Overspecification and the Cost of Pragmatic Reasoning about Referring Expressions

Peter Baumann (baumann@u.northwestern.edu)
Department of Linguistics
Northwestern University, Evanston, IL, USA

Brady Clark (bzack@northwestern.edu)
Department of Linguistics
Northwestern University, Evanston, IL, USA

Stefan Kaufmann (stefan.kaufmann@uconn.edu)
Department of Linguistics
University of Connecticut, Storrs, CT, USA

Abstract

In current approaches to pragmatic reasoning the comprehension and production of referring expressions is modeled as a result of the interlocutors' mutual perspective-taking under the additional assumption that speakers try to minimize their articulatory effort or production cost. The latter assumption is usually not tested and instead built into the experimental tasks of referential language games by artificially restricting the set of possible referring expressions available to identify a referent. We present two language game experiments: a production experiment, in which the speakers were allowed to freely choose a referring expression, and a comprehension experiment to replicate earlier findings with our stimuli. Our results show that while listeners easily perform pragmatic reasoning, speakers resort to overspecification when the effort of pragmatic reasoning becomes too high.
Keywords: Pragmatics; Language games; Referring Expressions; Language Production; Language Comprehension

Introduction

In a complex situation, like a room full of people, one may be able to identify a single person upon hearing a rather short description, like "the man with a hat", even in cases in which the man in question is not the only one with a hat and despite the fact that he certainly has many other characteristics besides wearing a hat. This amazing efficiency of human communication has made *reference* and the use of referring expressions a central topic in linguistic pragmatics. The most influential descriptive account for the efficiency of human communication was formulated by Grice (1975) in the form of his *Maxims of Conversation*, and since then, a number of proposals have been made to provide more quantitative models of pragmatic reasoning based on the Gricean maxims or more general principles of human cognition and interaction. The most prominent examples are game-theoretic models based on strategic reasoning (e.g. Benz & Van Rooij, 2007; Franke, 2011; Jäger, 2011) and Bayesian models grounded in social cognition (e.g. Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013). In both classes of models it is assumed that the speaker and hearer reason about each other's perspectives: the hearer is assumed to interpret a speaker's expression as referring to the referent for which the expression is 'optimal' under the perspective of the speaker, who in turn chooses the referring expression to be 'optimal' under the hearer's perspective, etc.

A trivial solution to this recursive reasoning process is for the speaker to choose a referring expression that explicitly mentions *all* features of the intended referent and is thus absolutely unambiguous in the given context. Since such an expression can hardly qualify as efficient, however, the above models make the crucial additional assumption that the speaker incurs a cost for producing an utterance, thus all things being equal, speakers have a preference for the most economic (i.e. shortest and least effortful) expression.

While theoretically appealing, production costs are notoriously hard to quantify, as the articulatory effort of speech production[1] is negligible (Moon & Lindblom, 2003; Locke, 2008). And anecdotally one may even be tempted to reject the notion of production costs altogether: people talk a lot.

However, we do not argue against a possible role for production costs in pragmatic reasoning. Instead, we show that speakers' behavior cannot be explained in terms of this factor alone: Under certain conditions, speakers do use costlier forms than would be required to identify the intended referent, which suggests that the process of pragmatic reasoning (often termed *implicature*) is itself effortful (like any other reasoning process) and thus incurs a cost for the speaker.

In language comprehension, it is well established that pragmatic reasoning can be effortful: in reading, sentences involving implicatures take longer to process than sentences without implicatures (Hamblin & Gibbs, 2003), and in referential language games, target identification is less accurate when the message involves an implicature than when it does not (e.g. Degen & Franke, 2012).

In language production, on the other hand, the speaker has the option of reducing the need for pragmatic inference by *overspecification*, i.e. saying more than is strictly necessary. And indeed, there is plenty of empirical evidence that speakers do make use of overspecification. In particular, research on the production of referring expressions (e.g. Koolen,

[1] Typological analyses of language efficiency (e.g. Piantadosi, Tily, & Gibson, 2012) are typically based on arguments involving processing restrictions in *comprehension* rather than production.

Gatt, Goudbeek, & Krahmer, 2011) has identified a number of properties of referents and their surroundings that can lead to a speaker making use of overspecification: in the above scenario of a room full of people one can easily imagine someone referring to the man in question by the expression "the bald man with a red hat", even if that man was in fact the only one wearing any hat.

However, in empirical research on pragmatic reasoning about referring expressions, in particular in referential language games (e.g. Degen & Franke, 2012; Frank & Goodman, 2012; Degen, Franke, & Jäger, 2013), the speakers' option of overspecification is often ignored. Instead the theoretical assumption that production costs influence the choice of referring expressions is built into the experimental tasks by artificially restricting the set of possible referring expressions available to identify a referent.

In this paper, we argue that such a restriction of the possible alternatives of referring expressions misses crucial features of human communication, since speakers in real-life situations are not constrained in their choice of forms. To show this, we present two referential language game experiments: a production experiment, in which we allow the speakers to freely choose a referring expression, and a comprehension experiment to replicate earlier findings with our stimuli. Our results demonstrate that while listeners easily perform pragmatic reasoning, speakers resort to overspecification when pragmatic reasoning is involved. This shows that also in language production, pragmatic reasoning can be effortful, and that this effort or cost of pragmatic reasoning can outweigh a possible production cost for the speaker.

Related Work

In this section we introduce referential language games as a common method for investigating pragmatic reasoning about referring expressions and then briefly review some relevant studies from the vast literature on the generation of referring expressions and their take on overspecification.

Pragmatic Reasoning in Language Games

Referential language games[2] (Wittgenstein, 1959; Lewis, 1969) have been used to empirically test recursive reasoning about referents and referring expressions in (simple) visual contexts.

As an example, consider the situation sketched by the three pictures shown in Figure 1. If in this context, a speaker said (1), a listener could infer that she is referring to the picture in the middle.

(1) My friend is the one wearing sunglasses.

This follows under the assumption that if the speaker had wanted to refer to the right picture, she would have used an expression like (2), which unambiguously identifies the intended referent.

[2] Also known as signaling games.

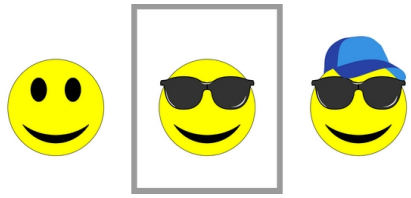

Figure 1: Example of the visual context of a language game.

(2) My friend is the one wearing a hat.

It has been shown that (adult) listeners can easily perform this kind of pragmatic reasoning (e.g. Stiller, Goodman, & Frank, 2011; Degen & Franke, 2012), which is similar to the calculation of scalar implicatures (Grice, 1975).

In the general form of a language game experiment, speakers and listeners see a visual display of several potential referents, from which the speaker picks (or is given) a referent to talk about, while the listener does not know the referent. The speaker then chooses a referential description, and if the listener can identify the intended referent, communication is successful, otherwise it fails.

Language games are particularly suitable for the study of pragmatic reasoning as they allow for an easy manipulation of the depth of (recursive) pragmatic reasoning required for successful communication by changing the distribution of features across the different referents. In Figure 1, the target referent (middle), shares one feature with each of its two competitors: like the competitor to its right, the target is wearing sunglasses, and like the competitor to its left, it is not wearing a hat. So upon hearing a sentence like (1), the hearer must employ pragmatic reasoning and strengthen the heard utterance to mean "the one wearing sunglasses, but no hat".

One interesting feature of language games is that it is possible to experimentally manipulate the set of possible referential descriptions or messages that the speaker has at her disposal, and this feature has been exploited in several studies of pragmatic reasoning: Degen and Franke (2012) and Degen et al. (2013) provide their speakers with a fixed set of four (visual or linguistic) messages describing only one feature and have them choose the best one to identify a given referent out of an array of three potential referents varying along two feature dimensions. Similarly, Frank and Goodman (2012) simulated a 'speaker condition' by asking their participants which one of two features they would choose to describe a simple geometric object out of an array of three objects varying along those two feature dimensions.

Speaker cost in language games In addition to restricting the set of possible referring expressions, one can also associate individual referring expressions with explicit or implicit production costs. The explicit approach was explored by Rohde, Seyfarth, Clark, Jäger, and Kaufmann (2012), who showed that when referring expressions are assigned an explicit dollar value, hearers take the cost of referring expres-

sions into account and reason about the speaker's set of possible expressions and their respective production costs. Degen et al. (2013) chose a more implicit approach to assigning speaker costs to referring expressions by embedding the referential language game into an artificial language learning task: in their artificial language, referring expressions differed in length, and during the learning phase participants had to enter the expressions by clicking on a virtual keyboard. This rather cumbersome method of entering their answers was chosen to increase the participants' awareness of possible production costs for speakers. In the following referential language game, participants drew more pragmatic inferences if an unambiguous alternative referring expression was of high cost than when it was of low cost.

In all referential language games reviewed so far, speakers were faced with a forced-choice task. One may argue that such a task is highly artificial and quite likely to miss crucial aspects of a real speaker's communicative situation with its potentially infinite choices of expression. If we put ourselves in the position of the speaker who wants to talk about the referent in the middle picture of Figure 1 we may ask: Would we actually use sentence (1) in this context? Or would we rather say something like (3)?

(3) My friend is the one wearing sunglasses and no hat.

Sentence (3) exemplifies one crucial aspect of human communication that is ignored in such forced-choice tasks: the speaker has the option of overspecification, i.e. of saying more than is minimally required for a successful identification of the intended referent.

Overspecification in the Generation of Referring Expressions

The fact that speakers have the option for overspecification and often make use of it has become a well-established fact in the psycholinguistic literature on referring expressions (e.g. Pechmann, 1989; Engelhardt, Bailey, & Ferreira, 2006). Numerous studies have shown that the properties (or features) of referents form preference hierarchies and that properties listed high on these hierarchies, such as color, tend to be used by speakers even if they are not necessary to identify an intended referent (Koolen et al., 2011).

Most of these studies involve rather complex visual contexts from which a specific referent is to be identified. This makes the comparison to the rather simple visual scenarios used in the study of pragmatic reasoning somewhat difficult. One notable exception is an experiment reported by Gatt, van Gompel, Krahmer, and van Deemter (2011): as in the above mentioned language games, the visual display consisted of three objects, which differed along two feature dimensions, namely color and size, and participants were given an open prompt to answer. The results showed that participants made significant use of overspecification, especially if the redundant feature was color. However, the authors of this study were primarily interested in the differential effects of color and size on the degree of overspecification, and so the experiment consisted only of conditions in which either color or size or both were unique features of the target object. Crucially, there was no 'ambiguous' condition, which would have involved pragmatic reasoning.

Experiment 1

Experiment 1 is a production experiment in which we remedy the shortcomings of previous work on pragmatic reasoning in referential language games by providing speakers with an open prompt instead of a forced choice, thus not imposing any artificial restrictions on the speakers' choice of referring expressions. However, like in the signaling games above, the speakers' choice of referring expressions was naturally restricted by the visual scenes, which are rather simple and vary only along two feature dimensions.

In contrast to the experiments reviewed in the section on overspecification of referring expressions, we used features for which no preference is established and which are generally low on any preference hierarchy, such as optional accessories of persons or animate beings.

Participants

Using Amazon Mechanical Turk, 96 participants were recruited for the experiment. All participants were naive to the purpose of the experiment and reported to be native speakers of English.

Materials and Design

We designed two sets of arrays of three pictures, like the one in Figure 1. The individual pictures differed by the presence or absence of two accessories, in this case a hat and a pair of sun glasses, and the accessories were chosen so that their most natural way of linguistic expression is by means of a post-nominal prepositional phrases (e.g. "smiley *with a hat*").

The pictures in the arrays were assembled according to two conditions: (1) the *implicature condition* (Figure 2, top) and (2) the *simple condition* (Figure 2, bottom). In the simple condition, the target picture (indicated by a grey frame) had one unique present feature (hat) in the context, by which it could easily be referred to (e.g. "smiley with a hat").

In the implicature condition, the target picture shared one present and one absent feature with each of its two competitors: like the competitor to its right, the target is wearing a hat, and like the competitor to its left, it is not wearing sunglasses. The target referent can be identified through pragmatic reasoning by just mentioning its one present accessory (e.g. "smiley with a hat"), as the competitor referent with the same accessory also has a second accessory (sunglasses), which could be used to uniquely identify it. Alternatively, the target referent could be identified by an overspecifying referring expression mentioning both features (e.g. "smiley with a hat and no sunglasses"). Crucially, overspecification here requires mentioning the *absence* of an accessory (e.g. "with no sunglasses").

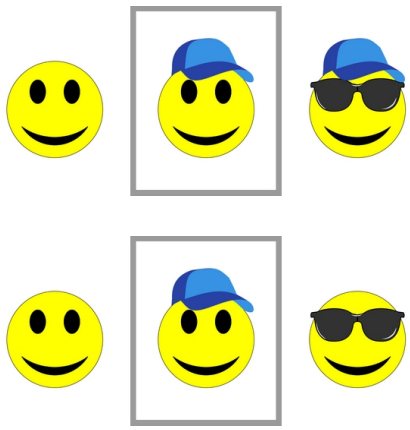

Figure 2: Sample stimuli of Experiment 1. Implicature condition (top): target picture (in grey frame) shares one present (hat) and one absent (sunglasses) feature with each of its two competitors. Simple condition (bottom): target picture has one unique present feature (hat).

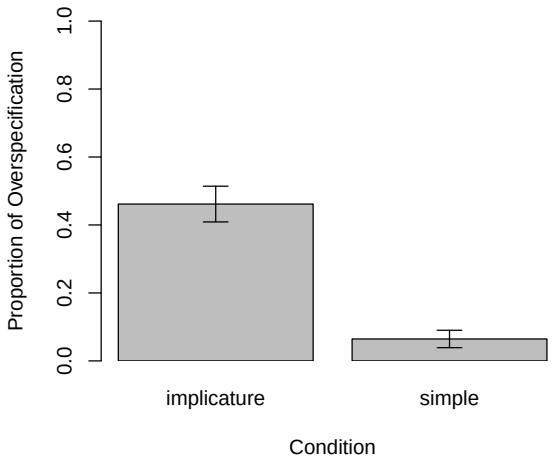

Figure 3: Results of Experiment 1: Proportions of overspecification by condition. Error bars are standard errors.

We employed a one-shot within-subject design: each participant completed only two trials (cf. Frank and Goodman (2012), who also used single trial experiments for a task very similar to ours): one in the implicature condition and one in the simple conditions. The two accessory dimensions and array sets were counter-balanced across participants and the order of the three pictures was randomized within the array.

Procedure

The two trials were presented on a single web page. Each trial consisted of a three-picture array with one picture marked by a grey frame as in Figure 2. Under the picture array there was a text line with the words "Pick the" followed by an open prompt and a period. The participants' task was to fill in the blank so that another person could identify the target picture with the grey frame.

Participants were instructed to imagine that they were communicating with another person over an instant messaging or chat system and that they wanted their partner to pick the picture with the grey frame out of the three pictures in the array. They were told that their partner saw the same three picture, but without the frame and in a possibly different order. This instruction was emphasized by the fact that the target picture appeared in a randomized position within each trial. In addition, participants were asked to complete the initial two words into a correct sentence of English.

Results

The individual answers were manually assessed for correctness, i.e. whether or not it was possible to identify the intended referent from the answer. Out of 192 responses, only 8 ($\approx 4.2\%$) did not allow for a unique identification of the target picture and were excluded. For the remaining 184 trials, we manually annotated which features (present or absent) were explicitly mentioned in the response. From this we calculated the *proportion of overspecification* as the number of trials with both features explicitly mentioned divided by the total number of trials. This proportion is shown in Figure 3. As expected, the proportion of overspecification is higher in the implicature condition (46.2%) than in the simple condition (6.5%). A logistic regression was fit to an indicator variable of overspecification per trial with the condition variable as a categorical predictor. It confirmed that the difference between the simple and the implicature condition is highly significant ($\beta = -2.52$, SE $= 0.47$, $p < .0001$).

Discussion

The fact that in the simple condition only 6.5% of the responses contained more information than necessary is in line with the claim that there may be a production cost for the speaker (as assumed by Degen et al., 2013) and that all things being equal speakers have a preference for the most economic (i.e. shortest and least effortful) expression. However, the effect of this preference is significantly reduced when pragmatic reasoning is involved, as in the implicature condition participants opted for overspecification in 46.2% of their answers. This suggests that speakers have an incentive for avoiding the need for pragmatic inference, which can easily outweigh the cost for articulatory effort (or longer expressions).

Experiment 2

Experiment 2 is a comprehension experiment. The main goal of this experiment is to replicate with our (visual) stimuli the well-established finding that participants are able to employ pragmatic reasoning in the identification of a referent based on a referring expression (e.g. Degen & Franke, 2012; Frank & Goodman, 2012; Stiller et al., 2011).

Participants, Material and Design

Using Amazon Mechanical Turk, 60 participants, all reported English native speakers, were recruited for the experiment.

The visual arrays were the same as in Experiment 1 (without the gray frame around the target picture) with the same two conditions: (1) the *implicature condition* (Figure 2, top) and (2) the *simple condition* (Figure 2, bottom).

Procedure

The two trials were presented on a single web page. Each trial consisted of the picture array and a sentence like (4).

(4) Pick the smiley with the hat.

In the simple condition, this sentence allows for an immediate unique identification of the target picture, while in the implicature condition pragmatic reasoning must be employed for a successful target identification.

Participants were asked to select the 'best' picture given the sentence and made their selection by clicking on a radio button under one of the pictures. Like in Experiment 1, participants were instructed to imagine that they are communicating with another person over an instant messaging or chat system and that they both saw the same three pictures, however in possibly different orders, and that they wanted to identify the picture their partner was referring to by the given sentence.

Results

The proportions of correct target choices are shown in Figure 4. As expected, participants were able to correctly identify the intended target in the implicature condition and chose it in 91.7% of the cases. This number is even higher than the one reported by Degen and Franke (2012)[3] and not considerably lower than 96.7% of correct responses in the unambiguous simple condition.

Discussion

As a replication of earlier findings (e.g. Degen & Franke, 2012; Frank & Goodman, 2012; Stiller et al., 2011), we showed that participants were able to perform simple pragmatic inference in a comprehension task involving the visual stimuli used in Experiment 1.

General Discussion

In models of pragmatic reasoning it is generally assumed that the speaker and hearer reason about each other's perspectives and that this recursive reasoning process leads to 'optimal' or 'rational' communication under the additional assumption that speakers try to minimize their articulatory effort or production cost. What is not included in this effort or cost calculation is the cognitive effort it takes for the speaker to find the optimal referring expression. In line with evidence showing

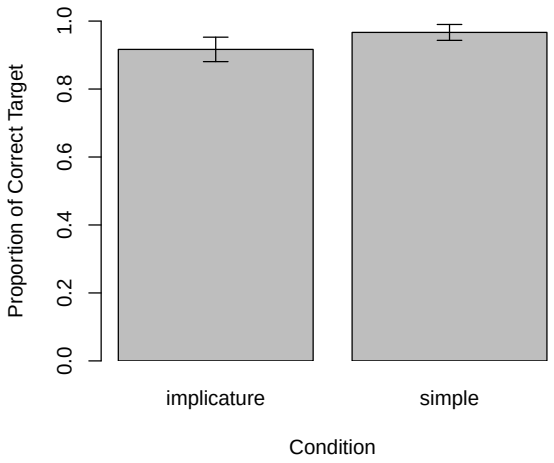

Figure 4: Results of Experiment 2: Proportions of correct target identification by condition. Error bars are standard errors.

that pragmatic reasoning is effortful in language comprehension, our results show that pragmatic reasoning is also effortful in language production, and that this cognitive effort of pragmatic reasoning should be taken into account when determining the speaker cost for a referring expressions: as a speaker has the option of overspecification, she can employ it whenever the effort of pragmatic reasoning becomes too high. This suggests that speakers readily incur an extra cost in order to obviate the need for pragmatic inference. We can think of different possible explanations for this behavior. One is that pragmatic inference requires an effort on the listener's part, and that cooperative speakers use costly forms to share the burden. Another possible explanation is that pragmatic reasoning, with its reliance on implicit assumptions about mutual beliefs about the situation, is a source of uncertainty and error, which speakers have an incentive to avoid. Finally, our results are also compatible with the assumption of audience design, in the sense that speakers choose a maximally explicit referring expression that minimizes the listener's burden of pragmatic inference and maximizes the likelihood of successful communication. Our present study was not designed to decide between these possibilities; we leave this question to future work. Here we only show that speakers use costly forms to short-cut pragmatic inference.

For the hearer, on the other hand, there is no way to avoid the cost of pragmatic inference (other than letting communication fail) once the speaker has used a form that requires it, and we showed in Experiment 2 that hearers generally make the effort of pragmatic reasoning. Based on this asymmetry, we argue that the model of the speaker that the hearer is reasoning about, must be different from the behavior of an actual speaker in the same communicative situation, as an actual speaker can always resort to overspecification, while a hearer needs to assume that the speaker chose the least effortful referring expression and perform all pragmatic inferences

[3] In the 'simple implicature condition' of the comprehension experiment, Degen and Franke (2012) report a proportion of target choice of $\approx 82\%$, which is averaged across six trials per participant.

licensed by that expression. This suggests that in their current form, models of pragmatic reasoning about referring expressions (e.g. Franke, 2011; Frank & Goodman, 2012) may be adequate models of how a hearer comprehends referring expressions, but they may be less adequate for the speaker or the production of referring expressions in general (for a similar observation, see Gatt, van Gompel, van Deemter, & Krahmer, 2013).

More generally, our results challenge the tacit assumption that utterances requiring pragmatic reasoning are ubiquitous (e.g. Breheny, Katsos, & Williams, 2006). Instead we would argue that such utterances may not be as common as previously assumed. These utterances need to be produced by a speaker, and a speaker always has the option of overspecification.

Acknowledgments

We thank the audiences of the 27th Annual CUNY Conference on Human Sentence Processing and the University of Connecticut LingLunch for valuable feedback, and three anonymous reviewers for their excellent comments.

References

Benz, A., & Van Rooij, R. (2007). Optimal assertions, and what they implicate: A uniform game theoretic approach. *Topoi*, *26*(1), 63–78.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? an online investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

Degen, J., & Franke, M. (2012). Optimal reasoning about referential expressions. In S. Brown-Schmidt, J. Ginzburg, & S. Larsson (Eds.), *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 2–11). Paris.

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 376–381). Berlin: Cognitive Science Society.

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554–573.

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

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*(1), 1–82.

Gatt, A., van Gompel, R., Krahmer, E., & van Deemter, K. (2011). Non-deterministic attribute selection in reference production. In K. van Deemter, A. Gatt, R. van Gompel, & E. Krahmer (Eds.), *Proceedings of the Workshop on Production of Referring Expressions: Bridging the Gap between Empirical, Computational and Psycholinguistic Approaches to Reference (PRE-CogSci'11).* Boston.

Gatt, A., van Gompel, R., van Deemter, K., & Krahmer, E. (2013). Are we Bayesian referring expression generators? In A. Gatt, R. van Gompel, E. G. Bard, E. Krahmer, & K. van Deemter (Eds.), *Proceedings of the Workshop on Production of Referring Expressions: Bridging the Gap between Cognitive and Computational Approaches to Reference (PRE-CogSci'13).* Berlin.

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

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech Acts.* Academic Press.

Hamblin, J. L., & Gibbs, R. W. (2003). Processing the meanings of what speakers say and implicate. *Discourse Processes*, *35*(1), 59–80.

Jäger, G. (2011). Game-theoretical pragmatics. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of Logic and Language* (p. 467-491). Amsterdam: Elsevier.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, *43*(13), 3231–3250.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge: Harvard University Press.

Locke, J. L. (2008). Cost and complexity: Selection for speech and language. *Journal of Theoretical Biology*, *251*(4), 640–652.

Moon, S.-J., & Lindblom, B. (2003). Two experiments on oxygen consumption during speech production: vocal effort and speaking tempo. In *XVth International Congress of Phonetic Sciences.* Barcelona, Spain.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*(1), 89–110.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.

Rohde, H., Seyfarth, S., Clark, B., Jäger, G., & Kaufmann, S. (2012). Communicating with cost-based implicature: A game-theoretic approach to ambiguity. In S. Brown-Schmidt, J. Ginzburg, & S. Larsson (Eds.), *The 16th Workshop on the Semantics and Pragmatics of Dialogue.* Paris.

Stiller, A., Goodman, N. D., & Frank, M. C. (2011). Ad-hoc scalar implicature in adults and children. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2134–2139). Boston.

Wittgenstein, L. (1959). *Philosophical Investigations*. Oxford: Blackwell.