The interactions of rational, pragmatic agents
lead to efficient language structure and use

Anonymous CogSci submission

Abstract

Languages display a diverse set of distributional regularities such as the relation between a word's frequency and rank in a corpus, the distribution of dependency lengths, or the presence of lexical properties such as ambiguity. We discuss a framework for studying how these properties emerge from in-the-moment interactions of rational, pragmatic speakers and listeners. Our work takes notions of lexicon-level efficiency as a starting point, connecting these ideas to notions of conversational-level efficiency. To do so, we derive 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 and that interactions between such agents provide a framework for examining efficient properties of language more broadly.

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 has been 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. (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 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 $\ell^*_{listener}$ 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.
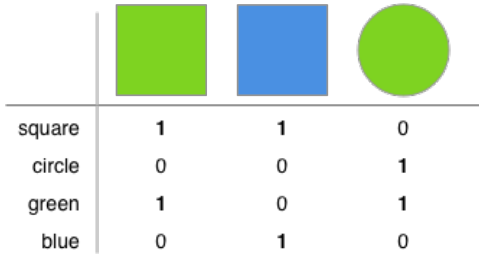
But how does such optimization take place? The example given by Zipf (1949), describes local, communicative interactions in terms of a *reference game*. Speakers intend to refer to some object in the world *m*. They choose some utterance *u* to transmit this intended meaning, $u \rightarrow m$. The listener attempts to reconstruct this intended meaning given the transmitted utterance, $m \rightarrow u$. Other functionalist projects have 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.

Zipf's conception of speaker and listener effort may be connected to speakers' pragmatic reasoning in the moment. Under a Gricean treatment of pragmatics, speakers and listeners are assumed to follow a set of conversational maxims in which they cooperate to transfer information (Grice, 1975). These maxims appear to emerge from efficiency concerns, however (Horn, 1984). A primary target of the current work is to formalize this connection, by providing a framework for deriving system-level efficiencies from local, interaction behavior. Our claim is that to understand an "efficient" property

of a system we need to consider *how* that property is *used* efficiently. Additionally, any quantitative description of "efficiency" should be a function of both the system and users.

In the current work we provide a case-study for the kind of framing presented above. We choose a property of languages that could, in principle, vary freely, but shows strong regularities across languages. The explanadum is why this regularity holds. We examine lexical ambiguity as our property, extending ideas by Piantadosi et al. (2011). We define a novel measure of efficiency that depends on the interactional behavior of interlocutors. For our model of speakers and listeners we adopt the Rational Speech Act (RSA) framework, which is nicely attested in experimental data on pragmatic interaction. With these basic ingredients we have a measure of efficiency given language and speaker-listener agents. Finally, we show that the property of interest is prevalent in languages that optimize this measure of efficiency and also how this property is used efficiently during local interaction. The primary contribution of this work is to show how the reference game frameowrk, in combination with formal models of communication, can be used to connect ideas about system-level efficiencies to in-the-moment language use.

Exploring efficient language design and use in rational pragmatic agents

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. 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). 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 $P(M)$[1]. We consider a set of contexts C with an associated prior $P(C)$. Contexts encode different meaning priors, formalized as separate conditional distributions over meanings e.g. $p(M|C = c_i) \neq p(M|C = c_j)$. Fi-

---

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

nally, we consider a set of communicative events $e \in E$ where $< u, m >= e$ is a tuple of utterance, meaning pairs. Note that in the case of single context $|C| = 1$ we simply have a prior over meanings $p(M)$, a simplification we will use in introducing our objective function.

Languages A language ℓ defines the set of semantic mappings between utterance and meanings. For example, Figure 1 contains four utterances $U = \{\text{blue}, \text{green}, \text{square}, \text{circle}\}$ and three meanings $M = \{\text{green-square}, \text{blue-square}, \text{green-circle}\}$. The boolean matrix describes the literal semantics of the language ℓ. 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. $|[[u_i]]| > 1$)[2]. For example, in Figure 1 both the words "blue" and "green" are ambigous so we would say that ℓ contains ambiguity.

Speakers and listeners The Rational Speech-act framework (RSA) is computational-level theory of pragmatic language use. RSA can be understood as a formalization of essential Gricean pragmatic principles – agents 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 appliations.

An RSA *speaker agent* defines a conditional distribution over utterances, mapping from intended meanings M to utterances U using ℓ in a given context c. 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 in a given context c. 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:

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

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

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 objective 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. Note that to simplify the introduction of this objective we assume there is just a single context ($|C| = 1$). In this case $p(C = c) = 1$ and there is only distribution over meanings ($p(m|c) = p(m)$). We pro-

---

[2]We use double brackets $[[\ldots]]$ to represent denotation

vide a more detailed description of this derivation in the supplementary materials, both with and without context C.

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

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).

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

Importantly, we assume that events $e = <u, m>$ are sampled according the to 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. $e \sim p(m)S(u|m)$). From the ingredients highlighted above, it is possible to derive the following objective between the speaker and listener joint distributions (see SI 2.2 for a complete derivation).

$$= -\mathbb{E}_{P_{speaker}}[log_2(P_{listener})] \\ = H_{cross}(P_{speaker}, P_{listener}; \ell) \quad (2)$$

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 $P_{listener}$. In other words, an "efficient" language ℓ minimizes the distance between what speakers and listeners think.

Simulating the communicative function of ambiguity

The task of understanding language is marked by a frequent need to handle various forms of ambiguity; lexical, syntactic, among others (Wasow, Perfors, & Beaver, 2005). The ubiquity of this property, however, has been argued to provide

evidence that langauges are not been optimized for communication[3] (Chomsky, 2002).

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 ℓ_1 and ℓ_2. In language ℓ_1, the low-cost utterance u_1 can be used to refer to both m_1 and m_2 (i.e. $[[u_1]]_{\ell_1} = \{m_1, m_2\}$ and $[[u_2]]_{\ell_1} = \emptyset$). With language ℓ_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 ℓ_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 ℓ_2 to ℓ_1. But, if context is disambiguating then the speaker can use u_1 to refer to either m_1 or m_2 and ℓ_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 ℓ^* 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?
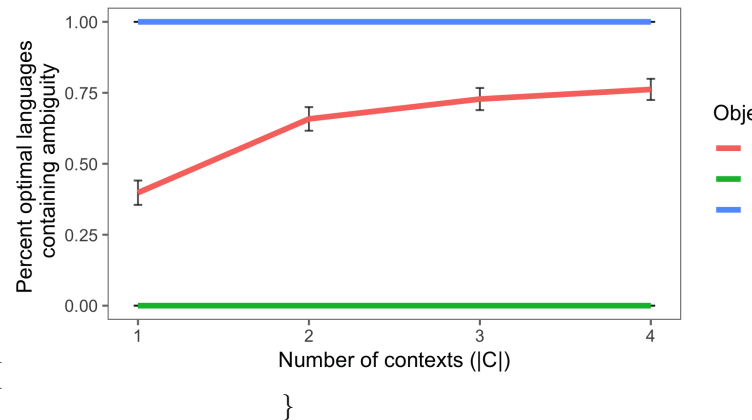
\begin{CodeChunk} \begin{figure*}[h]

{

}

\caption[Optimal languages are more likely to contain ambiguous items as the amount of contextual information increases]{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

---

[3]Chomsky (2002) famously claimed "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 2 in SI 2.2 .

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 objectives. Error bars represent 95% confidence intervals.}
\end{figure*} \end{CodeChunk}

Simulation 1: Optimal languages contain ambiguity when context is informative

We adopt the framework above, and (following Piantadosi et al. 2011), add *context* (*C*). 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 permutation of the meaning 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 (see section 3 SI).

We proceed by generating languages with different amounts of contextual support (varying the size of $|C|$). We search the space of languages, asessing if the ones which minimize our objective contain ambiguity. If context leads to more efficient commnication, then optimal languages should be more likley to be ambiguous as the amount of context increases.

Simulation set-up

We conduct $N = 2000$ simulations. For each simulation we enumerate the set of *valid* languages in which $|U| = |M| = 4$ (recall that U is our set of utterances and M our set of meanings). 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) and each form maps to at least one meaning (every column has some non-zero assignment)[4]. For a given simulation the goal is to find the language ℓ^* 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 $P(U) \sim \text{Dir}(1, |U|)$, $P(M|C = c) \sim \text{Dir}(1, |M|)$ (a separete conditional distribution over meanings for each context c), 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 $\{P(U), P(M|C), P(C)\}$. 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 $\ell^*_{speaker}$ would only optimize speaker effort and the optimal listener language $\ell^*_{listener}$ would only optimize listener effort. We define these objectives using the first and second half of equation 1 (see SI section 2.2. for more detail).

Results and Discussion

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 ℓ^*_{cross} under our speaker-listener cross-entropy objective contains ambiguity. We also plot the proportion of optimal speaker-only $\ell^*_{speaker}$ and listener-only $\ell^*_{listener}$ 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).

We explored the degree to which ambiguity improved efficiency during communication between rational, pragmatic agents. We showed 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.

While our results indicate that ambiguity is an efficient property of contextualized language these simulations assumed that speaker-listener agents could always disambiguate items in context. That is, in our four conditions both agents had perfect knowledge of the relevant conditional distributions ($P(M|C)$). This assumption may be too strong for describing much of day-to-day communication – we seldom begin a conversation with perfect knowledge of the current context of a conversation. We explore a case in which the listener has imperfect knowledge of context at the start of the conversation in Simulation 2.

---

[4]For the current set of simulations we consider languages in which there are four or five true values – in a "four" language ambiguity is not possible, but in a "five" language there are ambiguous mappings.

Simulation 2: Rational, pragmatic speakers use ambiguity efficiently

We shift time-scales in Simulation 2 – instead of examining the objective landscape of a space of languages, we now consider a single language that contains ambiguity and examine how that property is used over a discourse. If context is not informative with respect to meaning early in a discourse (position i), but is at later positions at position (position $i+k$), then the speaker should avoid using ambigous material early i, only using it 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 consider a single language ℓ, 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 $C = \{c_1, \ldots, c_k\}$, 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 infer both the current topic, $c_{current}$, as well as the particular meaning m of a given utterance u. That is, we consider updated listener and speakers ($L(m, c|u, D; \ell)$, $S(u|m, c, D; \ell)$). Note that both the speaker and listener can track the history of previous utterances D. Importantly, an agent can attempt to infer the current topic of conversation $c_{current}$ using the discourse history D (see SI section 3 for more detail).
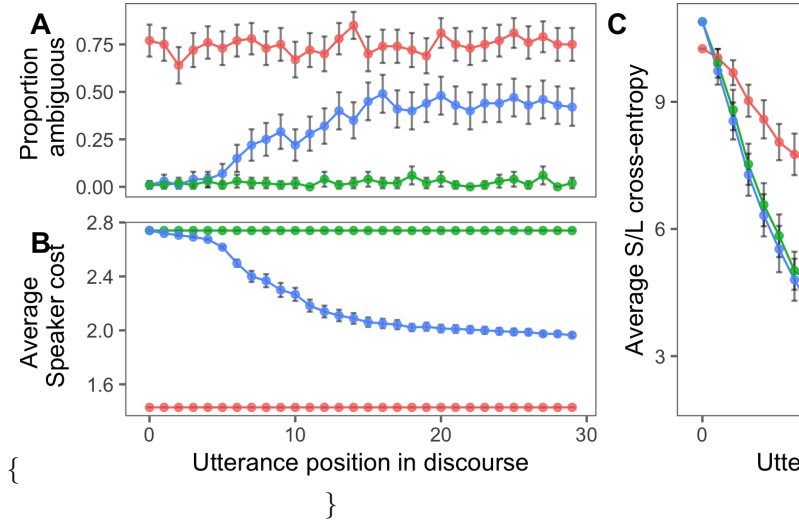
\begin{CodeChunk} \begin{figure*}[h]

{
}
\caption[(A) shows the empirical probability that our speaker used an ambiguous utterance as a function of discourse position]{(A) shows the empirical probability that our speaker used an ambiguous utterance as a function of discourse position. (B) shows speaker effort across the three models. (C) shows the Cross-Entropy objective under our three speaker models. Error bars represent 95% confidence intervals.} \end{figure*} \end{CodeChunk}

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 ℓ^5 with $|U| = 6$ and $|M| = 4$ in which two of the utterances u_5 and u_6 are ambiguous and that $p(u_5) = p(u_6) > p(u_1) = \cdots = p(u_4)$. That is, the two ambiguous utterances (u_5 and u_6) are less costly than the non-ambiguous utterances. (Note that use of this particular language is not essential – the results are broadly generalizable languages that contain ambiguity.)

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 discrouse, 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 listeners estimate of the current topic ($P(C = c|D)$) comes in line with the actual topic known to the speaker). We compare two baselines 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 langauge semantics (ℓ) and topic probabilities ($p(M|C = c_{current})$ (see SI section 3 for details).

---

[5] See SI for the matrix notation of this langauge.

Hypotheses

We assess whether 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 prefered 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 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 and Discussion

Figure 2, (A) shows the empirical probability that a speaker uses an ambiguous utterance as a function of discourse position. The No pragmatic baseline uses ambiguous utterances frequently and at a constant rate over the discourse and 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. (B) shows 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 (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.

In Simulation 2 we explored how ambiguity could be employed efficiently during discourse. We assumed that the listener did not know the current context (topic) *a priori*, but could infer it from the discourse history. We examined speaker production behavior, tracking use of low-cost, but ambiguous material over the discourse. Early in the discourse the *Full pragmatic* speaker avoided using ambiguous material, only using the material later when the context was known to the listener – 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. 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 (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 framework employed here – the analysis of the dynamics of repeated reference games between rational pragmatic agents – provides a tool for the analysis of functionalist theories, allowing linkage across timescales of analysis from conversation (Genzel & Charniak, 2002; &. J. Levy R., 2007) to typology (Regier et al., 2015). In theory, this framework provides extensive opportunity to vary properties of the languages and agents, however we have focused on only a small subset here. For example, examining how increasing or decreasing *rationality* (instantiated through the RSA α parameter) or *pragmatics* (instantiated through the RSA recursion depth) impacts these results is a clear avenue for future work. An important limitation of the current work is the broad assumption that language can be described through such reference games. The degree to which properties of non-referential language can be captured in such a framework remains an important target for future work.

> All simulation and analysis code are available in the public repository for the project: (link will be available upon acceptance)

References

Bennett, E., & Goodman, N. (2018). Extremely costly intensifiers are stronger than quite costly ones. *Cognition*.

Chomsky, N. (2002). An interview on minimilism. 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.

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 interpreation 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.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learnign and the evoluation 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 ahndbook 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.