Proceedings

Referring Phenomena in a Multimedia Context and their Computational Treatment

sponsored by the ACL Special Interest Group
on Multimedia Language Processing (SIGMEDIA)

Madrid, July 11th 1997

in conjunction with ACL/EACL-97
http://www.dfki.uni-sb.de/imedia/workshops/mm-references.html

Proceedings

Referring Phenomena in a Multimedia Context and their Computational Treatment

sponsored by the ACL Special Interest Group
on Multimedia Language Processing (SIGMEDIA)

Madrid, July 11th 1997

in conjunction with ACL/EACL-97
http://www.dfki.uni-sb.de/imedia/workshops/mm-references.html

Order additional copies from:

ACL
P.O. Box 6090
Somerset, NJ, 08875 USA
+1-908-873-3898
acl@bellcore.com

Organizers:

Elisabeth André, DFKI, Germany
Laurent Romary, CRIN, France
Thomas Rist, DFKI, Germany

Programme Committee:

Elisabeth André, DFKI, Germany
Doug Appelt, SRI International, USA
Jean Caelen, CLIPS-IMAG, France
Robert Dale, Microsoft Research Institute, Australia
John Lee, University of Edinburgh, UK
Luis Pineda, IEE, Mexico
Thomas Rist, DFKI GmbH, Germany
Laurent Romary, CRIN, France
Massimo Zancanaro, IRST, Italy
Bonnie Webber, University of Pennsylvania, USA

Preface

A growing number of research projects has started to investigate the use of referring expressions in multimedia systems. On the one hand, the use of multiple media has led to new problems, such as a proper treatment of cross-media references. On the other hand, it has turned out that many concepts already known from natural language processing, such as cohesion, take on an extended meaning in multimedia discourse. For example, a proper treatment of referring expressions in a multimedia discourse requires an explicit representation of the syntax and semantics of the graphical discourse. As theories of NL reference become more sophisticated, it is quite natural to investigate whether these theories also encompass other media, such as graphics and pointing gestures.

Several research projects have already started to transfer theories to the broader context of multimedia discourse. Examples of models that have been used for multimedia applications are Grosz and Sidner's theory of discourse structure, the centering model developed by Joshi and colleagues and Appelt's and Kronfeld's model of referring. However, there are researchers who doubt that linguistic phenomena, such as anaphora, also exist in multimedia dialogue. The reason they give is that there are no graphical devices for distinguishing between a reference-specifying and a predication-specifying part since objects and their properties are hardly separable once depicted.

The workshop will be centered around questions, such as "To what extent can linguistic models be applied to multimedia references?", "Which linguistic phenomena can also be observed in multimedia discourse?" and "Is a cross-modality theory of reference possible?". Papers were invited that deal with the following topics:

- computational models for the analysis/generation of referring expression in a multimedia discourse
- coordination/synchronization of multiple media, such as speech and pointing gestures
- deixis in multimedia environments
- cohesion and coherence in multimedia discourse
- representation of multimedia discourse
- encoding theories for text and graphics
- formal models of multimedia referring
- referring expressions in augmented/virtual realities
- empirical studies

Most topics were covered by the submissions which included full papers and short project descriptions. Each submission was reviewed by several members of the programme committee. The selected papers for this 1-day workshop give a good impression of the

broad spectrum of research and application. They address: empirical evaluations, computational approaches for the analysis and generation of referring expressions, case studies, models and formal frameworks.

We would like to take this opportunity to thank all the people who contributed to this workshop. In particular, we are grateful to the authors and the members of programme committee and to Harald Trost for his organizational support.

Elisabeth André Laurent Romary Thomas Rist

Contents

Common Myths about Multimodal Integration during Human-Computer Interaction

Sharon Oviatt
Center for Human-Computer Communication
Oregon Graduate Institute of Science & Technology
oviatt@cse.ogi.edu, http://www.cse.ogi.edu/CHCC

Computational work on multimodal system design has focussed on a number of phenomena that may best be described as "myths"- or misperceptions of the linguistic nature of typical multimodal constructions. Unfortunately, computational presumptions about multimodal constructions could lead to misdirected system building by the community at large. In this presentation, we will focus our discussion on three common myths about multimodal input, findings from recent empirical work that uncovered these myths and, finally- our own different view of what constitutes a prototypical multimodal construction for the case of combined speech and pen-based input.

To focus discussion, we will limit ourselves to a discussion of the following three myths:

1. Speech is the primary input mode in any multimodal construction that includes it, with gesture, gaze, touch, pen input, and other modes being secondary.

2. Speech + pointing is the dominant multimodal integration pattern.

3. Multimodal constructions involve redundant propositional content supplied by the input modes involved.

We also will discuss other input modalities that we view to be promising, and research strategies that can be used to build a scientific foundation of information on modes other than speech and pen. Finally, we will discuss general questions that we think need to be asked and answered by multimodal researchers in the field at large, as well as major outstanding challenges for multimodal/media system design.

The results to be presented and discussed (i.e., including simple graphic summaries of data relevant to the three myths articulated above) are included in the subsequent paper, which originally appeared in:

Oviatt, S. L., DeAngeli, A. & Kuhn, K. Integration and synchronization of input modes during multimodal human-computer interaction, in *Proceedings of Conference on Human Factors in Computing Systems: CHI '97*, New York, ACM Press, 415-422.

This and other related publications also can be viewed and printed from the "publications" section of the following URL: http://www.cse.ogi.edu/CHCC).