Referential domains and the interpretation of referring expressions in interactive conversation

Sarah Brown-Schmidt
Department of Brain and Cognitive Sciences
University of Rochester
sschmidt@bcs.rochester.edu

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

Abstract

This paper describes research investigating the on-line production and interpretation of referring expressions during interactive conversation. In particular, we focus on the interactive processes by which interlocutors establish shared referential domains. In a set of interactive, task-based dialogs, we show that referential domains constrain both the form of referring expressions, and their interpretation. We argue that various task-based factors strongly affect the referential domains used by interlocutors, and that understanding the mechanisms of reference interpretation will require a careful analysis of how these factors affect the referential domains used in interactive conversation.

1 Introduction

Although the generation and interpretation of definite reference has played a central role in real-time sentence processing research, little is known about how addressees interpret referring expressions on-line in interactive conversation. Much of the existing literature investigating the real time interpretation of referring expressions in spoken language comprehension focuses on the interpretation of noun phrases such as "the cube" in sentences like "Put the cube in the can". These sentences are embedded in tasks in which participants are instructed to manipulate a set of objects which are placed on a table in front of them. The instructions are typically pre-recorded, and the referential domain for interpreting the referring expressions is assumed to be the entire workspace. The experimental situations are typically non-interactive, in that the subject simply follows instructions, and does not converse with another person. Research in these constrained contexts suggests that addresses use multiple sources of information to restrict the domain of interpretation of referring expressions, including common ground (Hanna, Tanenhaus & Trueswell, in press), verb -based constraints (Altmann & Kamide, 1999), task relevant properties of objects (Chambers, et al. 2002) and contrast implied by use of a scalar adjective (Sedivy, et al. 1999; 2003).

The findings from constrained contexts indicate that pragmatic factors particular to the context in which a reference is uttered, are key to understanding how that reference is interpreted. However, detailed analyses of how these factors might arise during a conversation are less well understood. For example, the referential domain at the beginning of each instruction in a standard task is generally assumed to be the set of experimental items placed in front of the subject. However, in a natural discourse context, it is possible that the referential domain could include other objects in the room, such as the items on shelves, or that the referential domain could include only a subset of the experimental items. While experiments in constrained situations sug-

gest that linguistic and non linguistic factors both act to constrain this initial referential domain, it is not well understood how these factors are used during conversation.

We present data from two experiments that investigated the production and interpretation of referring expressions in an interactive task-based dialog between two naive participants. The first experiment shows that referential domains can be quite restricted and closely aligned between interlocutors. Speakers frequently used referential expressions that would be ambiguous if the domain were less restricted and addressees were not confused by these expressions, indicating that these potential entities were never considered as potential referents. We suggest that these effects result from domains becoming restricted and co-ordinated because of task-based factors. In the second experiment, we verified this observation, investigating the role of explicitly mentioning the referential domain before the onset of the referring expression. As in the first experiment, we found that when the referential domain was sufficiently restricted, listeners quickly interpreted the referring expressions without interference from other competing referents that were outside the domain.

2   Method

In both experiments, two naïve participants engaged in a referential communication task (Krauss & Weinheimer, 1966) in which they worked together to complete a task. The specific details of each experiment will be described in more detail below. In both tasks, the participants could not see one another, and were working with game pieces on physically separate, but matching workspaces. For both tasks, participants needed to instruct each other to move game pieces in order to successfully complete the task. We did not place restrictions on the way in which the participants spoke to one another. However, the characteristics of the task and the game pieces allowed us to investigate hypotheses about the interpretation of referring expressions, through naturally arising utterances. We employed a version of the visual-world eye-tracking methodology (Tanenhaus, et al., 1995) in which we obtained a record of one subject's eye-fixations with the use of a light-weight head-mounted eye-tracker.

Previous work using the visual-world eye-tracking methodology demonstrates that listener's fixations are closely time-locked to speech input. For example, in a task where a subject is asked to "Put the apple next to the frog", approximately 200ms following the onset of the word "apple", participants are more likely to look at the apple, than other unrelated objects in the scene (such as a can). A related finding was reported by Allopenna, et al. (1998). When participants hear an instruction such as "Click on the cloud" when viewing a computer screen which has pictures of a cloud, a clown, a dog, and a parrot, listeners are equally likely to look at the clown and the cloud upon hearing the onset of "cloud". This effect is due to the fact that "cloud" and "clown" begin with the same sequence of phonemes. When participants hear the disambiguating sounds in "cloud", they reliably look to the correct referent. This effect is commonly referred to as a "cohort effect", and words like "cloud" and "clown" are often referred to as cohort competitors in the spoken word recognition literature.

Our experimental methodology is partially based on the cohort effect. In designing our experiments, we used some game pieces which had pictures on them. We carefully selected easily nameable pairs of pictures that were cohort competitors, such as "cloud" and "clown"). Most of the pictures were selected from a database of normed pictures (Snodgrass & Vanderwart, 1980) and were easily recognized by our participants. Because the task required participants to refer to the game pieces, we expected to observe cohort effects during references to blocks with cohort competitors. Presumably in the Allopenna, et al. (1998) study, all four items pictured on the computer screen were included in the referential domain used by the listener. We predicted that if the referential domain was significantly restricted, that in some cases the target referent and the cohort competitor would be in different referential domains. In these situations, we expected that upon the reference to the target, the proportion of looks to the cohort competitor would be significantly reduced. By tracking the presence of cohort effect, we are able to gauge the size of the referential domain.

3   Experiment 1

In experiment 1, we monitored eye movements as pairs of participants, separated by a curtain, worked together to arrange blocks in matching configurations and confirm those configurations. We reported a more comprehensive analysis of this dataset in Brown-Schmidt, Campana & Tanenhaus (in press). Here we focus on the aspects of the data related to cohort effects and the circumscription of referential domains. During the task, participants placed 56 different blocks over the course of 2.5 hours. All 4 pairs of participants developed idiosyncratic ways of referring to the objects, and also developed strategies for completing the task. A popular strategy, for example, was to finish placing blocks in one area of the workspace before moving on to the next. Additionally, the partners tended to move from one area to an adjacent area, suggesting they had a preference to build off of structure they had already created.

Over the course of the experiment, each pair generated approximately 75 references to blocks with cohort competitors, like cloud and clown. While cohort competitors were only placed 3.5 inches apart, during the course of the conversation we did not observe a cohort effect. Upon hearing the word "cloud", listeners looked primarily at the target referent (the block with a picture of a cloud on it) and were no more likely to look at the clown than at an object in the scene with a completely unrelated name, such a penguin. This observation suggests that during the conversation, the cohort competitors were not included in the referential domain. However, we *did* observe a cohort effect during instructions which were not constrained by the task-related conversation. Periodically, participants needed to remove the eye-tracker to take a break. On one occasion when we put the tracker back on and recalibrated, we tested the calibration by asking the subject to look at different items on the board, using instructions like "Look at cloud, look at the lamb, look at the seal." Here we saw clear cases of the subject initially looking at the cohort competitor (e.g. clown, lamp) before looking at the intended referent (e.g. cloud, lamb). While the cohort effect appears large, the 15 trials of this

sort did not give us enough statistical power to replicate a standard cohort effect, but the pattern of fixations and mean differences between cohorts and targets are similar to those found in Allopenna et al. (1998).

These results suggest that listeners can use tightly circumscribed referential domains during reference interpretation imbedded in a dialog. Unlike the studies using more constrained contexts, the referential domain did not include all of the objects in the participants view- in some cases this would be a large number of blocks. Instead, it appears that strategies which partners mutually developed in order to complete the task, facilitated the use of small, task-relevant referential domains. These observations supported the primary result from this experiment which was that speakers tended not to modify a noun phrase, e.g., saying "the red block" rather than "the **vertical** red block" even when there was more than one red block in the scene. The situations under which speakers *did* choose to modify noun phrases was when the second red block was physically close to the intended referent and it fit the task constraints. When an unmodified NP was used, addressees' eye movements were primarily restricted to the intended referent, suggesting that non-linguistic factors guided the interpretation of these linguistically ambiguous references.

4   Experiment 2

In experiment 2, we created conditions where cohort competitors were more or less likely to be in the same referential domain. Conversational partners took turns instructing one another to click on objects on a computer screen as we monitored the eye movements of one partner and the speech of both partners. On each trial, each participant's screen contained an identical set of 14 pictures, separated into two domains which looked like 'islands'. At the beginning of each trial, a picture on one participant's screen became highlighted. This was a cue to tell their partner to click on this object. Participants were encouraged to speak freely in order to perform the task and no restrictions were placed on how they chose to describe any of the objects. Target objects always appeared with a cohort competitor and we ma-

nipulated whether the cohort appeared on the same or different island as the competitor. We predicted that in cases where the speaker specified the location of the target (e.g. "on the top island") *before* the onset of the referring expression, that this would establish that island as the appropriate referential domain. If the cohort competitor were on a different island than the target, and if the speaker chose to specify the location information before the noun phrase, then we predicted the cohort effect would be eliminated.

When the cohort competitor appeared on the same island as the target, we observed a standard cohort effect, replicating previous findings using pre-recorded instructions (Allopenna, et al. 1998). Approximately 52% of the time, participants specified which island the target was on before the onset of the noun phrase. In these constructions, when the cohort was on a different island than the target, the cohort effect was eliminated, suggesting that specification of the referential domain restricts attention to entities within that referential domain.

The results from this experiment suggest that our subject's explicit (and unscripted) establishment of the referential domain successfully constrained the interpretation of a subsequent referring expressions. We also observed that speakers tended to explicitly mention when the referential domain would *change*. On each trial, the speaker referred to two different objects. Half of the time, the second object was on a different island than the first. When the second object switched islands, speakers were more likely to explicitly ground which island the second referent was in. This strategy was likely to be helpful to listeners (we are currently analyzing the data to find out). Additionally, this adds support to the observation from Experiment 1 that participants tend to work on the task in a highly localized manner, only moving to a new area of the workspace when the previous area has been completed. We are interested in exploring whether these tendencies are specific to the kinds of tasks we selected, or are related to more general properties of discourse and expectancy for upcoming reference.

5.   Discussion

By combining the cohort competition effect, well documented in the word recognition literature, with a referential communication task, we were able to observe how participants with shared task-goals circumscribed referential domains. We found that referential expressions were interpreted with respect to a restricted referential domain, and that these referential domains were closely aligned between conversational interlocutors. These results replicate and extend previous studies demonstrating that referential domains are constrained by contextual and pragmatic factors (Chambers, et al, 2002; Hanna, et al. in press; Hanna & Tanenhaus, in press). Our results also demonstrate that it is possible to study real-time language processing in interactive conversation with the same precision as is typically achieved in controlled laboratory settings with scripted, pre-recorded language. We expect that a satisfactory understanding of the mechanisms of reference interpretation will require addressing the many factors that affect referential domains during interactive conversation.

Acknowledgement

References

Allopenna, P.D., Magnuson, J.S. & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419-439.

Altmann, G.T.M. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264.

Brown-Schmidt, S., Campana, E. & Tanenhaus, M.K. Real-time reference resolution by naïve participants during a task-based unscripted conversation. To appear in J.C. Trueswell & M.K. Tanenhaus (eds.), World-situated language processing: Bridging the language as product and language as action traditions. (MIT Press).

Chambers, C.G., Tanenhaus, M.K, Eberhard, K.M., Filip, H & Carlson, G.N. (2002). Circumscribing referential domains in real-time sentence comprehension. *Journal of Memory and Language*, 47, 30-49.

Hanna, J.E. & Tanenhaus, M.K. (in press). Effects of task constraints and speaker goals on addressee's referential domains in a collaborative task. *Cognitive Science*.

Hanna, J.E., Tanenhaus, M.K. & Trueswell, J.C. (in press). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*.

Krauss, R.M. & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4, 343-346.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109-147.

Sedivy, J.C. (2003). Informativity expectations and resolving reference: Some evidence from language processing and development. Paper presented at the City University of New York conference on Human Sentence Processing, 2003.

Snodgrass, J.G., & Vanderwart, M. (1980). *Journal of Experimental Psychology: Human Learning and Memory*, 6:3, 174-215.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (l995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 632-634.