Referential Grounding for Situated Human-Robot Communication

Vivien Mast, Daniel Couto Vale, Zoe Falomir, Mohammad Fazleh Elahi
SFB/TR-8 Spatial Cognition
University of Bremen
{vivien.mast, danielvale}@uni-bremen.de,
{zfalomir, fazleh}@informatik.uni-bremen.de

Abstract

We present a dialogue system and reference handling component for efficient and natural referential grounding dialogues from 2D images. Using a probabilistic representation of qualitative concepts, the system uses flexible concept assignment in reference handling for bridging conceptual gaps between the system and the user, and engages in clarification dialogues based on an evaluation of miscommunication risk.

1 Introduction

From her comfortable sofa, Mary asks her personal assistant robot Amanda: *Could you pass me that yellow book on my desk?* Amanda is not sure which book Mary meant, and asks: *Do you mean the one in front of the coffee cup?* Slightly annoyed, Mary replies: *No, not the green one, the yellow one.* Amanda confirms: *Oh, ok. I thought that was orange. I'll get it.* and brings the book to Mary.

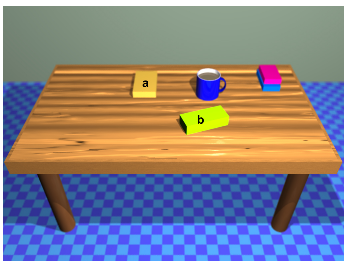

Figure 1: Intended referent (a) and distractor (b).

While standard algorithms for referring expression generation (REG) assume that objects can be defined by a fixed set of crisp properties (Krahmer and van Deemter, 2012; Dale and Reiter, 1995), humans carve up the world in multiple ways (Steels, 2008), depending on sensory differences (Roorda and Williams, 1999), exposure to a domain (Goldstone et al., 2012, p. 621), or situational conditions (Spranger and Pauw, 2012).

In order to bridge conceptual gaps between interactants and establish common ground, humans flexibly adapt the use of concepts (Clark and Brennan, 1991; Garrod and Anderson, 1987; Clark and Wilkes-Gibbs, 1986). As the conceptual gap in human-machine interaction is even larger, dialogue systems may benefit greatly from sophisticated grounding abilities.

Our implementation of the agent based dialogue system architecture and framework DAISIE (Ross and Bateman, 2009), using the ***P**robabilistic **R**eference **A**nd **GR**ounding* mechanism PRAGR (Mast and Wolter, 2013b; Mast and Wolter, 2013a), flexibly assigns properties during reference handling in order to maximize communicative success. We show how the DAISIE+PRAGR system is capable of engaging in grounding dialogues about images as they may be provided by a camera installed on the head of a mobile robot, by generating and resolving referring expressions (REs), and using probabilistic evaluations of the REG and reference resolution (RR) output for making reasonable dialogue decisions.

2 PRAGR

PRAGR is a probabilistic reference handling system for enabling dialogic grounding, described in detail by Mast and Wolter (2013b). PRAGR's core concepts are *acceptability*—the probability $P(D|x)$ that the interlocutor accepts D as a description of object x, and *discriminatory power*—the probability $P(x|D)$ that D discriminates x from its distractors, a value comparing the acceptability of D for the target to its acceptability for distractors. The stochastic model handles descriptions of arbitrary complexity, including relations.

In RR, given a description D, PRAGR selects as best referent x^* the object for which D has the highest acceptability. In REG, PRAGR aims for effectiveness in communication, given uncertain knowledge. Thus, it searches for the most *appropriate* description D^* which jointly maximizes acceptability and discriminatory power.

3 Layered Representation

In PRAGR's two-layer knowledge representation, the *perceptual layer* represents qualitative and metric perceptual properties of objects (e.g. defining shape points and hue, lightness and saturation) obtained in a first step of abstraction (Falomir et al., 2012). *Perceptual grounding* modules provide probabilistic mappings of objects to conceptual properties such as ORANGE in the *conceptual layer*. A *dialogic grounding module* may add or overwrite mappings on the conceptual level.

Perceptual grounding modules include a probabilistic model of projective terms adapted from Mast and Wolter (2013b), a crisp model of object type based on Qualitative Image Description (Falomir et al., 2012) and a fuzzy adaptation of the colour model by Falomir et al. (2013).

With these probabilistic feature models, PRAGR can consider gradual differences in discriminatory power and acceptability and provide the most appropriate description. It may call the same ball *the red ball* (Figure 2a) or *the orange ball* (Figure 2b), depending on present distractors, as acceptability of a property for distractors dampens discriminatory power.

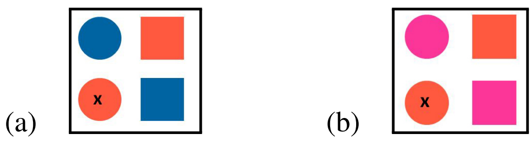

Figure 2: Context sensitivity of category assignment: (a) the red circle, (b) the orange circle.

4 DAISIE+PRAGR

The current update cycle of DAISIE's information state depends on 5 subprocesses for the automation of linguistic understanding and on 4 subprocesses for the automation of linguistic expression as shown in Figure 3. REs in the input are identified during experiential interpretation and enriched via co-reference resolution against the discourse history during textual interpretation. They are then queued for handling by the dialogue manager which directly accesses PRAGR. In the following dialogue: **H**uman: *Bring me the red box.*,

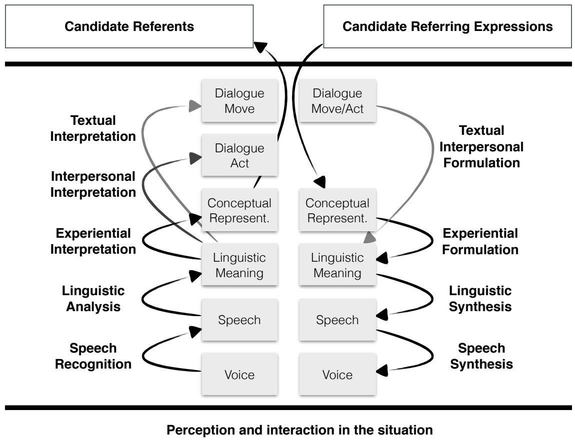

Figure 3: Architecture of DAISIE

System: *Which red box do you mean?*, H: *The one on the floor.*, the expression *one* is resolved to the conceptual representation [RED, BOX, SUPPORT(FLOOR)] before being passed to PRAGR which then provides an n-best list of potential referents with the evaluation values of the input expression. Based on this evaluation, the dialogue manager plans the next dialogue move. For example, if there is no substantial difference between candidates, an open clarification question (*Which red box do you mean?*) is generated. If one preferred candidate is found, depending on acceptability and discriminatory power, the system may generate an expansion (*Do you mean the one on the table?*) or a confirmation (*OK, I'm getting it.*).

In generation, conceptual representations of REs are selected and evaluated by the reference handling component, called as part of the dialogue move selection. If no sufficiently appropriate RE for an intended target can be found, the system may attempt to ground a potential reference object first, in order to use this for a follow-up reference to the intended target: S: *Can you see the low table to the left of the door?* H: *Yes.* S: *Your keys are in the small green box on that table.*

5 Summary

The proposed referential grounding dialogue system DAISIE+PRAGR is capable of flexibly using concepts in order to improve referential success in generation and understanding. Decisions of the dialogue manager about next dialogue moves are informed by the evaluation results of the reference handling component, thus enabling natural and efficient grounding dialogues in situated communication.

Acknowledgments

Funding by the Deutsche Forschungsgemeinschaft (DFG) for the SFB/TR 8 Spatial Cognition, project I5-[DiaSpace] and the European Commission through FP7 Marie Curie IEF actions under project COGNITIVE-AMI (GA 328763) is gratefully acknowledged.

References

Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition*, 13:127–149.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1 – 39.

Robert Dale and Ehud Reiter. 1995. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, 18:233–263.

Zoe Falomir, Lledó Museros, Gonzalez-Abril Luis, M. Teresa Escrig, and Juan A. Ortega. 2012. A model for the qualitative description of images based on visual and spatial features. *Computer Vision and Image Understanding*, 116(6):698–714.

Zoe Falomir, Lledo Museros, Luis Gonzalez-Abril, and Ismael Sanz. 2013. A model for qualitative colour comparison using interval distances. *Displays*, 34:250–257.

Simon Garrod and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27:181–218.

Robert L. Goldstone, Alan Kersten, and Paulo F. Carvalho. 2012. Concepts and categorization. In A. F. Healy and R. W. Proctor, editors, *Comprehensive handbook of psychology, Volume 4: Experimental psychology*, pages 607–630. Wiley, New Jersey.

Emiel Krahmer and Kees van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.

Vivien Mast and Diedrich Wolter. 2013a. Context and vagueness in REG. In *Proceedings of PRE-CogSci 2013*, December.

Vivien Mast and Diedrich Wolter. 2013b. A probabilistic framework for object descriptions in indoor route instructions. In T. Tenbrink, J. Stell, A. Galton, and Z. Wood, editors, *Spatial Information Theory*, volume 8116 of *Lecture Notes in Computer Science*, pages 185–204. Springer International Publishing, October.

Austin Roorda and David R Williams. 1999. The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520–522.

Robert J. Ross and John A. Bateman. 2009. Daisie: Information state dialogues for situated systems. In V. Matoušek and P. Mautner, editors, *Text, Speech and Dialogue*, LNCS, pages 379–386, Berlin, Heidelberg. Springer.

M. Spranger and S. Pauw. 2012. Dealing with Perceptual Deviation - Vague Semantics for Spatial Language and Quantification. In L. Steels and M. Hild, editors, *Language Grounding in Robots*, pages 173–192. Springer.

Luc Steels. 2008. The symbol grounding problem has been solved, so what's next? In M. De Vega, A.M. Glenberg, and A.C. Graesser, editors, *Symbols and embodiment: Debates on meaning and cognition*, pages 223–244. Oxford University Press.