

# Pragmatic interactions lead to efficient language structure and use

Anonymous CogSci submission

## 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 the way they do? Zipf (1949) 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  $\ell$  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. (2015) showed that languages appear to organize semantic domains (form-meaning mappings) in maximally efficient ways. Likewise, Piantadosi et al. (2011) 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 communica-

tion. 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  $\ell_{listener}^*$  would 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.

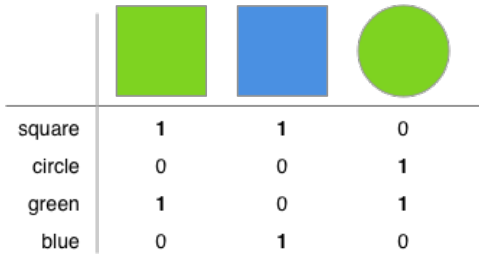
But how does such optimization take place? The example given by Zipf (1949), he appears to describe local, communicative interactions in terms of a *reference game*. Speakers intend to refer to some object in the world  $m$ . They choose some utterance  $u$  to transmit this intended meaning,  $u \rightarrow m$ . 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 (Piantadosi et al., 2011; Regier et al., 2015) and this simplification of the communicative act has proven productive in theoretical (Ferrer-i-Cancho, 2018), simulation-based (Kirby, Griffiths, & Smith, 2014) and empirical explorations (Hawkins, Franke, Smith, & Goodman, 2018) of efficient language structure and use.

This move to study efficient language use in terms of reference games has clear theoretical underpinnings as well. The linguist Lawrence Horn 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 the philosopher Paul Grice, 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” (Horn, 1984). Put differently, the system-level efficiencies we see in language are fundamentally related to the local-level efficiencies that occur during rational, pragmatic 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 (Grice, 1975). 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. (2011) 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.

The current work examines the question, “how does having *lexical ambiguity* make using language  $\ell$  more or less useful for communication”? 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 exploring 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






			
square	1	1	0
circle	0	0	1
green	1	0	1
blue	0	1	0

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

**Reference games** Zipf’s 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 we will refer to as *meanings* and are knowledgeable about the set of possible signals  $U$  (*utterances*) that can be used to refer to a given meaning (see Figure 1 for an example reference game). 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, operational-

ized as a prior over meanings  $P(M)$ <sup>1</sup>. We consider a set of contexts  $C$  with an associated prior  $P(C)$ . Contexts describe different orderings of the underlying meaning prior, formalized as different conditional distributions over meanings e.g.  $p(M|C = c_i)$ . Finally, we consider a set of communicative events  $e \in E$  where  $\langle u, m \rangle = e$  is a tuple of utterance, meaning pairs.

**Languages** A language  $\ell$  defines the set of semantic mappings between utterance and meanings. For example, figure 1 contains four utterances  $U = \{\text{blue, green, square, circle}\}$  and three meanings  $M = \{\text{green-square, blue-square, green-circle}\}$ . The boolean matrix describes the literal semantics of the language  $\ell$ . Note that will consider language to “contain ambiguous items” if there is at least one utterance  $u \in U$  which can apply to multiple meanings (i.e.  $|\llbracket u_i \rrbracket| > 1$ ). For example, in Figure 1 both the words “blue” and “green” are ambiguous so we would say that  $\ell$  contains ambiguity.

**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 (Grice, 1975). We refer the reader to Section 1 of our supplementary materials for a more detailed description of basic RSA speaker-listener definitions, and to Goodman & Frank (2016) for an overview of its applications. We limit 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  $\ell$ . That is, a speaker defines  $P_{\text{speaker}}(u|m; \ell)$ . 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  $\ell$ . We will use  $L(m|u; \ell)$  as shorthand. Note that both speakers and listeners can induce joint distributions over the set of all events  $E$ , although, importantly, these distributions may differ:

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

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

### Zipfian objective for linguistic system efficiency

Zipf (1949) proposed that the particular distributional properties found in natural language emerge from competing speaker and listener pressures. We operationalize this in equation (1) – the efficiency of a linguistic system  $\ell$  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  $E$ .

<sup>1</sup>The prior over meanings are analogous to the *need probabilities* assumed in previous work Regier, Kemp & Kay (2015).

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

We assume that speaker effort is the negative log probability of an utterance (surprisal) – intuitively, the number of bits needed to encode the utterance  $u$ . This particular formalization of speaker-cost is general enough to accommodate a range of instantiations that might in theory be related to production difficult via articulation effort, cognitive effort related to lexical access, etc (Bennett & Goodman, 2018).

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

We assume listener effort is the negative log probability of the listener’s conditional  $-\log_2(L(M|U))$ . This operationalization of listener effort is intuitively related to existing work in sentence processing in which word difficulty during comprehension is proportional to surprisal (Hale, 2001; R. Levy, 2008).

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

Finally, we assume that events  $e = \langle u, m \rangle$  are sampled according to the following generative model – some object occurs in the world with probability  $p(M = m)$ , the speaker attempts to refer to that object by sampling from her conditional distribution  $S(u|m)$  (i.e.  $e \sim p(m)S(u|m)$ ). Using these basic ingredients we arrive at the following objective:

$$= \sum_{u, m} P_{\text{speaker}}(u, m; \ell) [-\log_2(P_{\text{listener}}(u, m; \ell))] \quad (2)$$

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

$$\begin{aligned} &= \mathbb{E}_{P_{\text{speaker}}}[-\log_2(P_{\text{listener}})] \\ &= H_{\text{cross}}(P_{\text{speaker}}, P_{\text{listener}}; \ell) \end{aligned} \quad (3)$$

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  $P_{\text{speaker}}$  and listener  $P_{\text{listener}}$ . An “efficient” language  $\ell$  used by a set of a pair of speaker-listeners will have properties which minimize this objective. For a complete derivation and its components we refer the reader to Section 2.1 of our supplementary materials.

### Simulating the communicative function of ambiguity

Ambiguity is ubiquitous in natural language. The task of understanding language is marked by a 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, 2005). 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  $\ell_{\text{listener}}^*$  we described earlier and define in terms of equation 2 in section 2.2 for our supplementary materials.

Piantadosi et al. (2011) argue just the opposite, claiming 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. In particular, context (or common ground) can provide useful information for disambiguation.

We provide a toy example of their argument here. Say we have two objects in the world  $m_1$  and  $m_2$ , and two languages  $\ell_1$  and  $\ell_2$ . In language  $\ell_1$ , the low-cost utterance  $u_1$  can be used to refer to both  $m_1$  and  $m_2$  (i.e.  $[[u_1]]_{\ell_1} = \{m_1, m_2\}$ ). With language  $\ell_2$ , however, the low cost utterance  $u_1$  can only refer to  $m_1$  and a higher-cost utterance  $u_2$  can refer to  $m_2$  (i.e.  $[[u_1]]_{\ell_2} = \{m_1\}$  and  $[[u_2]]_{\ell_2} = \{m_2\}$ ). While it is cheaper for a speaker to use  $\ell_1$  (because she can always use a low-cost utterance), it is more difficult for a listener (because  $u_1$  is ambiguous). If context isn’t disambiguating then we might prefer  $\ell_2$  to  $\ell_1$ . But, if context is disambiguating then the speaker can use  $u_1$  to refer to either  $m_1$  or  $m_2$  and  $\ell_1$  should be preferred.

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  $\ell^*$  is most likely to contain ambiguous lexical items. In the second experiment, we explore an efficient language *use* aspect of the claim – 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 Piantados et al. (2011) is compatible with our current reference game setting with an addition – context. The authors argue that mapping two meanings  $m_1$  and  $m_2$  to a single utterance  $u_1$  is useful when they can be disambiguated *in context*. In our case we consider a context  $c_i \in C$  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 (section 3 of supplementary materials).

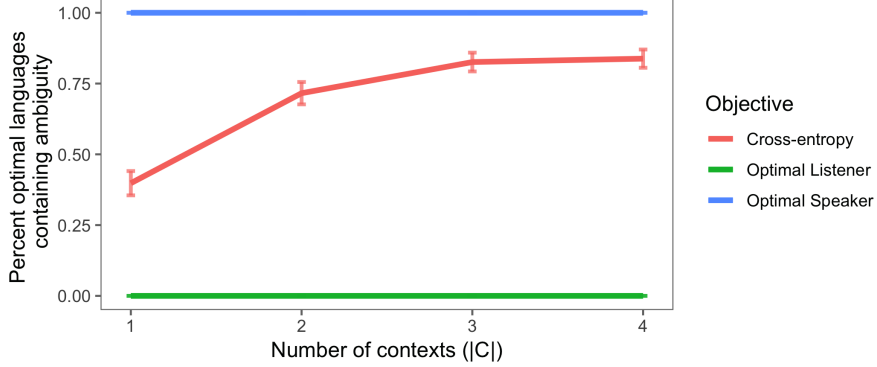


Figure 2: Optimal languages are more likely to contain ambiguous items as the amount of contextual information increases. Vertical axis shows the proportion of optimal languages that contain ambiguity. Horizontal axis shows the number of contexts in each condition (1-4). Red-line represents the optimal language under our Zipfian cross-entropy objective while the blue and red lines show optimal languages under speaker- and listener-only consideration.

### Simulation set-up

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  $\ell$  has some non-zero assignment) and each form maps to at least one meaning (every column has some non-zero assignment). For a given simulation the goal is to find the language  $\ell^*$  which minimizes our cross-entropy objective and then check to see if that language contains ambiguity.

Recall that language efficiency is both a function of the particular semantic mappings induced by that language, the speaker and listener agents ( $S$  and  $L$ ), as well as the utterance ( $P(U)$ ), meaning ( $P(M)$ ), and context priors ( $P(C)$ ). Rather than assume particular structure for our utterance, meaning, and context prior distributions, for each simulation we generate  $P(U) \sim \text{Dir}(1, |U|)$ ,  $P(M) \sim \text{Dir}(1, |M|)$ , 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. (2011), 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  $P(M)$ . Our second condition contains two-contexts ( $|C| = 2$ ) – we consider efficiency under both  $P(M|C = c_1)$  as well as  $P(M|C = c_2)$ . 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 (1949) proposed that the optimal speaker language  $\ell^*_{\text{speaker}}$  would only optimize speaker effort and the optimal listener language  $\ell^*_{\text{listener}}$  would only optimize listener effort. We define these objectives using the first and second half of equa-

tion 1 and with greater detail in the supplementary materials (Section 2.2).

### 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  $|C| \rightarrow \text{inf}$ ). (2) The optimal speaker language under our “Speaker-only” objective should always contain ambiguity (maximally re-using the lowest-cost utterance). (3) The optimal listener language should never contain ambiguous material.

### Results

Figure 1. plots the proportion of optimal languages under each objective, as a function of context condition. The red line shows that as the number of contexts increases, so does the probability that the optimal language  $\ell^*_{\text{cross}}$  under our speaker-listener cross-entropy objective contains ambiguity. We also plot the proportion of optimal speaker-only  $\ell^*_{\text{speaker}}$  and listener-only  $\ell^*_{\text{listener}}$  languages that contain ambiguity. In line with Zipf’s predictions, if languages are designed only to minimize speaker effort then optimal languages will assign lower-cost forms multiple meanings (all of the speaker-optimal languages contained ambiguity). Likewise, if languages are designed only to minimize listener effort then ambiguity should always be avoided (none of the listener-optimal languages contained ambiguity).

### Summary

Piantadosi et al. (2011) argued that it is useful to re-use low-cost linguistic forms for multiple meanings when they can be disambiguated in context. We explored the degree to which this property emerged during communication between rational, pragmatic agents, showing that as contextual information increased, the optimal languages under our Cross-Entropy objective were more likely to have ambiguous

lexical items. Importantly, we did not see this effect under alternative objectives which only emphasized speaker- or listener effort alone.

## Simulation 2: Rational, pragmatic speakers use ambiguity efficiently

Our first experiment made a fairly direct test of the communicative function of ambiguity theory. Sampling meaning, utterance, and context priors, we examined properties of the language  $\ell^*$  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 relevant conditional distributions ( $P(M|C_i)$ ).

This may be a strong assumption for describing much of day-to-day communication – we seldom begin a conversation with perfect knowledge of the current context of a conversation. Under Piantadosi’s framework, if context is not informative with respect to meaning at time position  $i$  in a discourse, but is informative at position  $i + k$ , then the speaker should avoid using ambiguous material at position  $i$ , only using it at a future point,  $k$  steps later in the conversation. In this toy example we can consider “context” as analogous to a “topic” of conversation. Intuitively, we can imagine a 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 setting, using an ambiguous pronoun early (position  $i$ ) in the article can lead to misunderstanding without sufficient context to support disambiguation. But if by position  $i + k$  there is sufficient contextual information it may be efficient for the writer to use ambiguous material.

### Simulation set-up

We consider such a scenario in simulation two – analyzing efficient *use* of ambiguity over discourse in a reference game setting. Unlike our first simulation, we now assume a single language  $\ell$ , which contains both ambiguous and unambiguous utterances. The ambiguous utterances are less costly than the unambiguous utterances. Crucially, we do not assume that the listener knows the particular topic ( $c_{current}$ ) of the conversation *a priori*. Rather, we assume that the listener has knowledge of the set of possible topics  $C = \{c_1, \dots, c_k\}$ , but *does not know which one is currently being used by the speaker*. Formally, this means the listener does not have access to the correct conditional distribution over meanings  $P(M|C = c_{current})$  at the start of the discourse. In this setting, a speaker minimizing production effort would like to use the low-cost material as frequently as possible. However, this will incur a disambiguation cost to the listener if they are uncertain of the current context  $c_{current}$  (topic of conversation).

Over the course of a discourse,  $D$  the listener tries to in-

fer both the current topic,  $c_{current}$ , as well as the particular meaning  $m$  of a given utterance  $u$ . That is we consider a modified 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  $D$ . Importantly, a listener can attempt to infer the current topic of conversation  $c_{current}$  using the discourse history  $D$ . Section 3 of the supplementary materials provides more detail on the speaker and listener models.

We conduct  $N = 600$  simulations, generating discourses of length  $|D| = 30$  utterances with three different speaker models ( $n = 200$  each). We consider a single language  $\ell$  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  $p(u_5) = p(u_6) > p(u_1) = \dots = p(u_4)$ . That is, the two ambiguous utterances ( $u_5$  and  $u_6$ ) are less costly than the non-ambiguous utterances.

### Speaker agents

We consider three types of speaker models. In our *Full pragmatics* model, we consider a speaker agent who reasons about her listener and also has complete recall of the set of utterances in the discourse  $D$ . This speaker believes that the listener may not know the current topic  $c_{current}$  at the start of the discourse, but can infer it with enough data. In this way, contextual information shared by the speaker and listener may increase over the discourse (as the listener’s estimate of the current topic ( $P(C = c|D)$ ) comes in line with the actual topic known to the speaker). We compare two baseline models. The first, a *Partial pragmatics* baseline describes a speaker who reasons about a listener, but assumes they have no access to discourse history. The second is a *No pragmatics* speaker who does not consider a listener at all, but produces utterances according to the underlying language semantics ( $\ell$ ) and topic probabilities ( $p(M|C = c_{current})$ ). We provide detailed descriptions of each model in the supplementary materials Section 3.

### Hypotheses

Under Piantadosi’s claim it is efficient to use low-cost, ambiguous material when context is disambiguating. In the current setting, this corresponds to speaker behavior in which ambiguous forms are avoided early in the discourse, but preferred later in the discourse once the listener is confident of the topic. We should expect this to be reflected in our *Full pragmatics* model. By contrast, our *Partial pragmatics* speaker does not believe that there is sufficient contextual information for the listener to disambiguate her utterances. This should lead to total avoidance of ambiguous material

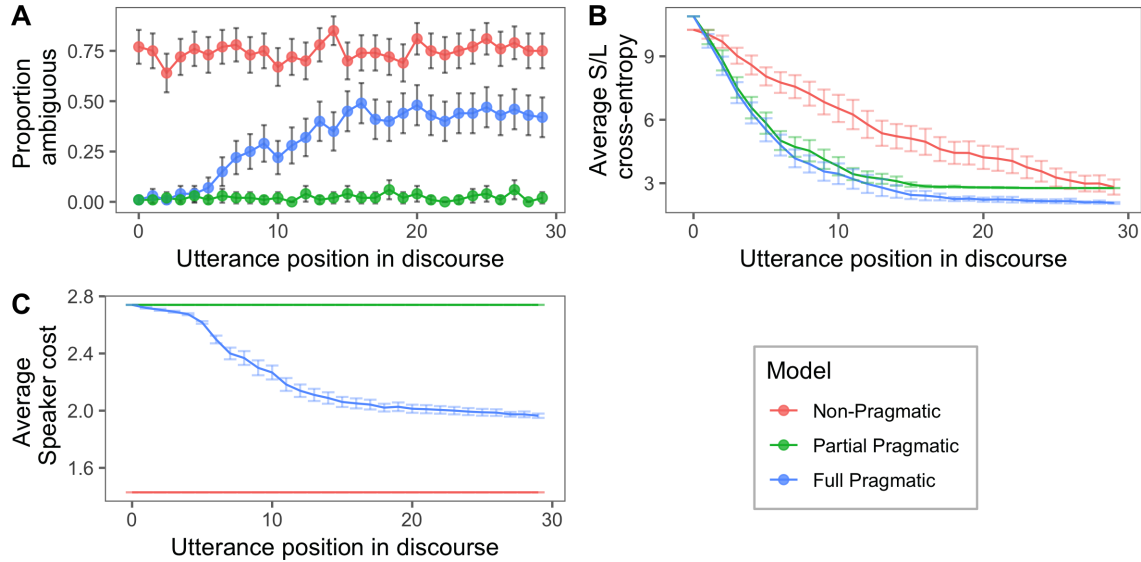


Figure 3: Left facet (A) shows the empirical probability that our speaker used an ambiguous utterance as a function of discourse position. As the discourse proceeds, only our Full Pragmatic speaker is more likely to use an ambiguous term. Plot B tracks the Cross-Entropy objective under our three speaker models. The objective declines in all cases as the listener infers the correct topic distribution. However the pragmatic speaker arrives at the lowest value, due to reduction in both speaker and listener effort. Plot C tracks speaker effort across the three models. Only the Full Pragmatic speaker shows declines in effort across the discourse as they increasingly use ambiguous items as the listener becomes more sure of the topic of conversation.

over the discourse. Finally, our *No pragmatics* speaker should use low-cost material greedily since she does not consider a listener’s need to disambiguate whatsoever.

## Results

Figure 2, plot A shows the empirical probability that a speaker uses an ambiguous utterance as a function of discourse position. Note that the No pragmatic baseline uses ambiguous utterances frequently and at a constant rate over the discourse, the Partial pragmatic baseline avoids ambiguous utterances entirely (reasoning that the utterances cannot be disambiguated), but the Full pragmatic model avoids ambiguous material at first, but employing them gradually increasingly as the discourse proceeds. Figure 2, plot B tracks the speaker-listener Cross-Entropy objective. Note that the objective decreases for all three models. This decrease in all three models is primarily driven by the listener updating his belief about the actual topic ( $P(C = c_{current} | D)$ ). Of particular importance is the value of the objective at the end of the discourse (position  $i = 30$ ). The difference in the Cross-Entropy between the Full Pragmatic and Partial Pragmatic models at the end of the discourse is driven by the reduction in speaker costs described in Figure 2, plot C. That is, while speaker effort remains constant in both *No pragmatic* and *Partial pragmatic* baselines, effort declines for the Full Pragmatic model as they are increasingly able to use ambiguous material later in the discourse.

## Summary

These simulations explored the efficient *use* of ambiguity over a discourse, in an alternative instantiation of the argument outlined by Piantadosi et al. (2011). In the first set of simulations we varied the set of languages, examining the properties of languages which minimized our cross-entropy objective under different context conditions. In the current simulation we assumed that the listener did not know the current context (topic) *a priori*, but could infer it from the discourse. We examined speaker production behavior, tracking use of low-cost, but ambiguous material over the discourse. We found that early in the discourse our *Full pragmatic* speaker avoided using ambiguous material, only using the material later when the context was known to the listener. This is consistent with Piantadosi’s claim – it is useful to use low-cost material, but only when context is disambiguating.

## General Discussion

How do the competing pressures of speakers and listeners give rise to the distributional forms found in natural language? Zipf (1949) 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 aspects of language structure and language use appear to be optimized for efficiency. 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 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 interactions of rational pragmatic agents as a framework for understanding efficient language structure and use. We focused on an argument on the communicative function of ambiguity put forth by Piantadosi et al. (2011). 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 (Regier et al., 2015), the emergence of compositionality (Kirby, Griffiths, & Smith, 2003) and convention formation, as well as efficient examples of language use including the efficient use of reduction and redundancy (Genzel & Charniak, 2002; & J. Levy R., 2007).

## References

- Bennett, E., & Goodman, N. (2018). Extremely costly intensifiers are stronger than quite costly ones. *Cognition*.
- Chomsky, N. (2002). An interview on minimalism. In *N chomsky, on nature and language*.
- Ferrer-i-Cancho, R. (2018). Optimization models of natural communication. *Journal of Quantitative Linguistics*, 25 (3), 207–237.
- Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998.
- Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th annual meeting on association for computational linguistics*.
- Goodman, N., & Frank, M. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829.
- Grice, P. H. (1975). Logic and conversation.
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the naacl*. 159-166.
- Hawkins, R., Franke, M., Smith, K., & Goodman, N. (2018). Emerging abstractions: Lexical conventions are shaped by communicative context. In *Proceedings of the 40th annual conference of the cognitive science society*.
- Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications*. Washington: Georgetown U. Press.
- Kirby, S., Griffiths, T., & Smith, K. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, 9, 587–607.
- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28, 108–114.
- Levy, & J., R. (2007). Speakers optimize information density through syntactic reduction. In *Proceedings of the twentieth annual conference on neural information processing systems*.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Piantadosi, S., Tily, H., & Gibson, E. (2011). The communicative function of ambiguity in language. *Cognition*, 122, 280–291.
- Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. M. & W. O’Grady (Ed.), *The handbook of language emergence* (pp. 237-263). Hoboken, NJ: Wiley-Blackwell.
- Wasow, T., Perfors, A., & Beaver, D. (2005). The puzzle of ambiguity. In CSLI (Ed.), *Morphology and the web of grammar: Essays in memory of steven g lapointe*. Stanford, CA: McGraw-Hill.
- Zipf, G. (1949). *Human behavior and the principle of least effort*. New York, NY: Prentice-Hall.