Pragmatic interactions lead to efficient language structure and use

Benjamin N. Peloquin
bpeloqui@stanford.edu
Department of Psychology
Stanford University

Noah D. Goodman
ngoodman@stanford.edu
Department of Computer Science
Stanford University

Michael C. Frank
mcfrank@university.edu
Department of Psychology
Stanford University

Abstract

We discuss a framework for studying how the distributional properties of linguistic systems emerge 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. These simulations suggest that rational pragmatic agents will produce communicatively efficient systems.

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

Introduction

Why do languages look they way they do? Zipf (1935) 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 properties 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.

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 "hard" for a speaker is likely different from what is "hard" for a listener. Zipf characterized this as follows – purely from the standpoint of speaker effort, 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, an optimal language $\ell^*_{listener}$ should map all possible meanings to distinct words, removing a listener's need to disambiguate. In this example, 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 were a result of these competing forces – the pressure to jointly minimize speaker and listener effort.

But at what level of does local, in-the-moment interaction 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).
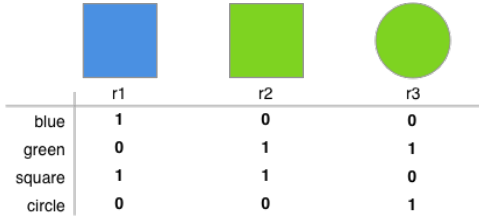
This move to study efficient language use in terms of reference games has clear theoretical underpinnings as well. The linguist Lawrence Horn (1984), highlighted the importance of Zipf's principles for explaining conversation-level phenomena. Horn suggested a link between Zipf's conception of speaker- and listener-effort and ideas related to conversational pragmatics by Grice (1975), arguing 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, the system-level efficiency we see in languages is fundamentally related to local-level efficiency during the in-the-moment interactions of rational, pragmatic speakers and listeners at the level of conversation.

In this work, we present a step toward formalizing the framing presented above – 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. In terms of our opening question, we are exploring the question, "how does having *lexical ambiguity* make using language ℓ more or less useful for communication"? In more general terms, this project is meant to demonstrate the ways in which a range of functionalist theories might make use of such a framework – e.g. we could replace *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. We move on to two case-studies examining the questions posed above, examining the efficiency of ambiguity from a language design and language use perspective.

Exploring efficient language- design and use in rational pragmatic agents

Figure 1: An example reference game with associated literal semantics (in our terminology 'language').

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 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 states, formalized as different conditional distributions over meanings $p(m|c)$. Finally, we consider a set of

communicative events $e \in E$ where $< u, m > = e$ is a tuple of utterance, referent pairs.

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. RSA can largely be understood as a formalization of essential Gricean pragmatic principles – interlocutors reason about one another and their shared context. For this reason, we adopt RSA as our representational framework to operationalize Gricean (rational and pragmatic) speaker-listeners. For a more extensive description of RSA as well as its applications 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 – the average number of bits required to communicate under one distribution, given that the "true" distribution differs. In our case, this is the difference between the joint distribution assumed by the speaker $P_{speaker}$ and listener $P_{listener}$. A good language ℓ used by a set of a pair of speaker-listeners will have properties which minimize this objective.

Simulations

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 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).

Piantadosi et al. (2012) argue just the opposite. They claim that ambiguity is an *efficient* property of any communication system in which *communication is contextualized*. Simply put, 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. Context can provide useful information for disambiguation.

In the following experiments we explore two aspects of Piantadosi's claim. First, we examine the efficient language *structure* aspect of their claim, exploring when the optimal linguistic system ℓ^* (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?

Simulation 1: Optimal languages contain ambiguity when context is informative

The argument outlined by Piantadosi et al. (2012) is 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 be disambiguated *in context*. In our case we consider a context c to specify a unique ordering of the need probabilities. That is, $p(m|c_1) \neq p(m|c_2)$, when there are two contexts $|C| = \{c_1, c_2\}$.

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,

$$\sum_{c \in C} p(c) \sum_{u,m} S(u|m,c;\ell)p(m|c) - log[L(m|u,c;\ell)p(u)] = \quad (8)$$

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

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

Simulation set-up

Basic ingredients We conduct $N = 2000$ 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 ($P(U)$), meaning ($P(M)$), and context priors ($P(C)$). Rather than assume particular structure for our utterance and meaning prior distributions, for each simulation we generate $P(M) \sim \text{Dir}(1,|M|)$, $P(U) \sim \text{Dir}(1,|U|)$ and $P(C) \sim \text{Dir}(1,|C|)$ 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 $n = 500$ simulations each (that is, 500 unique sets of $\{P(U), P(M), P(C)\}$. 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 the optimal speaker language should consist of a single low cost word $\ell^*_{speaker}$. We operationalize this using the *first half* of equation (1).

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

The optimal listener language $\ell^*_{listener}$, by contrast, should bijectively map all words to unique meanings. We operationalize this using the *second half* of equation (1).

$$\ell^*_{listener} = argmin_{\ell \in L} \mathbb{E}_{P_{speaker}(u,m;\ell)}(L(m|u;\ell)) \quad (10)$$

Hypotheses Piantadosi's claim 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 $H_{cross}(P_{Speaker}, P_{Listener}; \ell)$ for each context condition. The red line indicates that as the number of contexts increases, so does the probability that the optimal language ℓ_* under our speaker-listener cross-entropy objective contains ambiguity. We also plot the proportion of optimal speaker-only and listener-only languages that contain ambiguity. 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 utterance (full ambiguous). Likewise, if languages are designed only to minimize speaker listener effort then ambiguity should always be avoided (bijective form-meaning mappings).

More context leads to larger efficiency gains Piantadosi et al. (2012) framed their theory in terms of properties of conditional entropy. That is, $H(X|C) \leq H(X)$. This inequality is strict when C provides information about X (Cover & Thomas, XXXX). Intuitively, this means that they amount of uncertainty we have about X decreases when C tells us something about X. In our setting, knowing the particular context c_i is always informative with respect the distribution over meanings ($p(m|c_i)$). This means that if we have $|C_1| = 1$ and $|C_2| = 2$ then we expect $H_{cross}(P_{Speaker}, P_{Listener}|C_2) < H_{cross}(P_{Speaker}, P_{Listener}|C_1)$. To examine this in our simulations, we look at the difference in efficiency scores, comparing pragmatic agents who have access the conditional distributions $P(m|c)$ an non-pragmatic agents who don't and simply consider \$P(m) =

Additionally, the difference between these two quantities should increase with the difference in the size of $|C_i|$. E.g. if we also have $|C_3| = 3$ then . In our case, we can examine the difference between the unconditional and conditional cross-entropy measures as $|C|$ increases from $|C| = 1$ in condition 1 to $|C| = 4$ in condition 4. The right facet of figure one shows the difference between conditional and unconditional corr-entropy measures as the number of contexts $|C|$ increases.

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-
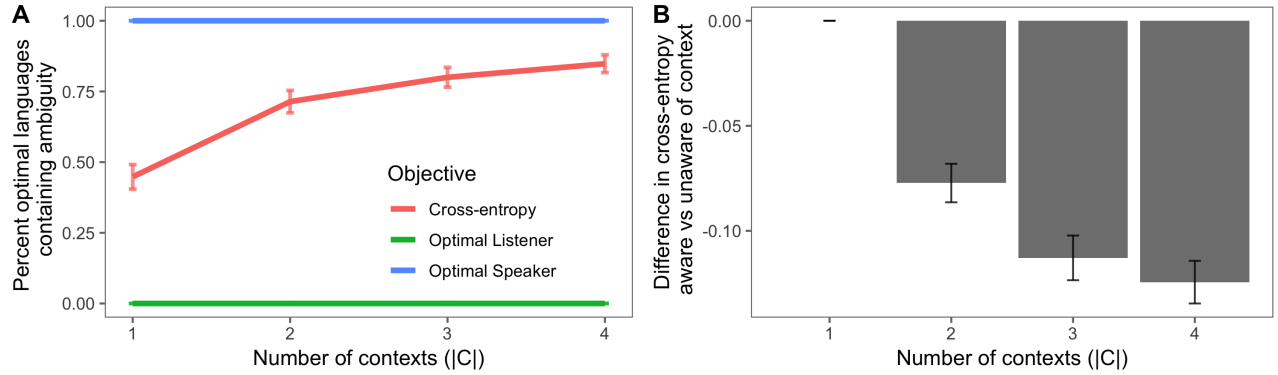
Figure 2: Left facet – optimal languages are more likely to contain ambiguous items as amount of contextual information increases. Vertical axis shows the proportion of optimal languages that contain ambiguity. Horizontal axis shows the number of contexts in each condition (1-4). Red-line represents 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.

ground (context) increases.

Simulation 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 simulation, 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 $p(m|c_i)$.

However, in day to day communication we often don't have perfect knowledge of the current context of a conversation. We can consider this notion of "context" as analogous to knowledge of the *topic* of conversation. Consider the scenario in which a reader is beginning a newspaper 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 as there isn't sufficient context to disambiguate it. Using such ambiguity is only warranted when the appropriate person or event is salient in the context.

We consider such a scenario in simulation 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. Crucially, we do not assume that a listener has access the conditional distributions $p(m|c)$ a priori. Rather, we assume that the listener has knowledge of the set of possible contexts $C = \{c_1, \ldots, c_k\}$,

but does not know which one is currently being used by the speaker.

Over the course of a discourse, D the listener tries to infer both the current topic, c, as well as the particular meaning of a given utterance. That is we consider a listener model $L(m,c|u,D;\ell)$ and speaker model $S(u|m,c,D;\ell)$. 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 only use the ambiguous material 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 $n = 200$ simulations, generating discourses of length $|D| = 30$ utterances. We consider a single language ℓ with $|U| = 6$ and $|M| = 4$ specified by the boolean 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 $p(u_5) = p(u_6) > p(u_{1,\ldots,4})$. That is, the two ambiguous utterances are less costly than the non-ambiguous utterances. In this setting, it is less costly for the speaker to use the ambiguous utterances. However, if the listener cannot disambiguate them, then they will incur a disambiguation cose. We assume that over the discourse both agents have
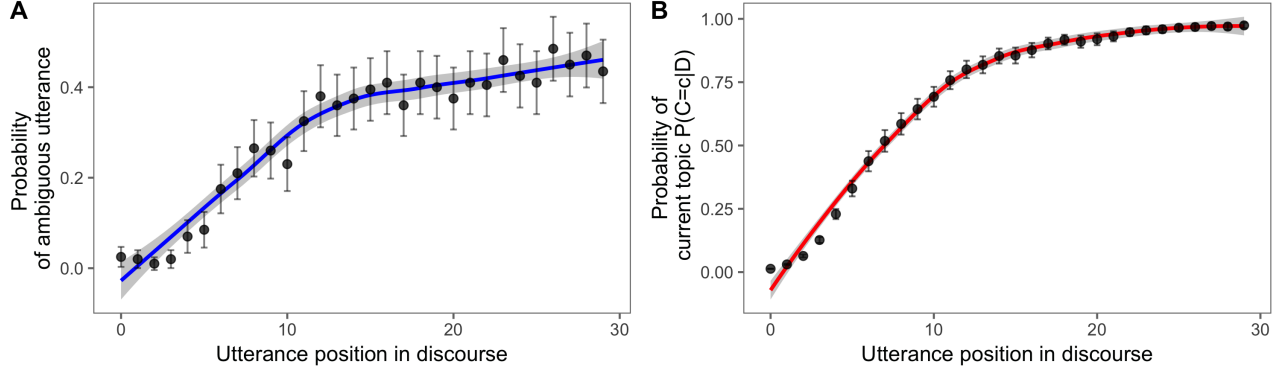
Figure 3: Left facet (A) shows the empirical probability that our speaker used an ambiguous utterance as a function of discourse position. As the discourse proceeds, the speaker is more likely to use an ambiguous term. Right face (B) shows the probability the listener assigns to the current topic of conversation.

complete recall of the set of possible utterances by the speaker D. This means that the listener can infer the current topic distribution $p(C = c|D)$ via $p(C = c|D) \propto \prod_{i=0}^{|D|} S(u_i|m_i)p(m_i|c)$. In this way, an efficient speaker should avoid ambiguity *early* in the discourse when the listener is unsure of the context, only using it later once they are confident the listener knows the context $p(m|C = c)$.

Results Figure 2, plot A shows the empirical probability that an RSA speaker used an ambiguous utterance as a function of discourse position. While at the start of the discourse the speaker is unlikely to use the utterance we see consistent use later, once the listener is sure of the particular context c_i this speaker is using.

The right facet of figure 2 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 simulations demonstrated the efficient *use* of ambiguous material over a discourse, in an alternative instantiation of the argument outlined by Piantadosi et al. (2012) – 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 at 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).

References