Incremental processing in the pragmatic interpretation of contrastive prosody

Chigusa Kurumada
kurumada@stanford.edu
Dpt. of Linguistics
Stanford University

Meredith Brown
mbrown@bcs.rochester.edu
Dpt. of Brain and Cognitive Sciences
University of Rochester

Sarah Bibyk
sbibyk@bcs.rochester.edu
Dpt. of Brain and Cognitive Sciences
University of Rochester

Daniel F. Pontillo
dpontillo@bcs.rochester.edu
Dpt. of Brain and Cognitive Sciences
University of Rochester

Michael K. Tanenhaus
mtan@bcs.rochester.edu
Dpt. of Brain and Cognitive Sciences
University of Rochester

Abstract

We present an eye-tracking experiment investigating the time course with which listeners derive pragmatic inferences from contextual information. We used as a test case the construction "It looks like an X" pronounced either with (a) a nuclear pitch accent on the final noun, or (b) a contrastive L+H* pitch accent and a rising boundary tone, a contour that can support a complex contrastive inference (e.g., It *LOOKS* like a zebra...(but it is not)). The contrastive intonational contour elicited higher proportions of fixations to non-prototypical target pictures (e.g., a zebra-like animal) during the earliest moments of processing the target noun. Further, when the display only contained a single related pair of pictures, effects of the contrastive accent on "looks" emerged prior to the target noun, indicating that efficient referential resolution is supported by rapidly generated inferences based on visual and prosodic context.

Keywords: Prosody, contrastive accent, pragmatic inferences, eye-tracking.

Introduction

Few, if any, would question the claim that addressees must make use of context to infer the intentions of a speaker (speaker meaning). Herb Clark (1992) gives a lovely example to illustrate the richness of context-based inferences. Clark describes a situation in which he addressed the utterance, "I'm hot", to his school-age son, Damon. After going through the plausible pre-compiled senses, Clark notes that none captures his intended (and immediately understood) meaning of his utterance, which could only be inferred from the specific context. Herb and Damon were playing poker and Damon was about to make a large bet. Herb was warning Damon that he should think twice about it.

Despite countless everyday examples of this sort, there is also a widely held view that pragmatic inference is external to the core mechanism of language comprehension. For example, this assumption underlies Levinson's (2000) influential proposal that common inferences might be pre-compiled as automatically generated defaults, by-passing the need for making a slow and resource intensive inferences (e.g., Neely, 1977; Posner & Snyder, 1975; Shiffrin & Schneider, 1977). This idea receives support from the hypothesis that the remarkable speed and ease of real-time language processing is possible, in part, because of its modularity in the processing system. A syntactic module, for example, performs computations on restricted inputs without appealing to slow and resource-demanding processes, such as inference (e.g., Fodor, 1983).

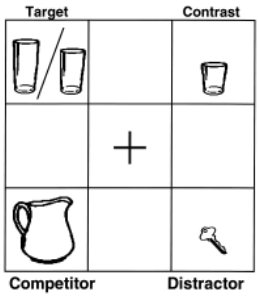


Figure 1: A sample visual display used in Sedivy et al. (1999) for an instruction "Pick up the tall glass"

This modularity hypothesis, however, lacks an explanation for cases in which expectations based on context can effectively constrain parsing decisions. In fact, there is now a large body of research demonstrating that listeners rapidly use information from the linguistic and visual context to resolve ambiguity (e.g., Altmann 1998; Chambers, Tanenhaus & Magnuson, 2004; Snedeker & Trueswell, 2003). In this constraint-based approach, the context of language use is integral to effective and incremental language processing in guiding expectations (e.g., Levy, 2008). Furthermore, Piantadosi, Tily and Gibson (2012) propose that inherent ambiguity in the linguistic signal is in fact a design feature of an efficient encoding system, given the assumption that listeners can integrate context information to inferentially resolve ambiguity.

Consistent with these accounts, a number of studies using online measures have shown that listeners can, and do, incorporate visual information to process linguistic input incrementally. For example, Sedivy et al. (1999) examined listeners' processing of prenominal adjectives during incremental language processing. They asked participants to manipulate objects based on spoken instructions such as "Pick up the tall glass". In Figure 1, the pitcher on the lower left is the tallest object, but the glass on the upper left is both tall by comparison to glasses in general, and taller than the other glass in the upper right-hand corner. Sedivy et al. found that the partial instruction "Pick up the tall —" elicited fixations to the tall member of the contrast pair (e.g., the tall glass) rather than

the other tall object (e.g., the pitcher) in the display. This suggests that listeners rapidly integrate context-specific contrast information to begin resolving referential ambiguity prior to the head noun.

Nonetheless, it remains to be understood how readily listeners can derive more complex inferences such as conversational implicature. For example, some experimental studies on the English quantifier *some* (but not all) have concluded that even the basic scalar implicature is indeed slow and costly, compared with computing its logical meanings (i.e., at least one, possibly all) (e.g., Bott & Novek, 2004; Huang & Snedeker, 2009). On the other hand, there is a recent body of work (e.g., Grodner et al., 2010, Degen & Tanenhaus, under review) suggesting that delays arise only when use of *some* in the particular context is less natural than another rapidly available alternative.

In this current study, we approach this problem by examining the time course of English speakers' comprehension of contrastive prosody. In English, the pitch accent L+H* is known to evoke an alternative set of referents and invites a contrastive inference (e.g., Katie did not win a $TRUCK_{L+H*}$ (but won a motorcycle), Ito & Speer, 2008). Previous work has found that the use of L+H* in an appropriate discourse context restricts the domain of reference during incremental language comprehension. For instance, the L+H* in "Give me the red ball. Now give me the $GREEN_{L+H*}$—" triggers anticipatory eye-movements to a green object of the same type as the preceding referent (i.e., a green ball).

While this contrast-evoking function of L+H* is known to be robust (Weber et al., 2006), previous experimental work has almost exclusively focused on prenominal adjectives highlighting color or size contrast. Moreover, studies so far have found incremental processing of contrastive prosody only when a member of the relevant contrast set was linguistically mentioned in prior discourse. These limitations make it difficult to scale up previous findings to cases where contrastive accent triggers complex, and hence allegedly costly, conversational implicatures.

To address this, we used a different linguistic construction, "It looks like an X", which can support two opposing pragmatic interpretations depending on its prosodic realization. A canonical declarative prosodic contour, with a nuclear pitch accent on the final noun (as illustrated in Figure 2, left panel, henceforth *Noun-focus prosody*), typically elicits an affirmative interpretation (e.g., *It looks like a zebra and I think it is one*). When the verb "looks" is lengthened and emphasized with a contrastive L+H* accent and the utterance ends with a rising L-H% boundary tone (Figure 2, right, *Verb-focus prosody*), it can trigger a negative interpretation (e.g., *It LOOKS like a zebra but its actually not one* (Kurumada, Brown, & Tanenhaus, 2012).

In the current study, we tested if and how the listeners develop the two different interpretations as they receive prosodic information. Specifically, we asked the following questions:

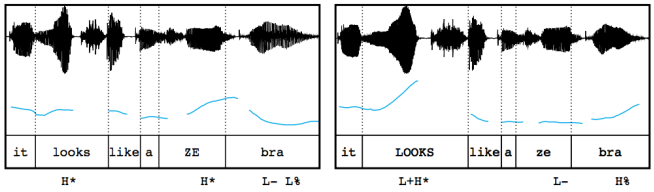


Figure 2: Examples of Noun-focus prosody (left) and Verb-focus prosody (right).

1. Can listeners integrate visually represented contrasts with prosodic information to guide pragmatic interpretation?

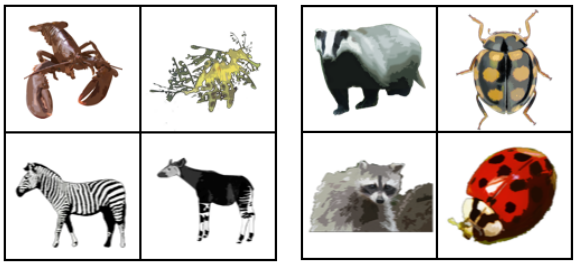
2. Do listeners process intonational contours and develop pragmatic expectations incrementally?

Experiment overview

We examined the time course of pragmatic intonation interpretation using the visual world paradigm (Tanenhaus et al., 1995). Participants listened to the construction "It looks like an X" produced with either Noun-focus or Verb-focus prosody, and they were asked to click on the corresponding referent in a four picture display. In each display, there was at least one pair of visually similar items (e.g., a zebra and an okapi; Figure 3-a, bottom row). We assumed based on previous work that Noun-focus prosody would bias responses toward the more prototypical member of each pair (e.g., a zebra for "It looks like a ZEBRA"), while Verb-focus prosody would bias responses toward the less prototypical member (e.g., an okapi for "It *LOOKS* like a zebra"). Thus, our first hypothesis is that listeners should integrate the contrasting relation between the prototypical and non-prototypical target pictures in their interpretation of the utterance intonation.

A previous study has shown that listeners can develop a similar contrastive inference in a visual search task (Dennison & Schafer, 2010). Their study used the construction "X *HAD* Y" (e.g., "Lisa *HAD* a bell..." (but she no longer has one)), but they found no evidence of incremental processing. They proposed that the contrastive inference requires both the pitch accent and the boundary tone, and hence occurs only after the sentence offset.

In the current study, we hypothesize that listeners can compute an implicature incrementally, based on the prosodic and visual context. We tested this hypothesis by comparing the time course of eye movements in a display with a single pair of contrasting items, to those in a display with two pairs. In the one-contrast display, we predicted that participants would be able to use the contrastive pitch accent to infer that the likely referent is a member of the contrast pair (more specifically, the less prototypical member) prior to the processing of the target word. In the two-contrast display, the target referent cannot be determined until it has been explicitly mentioned, which should result in effects of prosody emerging later, i.e., during the processing of the target word.



(a) 1-contrast condition (b) 2-contrast condition

Figure 3: Sample visual displays

Methods

Participants

Twenty-four students from University of Rochester were paid ($10) to take part in the experiment. They were native speakers of American English with normal or corrected-to-normal vision and normal hearing.

Stimuli

We selected 16 imageable high-frequency nouns and embedded them in the sentence frame "It looks like an X". A native speaker of American English recorded two tokens of each item with the Noun-focus and the Verb-focus prosodic patterns. The same speaker also recorded 44 filler items in which a target referent was described (e.g., "Can you find the one with white fur?").

We selected 16 more items to form pairs with the 16 target nouns. In each pair, the items were visually similar to each other (e.g., a zebra and an okapi) and one item (e.g., a zebra) was always more common. Hereafter, the picture from each pair that is more common (e.g., a zebra, Figure 3, bottom left) is referred to as the *prototypical* target picture, and the other (e.g., an okapi, Figure 3, bottom left) is referred to as the *non-prototypical* target picture.

Prototypicality and nameability norming To create visual stimuli, we ran two types of norming studies using Amazon Mechanical Turk, an online crowd-sourcing platform. In the first study, 40 subjects provided names and nameability ratings (on a seven-point rating scale) for each of the 240 images. In a second norming study, we presented 40 subjects with the images along with a label and collected ratings of referential fit for both adult-directed speech and, as a separate response, child-directed speech. The non-prototypical pictures (e.g., okapi) were always presented with the names of their respective prototypical items (e.g., zebra) in order to establish an empirical measure of prototypicality.

Based on this information we constructed 60 visual scenes (16 critical trials and 44 filler trials). Each of the scenes consisted of four pictures including one pair of target pictures described in an auditory stimulus. We created two types of visual scenes: a) 1 target pair + 2 singletons (*1-contrast condition*), and b) 1 target pair and 1 distractor pair (*2-contrast condition*) (Figure 3). Singletons in 1-contrast trials consisted of one easily nameable picture and one less-nameable picture to equate the complexity of the visual display across trials.

Procedure

Participants were first presented with a cover story in which a mother and a child are looking at a picture book. The mother is helping the child to identify various objects and animals by verbally commenting on them. Each trial began with the presentation of a visual display containing four pictures. After 1 second of display preview, participants heard a spoken sentence over Sennheiser HD 570 headphones and clicked on a picture described by the sentence. Their mouse-clickng responses were collected while their eye movements were tracked using a head-mounted SR Research EyeLink II system sampling at 250 Hz, with drift correction procedures performed after every fifth trial.

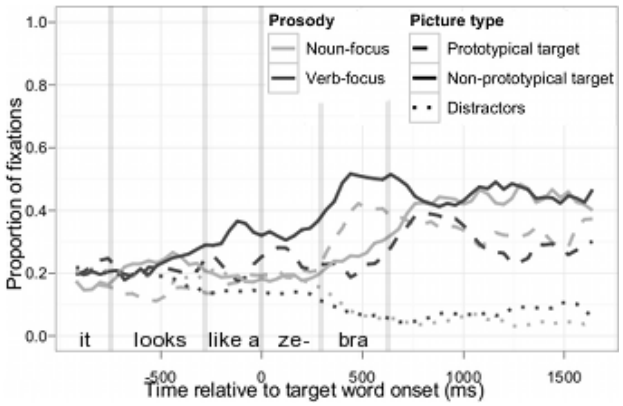
Eight lists were constructed by crossing the 1) item presentation order, 2) the location of the prototypical and the non-prototypical items on the display, and 3) the prosodic contour (Noun-focus vs. Verb-focus). All lists started with three example items to familiarize participants with the task. The mean duration of the experiment was 12 minutes.

Results and discussion

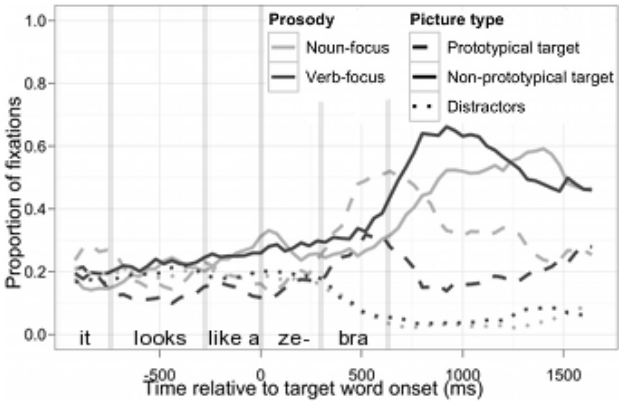
We analyzed three dependent measures to obtain converging evidence about the role of prosody and visual display characteristics in the processing of critical items: response choices in the picture selection task, response times, and proportions of fixations to different alternatives within the display. Each variable was assessed with multi-level generalized linear regression models implemented using the lmer function within the lme4 package in R (R Development Core Team, 2010; Bates et al. 2008)[1].

We first confirmed that participants selected a correct target picture in 96% of filler trials, indicating that participants did not have difficulty completing or attending to the picture selection task. We then analyzed their responses in the 16 critical trials to ask if they encoded the visual contrasts of items on the screen and associated them with the two prosodic contours. Participants selected the prototypical target picture 65.6% of the time in the Noun-focus prosody condition, but only 25.5% of the time in the Verb-focus prosody condition. A multilevel logistic regression model of responses confirmed that depending on the prosodic contour, participants reliably chose either a prototypical or a non-prototypical item

[1]Logistic regression models of response choices were fit by the Laplace approximation, whereas linear regression models of response times and fixation proportions were fit using restricted maximum likelihood estimation. Fixed effects included prosody condition (Noun- vs. Verb-focus), display type (one- vs. two sets of related pictures), and standardized trial number. Analyses of fixation proportions additionally included picture type. We also included random by-subject and item intercepts as well as slopes for the interaction between prosody condition and picture type. To minimize the risk of over-fitting the data, fixed effects were removed stepwise and each smaller model was compared to the more complex model using the likelihood ratio test (Baayen, Davidson, & Bates, 2008).



(a) 1-contrast condition



(b) 2-contrast condition

Figure 4: Proportions of fixation to pictures in response to the Noun-focus (solid line) and to the Verb-focus (dashed line). The lines are aligned at the onset of the final noun.

($\beta = 6.37$, $z = 4.394$, $p < .0001$). Thus, without any explicit mention of a contrasted item, participants encoded a relevant contrast set in the visual field and developed a contrastive inference based on the Verb-focus prosody.

Response times indicated that Verb-focus prosody elicited slower responses (mean RT=2204 ms) than Noun-focus prosody (mean RT=1969 ms, $\beta = -.242$, $t = -2.09$,$p < .05$). However, the effect of prosody was dependent on whether the prototypical or non-prototypical target picture was selected ($\beta = .509$, $t = 2.94$, $p < .005$). On trials with Noun-focus prosody, RTs were significantly faster when a prototypical picture was selected (mean RT=1762 ms) than when a non-prototypical picture was chosen (mean RT=2364 ms, $\beta = -.272$, $t = -3.20$, $p < .005$). On trials with Verb-focus prosody, however, there was a numerical trend in the opposite direction (mean RT=2540 ms for prototypical target responses vs. 2089 ms for non-prototypical target responses,($\beta = .201$, $t = 1.10$, $p > .10$). This finding suggests that responses deviating from the expected association between prosody and picture type were associated with greater processing difficulty, further supporting the hypothesis that participants were interpreting the prosodic contour based on the visual contrasts.

Next we analysed the eye-tracking data to examine the time course of processing the contrastive prosody. Our analysis focused on two regions, which were both defined with respect to the point at which segmental information from the target word would be expected to influence processing. The first region, which we termed the *pre-target region*, was defined as the region beginning at 200 ms after the offset of "looks" and ending at 200 ms after the onset of the target word. This roughly corresponds to the region indicated with the caption "like a" in Figure 4, shifted to the right by 200ms. Because it takes approximately 200 ms to program and execute an eye movement, fixations within this window should not be influenced by segmental information from the target word. The only information that would be expected to influence eye movements within this window, then, is information from preceding prosody (e.g. the contrastive accent on "looks").

The second region, the *early target-word region*, began at 200 ms after target word onset and ended at 200 ms after the offset of the first syllable of the target word. This roughly corresponds to the region indicated with the caption "ze-" in Figure 4, shifted to the right by 200ms. Fixations within this window were expected to reflect the integration of expectations based on preceding prosody and initial effects of incrementally presented target word information. Within each window, mean proportions of fixations to each picture were calculated and then transformed using the empirical logit function (Cox, 1970) for the purposes of linear regression analysis.

Pre-target fixations We analyzed logit-transformed proportions of fixations averaged across the pre-target region in two linear mixed-effects regression models. The first model examined effects of prosody contour, display type (i.e., one- vs. two contrast sets), and trial number on logit-transformed mean proportions of fixations to the distractor pictures vs. either member of the target contrast set (e.g. the zebra and okapi). The goal of this analysis was to assess prosody- and display-wise differences in anticipating the target contrast set. We predicted that Verb-focus prosody would bias participants to fixate members of the target contrast set in one-contrast trials but not in two-contrast trials.

Results from the regression analysis revealed that the predicted three-way interaction between prosody condition, picture type, and display type was significant ($\beta = .754$, $t = 1.98$, $p < .05$). Analyzing proportions of fixations by display type revealed that effects of prosody condition were dependent on picture type in one-contrast trials ($\beta = -.322$, $t = -2.61$, $p < .01$). Participants were no more likely in the Noun-focus condition to fixate the target picture (mean untransformed proportion of fixations=.209) and the distractor pictures (mean=.186, $\beta = .007$, $t = .068$, $p > .1$), but they exhibited a significant bias toward the target contrast set in response to Verb-focus prosody (mean=.245 vs. .167, $\beta = -.315$, $t = -3.34$, $p < .001$). This interaction was not sig-

nificant in two-contrast trials ($\beta = -.058, t = -.529, p > .1$).

The second linear mixed-effects regression model examined effects of prosody condition, display type, and trial number on logit-transformed mean fixation proportions to prototypical vs. non-prototypical target pictures. We predicted that the difference between fixations to non-prototypical pictures and fixations to prototypical pictures would be greater in response to Verb-focus prosody than Noun-focus prosody. Indeed, the regression model revealed a marginal two-way interaction between prosody condition and picture type ($\beta = .138$, $t = 1.71$, $p = .087$). In the Noun-focus prosody condition, fixations to bad target pictures (mean=.238) did not differ significantly from fixations to good target pictures (mean=.181, $\beta = -.124$, $t = -1.20$, $p > .1$). In the Verb-focus condition, however, participants were significantly more likely to fixate the bad target picture (mean=.300 vs. .190, $\beta = -.243, t = -2.69, p < .01$).

Taken together, these findings suggest that participants rapidly encode the visual attributes of and relations between potential referents, and rapidly integrate this visual information with the incoming prosodic input. When a single contrast set is present in the display, contrastive Verb-focus prosody biases listeners to fixate members of that set. This trend is illustrated in Figure 4. In the 1-contrast condition (Figure 4-a), the fixation proportions to the non-prototypical target based on the Verb-focus prosody begins to diverge in the pre-target region. On the other hand, such divergence is delayed in the 2-contrast condition (Figure 4-b).

Early target-word fixations For the early target-word region, we again analyzed logit-transformed mean proportions of fixations in two linear mixed-effects regression models, to compare effects of prosody condition, display type, and trial number on (a) target-picture fixations to distractor fixations, and (b) prototypical target-picture fixations to non-prototypical target-picture fixations.

For the analysis comparing logit-transformed mean proportions of fixations to target pictures vs. distractor pictures, we predicted that the three-way interaction between prosody condition, picture type, and display type would no longer be significant. Instead, the main prediction was that fixations to both target pictures would be significantly higher than fixations to distractors across trial types.

The results of the analysis indicated that neither display type nor prosody condition accounted for a significant proportion of variance in target vs. distractor fixations in the early target-word region. Instead, the main finding was that participants were significantly more likely to fixate target pictures (mean=.280) than distractor pictures (mean=.127, $\beta = -.240$, $t = -4.12$, $p < .0001$), reflecting their early use of incoming segmental information to restrict the referential domain to the two target pictures.

The analysis of fixations to the two target pictures was predicted to show that the difference between non-prototypical target picture fixations and prototypical target picture fixations would continue to be greater in the Verb-focus condition. In addition, display type was no longer predicted to significantly influence patterns of fixations, since the segmental information from the initial sounds of the target word should restrict the domain of reference to the target contrast set.

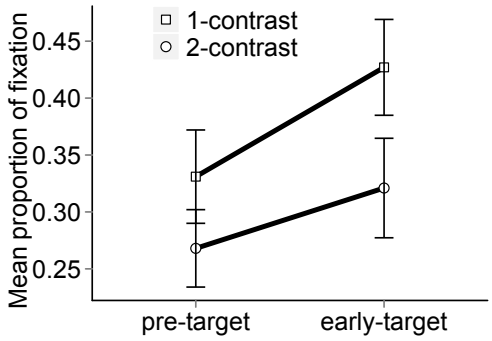


Figure 5: Mean fixation proportions to the non-prototypical item in response to the Verb-focus prosody. Error bars represent standard errors.

The second linear mixed-effects regression model indeed revealed a two-way interaction between prosody condition and picture type ($\beta = -.472$, $t = -4.02$, $p < .0001$). In the Verb-focus condition, participants were significantly more likely to fixate non-prototypical target pictures (mean=.374) than prototypical target pictures (mean=.222, $\beta = -.335$, $t = -3.97$, $p < .0001$). In the Noun-focus condition, however, there was a non-significant trend in the opposite direction, with more fixations to prototypical target pictures (mean=.293) than non-prototypical target pictures (mean=.231, $\beta = .138$, $t = 1.49$,$p > .1$). This interaction between prosody condition and picture type suggests that listeners rapidly integrated incoming segmental information from the target word with their pragmatic expectations for a prototypical vs. non-prototypical referent based on preceding prosodic information.

Figure 5 summarizes mean fixation proportions to the nonprotptypical item based on the Verb-focus prosody. Within the pre-target region, participants were looking at the non-prototypical item more when they were in the 1-contrast condition than in the 2-contrast condition. This trend was even more magnified when the segmentatal information of the target noun becomes available. This demonstrates that the contrastive pitch accent was processed incrementally under the constraints of the visual context.

Conclusion

The results show that participants generated complex pragmatic interpretations in an incremental manner. In a context with only one contrast pair, listeners began to launch eye movements to a less prototypical target picture even before segmental cues to the final noun become available. This is of particular interest because, unlike in previous studies, the contrastive accent in the current study was used with the verb.

The contrast was not simply based on individual visual features of objects (e.g., color or size); rather, it was mediated by the implicature based on different predicates. Namely, "It LOOKS like an X" is contrasted with "It IS an X", and therefore interpreted as "It is not an X". Our results demonstrate that such complex pragmatic reasoning can develop online.

The results also highlighted the facilitative roles of visual context in intonation interpretation. The early timing of the prosody effect in the 1-contrast condition suggests that listeners made use of visually represented contrast to guide their inference. This enabled us to demonstrate that inferences based on contrastive prosody do not require explicit previous mention of a contrasting item and can be made incrementally on the basis of partial prosodic contour as well as visual information. These findings together advance our knowledge about the remarkably rapid and robust inferential mechanisms supporting online language comprehension and pragmatic communication.

Acknowledgements

Thanks to Eve V. Clark, Christine Gunlogson and T. Florian Jaeger for valuable discussion, and to Dana Subik for support with participant testing. This research was supported by the NICHD grant HD27206 (MKT).

References

Altmann, G.T.M. (1998). Ambiguity in Sentence Processing. Trends in Cognitive Sciences, 2, 146–152.

Baayen, R., Davidson, D., & Bates. D., (2008). Mixedeffects modeling with crossed random effects for subjects and items. *JML, 59,* 390–412.

Bates, D.M., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes.

Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. JML, 51, 437–457.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *JEP:LMC, 30,* 687 –696.

Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Cox, D.R. (1970). *The analysis of binary data*. London: Methuen.

Degen, J., & Tanenhaus, M. K. (under review). Processing scalar implicature: A constraint-based approach.

Dennison, H. Y., & Schafer, A. (2010). Online construction of implicature through contrastive prosody. *Speech prosody 2010 conference.*

Fodor, J. A. (1983).*The modularity of mind: An essay on faculty psychology*. MIT Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). Some, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition, 116,* 42–55.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. *Cognitive Psychology, 58,* 376–415.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *JML, 58,* 541–573.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Prosody and pragmatic inference: It looks like speech adaptation. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

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

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. JEP: G., 106, 226–254.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122,* 280 –91.

Pierrehumbert, J. & Hirschberg, J. (1990) The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack (eds), *Intentions and plans in communication and ciscourse* (pp. 271–311). MIT Press.

Posner, M. I., & Snyder, C. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (pp. 669–682). New York: Academic Press.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557–580.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109–147.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. *Psychological Review, 84,* 127–190.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. JML, 48, 103–130.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49,* 367–392.