Syntactic Complexity induces Explicit Grounding in the MapTask corpus

Martin Tietze , Vera Demberg, Johanna D. Moore

[1]School of Informatics, University of Edinburgh, UK
mtietze@inf.ed.ac.uk, v.demberg@ed.ac.uk, j.moore@ed.ac.uk

Abstract

This paper provides evidence for theories of grounding and dialogue management in human conversation. For each utterance in a corpus of task-oriented dialogues, we calculated integration costs, which are based on syntactic sentence complexity. We compared the integration costs and grounding behavior under two conditions, namely face-to-face and a no-eye-contact condition. The results show that integration costs were significantly higher for explicitly grounded utterances in the no-eye-contact condition, but not in the face-to-face condition.

Index Terms: dialogue, syntactic complexity, grounding

1. Introduction

In this study we compared the sentence complexity of utterances comprising core speech acts which were either explicitly grounded or not. The dialogues between the two human participants took place in two conditions, either face-to-face or without eye-contact. To analyze sources of misunderstandings in dialogue, we applied the Syntactic Prediction Locality Theory (SPLT) [1] to the MapTask corpus[2]. SPLT relates syntactic sentence complexity to working memory resources via the integration costs associated with processing a sentence. Integration costs are the computational costs incurred when attaching syntactic structures together.

To determine whether sentence complexity is a source of potential misunderstanding in dialogue, we looked at the grounding behavior shown by the participants during the dialogues in the corpus. Grounding is the process of ensuring mutual understanding during a dialogue [3]. The form of grounding varies depending on the underlying task and the situation. It can be performed explicitly, e.g., "OK, got it", or implicitly by simply continuing with the dialogue, or even by gestural and facial expressions [4]. In [5] four classes of *dialogue acts* are used to describe the meaning of utterances. Utterances used for grounding are classified as *Grounding Speech Acts*, whereas utterances motivated by the underlying task, e.g., "Then turn left at the second corner.", are called *Core Speech Acts*.

2. Materials

For this study, we used the MapTask corpus [2], which is annotated at various levels. Here we used the annotations at two levels: the *utterance* level, which assigns transcribed words to utterances, and the *move* level, which assigns dialogue act labels to utterances. The move taxonomy of MapTask [6] does not distinguish between core and grounding speech acts, however the moves can be mapped directly to the dialogue acts in [5]. In this analysis 16 of the 128 dialogues in corpus were used, all from one pair of participants. Half of the dialogues were recorded in each of the eye-contact/no-eye-contact conditions. For the syntactic analysis, MINIPAR [7] was used to produce the necessary input to the algorithm that computes the integration costs. SPLT assigns costs to each word depending on its syntactic class and context. We computed the costs for each utterance containing a CSA and checked whether it was followed by a GSA. We used two metrics to calculate the cost for an utterance from its words. One metric sums the costs of all words in the utterance, and the other uses the maximum of the costs associated with the words in the utterance.

3. Results

All tests were performed using the *maximum* metric[1]. Comparing the utterance costs for each condition, we found no significant cost difference between the grounded and ungrounded utterances in the eye-contact condition ($t(889) = -0.82, p > 0.6$), but there is a significant difference in the no-eye contact condition ($t(779) = 2.69, p < 0.01$)). To test the effect of both grounding and eye-contact, we ran a 2×2 ANOVA. The results of the ANOVA showed clear effects of both groundedness ($p < 0.05$) and condition ($p < 0.01$) as well as an interaction between the two ($p < 0.01$).

4. Conclusions

In this study we showed that increased sentence complexity increases the likelihood of explicit grounding when the participants do not have eye-contact. This result is relevant for the design of spoken dialogue systems, for example, sentence complexity of system utterances could be reduced in the face of frequent misunderstandings. The lack of a difference in the eye-contact condition is an indication of the effectiveness eye gaze and facial expressions.

5. References

[1] E. Gibson, "Linguistic complexity: Locality of syntactic dependencies," *Cognition*, vol. 68, no. 1, pp. 1–76, 1998.

[2] A. H. Anderson, M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert, "The HCRC Map Task Corpus," *Language and Speech*, vol. 34, no. 4, pp. 351–366, 1991.

[3] H. H. Clark and E. F. Schaefer, "Contributing to discourse," *Cognitive Science*, vol. 13, pp. 259–294, 1989.

[4] H. H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.

[5] D. Traum and J. Allen, "A speech acts approach to grounding in conversation," in *Proceedings 2nd International Conference on Spoken Language Processing (ICSLP-92)*, 1992, pp. 137–140.

[6] J. Carletta, S. Isard, G. Doherty-Sneddon, A. Isard, J. C. Kowtkot, and A. H. Anderson, "The reliability of a dialogue structure coding scheme," *Computational Linguistics*, vol. 23, pp. 13–32, 1997.

[7] D. Lin, "Minipar: a minimalist parser," in *Maryland Linguistics Colloquium*. University of Maryland, 1999.

Supported by Marie Curie Early Stage Training Site EdSST (MEST-CT-2005-020568)

[1]The results for the *sum* metric were similar.