The strategic use of noise in pragmatic reasoning

Leon Bergen (bergen@mit.edu)[1], **Noah D. Goodman** (ngoodman@stanford.edu)[2]
[1]Department of Brain and Cognitive Sciences, MIT, Cambridge MA 02139
[2]Department of Psychology, Stanford University, Stanford CA 94305

Abstract

We combine two recent probabilistic approaches to natural language understanding, exploring the formal pragmatics of communication on a noisy channel. We show that nominal amounts of actual noise can be leveraged for communicative purposes. Common knowledge of a noisy channel leads to the use and correct understanding of sentence fragments. Prosodic emphasis, interpreted as an intentional action to reduce noise on a word, results in strengthened meanings.

Keywords: Pragmatics; Communication; Bayesian modeling

Recent work in cognitive science has investigated the inferential aspects of human communication: how people use general-purpose reasoning to go beyond the literal content of their language input. This work has generally taken one of two perspectives. Researchers have used *noisy channel* models (Levy, Bicknell, Slattery, & Rayner, 2009; Gibson, Bergen, & Piantadosi, 2013) to investigate how people infer meanings from language input that may have been corrupted by noise. Work on *probabilistic pragmatics* models (Frank & Goodman, 2012; Franke, 2009) has looked at how people use strategic reasoning to enrich the meanings that they communicate with their conversational partners. The current work will combine these perspectives, providing a unified explanation of two distinct phenomena: the use of sentence fragments to communicate full propositions, and the use of prosody to shift pragmatic interpretations.

People's communication channels are limited in various ways, and successful communication often requires complex inferences in order to overcome these limitations. When a speaker in a conversation tries to communicate something to the listener, their intended signal may be corrupted by speech errors (e.g., if they are excited or intoxicated), environmental noise (if they are in a loud setting, e.g. at a cocktail party), or perceptual noise (if the listener is not paying attention). For the listener to successfully understand the speaker's intended meaning, they must take into account these possible sources of noise, and infer whether they heard what the speaker actually intended. A growing body of experimental evidence suggests that people account for the possibility of noise when interpreting language; this has been successfully modeled as probabilistic inference of the original message given the received message. A different body of computational research has emphasized the implications for communication of probabilistic models of social inference. This work views the agents in a conversation as rational and cooperative, and as having mutual knowledge of these properties. These agents are able to draw sophisticated inferences about each others' intentions by using this knowledge, explaining pragmatic enrichment of sentence meanings.

Here we ask whether novel pragmatic inferences can be driven by the physical properties of the communication channel. Consider that agents in a conversation are typically mutually aware of the physical limitations of their communication system, e.g. that their utterances may be corrupted by noise. These agents can therefore take actions that exploit these limitations, and their conversational partners' knowledge of these limitations. Although noise is typically viewed as a problem for communication, we will argue that the *possibility* of noise is in many circumstances a resource, without which certain types of communication would not be possible.

Using this perspective, we will explore two sets of phenomena that on the surface look only loosely related (see Benz (2012) for the application of similar ideas to a different set of pragmatic phenomena). The first of these phenomena is the effective use sentence fragments to communicate full propositions (Merchant, 2004). Consider, for example, the typical manner in which wh-questions are answered:

(1) A: Who went to the movies?
B: Bob

The utterance *Bob* is used to express the proposition that Bob went to the movies. This proposition could have been expressed with the full sentence *Bob went to the movies*, but this would be redundant given the context; it is common knowledge between the two speakers that the response will describe who went to the movies. Thus, given an appropriate context, the use of sentence fragments allows people to remove extraneous words from their utterances. However, fragmentary sentences present a difficulty for probabilistic pragmatics models, which generally require a complete truth-functional literal meaning to initiate pragmatic reasoning.

The second phenomenon is the use of prosodic stress—a change in how a sentence is pronounced—to change the interpretation of the sentence. In particular, speakers can use prosody to place focus on particular parts of an utterance. Consider the following example:

(2) A: Who went to the movies?
B: BOB went to the movies.

Here capital letters are used to indicate that the speaker has placed prosodic stress on the word "Bob." A stressed utterance will be interpreted exhaustively (Groenendijk & Stokhof, 1984; Von Stechow, 1991): the speaker intends to communicate that Bob was the only person who went the movies. Without stress, this utterance is compatible with other people having gone to the movies in addition to Bob. One explanation of this relationship between prosody and meaning is that prosody simply carries conventional meaning which composes with lexical meaning. An alternative, which we will explore here, is that the meaning of prosodic emphasis comes as a pragmatic enrichment given the effect that prosody has on the physical acoustic channel of speech.

We will demonstrate how both of these phenomena emerge from pragmatic reasoning given the possibility of noise. Our results will be theoretical, showing that the machinery of noisy-channel and probabilistic pragmatics models can be combined and exploring the predictions of this combination.[1] Introspective data from examples will be strong enough to motivate our explorations, but our theoretical results also suggest experimental questions for future research.

Background

Noisy-channel models provide an account how people understand language despite the possibility of noise corrupting their language inputs. Under noisy-channel accounts, the comprehender has a model of the noise process that generated their observed utterance, and they determine the speaker's intended utterance by integrating this knowledge with their prior expectations about what the speaker is likely to say. This is formalized as a Bayesian decoding procedure. The listener has a prior distribution over utterances $P(u_i)$, which gives the probability that the speaker intends to use utterance u_i. The listener's model of the noise process is represented by the distribution $P_N(u_p|u_i)$, which gives the probability that the listener will perceive utterance u_p given that the speaker intended u_i. The posterior probability that the speaker intended u_i given that the listener perceived u_p is given by:

$$L(u_i|u_p) \propto P(u_i)P_N(u_p|u_i) \tag{1}$$

Probabilistic pragmatics models give an account of a different aspect of people's linguistic knowledge: how the participants in a conversation draw inferences that go beyond the literal meanings of the utterances that they hear. In several contemporary models (Frank & Goodman, 2012; Franke, 2009; Jager, 2010; Goodman & Stuhlmüller, 2013), this is formalized with a hierarchy of agents of increasing strategic sophistication. Here we will present a slightly modified version of the Rational Speech Acts model of (Frank & Goodman, 2012). The speaker has a meaning $m \in M$ that they want to communicate to the listener, where M is the set of possible meanings, or worlds. The set of utterances U specifies the grammatical utterances available to the speaker. The literal meaning $[\![u]\!]$ of an utterance u is the set of possible worlds in which the utterance is true: $m \in [\![u]\!]$ if and only if the utterance is true of the meaning.

The least sophisticated agent in the model, the listener L_0, does not reason about why the speaker would have chosen a particular utterance. Rather, they interpret an utterance according to its literal meaning by Bayesian inference:

$$L_0(m|u) \propto \begin{cases} P(m) & \text{if } m \in [\![u]\!] \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The term $P(m)$ is the prior distribution on meanings that the speaker may want to communicate.

The speaker S_n chooses utterances by considering which would most effectively communicate their intended meaning to the listener L_{n-1}. They want to minimize the information that the listener would still lack to recover meaning m after perceiving utterance u_p. This can be formalized as the surprisal: $-\log L_{n-1}(m|u_p)$ (Cover & Thomas, 1991). The speaker wants to minimize surprisal, thus has utility function:

$$U(u|m) = \log L_{n-1}(m|u_p) - c(u), \tag{3}$$

where $c(u)$ is the cost of utterance u. The distribution over utterances for the speaker S_n is defined using the Luce-choice rule, which describes the choice behavior of approximately rational agents (Sutton & Barto, 1998):

$$S_n(u|m) \propto e^{\lambda U_n(u|m)} \tag{4}$$

The term $\lambda > 0$ controls the speaker's degree of rationality.

Finally, listener L_n ($n \geq 1$) interprets an utterance u by considering which meanings would make the speaker S_n likely to choose u. This reasoning is captured with Bayes' rule:

$$L_n(m|u) \propto P(m)S_n(u|m) \tag{5}$$

These models are each intended to capture distinct phenomena, and as currently presented the models are not compatible with each other. The probabilistic pragmatics models assume that there is no noise in the transmission of the utterance from the speaker to the listener; if the speaker intends for the listener to receive utterance u, then this is what the listener will receive. On the other hand, the noisy-channel models do not say anything about how a comprehender can extract non-literal meanings from a speaker's utterance; from the perspective of these models, the speaker is a black box, not a rational agent with sophisticated intentions.

Neither type of model can, on its own, explain the phenomena illustrated above. The pragmatics models only deal with grammatical utterances whose literal meanings are specified in advance, ruling out the use of non-grammatical sentence fragments within these models (unless a semantics for these utterances is given as an input). In addition, without stipulating that prosody shifts the literal meaning of an utterance, these models cannot explain why prosody influences pragmatic interpretation (see Bergen, Goodman, and Levy (2012) for a discussion of related issues). Noisy-channel accounts, which do not provide an endogenous model of how the speaker chooses utterances, do not explain why the speaker would choose ungrammatical sentence fragments as utterances, or place prosodic stress on their utterances.

[1] All of the models discussed in the paper were implemented in the probabilistic programming language Church (Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2008). The code is available at `http://probmods.org/prosody.html`.

Ellipsis

As example 1 illustrates, people can use sentence fragments to communicate full propositions; indeed, this is often the most natural way to communicate. A basic question, then, is what linguistic knowledge allows for the successful interpretation of sentence fragments. One possibility is that this knowledge is conventionalized as part of a language's grammar or semantics, i.e., the language provides a procedure for interpreting fragments that are used in response to questions, or in other contexts. The language might contain a rule stipulating that the name of an agent is a grammatical response to a "who"-question, and that the answer to such a question can be obtained by substituting this name into the question.

It is unlikely, however, that such a grammatical rule would account for the productivity of sentence fragment usage. Consider the following examples (see Stainton (1998) for related cases):

(3) A: I think I've discovered the culprit.
B: The butler!

(4) A: I got laughed out of the seminar again today.
B: Your pragmatic theories.
A: Yeah, everyone thinks they're crazy.

In neither case is there an apparent grammatical marker which licenses the fragment's use as an utterance. Rather, in both examples, the first speaker's utterance raises a topic for discussion, and the second speaker offers a fragment that can be understood given this topic. These examples suggest that the interpretation of sentence fragments is not fixed by the grammar, but rather is driven by general inferential mechanisms.

We will be pursuing the hypothesis that people interpret sentence fragments through pragmatic inference, not grammatical rules. The speaker can successfully communicate using an ungrammatical, fragmentary utterance by exploiting the possibility—known to the listener—that noise may have corrupted the utterance. When a strategically naive listener hears a sentence fragment, they will believe that noise corrupted it, because sentence fragments are not grammatical parts of the language. They will therefore infer what sentence the speaker actually intended to use. A strategically sophisticated speaker, reasoning about the naive listener, will therefore choose a sentence fragment to express their intended meaning if it is parsimonious and likely to be interpreted correctly by the listener. A more sophisticated listener will expect such fragments, and so on. Neither speaker nor listener assign a literal meaning to the sentence fragment itself.

Model

This reasoning can be formalized by combining the Rational Speech Acts model of (Frank & Goodman, 2012) with a simple Bayesian model for interpreting noisy language input (Levy, 2008; Gibson et al., 2013). The model starts with the literal listener L_0 who interprets utterances according to their literal meanings. There is an important difference between the definition of this literal listener and the one in Equation 2. Because noise will sometimes corrupt the speaker's utterances, the listener L_0 needs to infer which utterance the speaker intended before assigning it a literal interpretation. The distribution $P_N(u_p|u_i)$ defines the noise process: the probability that the listener will perceive utterance u_p given that the speaker intends to use utterance u_i. The literal listener assigns a probability to meaning m by integrating the prior probability of m with the probability that m is literally consistent with the speaker's intended utterance:

$$L_0(m|u_p) \propto P(m) \sum_{u_i : m \in [\![u_i]\!]} P(u_i) P_N(u_p|u_i). \tag{6}$$

The speaker S_n (for $n \geq 1$) chooses utterances by maximizing the probability that the listener L_{n-1} will recover their intended meaning, while minimizing effort. The speaker needs to consider that their utterance may be corrupted by noise; whether the listener successfully recovers the speaker's intended meaning will depend on what utterance the listener perceives. Thus, the speaker's utility function U_n generalizes the one defined in Equation 3, and now takes the expectation over the surprisal of their intended meaning:

$$U_n(u_i|m) = \sum_{u_p} P_N(u_p|u_i) \log\left(L_{n-1}(m|u_p)\right) - c(u_i). \tag{7}$$

The speaker's distribution over utterances is defined using the Luce-choice rule, as in Equation 4.

The listener L_n (for $n \geq 1$) has a model of the speaker S_n, and, to a first approximation, interprets an utterance by reasoning about what intended meanings would have made this speaker likely to use the utterance. However, the listener accounts for the possibility that noise corrupted the utterance she received; combining Equations 5 and 1:

$$L_n(m|u_p) \propto P(m) \sum_{u_i} S_n(u_i|m) P_N(u_p|u_i). \tag{8}$$

Results

We will use example 1 to illustrate how the model assigns interpretations to sentence fragments. Suppose that it is common knowledge, following the question, that if anyone went to the movies, then it was either Alice or Bob who went. Thus the speaker either wants to communicate that Alice went to the movies, that Bob went, or that nobody went. For simplicity, we will assume that the utterances available to the speaker are of the form "X went to the movies," where X is "Alice," "Bob," or "Nobody"; it is straightforward to generalize the reasoning here to larger sets of alternative utterances. Finally, we will suppose that the noise process consists of three types of noise (Levy, 2008; Gibson et al., 2013): insertions of words into the sentence, deletions of words, and replacements of words, which are assumed to occur independently. Only deletions will be relevant for the current example.

Figure 1 shows model results: robust interpretation of "Bob" as meaning that Bob went to the movies, almost independent of the actual amount of noise. This can be explained as follows. The naive listener L_0 will interpret the

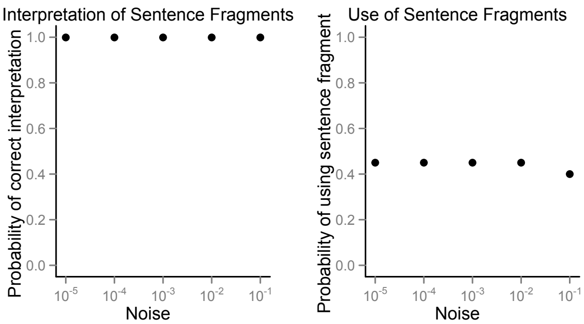

Figure 1: The left panel shows the probability the listener will assign the correct meaning to the sentence fragment "Bob." The right panel shows the probability the speaker will use a sentence fragment, as a function of the true noise rate.

two grammatical utterances according to their literal meanings. If the listener hears the sentence fragments "Alice" or "Bob," then they will believe that this was not the speaker's intent, but rather that noise corrupted the intended utterance. After hearing "Bob," the most probable inference is that the speaker intended to say "Bob went to the movies": this requires positing the smallest number of string edits, i.e. the deletion of the last four words. If the first word of the utterance is deleted, so that the listener hears "went to the movies" (or any subset of these words), then the speaker's intent will be completely lost, because either grammatical utterance will have been equally likely to produce the perceived utterance.

The speaker S_1 reasons about the listener L_0, and knows that this listener will usually interpret the utterance "Bob" as meaning that Bob went to the movies, and similarly for "Alice." If this speaker wants to communicate that Bob went to the movies, then choosing the fragment "Bob" is a reasonable strategy: it will usually be intepreted correctly, and requires less effort than "Bob went to the movies." This explains why sentence fragments will be pragmatically licensed in many contexts to communicate propositional meanings.

It is important to note that this reasoning will work even if the noise rate is arbitrarily close to 0, so long as it is positive. This is illustrated in Figure 1. When the listener L_0 hears an ungrammatical sentence fragment, they know that they perceived something that the speaker did not intend. Conditional on having heard this ungrammatical utterance, they know that noise must have corrupted the speaker's intended utterance, however *a priori* unlikely that is. As a result, the speaker S_1 may utter sentence fragments much more frequently than would be expected under the noise model. The frequency with which the speaker S_1 chooses a sentence fragment is determined by how successfully it communicates their intended meaning, and its cost relative to a grammatical sentence.

Prosody

In the previous section, we showed how participants in a conversation can use the possibility of noise to enrich the meanings that their utterances communicate. We now consider the possibilities for communication that are opened when the speaker can affect the noise rate of their utterances.

As example 2 illustrates, the interpreted meaning of an utterance is changed by prosodic stress. In that example, the speaker uses stress to signal that they have exhaustive knowledge about the question under discussion, namely who went to the mall. Several related uses of prosodic stress have been noted in the literature. Consider the following example:

(5) A: Who did John introduce at the party?
B: John introduced MARY to Bill.

By placing stress on "Mary," the speaker is able to indicate that Mary was the only person that John introduced to Bill; this leaves open the possibility that Mary was introduced to other people by John. This contrasts with the meaning that results from placing stress on "Bill":

(6) B: John introduced Mary to BILL.

In this case, the speaker intends to communicate that Bill was the only person that Mary was introduced to; it is still possible, however, that other people were introduced to Bill. Thus, prosodic stress in these cases is used to indicate the dimension along which the speaker has exhaustive knowledge.

Prosody can also be used to shift the interpretation of scalar items. Consider the following contrast:

(7) I passed the test.

(8) I PASSED the test.

The first example generates a weak scalar implicature: the speaker at least passed the test, and they do not know if they aced it. In contrast, the second example generates a strong implicature: the stress on the scalar item "passed" indicates that the speaker knows that they did not ace the test. More generally, prosodic stress communicates that the speaker knows that the stronger scalar alternatives to their utterance are false.

Why does prosodic stress have these effects on the pragmatic interpretation of utterances? We propose that these effects result from the acoustic signature of prosody. There are two main acoustic changes associated with prosodic stress: increased loudness and duration (Breen, Fedorenko, Wagner, & Gibson, 2010). For our purposes, the significance of these acoustic changes is that they will increase the robustness of the utterance to noise. An utterance that is louder and longer is less likely to get swamped by sounds in the environment, and more likely to be the focus of the listener's attention. Thus, by placing stress on part of an utterance, the speaker is choosing to decrease the noise rate for that part of the utterance. The speaker's choice to reduce the noise rate for a particular part of the utterance is an intentional action, and as a result can receive a pragmatic interpretation by the listener.

Model

As with the previous model, this version begins with a listener L_0 who interprets utterances literally. Before defining this listener, we will consider several changes to the speaker model. Previously, we assumed that the speaker was fully knowledgeable, so that the speaker's intended meaning could be modeled by a point distribution on meanings. The current model will relax this assumption, so that the speaker's knowledge state is modeled by a non-trivial distribution over meanings (Goodman & Stuhlmüller, 2013); the listener will try to infer this knowledge state. In particular, assume that the speaker made an observation o, resulting in a posterior distribution over meanings: $P(m|o) \propto P(o|m)P(m)$, where $P(o|m)$ is the probability that the speaker made observation o given that the true world state is m.

The noise rate of the speaker's communication channel is a function of the prosody that the speaker chooses. This is captured by the term $P_N(\cdot|u_i,s)$, which is the distribution over utterances that will be perceived by the listener, given that the speaker intends utterance u_i and uses prosody s. The probability that the listener accurately perceives u_i will be higher if the speaker chooses to apply prosodic stress to the utterance.

The speaker S_n wants to choose an utterance and prosody that will communicate their knowledge state most effectively, while simultaneously minimizing their effort. Because the speaker's knowledge state is no longer represented by a single meaning, but rather by a distribution over meanings, the informativeness of the speaker's utterance is no longer measured by surprisal, as it was in Equation 7. Instead, we will use KL-divergence to measure the distance between the speaker's belief distribution $P(\cdot|o)$ and the listener's posterior distribution after hearing the speaker's chosen utterance and prosody:

$$U_n(u_i,s|o) = -\sum_{u_p} P_N(u_p|u_i,s) KL(P(\cdot|o)||L_{n-1}(\cdot,o|u_p,s)) - c(u_i)$$

The speaker's distribution $S_n(u_i,s|o)$ over utterances is again defined using the Luce-choice rule (Equation 4).

The literal listener L_0 interprets an utterance u_p with prosodic stress s in a manner that generalizes Equation 6:

$$L_0(m,o|u_p,s) \propto P(m)P(o|m) \sum_{u_i:m\in[\![u_i]\!]} P(u_i)P_N(u_p|u_i,s) \quad (9)$$

Similarly, after perceiving utterance u_p and prosody s, the listener L_n infers the world state and speaker's observations by:

$$L_n(m,o|u_p,s) \propto P(m)P(o|m) \sum_{u_i} S_n(u_i,s|o)P_N(u_p|u_i,s) \quad (10)$$

Results

We will first consider how the model derives the interpretation of sentences like example 2. Assume that following the question in that example, there are three possible meanings that the speaker may want to communicate: Alice went to the movies alone, Bob went alone, or Alice and Bob both went. Assume for simplicity that at least one of them went. The speaker may, however, be ignorant about whether both Alice and Bob went. The utterances available to the speaker will be of the form "X went to the movies," where X can be "Alice," "Bob," or "Alice and Bob." The speaker can adopt one of two prosodic strategies: they can either use no prosodic stress, or place prosodic stress on the names of the agents in the sentence. The base rate of noise is assumed to be 1% and the use of prosodic stress is assumed to decrease the noise rate by a factor of 2 (though the model's predictions are invariant to a range of base noise rates and noise reduction factors).

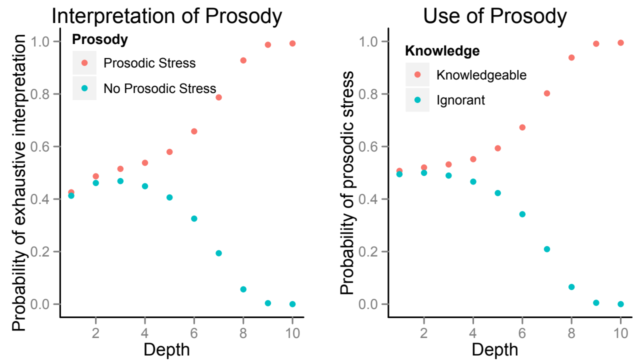

Figure 2: The left panel shows the probability that the listener will assign an exhaustive reading to "BOB went to the movies," as a function of recursion depth. The right panel shows the probability that the speaker will use prosodic stress when they have full or partial knowledge about the situation.

Figure 2 shows that at higher recursion depths—i.e., for more sophisticated speakers and listeners—the model interprets prosodic stress on "Bob" as indicating that the speaker knows that Alice did not go to the movies. This can be explained as follows. The listener L_0 is strategically unsophisticated, and interprets utterances literally. This listener will not gain information from prosodic stress, as prosodic stress is semantically inconsequential under the current account.

The pragmatic effects of prosody first emerge with the speaker S_1. If this speaker knows that only Bob went to the movies, then they have a very strong preference for the listener to hear "Bob went to the movies": if the listener accidentally hears "Alice went to the movies" she will infer that Alice went to the movies, when in fact the speaker knows that this is false. This speaker will therefore be relatively likely to place prosodic stress on the word "Bob," in order to decrease the noise rate and ensure that the listener perceives this accurately. In contrast, if the speaker knows that Bob went to the movies, but is uncertain about whether Alice also went, then they will be less likely to place stress on the word "Bob." If the listener accidentally hears "Alice went to the movies" in this case, it is still a problem for the speaker, as it will imply something that the speaker does not know. However, in contrast to the previous case, Alice going to the movies is at least compatible with this speaker's knowledge. Prosody is thus marginally less valuable for the ignorant speaker; this effect is amplified by further levels of pragmatic recursion.

Next consider the interpretation of prosody in example 5, where selective prosody exhaustifies only part of the interpretation. The modeling assumptions for this case straight-

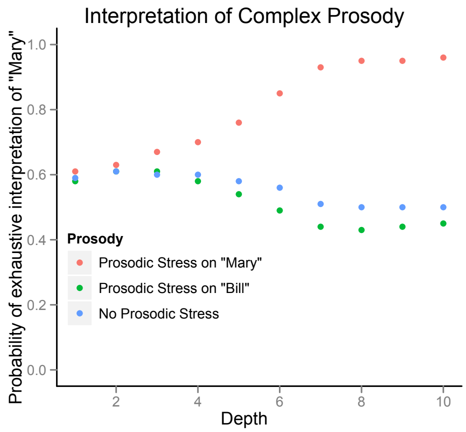

Figure 3: The probability that "John introduced Mary to Bill" will communicate that the speaker knows that only Mary was introduced to Bill, given different prosodic actions.

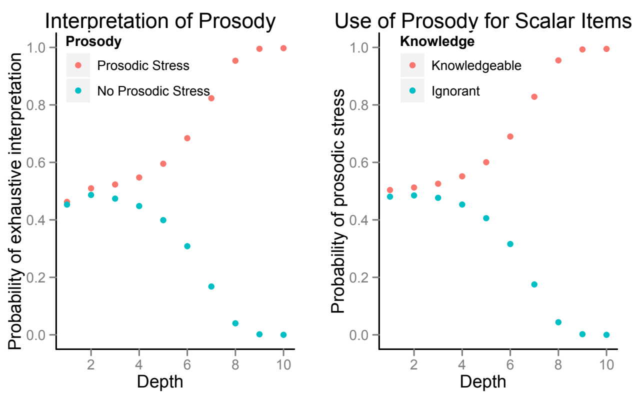

Figure 4: The left panel shows the probability that the listener will assign a strong (knowledgeability) implicature to "I PASSED the test," as a function of recursion depth. The right panel shows the probability that the speaker will use prosodic stress when they have full or partial knowledge.

forwardly extend those for the previous case. The pragmatic effects of prosody again arise at the level of the speaker S_1. If the speaker S_1 knows that Mary was the only person who was introduced to Bill, but is uncertain about whether Mary was introduced to anyone else, then they will be more likely to place prosodic stress on "Mary" than "Bill." Conversely if the speaker has complete knowledge of who was introduced to Bill. As a consequence, the speaker's choice of where to place prosody is informative about the dimension along which they have exhaustive knowledge, as shown in Figure 3.

Finally, we consider the model's derivation of strong implicatures from prosodic stress, as in example 8. Suppose that the two scalar alternatives available to the speaker are "passed" and "aced" (where "aced" entails "passed"). If the speaker S_1 knows that they passed the test, but did not ace it, then it is important for the listener to correctly hear their utterance. If the listener accidentally hears "aced" instead of "passed," then they will believe something that the speaker knows is false. This speaker will therefore be relatively likely to stress "pass." In contrast, if the speaker knows that they at least passed, but does not know if they aced the test, then it is not as important for the listener to correctly perceive their utterance. As a result, the speaker with partial knowledge will be less likely to place prosody on the scalar term. Figure 4 shows how prosody affects the interpretation and usage of scalar items.

Conclusion

We have provided a unified pragmatic account of two distinct linguistic phenomena: the use of sentence fragments to communicate full propositions, and the use of prosody to shift the interpretation of utterances. We have argued that both of these phenomena result from the strategic exploitation of the possibility of noise, which emerges naturally when noisy channel and probabilistic pragmatics models are integrated.

References

Benz, A. (2012). Errors in pragmatics. *Journal of Logic, Language and Information*, *21*(1), 97–116.

Bergen, L., Goodman, N. D., & Levy, R. (2012). Thats what she (could have) said: How alternative utterances affect language use. In *Proceedings of the thirty-fourth annual conference of the cognitive science society.*

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7-9), 1044–1098.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. John Wiley & Sons.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Franke, M. (2009). Signal to act: Game theory in pragmatics.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Goodman, N. D., Mansinghka, V. K., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2008). Church: a language for generative models. *Uncertainty in Artificial Intelligence*.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, *5*(1), 173–184.

Groenendijk, J., & Stokhof, M. (1984). *Studies in the semantics of questions and the pragmatics of answers* (Unpublished doctoral dissertation). University of Amsterdam.

Jager, G. (2010). Game-theoretical pragmatics. *Handbook of logic and language*, 467.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 234–243).

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090.

Merchant, J. (2004). Fragments and ellipsis. *Linguistics and Philosophy*, *27*(6), 661–738.

Stainton, R. J. (1998). Quantifier phrases, meaningfulness "in isolation", and ellipsis. *Linguistics and Philosophy*, *21*(3), 311–340.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). Cambridge Univ Press.

Von Stechow, A. (1991). Focusing and backgrounding operators. *Discourse Particles: Pragmatics & Beyond, Amsterdam: John Benjamins*, 37–84.