Adapting to a listener with incomplete lexical semantics

Sadhwi Srinivas (sadhwi@gmail.com)
Cognitive Science Department, Johns Hopkins University
Baltimore, MD 21218 USA

Barbara Landau (landau@jhu.edu)
Cognitive Science Department, Johns Hopkins University
Baltimore, MD 21218 USA

Colin Wilson (colin.wilson@jhu.edu)
Cognitive Science Department, Johns Hopkins University
Baltimore, MD 21218 USA

Abstract

Speakers involved in a communicative exchange construct an internal model of their addressees and draw upon the model to craft utterances that are likely to be understood. In many real-world situations (e.g., when talking to a non-expert, non-native speaker, or a child), this process of audience design involves identifying gaps in the lexical-semantic knowledge of the listener and selecting alternative expressions. We examine speaker adaptation to a listener with incomplete lexical knowledge in the spatial domain, specifically a failure to comprehend the basic terms *left*/*right*. Experimental and modeling results provide evidence of rapid adaptation that is modulated by the availability of alternative spatial terms. We consider how our approach relates to recent work in computational pragmatics, and suggest that adaptation to the lexical knowledge of the addressee is an important but relatively understudied topic for future research.

Keywords: language adaptation; audience design; spatial language; lexical semantics; computational pragmatics

Introduction

Speakers choose referring expressions on the basis of several factors, including their beliefs about the linguistic and conceptual knowledge of addressees (e.g., Pate & Goldwater, 2015; Brennan & Clark, 1996). For example, speakers tend to avoid or supplement proper names that, in their judgment, listeners do not know (e.g., Isaacs & Clark, 1987; Fussell & Krauss, 1992; Wu & Keysar, 2007; Kutlak et al., 2016). This is part of a more general pattern of *audience design* (Clark & Murphy, 1982) in which speakers construct internal models of specific addresses and use these models to facilitate communication. Another example of audience design is the formation of partner-specific conceptual pacts during a conversation. Speakers also show some ability to adapt their descriptions to a listener's viewing perspective when it is different from their own, presumably making it easier for the listener to identify intended referents in a scene (e.g., Schober, 2009). These adjustments to the needs of particular listeners in specific circumstances are analogous to the well-known Lombard effect, in which people tend to talk louder in the presence of ambient noise (Lombard, 1911), suggesting that audience design is a fundamental phenomenon that occurs at many linguistic and conceptual levels.

Empirical and computational work on audience design has largely adopted the assumption that discourse participants share knowledge of basic vocabulary items. For example, the Rational Speech Acts framework (Frank & Goodman, 2012) assumes that speakers and addressees have the same literal meanings for lexical expressions, and derives pragmatic usage from literal semantics through iterated probabilistic inference. However, the assumption of common word knowledge is not completely valid for many real-world scenarios. In the same way that experts addressing novices should avoid overly technical jargon, speech tailored to non-native or child listeners must regularly work around basic lexical-semantic gaps.

Recently, Ferrara et al. (2016) closely investigated the linguistic choices that parents made when communicating spatial information to their 3-4yo. children. The language used by parents to describe the location of items in simple spatial arrays differed significantly from that of adults addressing other adults in the same task. Most notably, parents avoided the horizontal axis terms *left/right*—terms that are known to be acquired relatively late by children in general, and that were not reliably understood by many of the particular children in the study—while they used many other spatial terms (e.g., the vertical axis terms *above/below*) in essentially the same way that adults do when taking to one another. These findings support the claim that parents have well-tuned internal models of their children's lexical-semantic knowledge and can design utterances for them by circumventing their lexical gaps.

In this study, we investigated whether adaptation to gaps in spatial language would occur in a minimal communicative setting. The parents in Ferrara et al.'s study had developed internal models of their children through extensive interaction with them. Here we sought to determine whether calibration to the addressee's lexical knowledge could develop much more rapidly, perhaps after only a few instances of communicative breakdown. Furthermore, parents presumably have access to a variety of top-down and bottom-up cues to gaps in children's spatial lexicons (e.g., general experience with child spatial language, instances where children explicitly ask for clarification of spatial descriptions). Here we sharply restricted the interaction among interlocutors, providing only ambiguous, bottom-up cues to the addressee's knowledge of the spatial lexicon.

Following Ferrara et al. (2016), our study took the form of a referential communication task. Participants provided spatial descriptions to a listener who had either full knowledge of lexical-semantics or full knowledge except for a gap in the horizontal terms *left/right*. This is the same gap observed in young children, and indeed comprehension of *left/right* can be demanding even for typical adults (Sholl & Egeth, 1981). If audience design is operative in this setting, participants should more often supplement or employ alternatives to the basic horizontal terms when addressing the listener who does not accurately comprehend them.

We further investigated whether and how accommodation of the listener's lexical gap was modulated by the stimulus array. For example, in describing the location of the object marked by the arrow in Fig 1A, alternatives other than *left/right* may not be obvious to the speaker. Contrast this with the arrangements of Fig 1B and 1C, in each of which a non-horizontal spatial relation is available to identify the target (i.e., *inside the box*, *below a triangle*). For arrangements such as Fig 1C, speakers might even prefer to refer to the vertical axis independently of any particular addressee (e.g., Logan, 1995; Ferrara et al., 2016).

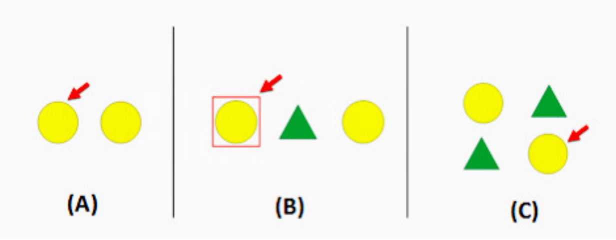

Figure 1: Examples of array types used in the spatial communication task

Finally, we were interested in whether the adaptation found experimentally could be accounted for with a simple inhibitory mechanism: one that penalizes the use of terms that have resulted in communication errors, leaving unchanged all other aspects of the speaker's system for generating referring expressions. We formalize this mechanism in a high-level computational model of referential communication, show that it matches the detailed pattern of spatial language with minimal free parameters, and discuss how it relates to and extends previous work in computational pragmatics. The detailed empirical findings presented here contribute to the understanding of the form and limitations of lexical audience design, and our model delineates a way by which previous computational models may be augmented to account for lexical differences among interlocutors.

Spatial communication experiment

Our experiment involved communication between speakers (the participants) and a simulated listener. In each trial, the speaker described the spatial location of a target object in a visual array (see examples in Fig 1). The listener responded by selecting one of the objects, and the participant was then shown the listener's choice alongside the original display. Apart from this minimal communicative interaction, participants were provided no information about or feedback from the listener.

There were two conditions that differed by listener type. In the **Full** knowledge condition, the listener comprehended all English spatial expressions without error. In the **Partial** knowledge condition, the listener was identical except that comprehension of *left/right* and minor variants of those terms was at chance. We varied the spatial arrangements across the stimulus arrays to elicit a range of linguistic expressions and, most importantly, to provide varying opportunities for adaptation in the Partial condition. We were further interested in whether adaptation would involve primarily avoiding *left/right* or supplementing those terms with other spatial information, as well as in the coarse-grained time course of adaptation. Rapid avoidance of the lexical-semantic gap would provide evidence of an adaptation mechanism that is quite sensitive to bottom-up feedback and that inhibits expressions that have resulted in communication errors.

Participants

This experiment was part of a series conducted online using Amazon's Mechanical Turk service. There were 48 participants (M_{age} = 33.9 years, 25 males), 24 in each condition (Full vs. Partial knowledge). Individuals received a small monetary compensation for participating.

Materials

The stimuli consisted of 32 spatial arrays similar to those of Ferrara et al. (2016). Target objects, marked by red arrows in the display, could not be uniquely identified by intrinsic properties such as shape and color. For example, *yellow circle* would not be a uniquely identifying description of the target in Fig 1C, but *yellow circle on the bottom* or *yellow circle below a triangle* would both be sufficient.

In all stimuli, the target could be identified by its position on the horizontal axis (e.g., *yellow circle on the right*). This maximized the potential contrast between participants in the Full condition, who could in principle describe all targets with *left/right*, and those in the Partial condition, who would have to employ other terms to communicate successfully.

The arrays varied in the alternative spatial descriptions that they afforded. In the **Horiz** type (10 items), the target could be identified only by its relative position on the horizontal axis (Fig 1A). Under the assumption that *left/right* are generally the most accessible or preferred terms for horizontal position, these arrays would be expected to provide the greatest challenge for adaptation.

All other array types contained alternative spatial relations that could be used in identifying descriptions: *proximity* of the target to another object; *containment* of the target within a bounding box (see Fig 1B); internal *vertical* or *horizontal orientation* of the target (e.g., a pencil pointing up or down);

or a *vertical* relation between the target and another object. We grouped the proximity, containment, and vertical/horizontal orientation arrays into a single type called **Horiz+Other** (18 total items). For such arrays, adaptation could involve using horizontal terms other than *left/right* or, perhaps more simply, reference to the other spatial relation (e.g., *circle in the box* instead of *circle on the left*).

The **Horiz+Vert** array type (4 items; see Fig 1C) was separated from the others on the basis of extensive evidence that reference to the vertical axis is both linguistically and non-linguistically privileged relative to the horizontal axis (e.g., Logan, 1995; Carlson-Radvansky & Logan, 1997; Fitneva & Song, 2009). Speakers could prefer vertical terms for these arrays, not because of any specific beliefs about the experimental addressee, but as a reflection of this general privilege (as was observed in Ferrara et al., 2016). Adaptation would then be difficult to measure for these arrays, as *left/right* may be independently dispreferred by participants in both the Full and Partial knowledge conditions.

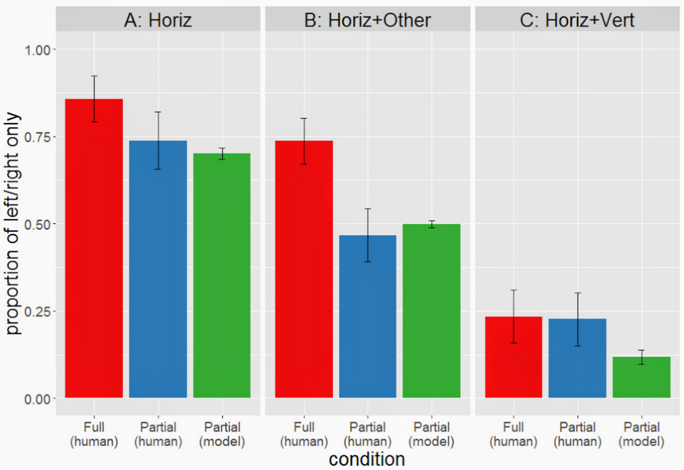

Figure 2: Proportion of responses using *left/right* only (i.e., without any additional spatial expression) produced by participants and predicted by the adaptation model. Errors bars show standard error of the mean.

Procedure

In each trial, the participant was presented with a single spatial array and asked to describe the target so that it could be uniquely identified by the listener. Participant descriptions were typed into a text box; no restrictions were placed on the terms used (i.e., responses were completely free). Following a brief delay, participants were provided feedback about the object selected by the listener (marked with a blue arrow in a copy of the array): this was either the target or the distractor object of the same shape and color.

In the Full knowledge condition, the simulated listener selected the target whenever the description contained at least one word from a large list of spatial expressions (determined through pilot testing). While participants in this condition could in principle have communicated "successfully" by using spatial terms that did not pick out the target, as we report below the vast majority of responses were felicitous and identifying. In the rare instance that a description contained no spatial term, the Full knowledge listener selected the target or the distractor with equal probability.

The simulated listener in the Partial knowledge condition behaved identically except when the description contained *left/right* (or variants such as *leftmost/rightmost*) without any other spatial terms. For such responses, the listener randomly selected the target or matched distractor with equal probability (i.e., the listener responded as if no spatial term was present in the description).

Three practice trials at the beginning of the experiment emphasized the goal of providing complete, unambiguous descriptions (e.g., *yellow circle at the bottom* instead of simply *circle* or *yellow circle* for Fig 1C). Each stimulus array was repeated 4 times over the course of the experiment, resulting in 128 trials per participant. The order in which the arrays appeared was pseudo-randomized by participant, with the constraint that there were an equal number of Horiz trials in each half of the experiment (20 per half). This ensured that a participant who consistently used *left/right* for this array type would receive, on average, at least ten instances of negative feedback during the first half of the experiment.

Results

Descriptions tended to be brief, containing a single spatial expression (average number of words per utterance: 2.83 in the Full condition, 3.13 in the Partial condition). Two manual coders determined that more than 95% of the responses across both conditions were sufficient to uniquely identify the target (given complete knowledge of spatial terms). The rare insufficient responses were produced sporadically across participants (i.e., not concentrated on any particular speaker).

Statistical analysis was performed on the sufficient descriptions. For these, communication errors could occur only in the Partial condition and only when *left/right* was used without any other spatial term. Accordingly, we focused on the way in which the rate of '*left/right*-only' descriptions (i.e., *left/right* alone or combined with only shape or color terms) varied across conditions and array types (see Fig 2).

A mixed-effects logistic regression was performed with *left/right*-only as the dependent variable and fixed factors of condition (Full vs. Partial), array type (Horiz, Horiz+Other, Horiz+Vert), and three experimental 'phases' (with approximately one-third of the total trials in each phase). Phase was included as a rough estimate of the time course of adaptation, which we informally gauged to occur quite rapidly (i.e., after only a few instances of negative feedback). All fixed effects were weighted sum-coded, and the model included random intercepts for participants and stimuli. In light of the number of comparisons involved and the relative novelty of our research question and design, a conservative level for significance was chosen ($p < .01$).

As anticipated earlier, the *left/right*-only rate was higher overall for the Horiz array type ($\beta = 2.24$, $p = .006$), with the rate for the Horiz+Other type effectively equal to the mean across all types ($\beta = 0.05$, $p > .8$). Note that this implies a much lower rate for the Horiz+Vert arrays (as expected from the vertical bias discussed earlier). There were no significant main effects of experimental condition or of phase. However, condition and array type interacted significantly, reflecting the fact that the difference in *left/right*-only rate for Horiz vs. Horiz+Other was much larger in the Partial condition (Condition × Horiz: $\beta = -0.28$, $p < .01$; Condition × Horiz+Other: $\beta = 0.39$, $p = .01$). In the Full condition, the rate of *left/right*-only descriptions was high for both Horiz and Horiz+Other. The Partial condition showed adaptation to the listener for the Horiz+Other array type that did not fully generalize to the Horiz arrays.

Finally, condition and phase interacted significantly (Condition × Phase1: $\beta = -0.5$, $p < .01$). This accords with the numerical finding that *left/right*-only rate increased slightly across phases in the Full condition (phase1: 69%, phase2: 71%, phase3: 73%) but decreased across the first two phases in the Partial condition (phase1: 57%, phase2: 47%, phase3: 42%). The small increase in the Full condition may be due to the sufficiency of *left/right* for all arrays: implicit self-priming may have elevated the frequency of these terms, or participants may have explicitly realized that there was little need to generate alternative expressions. The decline after the first phase in the Partial condition, and indeed the difference in the first phase across the two conditions, indicates that adaptation to the listener occurred rapidly (i.e., within approximately 43 trials, after an average of 12 instances of unsuccessful communication).

In summary, we observed fine-grained adaptation to the addressee with a lexical-semantic gap in the spatial domain. There was little opportunity to find adaptation in the case of Horiz+Vert arrays, given the general bias to use vertical expressions. But for Horiz+Other arrays, participants in the Partial knowledge condition began to avoid *left/right*-only descriptions, responding to listener errors and performing differently than participants in the Full knowledge condition, within the first third of the experiment. Adaptation in the Partial condition was significantly lower for the Horiz arrays, which did not provide an alternative spatial relation that could be used to identify the target.

The preceding analysis does not reveal whether participants in the Partial condition attempted to avoid using *left/right* altogether, or continued to use the problematic terms but supplemented them with additional spatial expressions. For the Horiz+Other arrays, we found that avoidance was the primary strategy. For the Horiz+Vert arrays, we observed more of a tendency to produce redundant descriptions (e.g., *top right* or *bottom left*). However, this tendency was observed in both the Full and Partial conditions, suggesting that it may reflect lexicalization o spatial collocations rather than any adaptation strategy.

A major remaining issue is why the Partial condition participants found it relatively difficult to avoid (or supplement) *left/right* for the Horiz arrays. Replicating Ferrara et al. (2016), we found that when participants did switch to alternative expressions these were mostly ordinal terms such as *first/second/last*. Ordinals would in fact have been sufficient to identify the target in all of the array types, making their relative infrequency as alternatives to *left/right* all the more striking. One hypothesis is that, when exposed to errors on *left/right*, participants (implicitly) concluded that the listener has imperfect understanding of reference to the horizontal axis in general. If correct, this hypothesis would imply that adaptation occurred at the level of spatial relations or axes rather than at the level of spatial expressions.

An alternative hypothesis is that ordinals—in contrast to proximity, containment, etc. terms—are strongly dispreferred relative to *left/right* for the purpose of describing our stimulus items. Under this hypothesis, the difficulty of adaptation for the Horiz arrays in the Partial condition should be mirrored by avoidance of ordinals in the Full condition. More generally, the probability of switching from *left/right* to another spatial expression in the Partial condition may closely track independently-established relative frequencies of terms used to describe our arrays. We formalized this hypothesis in the computational model of spatial language use and adaptation developed below.

Computational model of adaptation

The model has two main components: baseline (or pre-adaptation) preferences for spatial term usage, and a mechanism for modifying the preferences in response to errors made by the listener. Our goal in this paper is not to explain the baseline preferences, but rather to estimate them from empirical usage frequencies. The estimates take the form of numerical (dis)preferences (or 'weights') assigned to various spatial and non-spatial term types (or 'attributes'). Once the baseline has been established, we show that a single inhibition parameter (i.e., a penalty for using *left/right*) suffices to closely match the detailed adaptation pattern of the experiment. A uniform penalty for *left/right* has different effects across the array types because the viable alternative attributes for each type vary independently in their weights.

Baseline preferences

The baseline model assigns probabilities to a large set of array-specific *sufficient descriptions*. Each description contains one or more binary-coded attributes indicating the presence of spatial and other terms. Specifically, the spatial attributes we considered are horizontal (**horiz**: *left/right*), vertical (**vert**: *above/below/up/down/top/bottom*), proximity (**prox**: *close to/next to/near/far/beside*), containment (**cont**: *inside /outside/within*), vertical orientation (**v.o.**: *pointing up/down, facing up/down*), horizontal orientation (**h.o.**: *facing towards/ away from*), and ordinal (**ord**: *first/second/last*). The non-spatial attributes are **shape** (*circle/pencil*) and **color** (*yellow/green*) and. This coding abstracts away from minor syntactic permutations (e.g., *circle to the right* vs. *right circle*) and lexical variation (e.g., *the circle on the right* vs. *rightmost circle*). The set of sufficient descriptions for each array was formed by considering the spatial relations that could be used to identify the target and fully crossing these with one another and all possible shape and color combinations (e.g., Table 1 lists the relations that are relevant for the Horiz+Other arrays).

Relative frequencies of the sufficient descriptions for each array were determined, in part, from the results of the Full knowledge condition above. However, because all targets in that condition could be successfully identified with *left/right*, it is plausible that this data overestimates the frequency of **horiz** (e.g., due to participant self-priming). More generally, we were concerned that Full condition data may provide a somewhat skewed estimate of the relative accessibility of different sufficient descriptions. Therefore, we conducted an additional experiment in which each participant provided up to five descriptions of an array. This experiment was performed by 19 undergraduates at the Johns Hopkins University, each completing 32 trials (one per array) for a small amount of course credit. The total frequency of a description for a given array was equal to the sum of its frequencies in the Full condition and in this experiment.

A conditional log-linear (or maximum entropy) probability distribution over descriptions was defined by assigning a weight w_i to each binary attribute f_i (e.g., Jurafsky & Martin, 2009). The conditioning information was the array, which determines the set of alternative sufficient descriptions. Weights were tied across array types and fit by maximum likelihood to the array-specific description frequencies. The resulting weights were as follows: **horiz** (-1.14), **vert** (-0.43), **prox** (-2.73), **cont** (-2.3), **v.o.** (-2.57), **h.o.** (-5.78), **ord** (-3.06), **shape** (1.0) and **color** (-0.68), where a higher weight indicates a greater preference for the attribute. Note that the model assumes independence of attributes, an idealization that we show to be largely effective but which is not inherent to the maximum entropy formalism.

Modeling adaptation

Prior to experience with the listener, participants in the Partial condition should have the baseline attribute weights. After failed instances of communication with *left/right*, the weights could in principle be modified in various ways (e.g., by large changes after single errors, or much more gradually over the course of the entire experiment). Given the rapid adaptation found in the experiment, and in order to restrict the number of free parameters, we implemented adaptation as a single array- and speaker- independent decrease in the weight of **horiz** subsequent to the first listener error.

The error-driven penalty against **horiz** was fit by maximum likelihood to the Partial data, with the weights of all other attributes fixed at their baseline values. The best-fitting penalty (≈ -0.70) was sufficient to make alternative spatial expressions more probable than *left/right* in the Horiz+Other arrays. However, *left/right* remained the most probable expression for Horiz arrays. Because the penalty was uniform across all array types, this and other asymmetries must reflect the relative frequencies of alternative spatial expressions in the baseline data. In this sense, the model derives the nuanced pattern of audience design in the Partial condition from independently-established usage patterns and a minimal assumption about the mechanism of adaptation (see Fig 2).

Detailed results

We examined the predictions of the model in more detail for the four subtypes of Horiz+Other arrays. Using the weights and **horiz** penalty above, we generated predicted frequencies of the sufficient descriptions for each subtype by sampling responses for 24 simulated participants.

Collapsed over shape and color, the predicted frequencies of the various spatial attributes were highly correlated with the actual frequencies across the subtypes (r=0.96). In particular, for the Horiz+Other arrays offering containment, proximity and vertical orientation as alternatives to the horizontal relation, the experiment revealed that participants adapted by switching to these alternatives in the Partial condition, thereby increasing the frequencies of these features over the horizontal. However, in the arrays where horizontal orientation of the target was available as an alternative, participants continued to use *left/right* for identifying the target. Table 1 shows that the model captured this difference and other variations in attribute frequency.

Table 1: Proportion of responses containing each relevant spatial attribute (in bold) produced by participants in the Partial knowledge condition, and predicted by the adapted model (in parentheses), for the Horiz+Other array subtypes.

Proximity		containment	
horiz	.48±.08 (.53)	**horiz**	.44±.08 (.39)
ord	.20±.08 (.21)	**ord**	.13±.06 (.16)
prox	.21±.06 (.20)	**cont**	.26±.06 (.34)
vertical orientation		horizontal orientation	
horiz	.33±.09 (.45)	**horiz**	.73±.08 (.69)
ord	.17±.07 (.19)	**ord**	.23±.08 (.25)
v.o.	.36±.09 (.28)	**h.o.**	.00 (.02)

Model limitations

While highly successful relative to our original goals, the model contains a number of simplifications that could be addressed in future iterations. The attribute weights and **horiz** penalty were assumed to be identical for all participants (and trials), but there may be substantial individual (or even trial-level) variation in preferences for referential descriptions. The assumption of independent attribute weights was for the most part viable, but some form of interaction is required to account for frequent **vert+horiz** collocations (e.g., *top right*). Our focus was on spatial expressions, but use of shape and color terms would also be of interest, especially when these are used redundantly. The assumption that the inhibition of **horiz** applies after the first communication error, rather than coming into effect more gradually, was also an idealization.

Finally, no attempt was made to predict the baseline preferences for spatial terms or attributes (e.g., the strong preference for *left/right* over *first/last* in the Full knowledge condition). This raises the more general question of what cognitive representations and processes lead speakers to select particular utterances from a set of sufficient referential descriptions, only some aspects of which are due to audience design.

General discussion

In this paper, we experimentally tested whether speakers adapt their language to listeners with a lexical-semantic gap. Such situations may arise commonly, both when experts talk to novices and when adult speakers of a language address second language learners or children. Inspired by previous work with children, we focused on the case in which the listener commands all spatial terms other than *left/right*.

We found that participants were able to rapidly identify the listener's lexical gap, and to avoid it in cases where other alternatives were readily available. Specifically, when the target object could be identified with another spatial relation, participants mostly switched to using that relation. However, adaptation occurred to a lesser extent when the target could only be identified by its horizontal relation. This pattern of results suggests that spatial language elicited in the experiment was shaped by audience design, but that other factors prevented complete adaptation to the listener.

We formalized those factors with a computational model that assigns probabilities to sufficient descriptions with independent attribute weights. The weights were fit to utterances from the Full knowledge condition, supplemented by data in which participants provided multiple descriptions of each array. This model may reflect the endpoint of iterative pragmatic reasoning, as in the RSA framework (Frank & Goodman, 2012), but is closer in practice to the approach of Monroe & Potts (2015), who remedy limitations of that framework by setting attribute weights empirically. Adaptation was then modeled in a simple form, as an error-driven inhibition of *left/right* that applied uniformly to all array types (and participants). Despite its simplicity, the model correctly predicted the different types and degrees of adaptation observed across arrays in the experiment.

While some previous models have addressed adaptation from bottom-up information about the listener (e.g., Janarthanam et al., 2010), none have considered gaps in basic lexical knowledge. Indeed, much work in theoretical and computational pragmatics assumes a generic addressee with the same lexical semantics as the speaker. The model developed here could be applied to other cases in which listeners have idiosyncratic gaps in technical or non-technical vocabulary. Adaptation to the lexical knowledge of the listener is an important aspect of cooperative communication.

References

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482-1493.

Carlson-Radvansky, L. A., & Logan, G. D. (1997). The influence of reference frame selection on spatial template construction. *Journal of Memory and Language*, *37*(3), 411-437.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*, 287-299.

Ferrara, K., Silva, M., Wilson, C., & Landau, B. (2016). Spatial language and the Embedded Listener Model in parents' input to children. *Cognitive Science*, *40*, 1877-1910.

Fitneva, S. A., & Song, Y. (2009). The comprehension of "left" and "right" in a referential communication task. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2687-2691). Austin, TX: Cognitive Science Society.

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

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*(3), 378-391.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26-37.

Janarthanam, S., & Lemon, O. (2010, September). Adaptive referring expression generation in spoken dialogue systems: Evaluation with real users. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 124-131). Association for Computational Linguistics.

Jurafsky, D., & Martin, J. (2009). *Speech and Language Processing : An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Upper Saddle River: Pearson Prentice Hall.

Kutlak, R., van Deemter, K., & Mellish, C. (2016). Production of referring expressions for an unknown audience: A computational model of communal common ground. *Frontiers in Psychology*, *7*: 1275.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, *28*(2), 103-174.

Lombard, É. (1911). Le signe de l'elevation de la voix. *Annales Maladies de l'Oreille et du Larynx*, *37*, 101-119.

Monroe, W., & Potts, C. (2015). Learning in the rational speech acts model. arXiv preprint arXiv:1510.06807.

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, *78*, 1-17.

Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. In K.R. Coventry, T. Tenbrink, & J.A. Bateman (Eds.), *Spatial Language and Dialogue* (pp. 23-39). Oxford: Oxford University Press.

Sholl, M. J., & Egeth, H. E. (1981). Right-left confusion in the adult: A verbal labeling effect. *Memory & Cognition*, *9*(4), 339-350.

Wu, S., & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive Science*, *31*(1), 169-181.