Information theoretic factors in marking linguistic focus:
A laboratory-language approach

Gareth Roberts (gareth.roberts@ling.upenn.edu)
University of Pennsylvania, Department of Linguistics, 3401-C Walnut Street,
Philadelphia, PA 19104 USA

Jon Scott Stevens (stevens.400@osu.edu)
The Ohio State University, Department of Linguistics, 118G Ohio Stadium East, 1961 Tuttle Park Place,
Columbus, OH 43201 USA

Abstract

We present an experimental study investigating the role of information-theoretic factors in determining patterns of redundancy and focus in language and other communication systems. Pairs of participants played a simple communication game using a non-linguistic visual medium to send messages to each other. We manipulated noise, effort, and time pressures and measured message length, redundancy, and accuracy. Participants behaved as predicted based on an information-theoretic model, with message length and redundancy varying according to circumstance, but accuracy remaining constant.

Keywords: communication; focus; information theory; language; redundancy; signaling game

Introduction

Questions invite answers. However, the invitation is not entirely open: There are important constraints on the form that an answer can take. Most obviously, it must be relevant (Grice, 1975; Wilson & Sperber, 2004). In response to the question of who invented the printing press, for instance, "My sister loves cheesesteak" is not an acceptable answer unless some double meaning is understood. Perhaps less obviously, the length and syntactic completeness of an answer are also constrained. Fragment answers—that is, answers that are not whole sentences—are permitted and may even be preferred in some circumstances. At the same time, prosodic and morphosyntactic mechanisms (e.g., pitch fluctuations; cleft constructions such as "it was . . . who") may be employed to add redundancy to certain elements in the sentence. Patterns of message length and redundancy should not be assumed to be random. In this paper we present a set of communication game experiments investigating how they vary according to information-theoretic factors.

As an example of how such patterns vary in English, consider the three answers to the printing press question that are given in (1). While (1a) is an acceptable answer, (1b) and (1c) are just as good, and would even be preferable in many contexts.

(1)   Who invented the printing press?

    a. Gutenberg invented the printing press.

    b. Gutenberg did.

    c. Gutenberg.

As evidenced by the acceptability of (1a), fragment answers are not obligatory. Indeed, whole-sentence answers

may help establish that the respondent heard and understood the question correctly; in some cases they may also serve to aid parsing of the what Schmitz (2008) calls the "critical" element (in this case *Gutenberg*). An answer like "The printing press was invented by Gutenberg", for instance, makes very clear through syntactic and prosodic means when precisely the listener should expect to get the answer. An utterance that does contain such unnecessary material is constrained in important ways; in particular, *focus* must be marked on the critical element. This is realized in English and many other languages by a pitch accent on that element (i.e., the element is distinguished from other parts of the sentence through variation in pitch, length and intensity; for discussion of this well-documented phenomenon, see Ladd, 1996; Rooth, 1992). It is important to note that this focus is constrained by the grammar of the language and is more or less obligatory—distinguishing the wrong element, distinguishing the right element by the wrong means, or distinguishing no element at all, is likely to be perceived as odd. Example (2) illustrates this pattern, with small caps denoting pitch accent.

(2)   Who invented the printing press?

    a. ✓ GUTENBERG invented the printing press.

    b. ✓ GUTENBERG did.

    c. ✗ Gutenberg invented the PRINTING PRESS

    d. ✗ GUTENBERG invented the PRINTING PRESS

    e. ✗ Gutenberg invented the printing press

It should be noted that similar patterns pertain even when no explicit question has been asked. For example, a restaurant worker who notices a patron wandering around might say to them something like "Over there", "They're OVER THERE", or "The toilets are OVER THERE", even if the patron does not actually ask for directions.

Stevens (2016), following Schmitz (2008) and Bergen and Goodman (2015), argues for a theory of focus based partly in information theory (Shannon & Weaver, 1949), whereby focus and redundancy are presented as a solution to *noise*, the random deletion/alteration of parts of a signal. Given that elements in messages may be lost as a result of mishearings, attention failures, and the like, focus is a means of emphasizing those elements that the speaker considers most important to transmit accurately. This can be seen as a process of adding

redundancy to important parts of the linguistic signal, compensating for the effects of noise. At the same time, depending on pressures of time and effort, non-focused elements (the parts of the signal that are easily inferred from context) can be reduced or even elided.

Stevens (2016) followed Rooth (1992) and many others in arguing that the inferrabillity of semantic material from the discourse context follows from the set of *alternatives* that is evoked by that context. For example, "Who invented the printing press?" evokes a set of possible answers {x invented the printing press} which contains propositions like "Gutenberg invented the printing press", "Edison invented the printing press", etc. These alternatives all share overlapping semantic material, namely "invented the printing press" which is inferrable and thus less important in terms of information transmission. The material in need of protection from noise, in other words, is the material that does not overlap among all contextually available alternatives. We will therefore refer to redundancy on this material as *non-overlapping redundancy*, and redundancy on inferrable material as *overlapping redundancy*.

Predictions of an information-theoretic model

If patterns of focus and redundancy in language are indeed a reflex of general information-theoretic concerns, then we expect that analogs to it will arise in any communication system that shares the goal of signaling the selection of one object from among a set of alternatives. For such systems we can thus make the following predictions. First, overall message length should vary according to time and effort costs: Messages should be longer if effort costs are low and time is not pressing. This might seem trivial, but the key point is that messages will not simply be minimally short in all cases. Because redundancy can be useful, messages will in fact be longer than they strictly need to be if time and effort constraints allow. Second, longer messages should differ from shorter messages not only with respect to length, but also with respect to the proportion of overlapping redundancy; that is, overlapping redundancy is a greater luxury than non-overlapping redundancy, and should be dispensed with more readily. Third, the distribution of effort in a message should take noise into account. In particular, non-overlapping redundancy should be higher when noise is higher, both in an absolute sense (more redundancy overall) and in a relative sense (more non-overlapping redundancy than overlapping redundancy). Fourth, unless noise and time pressures become so great that they simply prevent accurate communication, we should expect communicative accuracy to remain relatively constant, regardless of the variation between messages predicted above. This is because that variation is (according to this account) designed to help maintain accuracy under different conditions.

There is evidence from natural language that speakers are sensitive to time and effort pressures. Corpus analysis suggests, for instance, that syntactic reduction is used to optimize information density (Jaeger, 2010), while work in experimental pragmatics demonstrates that referring expressions get shorter over time as interlocutors establish common ground (Krauss & Weinheimer, 1964; Clark, 1996).

In general, however, it is hard to fully test the information-theoretic basis of focus using natural-language data alone. There is more than one reason for this. A common difficulty in testing predictions in natural language is that the relevant explanatory factors are hard to manipulate (cf. Roberts, in press). To a small extent this is true here: While it is relatively straightforward to introduce noise and time pressures, for example, it is harder to make the act of producing natural-language utterances more effortful without imposing awkward physical constraints on participants. However, this is not the most serious obstacle in this case—in fact there is a history of research in phonetics in which awkward physical constraints are put on participants (e.g., Lindblom, 1990). The more serious problem is that, in established natural languages, there are constraints on such phenomena as pitch accent, focus, and utterance length that are encoded in the grammar. This means that while the patterns of interest might have arisen initially as a result of the information-theoretic factors discussed, we should expect them to an important extent to be "fossilized" as part of the language, incorporated into relatively constrained grammatical constructions, and less sensitive to contingent factors. Even if the information-theoretic account is right, therefore, behavior in established languages may be somewhat weak evidence for it; better evidence would be furnished by observing the establishment of a novel communication system. Furthermore, as stated above, we should expect our claims and predictions to apply across communication systems, linguistic or otherwise. It would therefore be advantageous for our purposes to test predictions in some non-linguistic communicative medium. Fortunately, the last decade or so has seen the development of a line of research in which participants communicate in the lab using "laboratory languages"—either artificial languages taught to the participants, or novel communication systems developed collaboratively over the course of the experiment (Galantucci, Garrod, & Roberts, 2012; Roberts, in press). This approach allows researchers to investigate principles of communicative behavior while abstracting away as far as possible from the established natural languages that the participants bring with them into the lab. A common kind of study involves the use of visual communication systems. For instance, Garrod, Fay, Lee, Oberlander, and MacLeod (2007) had participants play a Pictionary-like game, while Galantucci (2005) made participants communicate by drawing on a pad that simulated continuous vertical motion, preventing the use of most established conventions. This kind of approach is particularly useful for investigating the challenges involved in establishing reference and constructing a new system from scratch. However, we were interested in the pressures acting on a system in which reference is in principle relatively straightforward, as is the case for a native speaker of a natural language, but where there are no established grammatical conventions as

in language. For this reason we had participants play a very simple communication game in which they had to fill in cells in a grid to convey a line that was drawn over the grid (Figure 1). Under ideal conditions, this task is rather trivial. We made conditions less ideal by manipulating noise, time pressure, and the effort required to produce a signal. We then measured the redundancy in the signals that were produced and the success rate in interpreting them.

Method

Participants

One hundred and twenty University of Pennsylvania undergraduate students participated in pairs for course credit or $5.

Procedure

In each trial a pair of participants played a simple cooperative signaling game. Each sat in a separate cubicle with a computer; neither participant could see or hear the other. The game consisted of a series of turns in each of which one player was nominated as Sender and the other as Receiver; players alternated roles, with the Sender in the first turn being selected at random. At the start of a turn the Receiver saw a white screen with the message "You are waiting on a message transmission from the other player". The Sender saw a screen as in Figure 1. On the left were two 7×7 grids, over each of which a different line figure was drawn. Every line figure consisted of a continuous line drawn between the center points of eleven contiguous grid cells. One of the two line figures was selected in green, while the other was white.[1]

On the right of the screen was an empty 7×7 grid, slightly larger than the other two. Beneath the two leftmost grids there was also a button marked *Send*. The Sender's task was to communicate to the Receiver which of the two grids was selected in green by clicking on cells in the rightmost grid. At the moment of the Sender's first click a timer would start. Once the timer stopped (after either 5 or 30 seconds, depending on condition; see Section *Experimental conditions*) the grids would disappear and be replaced by the message "You are waiting on a guess from the other player." Clicking the Send button would have the same result. Once the Sender's turn had come to an end in one of those two ways, the Receiver's screen would change to display the two line figures that had been displayed to the Sender (in a random order, but each in the same orientation as for the Sender) as well as a third 7×7 grid in which certain cells might be colored black, and a button marked *OK* (Figure 2). The black cells would always be cells that the Sender had clicked; however, not all cells that the Sender had clicked would necessarily be sent. The means of deciding which cells would be sent depended on the condition (Section *Experimental conditions*). The Receiver's task was to select which of the two line figures they

---

[1]Given that the orientation of the two figures was never varied and that the two line figures were never identical, it is important to note that there was always at least one cell that would, on its own, allow the two line figures to be distinguished.

thought the Sender was trying to communicate. Both players were then told whether the Receiver chose correctly. Then a new turn began. There were 48 turns in total, which were preceded by two practice turns. In half the turns (the *Overlap turns*) the two line figures overlapped by five squares (as in Figures 1 and 2). In the other half (the *Filler turns*) there was no such overlap, such that any cell through which a line figure passed would serve to distinguish it from its competitor.

The only differences between the practice turns and the other turns were that the players' success in the practice turns did not count toward their final score, and that, following the practice turns, they were told to ask questions if they had any. If players scored over 80% in the non-practice turns, they were rewarded with $2 each.
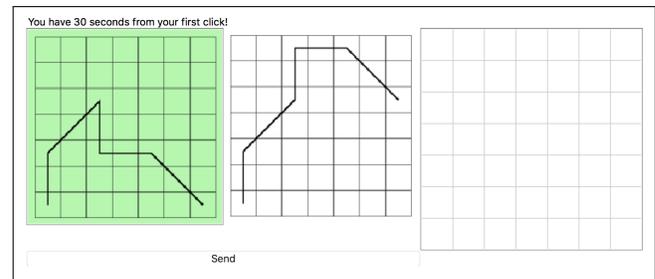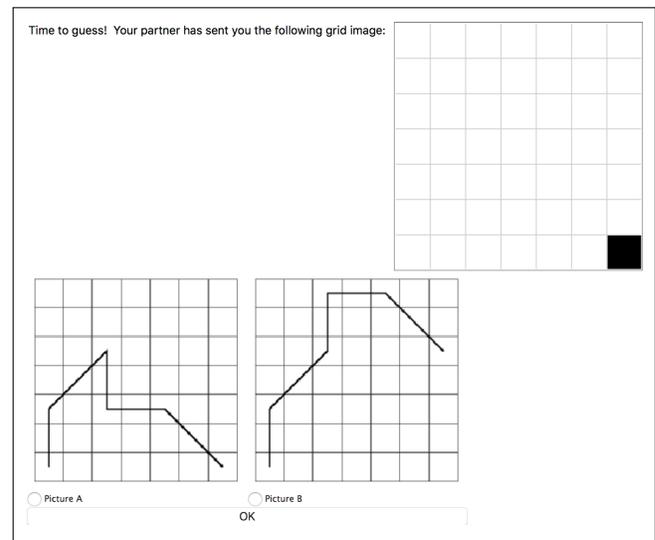

Figure 1: Sender's screen.

Figure 2: Receiver's screen.

Experimental conditions

There were six between-subjects conditions in total (Table 1). Conditions differed from each other with respect to the time available to the sender (either 5 seconds or 30 seconds) and with respect to the means with which the sender could ensure that a message be sent. In the *Effort conditions* clicking a cell a specific number of times would guarantee that it was sent

Table 1: Experimental conditions

Time limit	Effort	Noise
30 seconds	High effort	High noise
30 seconds	Low effort	Low noise
5 seconds	Low effort (5s)	Low noise (5s)

to the Receiver. The Sender had to click fifteen times on each cell in the *High effort condition* and five times in the *Low effort conditions*. Once a cell had been clicked the requisite number of times, it turned black. In the Effort conditions, a cell that turned black for the Sender was sure of being seen by the Receiver, but a cell that had not been clicked enough remained white and was sure not to be sent. The same was not true in the *Noise conditions*. In these conditions, each click on a cell would make it appear darker. Any cell that had been clicked once had a chance of being sent. Clicking more on the cell not only made the cell darker, but increased that chance (and the darkness of the cell increased proportionally to the probability that the cell would be sent). The probability of a given cell being sent was calculated as $1 - (1 - d)^n$, where d is a decay parameter between 0 and 1, and n equals the number of times the Sender clicked on the cell in question. Two values were used for the decay parameter. In the *High noise condition*, it was set at 0.1. In the *Low noise conditions* it was set at 0.4. This meant that it would take many more clicks in the High noise condition than in the Low noise conditions to feel confident that the cell would be sent. For instance, 4 clicks in the latter condition would result in an 87% chance of the cell being sent, but a 34% chance in the former. In all conditions the number of cells clicked on can be taken to correspond roughly to utterance length, while in the Noise conditions, cell darkness can be taken to be analogous to greater effort in marking prominence.

The length of time available for the Sender to click on cells was varied, being set at either 5 or 30 seconds. For the Low effort and Low noise conditions, there was both a 5-second condition and a 30-second condition. For the High effort and High noise conditions, however, we ran only 30-second conditions; this is because 5 seconds was not enough time for participants to send more than one cell in the High effort condition, or to have a good chance of doing so in the High noise condition, meaning that the results would be rather trivial and most of our measures could not be calculated.

Measures

We measured the following variables.

Message length. This was calculated by counting how many squares had been clicked on (for Noise conditions) or turned black (for Effort conditions). Since one cell would be sufficient to distinguish between any two line figures (assuming it arrived), any message consisting of more than

one cell can be considered to contain redundancy.

Click effort. This was calculated by counting the number of clicks made by the Sender.

Overlapping redundancy. This was calculated by counting how many cells the Sender clicked on (for Noise conditions) or turned black (for Effort conditions) that overlapped with both line figures, and would therefore not differentiate the target figure from its competitor. (Note that this could be measured only for the Overlap turns, and not the Filler turns.).

Accuracy. This was calculated by counting the number of turns in which the Receiver chose the correct line figure.

Results

Data reported in the following sections come from Overlap turns only, for the sake of consistency between the measures. Unless otherwise stated, all models reported are mixed effects linear regression models with random intercepts for both subject and item, using the Satterthwaite approximation of degrees of freedom to obtain a p-value from a t-value.

Variation in message length

We predicted that overall message length should vary according to time and effort costs, with messages being longer than strictly necessary if time and effort constraints allowed. This was supported by the data. One cell would have been sufficient in every condition to signal which line figure to choose; however, messages were longer than this in every condition (Figures 3 and 4). In all but the High noise and High effort conditions, message length also remained relatively constant throughout a game, though it was lower in the Low noise (5s) than in the Low noise condition ($\beta = -4.83, SE = 0.75, t = -6.48, p < 0.001$) and in the Low effort (5s) condition than in the Low effort condition ($\beta = -7.64, SE = 0.65, t = -11.83, p < 0.001$), suggesting that the shorter message lengths in the 5s conditions were due to time constraints. In both the High noise condition and the High effort condition, message length began high and fell over the course of the game in a rather linear fashion, converging in the High effort condition with the Low effort (5s) condition. This suggests that effort came to exercise increasing pressure as time went on. It is likely also the case that it partly reflects participants' growing familiarity with the game, although if this were the main explanation, one would expect a sharper fall by the middle of the game rather than a linear decline. Message length in the noise conditions did not quite converge (i.e., the slope was less steep), likely because less effort was necessary in this condition.

Message length and overlapping redundancy

We predicted that shorter messages should differ from longer messages with respect to the proportion of overlapping redundancy. This was confirmed. The proportion of clicks devoted
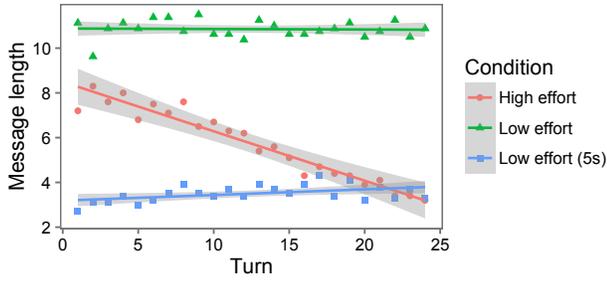
Figure 3: Message length in effort conditions.
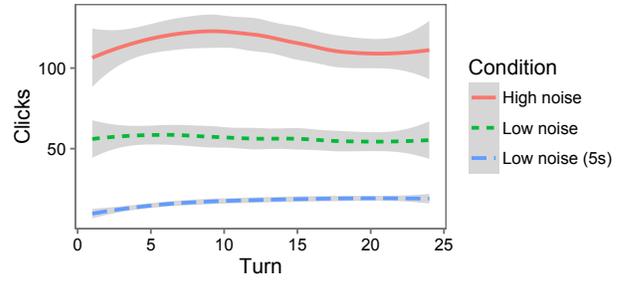

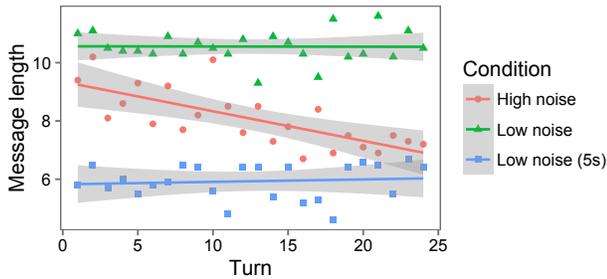Figure 6: Click effort in noise conditions.

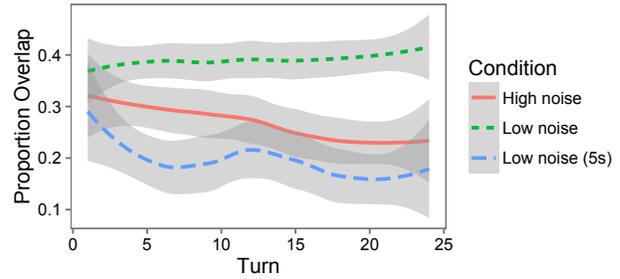Figure 4: Message length in noise conditions.

Figure 7: Proportion of clicks on overlapping redundant cells in Noise conditions.

to overlapping redundancy (as opposed to non-overlapping redundancy) correlated with click rate ; $r = 0.29, p < 0.001$, but the correlation was stronger ($r = 0.56$) in the High effort and High noise conditions alone, where there was greater pressure on participants (Figure 5).

Distribution of effort

We predicted that the distribution of effort would take noise into account and that non-overlapping redundancy would be higher when noise is higher, both in an absolute sense and relative to overlapping redundancy. This was confirmed by results. First, click effort was lowest in the Low noise (5s) condition, but higher in the High noise condition than in the Low noise condition ($\beta = 58.57, SE = 7.56, t = 7.75, p < 0.001$), in both of which rounds lasted 30 seconds (Figure 6). Sec-

Figure 5: Correlation between click effort and overlapping redundancy (High effort and High noise conditions only).

ond, for the proportion of overlapping redundancy, there was a significant interaction of turn and condition for the High noise and Low noise conditions ($\beta = 0.006, SE = 0.002, t = 3.07, p < 0.01$): Although all the noise conditions began at roughly the same place (Figure 7), the proportion of overlapping redundant cells clicked in the Low noise condition remained relatively constant, but decreased on the High noise condition. There is also a main effect of condition if the interaction terms are excluded from the model ($\beta = 0.11, SE = 0.04, t = 2.81, p < 0.01$). The proportion of clicks on overlapping redundant cells was significantly lower in the Low noise (5s) than in the High noise condition ($\beta = -0.10, SE = 0.04, t = -2.44, p < 0.05$).

Accuracy rates

We predicted that accuracy would remain similar between conditions, regardless of differences in message length, click effort, and redundancy. This was broadly true. Overall mean accuracy was very high (97%), and did not differ significantly between conditions, with one exception (Figure 8): It was lower in the Low noise (5s) condition ($\beta = -0.07, SE = 0.02, t = -2.81, p < 0.01$). This likely represents an underestimation of noise by participants; this was the only condition in which senders had little time to send a signal, but could not be sure before sending it how much the receiver would see.

Discussion

We modeled the emergence of a system similar to linguistic focus by having participants play a simple non-linguistic
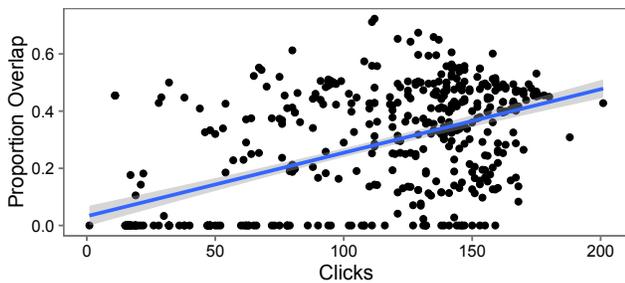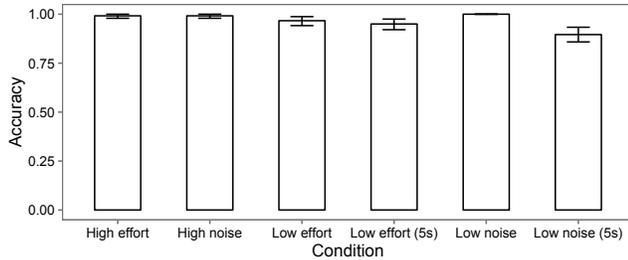
Figure 8: Accuracy rates by condition. Error bars show 95% confidence intervals.

communication game, in which we manipulated noise, effort, and time pressures. We made four predictions consistent with an information-theoretic account of focus: that message length should vary according to time and effort costs, that longer messages should differ primarily with respect to redundant material that is shared with alternatives, that the distribution of effort in the message should take noise into account, and that accuracy should be stable in spite of variation in other measures. These predictions were confirmed. The patterns we observed also resemble patterns in natural language. While message length varied considerably between conditions, accuracy was maintained, largely due to distribution of effort being skewed toward protecting non-overlapping material from noise. On the one hand this is consistent with the natural-language examples given in the introduction to this paper. On the other hand, it is also consistent with findings from experimental pragmatics studies in which participants repeatedly refer to a set of unfamiliar shapes; referring expressions are reliably shorter at later stages of such interactions (Krauss & Weinheimer, 1964; Clark, 1996). This is typically explained in terms of the establishment of common ground—participants develop a shared perspective on the objects in question. On the surface, this does not obviously apply so well to our study, but the fundamental mechanism is the same. The point in both cases is that message length varies as a result of time and effort pressures on the one hand, and the sender (or speaker)'s confidence that the message will be understood by the receiver, on the other.

Overall, our results lend support to the view that linguistic focus may have emerged as a response to information-theoretic pressures. We also consider that the experimental approach we have taken may prove fruitful for future work.

Acknowledgments

References

Bergen, L., & Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, *7*(2), 336–350.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*(5), 737–67.

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental Semiotics. *Language and Linguistics Compass*, *6*(8), 477–493.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–87.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, vol. 3: Speech acts* (pp. 41–58). New York: Academic Press.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction—a preliminary study. *Psychonomic Science*, *1*(5), 113–114.

Ladd, D. R. (1996). *Intonational phonology*. New York: Cambridge University Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (Vol. 55, pp. 403–439). Dordrecht: Kluwer Academic Publishers.

Roberts, G. (in press). The linguistic *Drosophila*: Experiments in language change. *Linguistics Vanguard*.

Rooth, M. (1992). A theory of Focus interpretation. *Natural Language Semantics*, *1*, 75-116.

Schmitz, H.-C. (2008). *Accentuation and interpretation*. New York: Palgrave MacMillan.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Champaign, IL: University of Illinois Press.

Stevens, J. (2016). Focus games. *Linguistics and Philosophy*, *39*(5), 395–441.

Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford: Wiley Blackwell.