



map. They communicate to reproduce IG's route on IF's map. Players cannot see each other's maps. They use landmarks for navigation. Although most landmarks are identical on both maps, some differ by: (1) being absent on one of the maps or present on both; (2) having clearly different attributes; (3) being affected or not by 'ink damage' that obscures the colour of some landmarks on IF's map.

Our Map Task (Fig 1) has three experimental variables: (1) homogeneity (whether the landmarks are of one or different kinds, e.g. only aliens, or aliens and fish); (2) orderliness (whether the 'ink blot' obscures a continuous stretch of the route); (3) animacy. These are varied factorially so that each pair of participants (dyad) completes a set of 8 map pairs. There are 32 dyads.

### 3 Data

Currently 210 of the 256 dialogues are transcribed and used here. Each dialogue is about 10 minutes long. Overall the 210 transcripts contain 184,711 words of which 5,251 are colour terms.

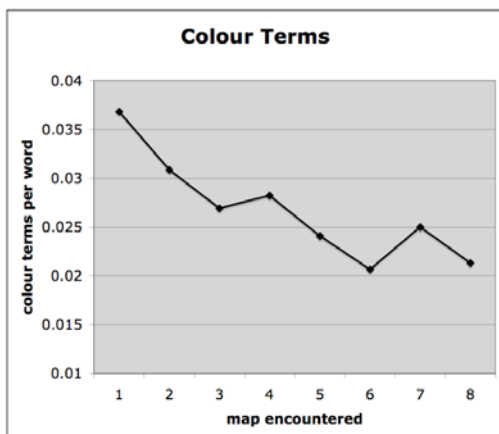


Figure 2: Use of colour terms per word over time.

Fig. 2 shows the mean number of colour terms per spoken word across the 8 map pairs that each dyad encounters. As the task environment affords no other occasions to use colour terms, we make the simplifying assumption here that all colour terms are used for referring to landmarks. The mean number of colour terms decreases over the course of the 8 maps. There is a significant negative correlation ( $r = -0.172$ ,  $p < 0.01$ ) between the rate of colour terms used and the number of the encountered map.

A 3-way repeated measures ANOVA showed that of the 3 experimental variables only landmark homogeneity affected the use of colour terms: ( $F_1(1,20) = 12.26$ ,  $p = 0.02$ ) on average the mixed landmark condition attracts fewer col-

our terms per word (0.024) than the uniform landmark condition (0.032).

### 4 Conclusions

The participants in our Map Task pick up the fact that colour is an unreliable attribute in referring to the landmarks on the maps. The adaptation is not a sudden change in behaviour but is a gradual adaptation to the properties of the items they have to refer to.

The effect of homogeneity is most likely due to the difficulty of the maps with landmarks of just one kind: the type attribute does not distinguish such landmarks; colour must be used to identify the target landmark.

The main result is that the use of colour terms changes over time during a task, which is not accounted for in Jordan and Walker's (2005) model, and to our knowledge such a model does not exist yet. For an adequate dialogue model it is insufficient to simply let the computer choose a level of colour terms (observed in a suitable corpus), because that would be unnatural. In such models the referring expressions in the first maps would not be natural, because they would use the colour attribute less often than humans (analogously too often in the last maps). One goal of our current project is to develop such a model.

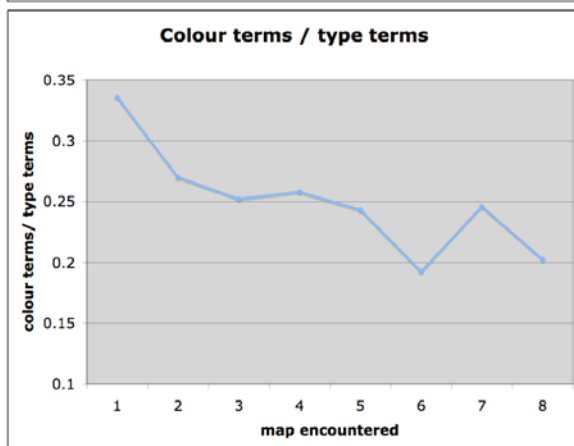
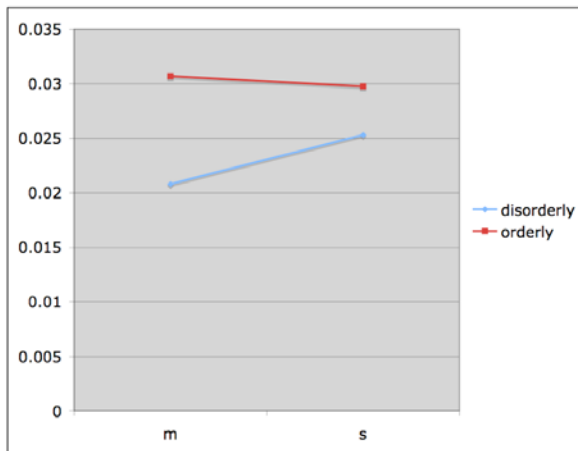
### Acknowledgements

This research was supported by the National Science Foundation (IIS-0416128). Thanks to the Memphis team for providing transcriptions: Max Louwerse, Gwyneth Lewis and Megan Zirnstein.

### References

- Anderson, A., et al (1991). The HCRC Map Task Corpus. *Language and Speech*, 34: 351–366.
- Brennan, S.E. and Clark, H.H. (1996) Conceptual Pacts and Lexical Choice in Conversation. *JEP: Learning, Memory, and Cognition*, 22:1482–1493.
- Dale, R. and Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.
- Guhe, M., Steedman, M., Bard, E.G., Louwerse, M.M. (2006) Prosodic marking of contrasts in information structure. In: *Proceedings of BranDial'06*. 179–180.
- Jordan, P. W. and Walker, M. A. (2005). Learning content selection rules for generating object descriptions in dialogue. *Journal of Artificial Intelligence Research*, 24:157–194.

## Additional Material



When the participants notice (in most cases on a subconscious level, although there are explicit discussions of the reliability of the colour attribute in our corpus) that a particular attribute of the domain (here: colour) is an unreliable disambiguator (remover of distractors) (or even may cause problems in the case of mismatching colours) they adapt the way they produce referring expressions and do use colour less often.

The use of a particular attribute depends on the overall reliability as well as on the local context: even though colour may be unreliable, it may still be required in some context in which it is the only/correct disambiguator.

While in the existing algorithms the adaptation involves an adaptation to the linguistic corpus, we are interested in the adaptation that the participants perform while carrying out the task. Thus, this is adaptation in quite a different sense, because it involves the adaptation of the way referring expressions are used to an attribute of the task environment. Furthermore, the adaptation in the existing algorithms is an adaptation to what

the dialogue partner said recently, not to the success of using a particular attribute.

These are all approaches how an algorithm is trained to reflect properties of a corpus. They do not provide a cognitive model of how an algorithm can be adapted during the dialogue to the success rate of communication.

Is metacognition needed to notice that the colour feature is unreliable?

In algorithms of the Dale and Reiter variant the selection of attributes and the ordering in which these attributes are tried during generation are critical.

Jordan and Walker (2005) point out that the algorithm is trained for a particular data set, and it is unlikely that the same algorithm will work as well on a different data set than on the one it was trained on. The MapTask is a different task than COCONUT.

The two existing papers do not, however, reflect the participants' discoveries about properties of the reliability of a particular feature in the domain. (But the model should be easily adaptable?)