Title

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

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Peloquin, Benjamin N.
Goodman, Noah D.
Frank, Michael C.

Publication Date

2019

Peer reviewed

The interactions of rational, pragmatic agents 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

Despite their diversity, languages around the world share a consistent set of properties and distributional regularities. For example, the distribution of word frequencies, the distribution of syntactic dependency lengths, and the presence of ambiguity are all remarkably consistent across languages. We discuss a framework for studying how these system-level properties emerge from local, in-the-moment interactions of rational, pragmatic speakers and listeners. To do so, we derive a novel objective function for measuring the communicative efficiency of linguistic systems in terms of the interactions of speakers and listeners. We examine 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 structure and use 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 efficiency-based framing has produced a rich set of theoretical and empirical targets exploring semantic typology (Regier, Kemp, & Kay, 2015), properties such as ambiguity (Piantadosi, Tily, & Gibson, 2011) and compositionality (Kirby, Griffiths, & Smith, 2014), and the efficient use of reduction and redundancy in production (Genzel & Charniak, 2002; Levy & Jaeger, 2007).

The approaches above typically posit efficiency measures that are motivated by information-theoretic principles, but they typically do not ground out in language use by interacting agents. In this work, we derive a novel objective function from first principles of *rational language use* and show how optimizing this objective can lead to communicatively efficient systems. We also demonstrate that assumptions about interlocutors impact whether language properties are used efficiently. In this way, we integrate questions of language design and language use in a single framework.

Functionalist theories commonly frame language efficiency in terms of a fundamental effort-asymmetry underlying everyday communication: what is "hard" for a speaker is likely different than what is "hard" for a listener. Zipf described this as follows: purely from the standpoint of speaker effort, an optimal language $\ell^*_{speaker}$ would tend toward a vocabulary of 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 optimization based on these competing forces – the pressure to jointly minimize speaker and listener effort.

But how does this 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 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 et al., 2014) and empirical explorations (Hawkins, Franke, Smith, & Goodman, 2018) of efficient language structure and use.

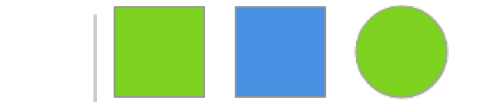
Adopting reference games as a basic unit of analysis suggests that optimization may take place at the level of conversation. Importantly, Zipf's conception of speaker and listener

effort should be connected to how language is used; in particular, whether interlocutors engage in pragmatic reasoning during conversation. Under a Gricean treatment of pragmatics, speakers and listeners 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). We formalize this connection – showing how system-level efficiencies can emerge from local interaction behavior of pragmatic agents. Our claim is that to understand an "efficient" property of a system it is essential that we consider how that property is *used* efficiently.

We provide a case study for this approach, in which functionalist regularities emerge from the dynamics of pragmatic communication. We choose a property of languages that could, in principle, vary freely, but shows strong regularities across languages. The explanandum is why this regularity holds. We examine 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 speaker and listener agents. We adopt the reference game as our primary unit of interaction and model language users with the Rational Speech Act (RSA) framework – a computational model of language use, which is supported by experimental data on interaction. Using these ingredients, we show that the property of interest (ambiguity) is prevalent in languages that optimize our measure of efficiency (Simulation 1). Further, we show how ambiguity is used efficiently during local, in-the-moment interactions (Simulation 2). Put differently, these simulations examine efficiency from two angles – in the first we vary languages, fixing agents, and search for efficient language designs. In the second we vary agents, fixing language, and examine efficient use.

The contributions of this work are twofold – we derive a novel measure of linguistic efficiency and also show how the reference game framework, 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



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. We adopt that formulation here, assuming these communication games as our basic unit of analysis. In this framework, speakers and listeners are aware of a set of objects M (*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 $P(C)$. Each context $c \in C$ describes a different distribution over meanings e.g. $p(M|C = c_i) \neq p(M|C = c_j)$. Finally, we consider a set of communicative events $e \in E$ where $< u, m, c >= e$ is an utterance-meaning-context triple.

Languages A language ℓ defines the set of semantic mappings between utterances and meanings. For example, Figure 1 contains four utterances $U = \{$"blue","green","square","circle"$\}$ and three meanings $M = \{\text{green-square}, \text{blue-square}, \text{green-circle}\}$. The boolean matrix describes the literal semantics of the language. We define a language as "ambiguous" if there is some utterance $u \in U$ which can apply to multiple meanings (i.e. $|[[u_i]]| > 1$)[2]. In Figure 1 both the words "square" and "green" are ambiguous so we would say that ℓ contains ambiguity.

Speakers and listeners The Rational Speech Act framework (RSA) is a computational-level theory of pragmatic language use, which has produced good fit to human communication behavior across a range of language phenomena (Frank & Goodman, 2012; Goodman & Frank, 2016). RSA is a formalization of essential Gricean pragmatic principles – agents reason about one another and their shared context (Grice, 1975). We adopt RSA as our representational framework to model Gricean (rational and pragmatic) speakers and listeners in the reference game setting (see SI).

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 $P_{speaker}(u|m,c;\ell)$. We will use $S(u|m,c;\ell)$ 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 (i.e. $L(m|u,c;\ell)$). Speakers and listeners can induce joint distributions over utterance-meaning pairs, although, these distributions may differ:

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

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

Zipfian objective for linguistic system efficiency

Zipf (1949) proposed that the particular distributional properties found in natural language emerge from competing

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

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

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 $e \in E$.

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

We assume that speaker effort is related to the surprisal of an utterance in a particular context[3] – 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 cost instantiations, such as production difficulty via articulation effort, cognitive effort related to lexical access, or others (Bennett & Goodman, 2018).

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

We assume listener effort is the semantic surprisal of a meaning given an utterance. This operationalization of listener effort is intuitively related to existing work in sentence processing in which word comprehension difficulty is proportional to surprisal (Hale, 2001; Levy, 2008).

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

Importantly, we assume that events $e =< u,m,c >$ are sampled according to the following generative model – some context occurs in the world with probability $P(C = c)$. Within this context, an object m occurs with probability $p(m|c)$. The speaker attempts to refer to that object by sampling from her conditional distribution $S(u|m,c;\ell)$ (i.e. $e \sim p(c)p(m|c)S(u|m,c;\ell)$). From these ingredients it is possible to derive the following objective between the speaker and listener distributions (see SI 2.1 for complete derivation).

$$= \mathbb{E}_{c\sim P(C)}[H_{cross}(P_{speaker}, P_{listener}|c;\ell)] \quad (2)$$

From an information-theoretic perspective this objective is intuitive: H_{cross} denotes the Cross-Entropy (CE), 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, we have an expectation over this term – the expected difference between the distributions assumed by the speaker $P_{speaker}$ and listener $P_{listener}$ given a set of contexts C[4]. In other words, an "efficient" language ℓ minimizes the distance between what speakers and listeners think.

[3]In the current set of simulations we consider utterance costs as independent from context (i.e.. $p(u|c)p(c) = p(u)p(c)$).

[4]Note that in the single context case $|C| = 1$ this objective is simply the speaker-listener Cross-Entropy.

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 languages have not been optimized for communication (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.

As an example, say we have two objects (m_1 and m_2), two utterances (u_1 and u_2), differing in cost, and two languages (ℓ_1 and ℓ_2), describing different utterance-meaning mappings. In language ℓ_1, the low-cost u_1 can be used to refer to both m_1 and m_2 ($[[u_1]]_{\ell_1} = \{m_1, m_2\}$), but the high-cost u_2 cannot be used at all ($[[u_2]]_{\ell_1} = \emptyset$). By contrast, in language ℓ_2, u_1 can only refer to m_1 and u_2 can only refer to m_2 ($[[u_1]]_{\ell_2} = \{m_1\}$ and $[[u_2]]_{\ell_2} = \{m_2\}$). While it is cheaper for a speaker to use ℓ_1 (because speaking it is always lower cost), it is more difficult for a listener (because u_1 is ambiguous). Crucially, 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 to ℓ_2.

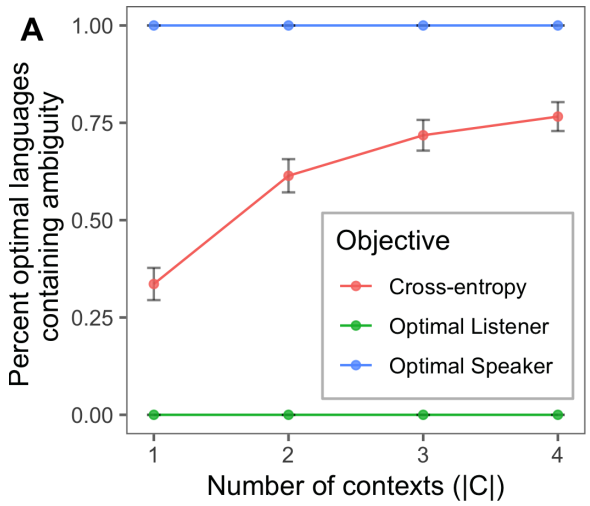
In the following simulations we explore two aspects of Piantadosi's et al.'s claim. In Simulation 1, we examine the efficient language *structure* aspect of their claim, exploring when the optimal linguistic system ℓ^* is most likely to contain ambiguous expressions. In Simulation 2, we explore an efficient language *use* aspect of the claim – under what assumptions will agents use ambiguity efficiently in a conversation?

Simulation 1: Optimal languages contain ambiguity when context is informative

We show that ambiguity is an efficient property under our CE objective in the reference game setting. We proceed by generating languages with different amounts of contextual support (varying the size of $|C|$). We search the space of languages, examining whether ones which minimize our objective contain ambiguity. If context leads to more efficient communication, then optimal languages should be more likely 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 $|U| = |M| = 4$ (U is our set of utterances and M our set of meanings). Recall that languages are boolean matrices and a language $\ell \in L$ is "valid" so long as each possible meaning $m \in M$ 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 row has some non-zero assignment).



B

	m_1	m_2	m_3	m_4
u_1	0	0	**1**	0
u_2	0	**1**	0	0
u_3	**1**	0	0	**1**
u_4	**1**	0	0	0

C

	m_1	m_2	m_3	m_4
u_1	0	0	0	**1**
u_2	0	**1**	0	0
u_3	0	0	**1**	0
u_4	**1**	0	0	0

Figure 2: Panel (A) Vertical axis shows the proportion of optimal languages containing ambiguity. Horizontal axis shows the context-size (1-4) in each condition. Optimal language under CE objective (red). Speaker-optimal (blue). Listener-optimal (green). Error bars represent 95 percent confidence intervals. Panel (B), example CE-optimal language (ambiguous) from a four-context simulation. Panel (C), example CE-optimal language (unambiguous) from a single-context simulation.

For a given simulation, the goal is to find the language ℓ^* which minimizes our objective and then check to see if that language contains ambiguity.

We define language efficiency as 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 each simulation we generate $P(U) \sim \text{Dir}(1,|U|)$, $P(M|C=c) \sim \text{Dir}(1,|M|)$ (a separate 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 We want to assess the impact of *context* on the presence of ambiguity in optimal languages. 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 *one-context* condition ($|C| = 1$) – only a single distribution over meanings $P(M)$. In our *two-context* condition ($|C| = 2$), we consider efficiency under both $P(M|C = c_1)$ as well as $P(M|C = c_2)$. *Three-* and *four-context* conditions corresponding accordingly.

Baselines For comparison, we examine properties of optimal languages under two additional objectives. Zipf (1949) proposed that the speaker-optimal language $\ell^*_{speaker}$ would minimize speaker effort and the listener-optimal language $\ell^*_{listener}$ would minimize listener effort. We define these objectives using the first and second half of equation 1 (see SI 2.2.).

Results and Discussion

In Simulation 1 we explored the degree to which ambiguity is an efficient property of languages when communication is contextualized. Figure 2, panel (A) plots the proportion of optimal languages under each objective as a function of number of contexts. The red line shows that as the number of contexts increases, so does the probability that an optimal language ℓ^*_{cross} contains ambiguity (at least one utterance maps to two meanings) under our CE objective. For comparison we also plot the proportion of speaker-optimal $\ell^*_{speaker}$ (blue line) and listener-optimal $\ell^*_{listener}$ (green line) languages that contain ambiguity. In line with Zipf's predictions, if languages are designed only to minimize speaker effort then optimal languages always contain ambiguity. If languages are designed to minimize listener effort then ambiguity is always avoided.

While our results indicate that ambiguity is an efficient property of contextualized language use, these simulations assumed that 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 interact with others with perfect knowledge of the current context (or topic) at the start of a conversation. To explore how ambiguity may be *used* efficiently in our framework, we next examine a case in which the listener has imperfect knowledge of context at the start of the conversation, but may infer it from the discourse history.

Simulation 2: Rational, pragmatic speakers use ambiguity efficiently

In Simulation 1 we showed that efficiency defined in terms of pragmatic agents leads to a preference for languages that contain ambiguity. In Simulation 2 we assume a single fixed language ℓ, which contains ambiguity, and instead vary the types of agents using ℓ. We will show that efficient *use* of ambiguity depends on an agent's ability to use context for disambiguation. More generally, Simulation 2 is intended to

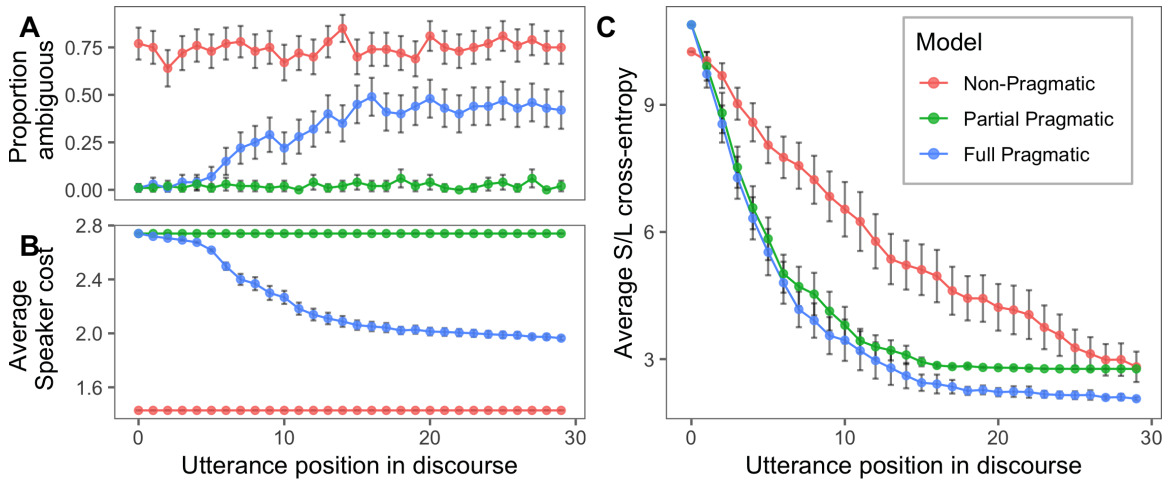


Figure 3: (A) shows the empirical probability that our speaker used an ambiguous utterance as a function of discourse position. (B) shows speaker effort across the three models. (C) shows the Cross-Entropy objective under our three speaker models. Error bars represent 95 percent confidence intervals.

demonstrate how we can assess both questions of efficient use (current simulation) as well as design (previous simulation) in the same framework.

Imagine a scenario in which a reader is beginning a news article. While they may have some knowledge about the article's topic (perhaps from the title), they may not have complete knowledge of its contents, including the persons or events involved. In this setting, using a low-cost, but ambiguous referring expression (say a pronoun like "he") early may lead to misunderstanding if context is not informative. But, if by a later position enough contextual information has accumulated, it may be efficient to use the ambiguous expression. We pursue this general framework in Simulation 2 – examining when in a discourse using ambiguity is efficient. We will consider "context" as analogous to a "topic" of conversation.

Simulation set-up

We consider a single language ℓ, which contains both ambiguous and unambiguous utterances. We assume ambiguous utterances are lower cost. Crucially, we do not assume that the listener knows the particular topic ($c_{current}$) of the conversation *a priori*. Rather, 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.

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 agents who 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.

We conduct $N = 600$ simulations, generating discourses of length $|D| = 30$ utterances, comparing three speaker models ($n = 200$ each). We consider a single language ℓ[5] with $|U| = 6$ and $|M| = 4$ in which two of the utterances are ambiguous and lower cost than the unambiguous utterances. (Note that use of this particular language is not essential – the results are broadly generalizable to languages that contain ambiguity, but exploring this space is computationally expensive.)

Speaker agents

We vary the degree to which agents can use context for disambiguation. We consider three types of speaker models. Our *Full pragmatics* agent, models a speaker 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 over the discourse. 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 the discourse history. The second, a *No pragmatics* baseline speaker does not consider a listener at all, but produces utterances according to the underlying language semantics (ℓ) and topic probabilities ($p(M|C = c_{current})$ (see SI 3).

Hypotheses

We are interested in how each speaker-model uses ambiguity over the discourse. A speaker strategy that is mutually efficient for both agents should avoid ambiguity until sufficient contextual information has accumulated. We should expect this to be reflected in our *Full pragmatics* model who reasons about the listener and discourse history. By contrast, a speaker-optimal model who does not consider the listener should greedily use ambiguous utterances (*No pragmatics* model), while a listener-optimal model should avoid ambiguity entirely (*Partial pragmatics* model).

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

Results and Discussion

Figure 3, panel (A) shows the empirical probability a speaker uses an ambiguous utterance as a function of discourse position. The *No pragmatics* baseline uses ambiguous utterances frequently and at a constant rate over the discourse. The *Partial pragmatics* baseline avoids ambiguous utterances entirely. But, the *Full pragmatics* model avoids ambiguous material only at the start of the discourse, employing it increasingly as the discourse proceeds. Panel (C) tracks our CE objective for each model over the discourse. Note that the objective decreases for all three models, primarily driven by the listeners updating their beliefs about the actual topic ($P(C = c_{current}|D)$). However, the objective declines more quickly under the *Full* and *Partial pragmatics* speakers as listener agents are better able to infer the correct context. Additionally, the difference in CE between the *Full-* and *Partial pragmatics* models at the end of the discourse is driven by the reduction in speaker costs. Panel (B) tracks speaker effort, which remains constant in both *No pragmatics* and *Partial pragmatics* baselines. But, effort declines in the *Full pragmatics* model as speakers increasingly rely on ambiguous material later in the discourse.

General Discussion

How do the competing pressures imposed by speakers and listeners give rise to the distributional regularities 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), deriving a novel speaker-listener Cross-Entropy objective for measuring the efficiency of linguistic systems from first principles of efficient language use. In Simulation 1 we showed that optimal languages are more likely to contain ambiguous material when context is informative. In Simulation 2 we showed how rational pragmatic agents use ambiguous material efficiently in conversation.

A limitation of the current work is an analysis of exactly how the CE objective compares to existing measures. For example, previous work has described competing speaker-listener pressures in terms of a trade-off of simplicity and informativeness (Kemp & Regier, 2012) or expressivity and compressibility (Smith, Tamariz, & Kirby, 2013) to explain linguistic regularities. Future work should assess the degree to which we can derive the same properties as previous studies using our current framework. More generally, we hope that this framework can serve as a domain general tool to assess the range of functionalist theories examining efficient language-structure and use.

SI and simulations: `https://bit.ly/2RBSGcU`, `https://github.com/benpeloquin7/zipf_principles`

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.

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*.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learnign and the evoluation of language. *Current Opinion in Neurobiology*, *28*, 108–114.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106(3)*, 1126–1177.

Levy, R., & Jaeger, T. (2007). Speakers optimize information density through syntactic reduction. In *Proceedings of the twentieth annual conference on neural information processing systems*.

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.

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the 36th annual conference of the cognitive science society*.

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.