 0.8$ and $sim(\texttt{mad}_1, \texttt{mad}_3) = 0.75$, $\texttt{mad}_1$ is in the sense class of $\texttt{mad}_2$.

Before the coordination occurs, there is a discrepancy in the discourse lexicon for *mad*:[1]

$$|mad|_{\mathsf{DL}_0^A} = \{\mathtt{mad}_2\}$$
$$|mad|_{\mathsf{DL}_0^B} = \{\mathtt{mad}_3\}$$

Note that B only considers the possibility that Hubert is mad (as in angry) *about* something, not that is he is *just* angry. Now B can infer something about the state of A's discourse lexicon when she asks "what is he mad *about*". Namely he infers that she has narrowed the interpretation set of *mad* to senses where *mad* has an object (i.e., $\mathtt{mad}_2$). When he B says "I mean the *other* kind of mad, he intends to coordinate a meaning that includes only the (shared lexicon) sense class of *mad* to which $\mathtt{mad}_2$ does not belong. In particular,

$$\delta^{\text{NLS}}(mad, \textit{other kind of mad}) = \langle\{\mathtt{mad}_3\}, \{\mathtt{mad}_3\}\rangle$$

[1] As in the previous example, this discrepancy stems from a discrepancy in DL_0^{i-} since DL inherits from SL.

Bibliography

Achermann, B. and Bunke, H. (2000). Classifying range images of human faces with Hausdorff distance. In *Proceedings of the 15th International Conference on Pattern Recognition*, volume 2, pages 809–813.

Aloni, M. (2001). *Quantification Under Conceptual Covers.* ILLC Dissertation Series.

Armstrong, S. L., Gleitman, L. R., and Gleitman, H. (1983). What some concepts might not be. *Cognition.*

Barabási, A.-L. and Bonabeau, E. (2003). Scale-free. *Scientific American.*

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, pages 101–140.

Barsalou, L. W. (1992). Frames, concepts and conceptual fields. *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organizaion*, pages 21–74.

Barwise, J. and Perry, J. (1981). Situations and attitudes. *The Journal of Philosophy*, pages 668–691.

Boas, H. and Sag, I. A. (2012). *Sign-Based Construction GRammar.* Center for the Study of Language and Information.

Brennan, S. E. and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482–1493.

Brochhagen, T. (2015). Minimal requirements for productive compositional signaling. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

Carston, R. (2002). Metaphor, ad hoc concepts and word meaning–more questions than answers. *UCL Working Papers in Linguistics*, 14:83–105.

Catteeuw, D. and Manderick, B. (2014). The limits of robustness of reinforcement learining in lewis signalling games. *Connection Science*, 26(2):161–177.

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

Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13(2):259–294.

Clark, H. H. and Wilkes-Gibbs, D. (1986). Refering as a collaborative process. *Cognition*, 22:1–39.

Cooper, R. (2005a). Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3(2–3):333–362.

Cooper, R. (2005b). Records and record types in semantic theory. *Journal of Logic and Computation*, 15(2):99–112.

Cooper, R. (2008). Type theory with records and unification-based grammar. *Logics for Linguistic Structures*, 201:9.

Cooper, R. (2012). Type theory and semantics in flux. *Handbook of the Philosophy of Science*, 14:271–323.

Cooper, R., Dobnik, S., Lappin, S., and Larsson, S. (2015). Probabilistic type theory and natural language semantics. In preparation.

Cooper, R. and Ginzburg, J. (2015). TTR for natural language semantics. In Fox, C. and Lappin, S., editors, *Handbook of Contemporary Semantic Theory*. Blackwell: Oxford, second edition edition. To appear.

Cooper, R. and Larsson, S. (2009). Compositional and ontological semantics in learning from corrective feedback and explicit definition. In *Proceedings of DiaHolmia: 2009 workshop on the semantics and pragmatics of dialogue*, pages 59–66.

Duvignau, K., Fossard, M., Guame, B., Pimenta, M. A., and Elise, J. (2007). Semantic approximations and flexibility in the dynamic construction and "deconstruction" of meaning. *Linguagem em (Dis)curso*, 7(3):371–387.

Fernández, R. (2014). *Dialogue*. Oxford University Press, 2nd edition.

Fernández, R. and Larsson, S. (2014). Vaugeness and learning: A type-theoretic approach. In *Proceedings of the Third joint Conference on Lexical and Computational Semantics*, pages 151–159.

Franke, M. (2014). Creative compositionality from reinforcement learning in signaling games. *Erica A. Cartmill, Sean Roberts, Heidi Lyn and Hannah Cornish/The Evolution of Language: Proceedings of EvoLang*, 10:82–89.

Frost, R. and Bentin, S. (1992). Processing phonological and semantic ambiguity: Evidence from semantic priming at different soas. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(1):58.

Ganter, B. and Wille, R. (1996). *Formal Concept Analisis: Mathematical Foundations*. Springer.

Garrod, S. and Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, (53):181–215.

Garrod, S. C. and Anderson, A. (1987). Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition*, 27:181–218.

Ghomeshi, J., Jackendoff, R., Rosen, N., and Russell, K. (2004). Contrastive focus reduplication in English (The salad-salad paper). *Natural Language & Linguistic Theory*, 22(2):307–357.

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

Grice, H. P. (1961). The causal theory of perception. *Proceedings of the Aristotelian Society Supplementary Volume*, 35:121–152.

Grice, H. P. (1975). Logic and conversation. pages 41–58. Academic Press.

Groenendijk, J. and Stohof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100.

Harris, Z. (1954). Distributional strucutre. *Word*, 10(23):146–162.

Holme, P. and Kim, B. J. (2002). Growing scale-free networks with tunable clustering. *Physical review E*, 65(2):026107.

Hopper, R., Knapp, M. L., and Scott, L. (1981). Couples' personal idioms: Exploring intimate talk. *Journal of Communication*, 31(1):23–33.

Horn, L. (1993). Economy and redundancy in a dualistic model of natural language. *Sky*, 1993:31–72.

Kamp, H. (1981). A theory of truth and semantic representation. *Formal Methods in the Study of Language*.

Kamp, H. and Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57(2):129–191.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, pages 458–508.

Larsson, S. and Cooper, R. (2009). Towards a formal view of corrective feedback. In *Proceedings of the EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 1–9. Association for Computational Linguistics.

Lasersohn, P. (1999). Pragmatic halos. *Language*, pages 522–551.

Lassiter, D. (2011). *Vaugeness and Probabilistic Linguistic Knowledge*. Springer.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, 20(1):1–30.

Lewis, D. (1969). Convention: A philosophical study.

Mills, G. J. and Healey, P. G. (2008). Semantic negotiation in dialogue: the mechanisms of alignment. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 46–53. Association for Computational Linguistics.

Molina, C. (2011). Curricular insights into translingualism as a communicative competence. *Journal of Language Teaching and Research*, 2(6):1244–1251.

Osherson, D. N. and Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*.

Palmer, M., Dang, H. T., and Fellbaum, C. (2007). Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering*, 13(2):137–163.

Partee, B. H. (2011). Formal semantics: Origins, issues, early impact. *Baltic International Yearbook of Cognition, Logic and Communication*, 6(1):1–52.

Paul, H. (1891). *Principles of the History of Language*. London; New York: Longmans, Green.

Peacocke, C. (1992). *A Study of Concepts*. The MIT Press.

Pedersen, T., Patwardhan, S., and Michelizzi, J. (2004). Wordnet:: Similarity: measuring the relatedness of concepts. In *Demonstration papers at HLT-NAACL 2004*, pages 38–41. Association for Computational Linguistics.

Pickering, M. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

Pollard, C. (1994). *Head-Driven phrase structure grammar*. University of Chicago Press.

Pustejovsky, J. (1995). *The Generative Lexicon*. The MIT Press.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of IJACI-95*.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In Moore, T., editor, *Cognitive development and the acquisition of language*. New York: Academic Press.

Roth, A. E. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior*, 8(1):164–212.

Schelling, T. (1960). *The Strategy of Conflict*. Cambridge, Massachusetts: Harvard University Press.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., and Biekowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14(4):489–537.

Shieber, S. M. (1986). *An Introduction to Unification-Based Approaches*. Microtome Publishing.

Smith, K., Tamariz, M., and Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the 35th annual conference of the Cognitive Science Society*, pages 1348–1353.

Song, M. and Lee, C. (2011). Cf-reduplication in english: Dynamic prototypes & contrastive focus effects. In *Proceedings of SALT*, volume 21, pages 444–462.

Sperber, D. and Wilson, D. (1986). Relevance: Communication and cognition.

Stalnaker, R. (1978). Assertion. In Cole, P., editor, *Pragmatics*, volume 9 of *Syntax and Semantics*. New York Academic Press.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(re) consideration of context effects. *Journal of verbal learning and verbal behavior*, 18(6):645–659.

Turney, P. and Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artifical Intelligence Research*, 27:141–188.

Whitton, L. (2006). What can be contrasted in contrastive reduplication. *Third Annual QP Fest, Stanford*.

William Onifer, D. A. S. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of the frequency of meaning and contextual bias. *Memory & Cognition*, 9(3):225–236.

Wilson, D. and Sperber, D. (2008). *Relevance Theory*. Oxford: Blackwell.

Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pages 189–196. Association for Computational Linguistics.