The interactions of rational pragmatic agents provide a framework for understanding efficient language structure and use

Author 1
bpeloqui@stanford.edu
Department of Psychology
Stanford University

Author 2
author1@university.edu
Department of Psychology
Some University

Abstract

We discuss a framework for studying the distributional properties of linguistic systems as emerging from in-the-moment interactions of speakers and listeners. Our work takes Zipfian notions of lexicon-level efficiency as a starting point, connecting these ideas to Gricean notions of conversational-level efficiency. To do so, we begin by deriving an objective function for measuring the communicative efficiency of linguistic systems and then examining the behavior of this objective in a series of simulations focusing on the communicative function of ambiguity in language.

Keywords: Communicative efficiency, Rational Speech Act theory, computational modeling, information theory, agent-based simulation

Introduction

Why do natural languages look they way they do? While Zipf (1935) presented his "Principle of Least Effort" as a domain general framework for understanding human behavior, he took language as his central case-study. He proposed that distributional properties found in natural language were evidence of speaker-listener effort minimization. In his own words, "we are arguing that people do in fact act with a maximum economy of effort, and that therefore in the process of speaking-listening they will automatically minimize the expenditure of effort." Evidence for this claim was largely derived at the level of the lexicon. Zipf argued that the particular relationship between a word's frequency and its rank, length, and denotation size could be explained as an emergent property of speaker-listener effort minimization.

Zipf articulated what is now considered a *functionalist* approach to language science – analyzing language structure and use in terms of efficiency. Such an approach might reframe our opening question as follows – how does having **property x** make using language ℓ more or less useful for communication? This reframing provides an opportunity to study both language *structure* (as Zipf primarily did) or *use* in terms of efficiency. For example, Regier et al. (201X) showed that languages appear to organize semantic domains (form-meaning mappings) in maximally efficient ways. Likewise, Piantadosi et al. (2012) argued that lexical ambiguity is an efficient property of any communication system when communication is contextualized. We will return to Piantadosi's argument later in the paper.

Zipf's original work and subsequent functionalist projects describe a varied set of linguistic properties and human behaviors. Common to many is a characterization of a fundamental effort-asymmetry underlying everyday communication. Simply put, what is effortful as a speaker is likely different from what is effortful as a listener. Zipf characterized this as follows – purely from the standpoint of speaker effort, what Zipf called "Speaker's Economy," an optimal language $\ell^*_{speaker}$ would tend toward a vocabulary of just a single, low-cost word. Given such a language, the full set of potential meanings would be conveyed using only that word, i.e. $\ell^*_{speaker}$ would be fully ambiguous and all possible meanings would need to be disambiguated by a listener. From the standpoint of listener effort, what Zipf called "Auditor's Economy," an optimal language $\ell^*_{listener}$ would bijectively map all possible meanings to distinct words, eliminating a listener's need to disambiguate. Here the asymmetry is operationalized by different costs – speaker effort is related to *production cost* and listener effort to *understanding or disambiguation cost*. Clearly natural languages fall between the two extremes of $\ell^*_{speaker}$ and $\ell^*_{listener}$. Zipf proposed that the particular lexicon-level properties he observed emerged from the competition of these forces – the pressure to jointly minimize speaker and listener effort.

This formulation begs the question – at what level of analysis do local, in-the-moment interactions of speakers and listeners lead to optimized structure at the level of the lexicon? The example given by Zipf (1935) appears to describe local interaction in terms of a reference game. Speakers intend to refer to some object in the world *m*. They choose some utterance *u* to transmit this intended meaning, $u \rightarrow m$. The listener attempts to reconstruct this intended meaning given the transmitted utterance $m \rightarrow u$. Other functionalist projects have also assumed this basic reference game setting (Regier, et al. (201X); Piantadosi et al. (2012); Kirby, Smith & Folds (201X), OTHERS) and this simplification of the communicative act has proven productive in both theoretical (Ferrer-i-cancho, et al., 201X), simulation-based (iterated learning stuff) and empirical explorations of efficient language structure and use (Some of Robert's work here).

This move to study efficient language use in terms of reference games has clear theoretical underpinnings as well. A half century after Zipf, the linguist Lawrence Horn (1984), highlighted the importance of Zipf's principles for explaining conversation-level phenomena. Horn suggested a direct link between Zipf's Speaker and Listener economy and, what was at the time, more recent work on conversational pragmatics by Grice (1975). Horn highlighted that the interaction of Zipf's forces were "largely responsible for generating Grice's conversational maxims and the schema for pragmatic inference derived therefrom." Put differently, system-level ef-

ficiency we see in languages is deeply related to local-level efficiency during the in-the-moment interactions of rational pragmatic speakers and listeners.

In this work, we present a step toward formalizing Horn's observation – connecting Zipfian notions of efficient language structure to Gricean notions of rational conversation. To do so minimally requires three basic ingredients – (1) a language property we'd like to explain (2) a framework for describing in-the-moment interactions of speaker-listeners and (3) some measure of linguistic efficiency. For the current project we focus on the "communicative function of ambiguity" for our property – translating the functionalist theory pursued by Piantadosi et al. (2012) into our framework. To model local, conversational interactions (2) we adopt the Rational Speech-act framework (RSA) (Frank & Goodman, 2012; Goodman & Frank, 2016). Finally, (3) we derive an objective function for measuring language efficiency in the reference game setting. Given this framework we can ask the question, "how does having *lexical ambiguity* make using language ℓ more or less useful for communication"? This project is meant to demonstrate the ways in which a range of functionalist theories might make use of such a framework – in theory replacing *lexical ambiguity* with a range of behaviors and properties identified by functionalist approaches to language science.

We begin with a high-level introduction to the modelling framework introducing the basic ingredients we will need to represent language as repeated reference games. Following this introduction we derive a simple objective function for measuring the efficiency of linguistic systems in this setting. Subsequently, we move on to two case-studies examining the questions posed above, framing results in terms of our efficiency measure.

Exploring efficient language- design and use in rational pragmatic agents

Reference games Zipf's canonical (1935) example of optimal speaker- and listener-languages took the form of a reference game (similar to those described by Wittgenstein). We adopt that formulation here, assuming these communication games as our basic unit of analysis. In this setting, speakers and listeners are aware of a set of objects M, which will refer to as *meanings* and are knowledgeable about the set of possible signals U (*utterances*) that could be used to refer to a given meaning. Utterances may have different relative costs, operationalized via a prior over utterances $p(u)$. Similarly, meanings differ in the relative degree to which they need to be talked about, operationalized as a prior over meanings $p(m)$. Note that the prior over meanings are analogous to the *need probabilities* assumed in previous work (Regier, CITATION). We consider a set of contexts C which describe different need probability distributions over our set of meanings $p(m|c)$. Finally, we consider a set of communicative events $e \in E$ where $< u, m >= e$ is a tuple of utterance, meanings pairs. The set E all possible utterance-meaning mappings.

Languages A language ℓ defines the set of semantic mappings between utterance and meanings. For example, in a world with three utterances $U = \{u_1, u_2, u_3\}$ and three meanings $M = \{m_1, m_2, m_3\}$ the boolean matrix

	m_1	m_2	m_3
u_1	1	1	0
u_2	0	1	0
u_3	0	0	1

describes the literal semantics of ℓ. E.g. the language describes semantic mappings $[\![u_1]\!] = \{m_1, m_2\}, [\![u_2]\!] = \{m_2\}, [\![u_3]\!] = \{m_3\}$.

Speakers and listeners The Rational Speech-act framework (RSA) is computational-level theory of pragmatic language use. This framework has proven productive in describing a range of phenomena in natural language understanding and generation including hyperbole, metaphor, scalar and ad-hoc implicature, among others (CITATIONS). RSA can largely be understood as a formalization of essential Gricean pragmatic principles – interlocutors reasoning about one another and their context. For this reason, we adopt RSA as our representational framework to operationalize rational speaker-listeners. For a more extensive description of RSA as well as its application we refer readers to Goodman & Frank (2016), limiting the current discussion to essential, high-level details of the framework.

An RSA *speaker agent* defines a conditional distribution over utterances, mapping from intended meanings M to utterances U using ℓ. That is, a speaker defines $P_{speaker}(u|m;\ell)$. We will use $S(u|m;\ell)$ as short-hand throughout. A *listener agent* defines a conditional distribution over meanings, mapping from utterances U to meanings M using ℓ. We will use $L(m|u;\ell)$ as shorthand. Note that both speakers and listeners can induce joint distributions over the set of all signaling events E, although, importantly, these distributions may differ:

$$P_{speaker}(u, m; \ell) = S(u|m; \ell)p(m)$$

$$P_{listener}(u, m; \ell) = L(m|u; \ell)p(u)$$

In general, we would like to consider the efficiency of a system ℓ in terms of these joint distributions.

Zipfian objective for linguistic system efficiency

Zipf proposed that the particular distributional properties found in natural language emerge as a result of competing speaker and listener pressures. We operationalize this in equation (1) – the efficiency of a linguistic system ℓ being used by speaker and listener agents S and L is the sum of the expected speaker and listener effort to communicate over all possible communicative events E.

$$\text{Efficiency}(S, L, \ell) = \mathbb{E}_{e \in E}[\text{speaker effort}] + \mathbb{E}_{e \in E}[\text{listener effort}] \tag{1}$$

Let speaker effort be the negative log probability (surprisal) of a particular utterance. Intuitively, the number of bits needed to encode the utterance u.

$$\text{speaker effort} = -log_2(p(u))$$

Let listener effort be the negative log probability a listener disambiguates an intended meaning m given an utterance u Intuitively, the number of guesses a listener would need to discover the intended meaning m given an utterance u.

$$\text{listener effort} = -log_2(L(m|u;\ell))$$

Rewriting (1) we have

$$\text{Efficiency}(S,L,\ell) = \mathbb{E}_{e \in E}[-log_2(p(u))] + \mathbb{E}_{e \in E}[-log_2(L(m|u;\ell))] \quad (2)$$

In general, these expectations are each taken over the set of possible communicative events $e \in E$ weighted by their probability, $p(e)$. Recall this is the set of all utterance, meaning pairs $<u,m> = e \in E$.

$$= \sum_{e \in E} p(e)[-log_2(p(u))] + \sum_{e \in E} p(e)[-log_2(L(m|u;\ell))] \quad (3)$$

We assume that the particular joint distribution over utterance-meaning $<u,m> = e$ pairs follows from a simple generative model. First, some meaning is sampled with probability $p(m)$. Our speaker attempts to convey this intended meaning to a listener by encoding it in the utterance u via the conditional $S(u|m;\ell)$. Combining these terms leads to the *speaker's joint distribution over events* which we can write as $P(e) = P_{speaker}(u,m;\ell) = S(u|m;\ell)p(m)$.

$$= \sum_{u,m} p_{speaker}(u,m;\ell)[-log_2(p(u))] + \sum_{u,m} p_{speaker}(u,m;\ell)[-log_2(L(m|u;\ell))] \quad (4)$$

Simplifying we arrive at (5):

$$= \sum_{u,m} P_{speaker}(u,m;\ell)[-log_2(L(m|u;\ell)p(u))] \quad (5)$$

Note that $L(m|u;\ell)p(u)$ is the listener-based joint distribution over all communicative events ($P_{listener}(u,m;\ell)$).

$$= \sum_{u,m} P_{speaker}(u,m;\ell)[-log_2(P_{listener}(u,m;\ell))] \quad (6)$$

This is the simply the cross-entropy between the speaker an listener joint distributions.

$$= \mathbb{E}_{P_{speaker}}[-log_2(P_{listener})]$$
$$= H_{cross}(P_{speaker}, P_{listener}) \quad (7)$$

From an information-theoretic perspective this objective is intuitive. Cross-entropy gives us a measure of dissimilarity between two distributions – informally, the the average length of communicating an event from one distribution with the code from another. In our current setting, the optimal linguistic system being used by a speaker and listener agent will induce a particular set of form-meaning mappings which brings the speaker and listener joint distributions over communicative events as *close* together as possible.

Optimal speaker and listener languages in the reference game setting

In terms of Zipf's initial example about an optimal speaker or listener language we might consider just the first or second term of equation (1). That is, given our formulation we should expect optimal speaker languages $\ell_{speaker*}$ to minimize:

$$\ell_{speaker*} = argmin_{\ell \in L} \mathbb{E}_{P_{speaker}(u,m;\ell)}(p(u)) \quad (8)$$

Likewise we might expect optimal listener languages $\ell_{listener*}$ to minimize:

$$\ell_{listener*} = argmin_{\ell \in L} \mathbb{E}_{P_{speaker}(u,m;\ell)}(L(m|u;\ell)) \quad (9)$$

Two case-studies on efficient language use and design

We now have a metric for exploring questions of language optimality in the reference game setting (the speaker-listener cross-entropy). In the following sections we examine the behavior of this objective with respect to the communicative function of ambiguity.

Ambiguity is ubiquitous in natural language. At a minimum, the listener task in everyday communication is marked by the frequent need to handle lexical ambiguity (words often have multiple meanings) as well as syntactic ambiguity (sentences often have multiple parses), in addition to many other instances of ambiguity (Wasow, Perfors & Beaver, 201X). Chomsky (2002) famously claimed that the presence of ambiguity in natural language provides evidence that language has not been optimized for communication, stating "If you want to make sure that we never misunderstand one another, for that purpose language is not well designed, because you have such properties as ambiguity." Note that this analysis is analogous to the Zipfian optimal Listener language $\ell_{listener}^*$ we described earlier and define in terms of equation (9).

In their paper, "The communicative function of ambiguity," Piantadosi et al. (2012) essentially argue the opposite. They claim that ambiguity is an *efficient* property of any communication system in which *communication is contextualized*. We provide the heart of their argument here, however we refer to Piantadosi et al. (2012) for more detail:

In an unambiguous linguistic system, m_1 is mapped to $\ell_1 \in L$ and m_2 is mapped to $\ell_2 \in L$, with $\ell_1 \neq \ell_2$. Suppose that ℓ_1 is easier than ℓ_2... If ℓ_1 is easier than ℓ_2, and m_1 and m_2 are well-disambiguated by context, then we can always create a linguistic system which is easier overall by mapping

m_1 and m_2 both to ℓ_1. *This costs the same in terms of effort every time we communicate m_1, but saves effort every time we communicate m_2*

In words, the Piantadosi argument might be summarised as – it is useful to have a language that re-uses low-cost material (has ambiguity) so long as the cost of disambiguating the material is low. They argue that disambiguation cost is low when context if informative.

In the following experiments we explore two aspects of Piantadosi's claim. First, we examine the efficient language *structure* aspect of the question exploring when the optimal linguistic system (under our cross-entropy objective) is most likely to contain ambiguous lexical items. In the second experiment we explore an efficient language *use* aspect of the question – at what point in a conversation is it useful for a speaker to use ambiguous lexical material?

Experiment 1: Optimal languages contain ambiguity when context is informative

The argument outlined by Piantadosi et al. (2012) is clearly compatible with our current reference game setting with just a single addition – context. The authors argue that mapping both m_1 and m_2 to l_1 is useful when they can disambiguated *in context*. In our case we consider a context c to specify a re-ordering of the referent need probabilities. That is, $p(m|c_1) \neq p(m|c_2)$. In the two context case.

This in turn, leads to an update to our linguistic efficiency objective as we now would like to consider the average speaker-listener effort over all contextualized communicative events. That is,

$$\sum_{c \in C} p(c) \sum_{u,m} S(u|m,c;\ell)p(m|c) - log[L(m|u,c;\ell)p(u)] =$$

$$\mathbb{E}_{c \sim P(c)}[\mathbb{E}_{P_{speaker(u,m,c;\ell)}}[-log(P_{listener}(u,m,c;\ell))]]$$

(10)

Note that in the case that $|C| = 1$, our objective simplifies to our original equation (7). With this update we can compare linguistics systems L while varying the degree to which context contains useful information by varying the size of $|C|$.

Simulation set-up

Basic ingredients We conduct $N = 400$ simulations. For each simulation we enumerate the set of *valid* languages in which $|U| = |M| = 4$. Note that a language $\ell \in L$ is "valid" so long as each possible meaning in $m \in M$ can be referred to by at least one form $u \in U$ (every column of ℓ has some non-zero assignment). For a given simulation the goal is to find the language ℓ^* which minimizers our cross-entropy objective.

Recall that language efficiency is both a function of the particular semantic mappings induced by that language, the speaker and listener agents, as well as the utterance and meaning priors. Rather than assume particular structure for our utterance and meaning prior distributions, for each simulation

we generate $p(m) \sim \text{Dir}(1, |M|)$ and $p(u) \sim \text{Dir}(1, |U|)$ where $\text{Dir}(1, k)$ specifies the uniform Dirichlet distribution over a k-dimensional probability vector.

Context Following the argument given by Piantadosi et al. (2012) we want to assess the impact of *context* on our objective. To do so we consider four conditions with $n = 100$ simulations each. Our first is a *single-context* condition ($|C| = 1$) – there is a only a single context describing $p(m|c_1)$. Our second condition contains two-contexts ($|C| = 2$) – we consider efficiency under both $p(m|c_1)$ as well as $p(m|c_2)$. The third and fourth condition correspond accordingly with $|C| = 3$ and $|C| = 4$, respectively.

Baselines For comparison, we also examine properties of optimal languages under two additional objectives. Zipf (1935) proposed that a language optimized for speaker-ease should consist of a single word. We operationalize this using the *first half* of equation (1).

$$\text{Speaker-only objective} = \mathbb{E}_{u,m \sim P_{speaker}}[\text{speaker effort}]$$

$$= \sum_{u,m} P_{speaker}(u,m;\ell)[-log(p(u))] \quad (11)$$

Zipf also proposed that the optimal listener language should bijectively map all words to unique meanings. We operationalize this using the *second half* of equation (1).

$$\text{Listener-only objective} = \mathbb{E}_{u,m \sim P_{speaker}}[\text{listener effort}]$$

$$= \sum_{u,m} P_{speaker}(u,m;\ell)[-log(L(m|u;\ell)] \quad (12)$$

Hypotheses This leads to the following set of hypotheses: (1) The probability that an optimal language contains ambiguity under our Cross-Entropy objective should increase with the number of contexts (as $|C| \to \inf$). (2) The optimal speaker language under our Speaker-only objective should always map all meanings to the single, lowest cost utterance. (3) The optimal listener language should never contain ambiguous material.

Results Figure 1. plots the proportion of optimal languages under $H_{cross}(P_{Speaker}, P_{Listener})$ for each condition. We find that as the number of contexts increases, so does the probability that the optimal language $\ell*$ contains ambiguity. For comparison we also include the optimal language under an objective that only considers speaker or listener effort (equations 11 and 12). In line, with Zipf's predictions, if languages are designed only to minimize speaker effort then optimal languages will assign all meanings to a single, low-cost utterance. Likewise, if languages are designed only to minimize speaker listener effort then ambiguity should always be avoided.
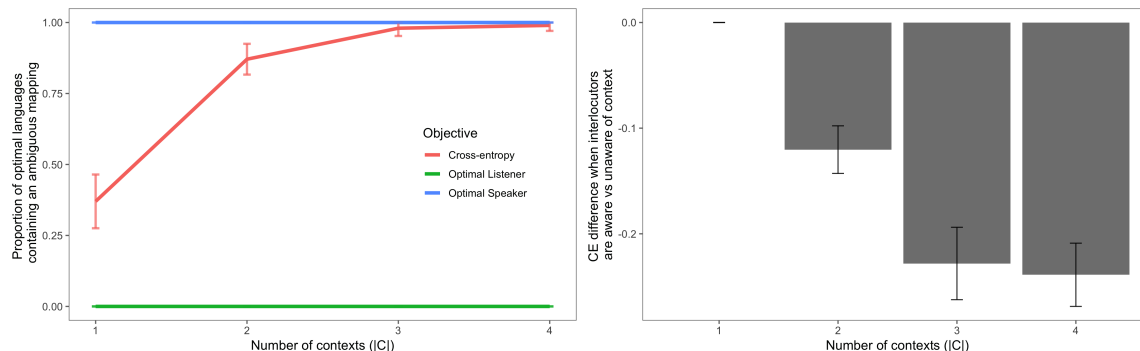
Figure 1: Left facet – optimal languages are more likely to contain ambiguous items as amount of contextual information increases. Vertical axis shows the proportion of optimal languages that contain ambiguity. Horizontal axis shows the number of contexts in each condition (1-4). Red-line represents the optimal langauge under our Zipfian cross-entropy objective while the blue and red lines show optimal languages under speaker- and listener-only consideration. Right facet – Gains in efficiency increase with the amount of contextual information (as the number of contexts increases). Vertical axis shows the difference in cross-entropy when the number of contexts is 1 and when the number of contexts is greater than 1. Horizontal axis shows each of the four conditions (the four context sizes). Note that the difference for condition 1 is zero as there is only a single context.

More context leads to larger efficiency gains Piantadosi et al. (2012) framed their theory in terms of conditional entropy following the relation between conditional and unconditional entropy. That is, $H(X|C) < H(X)$ when C provides information about X (CITATION). In our setting, since knowing the particular context c_i is always informative with respect the distribution over meanings ($p(m|c_i)$) this should mean that as the amount of contextual information increases, the difference between the conditional and unconditional measures of information should increase. In other words, $H_{cross}(P_{Speaker}, P_{Listener}|C_1) - H_{cross}(P_{Speaker}, P_{Listener}) = 0 < H_{cross}(P_{Speaker}, P_{Listener}|C_2) - H_{cross}(P_{Speaker}, P_{Listener})$ where $C_1 = |C| = 1$ and $C_2 = |C| = 2$. Figure 2 shows that as the number of contexts increases (the amount of information contained in the common ground, the difference in efficiency also increase.

Summary Piantadosi et al. (2012) argued that it is useful to re-use low-cost linguistic forms for multiple meanings when they can be disambiguated in context. Using our speaker-listener cross-entropy measure of efficiency we showed that optimal languages are more likely to have ambiguous items when context is informative. Further, we showed that the impact of being able to disambiguate language (having access to $p(m|c_i)$) is increasingly efficient as the amount of common-ground (context) increases.

Experiment 2: Rational-pragmatic speakers use ambiguity efficiently

Our first experiment made a fairly direct test of the communicative function of ambiguity proposed theory presented by Piantadosi et al. (2012). Sampling a set of need probabilities and utterance costs we explored the space of possible languages, examining the properties of the language ℓ^* which minimized our objective. Results indicated that ambiguity is

an efficient property when context is informative. This experiment, however, assumed that our speaker and listener agents could always disambiguate items context. That is, in our four conditions both agents had perfect knowledge of the conditional distributions $p(m|c_1), \ldots, p(m|c_4)$. Consider the scenario in which a reader is beginning a news article. While they may have some knowledge about the article's topic (perhaps from the title), they may not have complete knowledge of its contents, including the persons or events involved. In this scenario, using an ambiguous pronoun early in the article lead to misunderstanding. Using such ambiguity is only warranted when the appropriate person or event is salient in the context.

We consider such a scenario in experiment two – analyzing efficient *use* of ambiguity over discourse. In this scenario, we assume a single language ℓ which contains ambiguous items, known to both the speaker and listener. The topic of conversation $t \in T$ defines the particular meaning distribution at hand – $p(m|T = t)$. This conditional distribution is known only to the speaker a priori. Over the course of a discourse, D the listener tries to infer both the current topic as well as the meanings. That is we consider a listener model $L(m, t|u, D; \ell)$ and speaker model $S(u|m, t, D; \ell)$. Note that both the speaker and listener can track the history of previous utterances D.

Hypothesis In this setting, if speakers are *using ambiguity efficiently* they should use the ambiguous material only when they know the listener can disambiguate the items in context. Since listeners accumulate information about the current topic distribution ($p(m|t)$) over as the discourse progresses, we should expect to see more frequent use of ambiguous items later in the discourse.

Basic ingredients We conduct $n = 50$ simulations, generating discourses of length $|D| = 50$. We consider a single language ℓ with $|U| = 6$ and $|M| = 4$ specified by the boolean
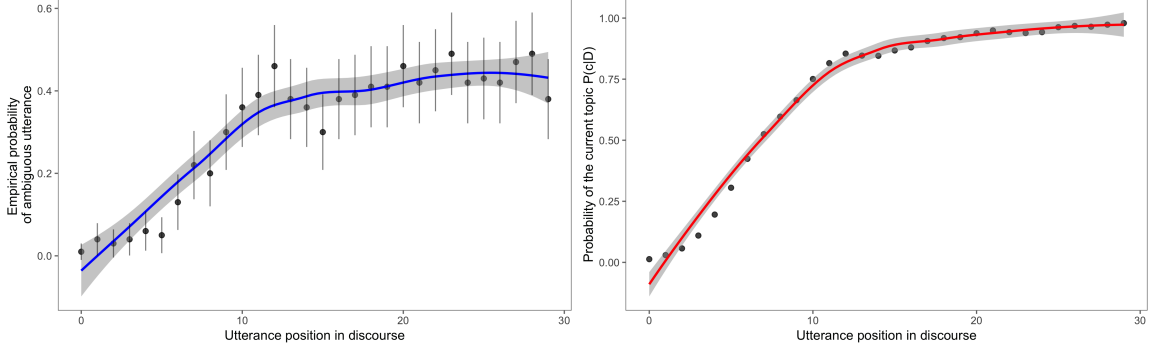
Figure 2: Left facet shows the empirical probability that our speaker used an ambiguous utterance as a function of discourse position. As the discourse proceeds, the speaker is more likely to use an ambiguous term. Right face shows the probability the listener assigns to the current topic of conversation. That is, P(c—D) which is inferred by reasoning about the discourse history.

matrix:

	m_1	m_2	m_3	m_4
u_1	1	0	0	0
u_2	0	1	0	0
u_3	0	0	1	0
u_4	0	0	0	1
u_5	1	1	0	0
u_6	0	0	1	1

We assume that $p(u_5) = p(u_6) > p(u_{1,...,4})$. That is, the two ambiguous utterances are less costly than the non-ambiguous utterances. Over the discourse both agents have complete recall of the set of possible utterances by the speaker D. Note that the listener can attempt to infer the current topic distribution $p(C = c|D)$ via $p(C = c|D) \propto \prod_{i=0}^{|} D|S(u_i|m_i)p(m_i|c)$.

Results The left facet of figure 3 shows the empirical probability that the speaker used an ambiguous utterance as a function of discourse position. The right facet shows the listener belief about the current topic $c \in C$ given the discourse history ($p(c|D)$). Note that the trajectory of these items are close – the speaker is more likely to use ambiguous material only when the listener is able to disambiguate that material. Figure 4 shows the listener belief about the current context. Additionally we can track the behavior of our cross-entropy objective and it's component speaker and listener elements over the discourse. Overall we see declines in the speaker-listener cross-entropy. These are derived primarily from the listener learning about the current context of conversation (i.e. the current c defining $p(m|c)$), however we also see declines in speaker effort, as they are able to use loss-costly material later in the discourse.

Summary These demonstrated the efficient *use* of ambiguous material over a discourse, in line with the argument of Piantadosi (2012). While the speaker may prefer to use the less costly, ambiguous utterances at all points throughout the discourse, using the ambiguous items early leads to higher listener-level cost. Only once the context is known to the listener does it makes sense to use the lower-cost material.

General Discussion

How might the competing pressures of speakers and listeners give rise to the distributional forms found in natural language? Zipf (1935) proposed that the asymmetry between speaker and listener costs gives rise to a range of properties the the level of the lexicon. Subsequent functionalist approaches to language science have provided additional evidence that various aspects of language structure and language use appear to be optimized for efficiency. But, it is not clear at what level this optimization occurs. In Zipf's famous example of an optimal speaker and listener language he appears indicate that such optimization occurs in in-the-moment local interactions between speakers and listeners. We explore this idea in our current project, following theoretical work by Lawrence Horn (1984) who proposed the speaker-listener pressures identified by Zipf can also be interpreted in terms of the maxims governing rational conversation proposed by Grice (1975). To so, we explored the degree to which the interactions of rational pragmatic agents can provide a framework for understanding efficient language structure and use. We focused on an argument on the communicative function of ambiguity put forth by Piantadosi. We derive a novel speaker-listener cross-entropy objective for measuring the efficiency of a language in a reference game setting showing that optimal languages are more likely to contain ambiguous material when context is informative. Further we show that rational pragmatic agents will *use* ambiguous material efficiently, only when such use is supported by context. The intent is that this project provides a demonstration for the types of functionalist theories that can be explored within this framework. Among other language properties open to this type of analysis are work examining semantic typology (CITATIONS) and the emergence of compositionality (CITATIONS), as well as efficient examples of language use including the efficient use of reduction and redundancy at the level of syntax (Levy & Jaeger, 2007) and discourse (Genzel & Charniak, 2002).

Conclusion

Acknowledgements

References