Reference resolution as a facilitating process towards robust multimodal dialogue management: A cognitive grammar approach

Ashwani Kumar, Susanne Salmon-Alt, Laurent Romary

Reference Resolution as a facilitating process towards robust Multimodal Dialogue Management : A Cognitive Grammar Approach

Ashwani Kumar **Susanne Salmon-Alt** **Laurent Romary**
Laboratory LORIA
Campus Scientifique, B.P. 239
54506 Vandoeuvre-lès-Nancy, France
{ashwani.kumar, susanne.alt, laurent.romary}@loria.fr

ABSTRACT

This paper tries to fit a novel reference resolution mechanism into a multimodal dialogue system framework. Essentially, our aim is to show that a typical multimodal dialogue system can actually benefit from the cognitive grammar approach that we adopt for reference resolution. The central idea is to construct and update reference and context models in a manner that imparts adequate level of under-specificity to multimodal semantics. Context-independent semantic representations are constructed based upon the surface structure of the referring expressions and syntactic constraints within an utterance. The reference resolution algorithm assimilates these semantic representations into a coherent context model, resulting in the profiling of the intended referent. The resolution model is built upon discursive, perceptual and conceptual cues, thus successfully accounting for multiple modalities and a multi-dimensional application domain model.

1 Introduction

The complexity of reference resolution is due, in part, to the variety of referring expressions, including indefinites, definite descriptions, pronominal reference and ellipses or one-anaphora. The problem is aggravated by the apparent variety of mechanisms required to deal with even one of these types of referring expressions. For example, the referent of a definite description may be linked to a prior discourse entity with the same head, associated to a prior entity from which it can be inferred, or extracted from a larger situation (Poesio and Vieira, 1998). As a result, much current work on reference centers around pronominal reference (cf. Centering Theory: Grosz et al., 1995; McCoy and Strube, 1999). The treatment of other types of referring expressions is often seen as an extension of or variation on the basic co-referential mechanism (DRT and it extensions: Kamp and Reyle, 1993; Bos et al., 1995).

Additionally, the interpretation of referring expressions is based on both discourse and perceptive context. For example, “another one” cannot be understood without previous discourse mention of, let’s say “a romantic song”. The need of perceptive information is evident for expressions like “the last two song writers” referring to a list displayed on a screen. What we need then is a unified framework to represent and update dynamically the information provided by both discourse and perceptive context and to constrain the access to this information. DRT, for example, provides access to all previously mentioned entities, while Centering Theory considers the previous discourse unit only. On the other hand, within the list of identified potential referents, Centering Theory provides a precise account of relative salience, whereas DRT specifies only general syntactic constraints to narrow the list. Some recent models attempt to apply more precise selectional criteria to global discourse (Asher, 1993; Hahn and Strube, 1997). However, all rely on some prior segmentation, implicitly assuming that discourse structure informs reference resolution, and ignoring the possibility of determining structure based on referential devices or on perceptive information.

Finally, we notice a gap between the predictions made by approaches in analysis and the generation of referring expressions. In an example like “Select a song and play it / the song”, DRT-like models do not predict any difference between pronominal and nominal anaphora whereas a generation model based on Dale and Reiter (Dale, 1992; Dale and Reiter, 1996) would largely prefer the pronoun.

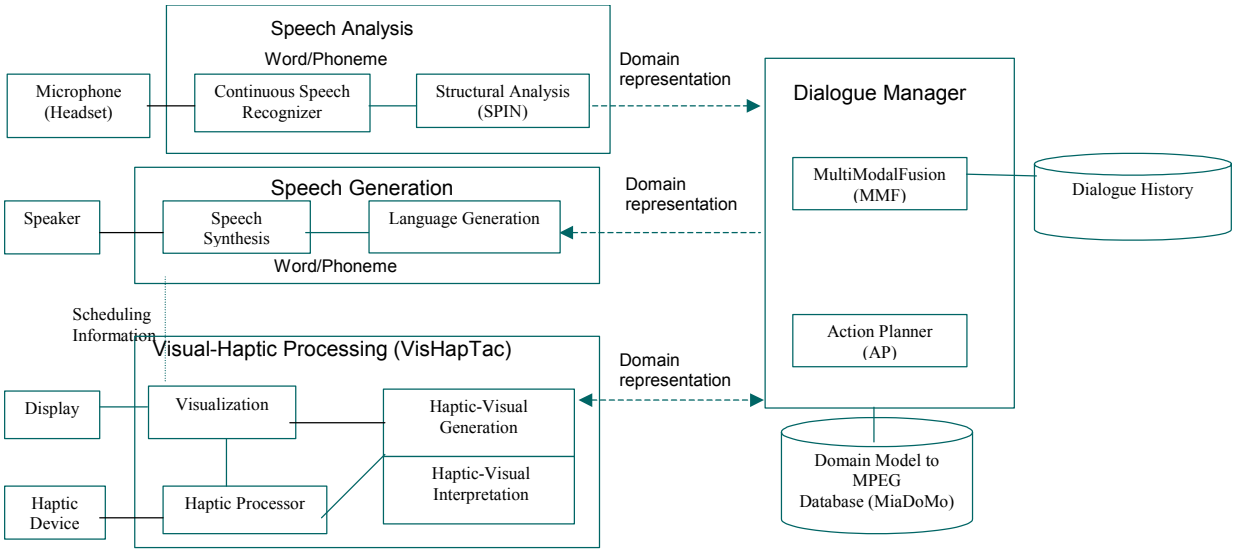

Figure 1: MIAMM System Architecture

For all these reasons, we concentrate on a model for reference resolution that attempts to overcome the diversity of resolution mechanisms. It is based on the fundamental assumption that all reference (independent of the type of referring expression) is accomplished via access to and restructuring of *Reference Domains (RD)* rather than by direct linkage to the entities themselves. It includes the same updating mechanisms for both discourse and perceptive information and is intended to be predictive in both language analysis and language generation – a particularly important feature for a model to be integrated into a dialogue system framework.

There have been efforts towards characterizing relationship between referential and discourse structures (Schauer, 2000; Seville, 1999). However, there does not seem to be much work on how the reference model could be used for robust multimodal dialogue processing. In this paper, we advocate a model which closely hinges the dialogue model on the reference model. The rest of the paper is structured as follows. The MIAMM framework is briefly described in Section 2. Section 3 describes the basic cognitive grammar hypothesis that we incorporate in our reference model. Section 4 describes our reference and dialogue framework in detail.

2 MIAMM Framework

The main objective of the MIAMM[1] project (Reithinger et al., 2002) is to provide an integrated and comprehensive framework (cf. Figure 1) for the design of modular multimodal dialogue systems that allow fast and natural access to multidimensional application databases. The MIAMM emulator integrates a speech interface with a graphic interface, which consists of haptic-tactile buttons and a visual display. The user can interact with the device using speech and/or the haptic buttons to search, select and play tunes from an underlying musical database

The application domain model, MiaDoMo realizes an intelligent uniform interface between the dialogue module and the various musical databases to be accessed for content selection. Requests could be as simple as “some country music” or as complex as “the soprano-alto duet piece by Charpentier” or “some Mozart-style happy orchestral music”. Essentially, the domain model is multidimensional in the sense that the application objects can have associated attributes along multiple discrete, as well as continuous dimensions. For example, information related to a musical band can be stored in the form of discrete dimensions such as band name, member artist names, genre objects. A main task of the visualisation functionality is to make it easy for the user to navigate in the visualisation using speech, pointing and haptic interaction., various albums produced, etc. and/or in the form of continuous dimensions such as temporal duration. Therefore, a query to MiaDoMo results in an information matrix which resides in the dynamic memory of the Visual-Haptic (VisHapTac) processor. The various dimensions of the data model as represented in the information matrix define the visualisation space

[1] www.miamm.org

in which the users can navigate. The restrictions of the display require for condensation and concentration of the visible objects. A main task of the visualization functionality is to make it easy for the user to navigate in the visualization using speech, pointing and haptic interaction.

The MIAMM framework allows vague and incomplete multimodal inputs as well as information aggravation and re-structuring along various dimensions. Such a complex scenario entails heavy usage of referring expressions, anaphoric as well as deictic. Apart from the common indefinites, definites and demonstratives, bridging expressions such as "the swing", referring to the musical city currently focussed on the map visualization, are quite frequent. At the dialogue level, the multimodal utterances are terse and potentially ambiguous in nature. However, for the sake of simplicity and naturalness, we are mostly concerned with task-oriented mixed-initiative dialogues.

3 From Cognitive Grammar to Reference Resolution

Cognitive Grammar (Langacker, 1986, 1991) situates linguistic competence within a more general framework of cognitive faculties by assuming that language is neither self-contained, nor describable without reference to cognitive processing. As a fundamental assumption of Cognitive Grammar, sense cannot be represented by logical forms. The first reason is that semantic structures are characterised relative to open-ended knowledge systems. The second reason is that an expression's meaning cannot be reduced to an objective characterization of the situation described: equally important for semantics is how the speaker chooses to construe the situation. Therefore, Cognitive Grammar assumes conceptual rather than truth-conditional semantics, considering that meaning consists of a process of conceptualisation, i.e. activation of conceptions in a hearer's mind.

More precisely, the conceptualization of an expression is said to impose a particular image on its domain, where *domain* is defined as a cognitive structure which is presupposed by the semantics of an expression. As an example, the definite in "Select **the first song** and play it" presupposes as its domain an ordered set of songs. In MIAMM, this domain could be a visual representation for a list of songs, displayed on the screen.

The particular image imposed by the expression results in profiling a substructure of the domain, namely that substructure which the expression designates. In the aforementioned example, this would be the part of the list representing the intended referent. As a result of the interpretation process, the profiled subpart of a domain is hypothesized to be more prominent or more activated than the rest of the domain.

Since an expression is always said to be interpreted within a limited domain, our context model – or *multi-modal dialogue history* – is built upon the notion of *reference domains* (Salmon-Alt, 2001). These domains are identifying representations for subsets of contextual entities to which it is possible to refer, including individual objects and collections of objects. A first important point is that these domains are not primarily linguistic constructs, since they are created and updated dynamically *via* discursive information, visual perception, haptic events and conceptual knowledge. The second important characteristic of our domains is that they present the entities from a particular cognitive viewpoint, for which we assume in the following that it is the most likely to be activated for referential access to the entity. In this sense, our model predicts optimal use of referring expressions, whereas fallback strategies can always be applied to failing interpretations. Taking up again the previous example "Select **the first song** and play it", we will, for instance be able to predict that in this context of an activated domain of *songs*, a one-anaphora such as "Delete **the last one"** has to be interpreted preferentially as referring to a song, even if the visual interface displays at the same time, for example, a list of song authors.

Based on a context modeled by *Reference Domains*, the interpretation process for referring expressions is seen as an extension of the hypotheses of Cognitive Grammar about the representation of grammatical meaning in terms of abstract symbolic schemas. More precisely, we assume that the semantics of a given expression can be represented by a schema which corresponds to an under-specified reference domain. This under-specified domain is calculated by combining abstract schemas for nouns, modifiers and determiners, taken from a lexical knowledge base.

The interpretation process consists of two steps:

1) The under-specified domain has to be matched to suitable *RDs* from the context model;
2) A restructuring operation updates the domain by profiling a substructure of the same domain as the referent.

An interesting point here is the fact that the same mechanism acts for linguistic expressions and for gestures: for example, a pointing gesture to a particular CD cover on the screen highlights this entity as the referent, whereas the other CD covers are considered as the reference domain. In this way, an expression like "Delete the other ones" will then be interpreted as referring to the rest of covers, even if there are other visual elements on the screen, (for example, portraits of song writers).

4 From Reference Resolution to Dialogue Management

4.1 Dialogue Functional Specification

In line with the Cognitive Grammar hypothesis, we assert that the dialogue functional behavior is essentially guided by the underlying processes which a multimodal system undertakes for input interpretation, fusion, fission and output generation. The system should be able to map the communicative behaviors within a multimodal utterance onto the communicative functions and vice versa (Cassell, 2001). This requires specification of how the interpretation or generation of the utterance changes the system's information state such as domain model, discourse model, user model, and task model (Bunt et al., 2002). The interpretation of a multimodal input, such as a spoken utterance combined with a haptic gesture, will often have stages of modality-specific processing, resulting in representations of the semantic content of the interactive behavior in each of the separate modalities involved. Other stages of interpretation combine and integrate these representations, and take contextual information into account, such as information from the domain model, the discourse model and the user model. Therefore functionally, the multimodal dialogue strategy ought to be incremental so as to account for low-level modality processing as well as high-level unified semantics.

4.2 Underspecified Semantics

In MIAMM, we adhere to multi-level approach to semantics. Various modalities and their respective functional behaviors vary significantly as regards to the distribution of semantics across the modality channels. Besides, the modularity constraint within the system architecture (cf. Figure 1) provides that every module does not have access to every available static or dynamic knowledge resource. Essentially, due to these reasons any conventional ambiguity resolution/multiple hypothesis algorithm (Alexandersson, 2001) will lead to various possible readings, resulting in the combinatorial explosion problem. Therefore, we resort to the Underspecification (van Deemter and Peters, 1996; Pinkal 1999) approach towards semantic specification. The choice fits perfectly with our integrated framework of reference resolution and incremental dialogue processing as our approach tries to specify the multimodal semantics in a context, which builds up incrementally.

The context-independent syntactic-semantic representations from SPIN (cf. Figure 1) and visualization representations from VisHapTac are encoded in MMIL (MultiModal Interface Language, Romary 2002). MMIL serves as the central representation format within the MIAMM architecture as it accounts for the transmission of data between the dialogue manager and both the multi-modal inputs and outputs and the application. It also forms the basis for the content of the dialogue history in MIAMM, both from the point of view of the objects being manipulated and the various events occurring during a dialogue session. MMIL incorporates FOL-type binary predicate-based semantics into a flat XML structure, maintaining two primitive levels of representation – events and participants – (Kumar et al., 2003). For example, the underspecified semantic representation for a simple referring expression, "the song" encoded in MMIL, will look like as follows:

```
<participant id = "p1">
        <objType>tune</objType>
        <individuation>singular</individuation>
        <refType>definite</refType>
        <refStatus>pending</refStatus>
</participant>
```

4.3 Specifying Semantics, Incorporating Pragmatics and Resolving References

There have been various attempts towards characterizing reference resolution models such

as coreference model (Tetreault, 2001), sense model (Hobbs, 1988), extensions model (Allen et al., 1996). However, none of these models account for the complete range of reference phenomena found in conversational language. Byron (2002) construes the ideal resolution model as a mapping from *initial* referring expression (RE) descriptions to *final* logical term descriptions. The under-specified representations as described in the previous section resemble the context independent description structures. However, an important distinction is that our aim is not to construct *final* logical term descriptions. What we are aiming at instead is the *maximum*[2] possible resolution, which may not always result into *final* logical forms, at the same time maintaining certain degree of under-specificity as it is necessitated for the continuity of the dialogue progress. For example, the user can command, "play me the pop one", while there exists only jazz tunes in the context. In this case an effort to construct a *final* logical term for the referring expression "the pop song", would not lead to the desired response from the system. Instead, in such cases of perceptual mismatch, we maintain the level of under-specificity, while informing the Action Planner (AP) about the level of feature mismatch. AP then initiates proper meta-communication with the user, presenting him with the choice to play the jazz tunes. MMIL also has this nice additional feature of percolating lexical information at various levels of processing, which provides for the fall back strategy of lexical semantic specification. Essentially, our algorithm models resolution as a contextual (dynamic) and conceptual (static) mapping from the underspecified RE representations to a maximally specified cognitive description, which in our case are *Reference Domains* (cf. Figure 2).

The semantic representations as introduced in the previous section are assimilated into the type-theoretic models of RDs (Salmon-Alt, 2001). These domains are minimally identified by an *Id*, which serves as a domain index and *type*, which is extracted from the conceptual hierarchy accessible to the dialogue manager. The important features of a reference domain are its partitions, which reflect the cognitive viewpoint towards the domain. More specifically, these partitions in conjunction with focus and salience criteria define the accessibility criteria for appropriate referent profiling. The partition types are discursive cues such as role properties, perceptual information such as Haptic SelectionStatus, and/or conceptual cues such as domain level information.

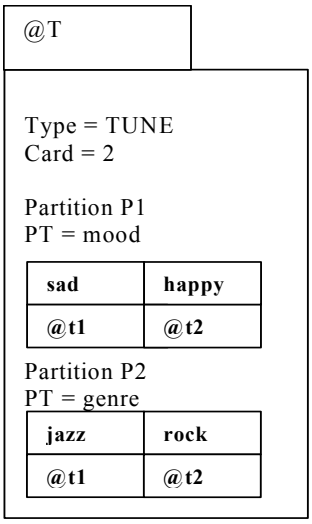

Figure 2 – Reference domain for a group of two tunes (@T)

The data structural representation for RD has the following form:

```
<ParentRefDomainObject>
        domainId
        domainType
        cardinality
        <Partition P1>
                PartitionTypeAttribute1
                <SetOf<PartTypeVal -
                ChildRefDomainObject>>
        <Partition P2>
                PartitionTypeAttribute2
                <SetOf<PartTypeVal -
                ChildRefDomainObject>>
        …
```

It is important to note here that depending upon the current discursive or perceptual state, there might not exist any partition within a RD. This is crucial so as to limit the accessibility of possible referents, as well as to provide fine-grained semantic resolution. During the dialogue progress, if certain partition is rendered out of scope by the resolution algorithm, it is deleted so that it does not lead to wrong extraction of the referent. Similarly, a tune set might not have any partition to begin with. However, by the usage of a discursive trigger like "the one by Madonna", it can be further resolved at the level of artist resulting in a new partition.

The MultiModalFusion (MMF) component of the Dialogue Module, maintains an incremental dialogue state by constructing under-specified RD representations for the referring expressions within the current utterance and by composing

[2] *maximality* refers to the most basic level of attributes associated with any object.

them with the existing context structure. The typical compositional operations are carried upon in the following stages:

1) Grouping: The under-specified RDs within a multimodal utterance are first evaluated for grouping. Based upon discursive triggers such as prepositions, conjunctions, disjunctions and/or perceptual triggers such as haptic gesture resulting in an item selection, RDs are grouped together if they match type, cardinality and temporal proximity constraints. For example, for the utterance "download the one by Madonna and this one + [haptic selection]", to begin with, the interpretation process results in 3 under-specified RDs: first for the definite RE, second for the demonstrative RE and third for the haptic event. Using the demonstrative cue and the temporal proximity, the resolution groups 2nd and 3rd RD, resulting in a further-specified RD, which is then composed with the 1st RD owing to the discursive trigger, i.e. the conjunctive *and*. The grouped RD has zero or one partition depending upon whether the 2 RDs have the same or different artists. It is to be noted that in this particular case the demonstrative is resolved at an early stage while the definite is still pending.

2) Assimilation: Depending upon the type of referring expressions, the context model tries to assimilate the under-specified RDs in differing but coherent ways (Salmon-Alt, 2001; Kumar, 2002). Essentially, owing to structural recursiveness and compositional nature, these RDs lead to a directed acyclic graph like context structure. The leaf RDs are at the level of maximum possible resolution at any stage of the dialogue processing. Firstly, a suitable node in the graph is selected. This selection is usually guided by two algorithms: the first one goes through the contextual domains, according to their activation level and starting with the most activated one, while the second one is intended to test the compatibility depending upon type, individuation, partition types etc. Secondly, the intended referent is extracted by profiling the sub-structure, resulting in re-structuring of the domains. For example, within an existing context of a tune list on the graphic display, the reference interpretation process for the speech utterance, " play the third song", would involve finding a node within the context structure representing an RD of tunes and having an index based partition of the member tunes.

In the following section, we provide a sample dialogue processing illustrating how this reference mechanism is useful for the MIAMM multimodal dialogue system framework.

4.4 Facilitating Dialogue Management

The following is a typical mixed-initiative dialogue within our framework:

(1)

U[1]: Play me the list I listened to this morning.

S[2]: Which one do you want to listen? + [displays a list of 2 tune-list items]

U[3]: the first one/ the one by Madonna.

S[4]: [plays the tune list]

U[5]: Save it/ * Save this/ *Save the list.

S[6]: [Saves]

To begin with, there exists a multimodal dialogue history for the system's perusal in the form of a stack of context structures, while the discursive and perceptual current context is empty. The definite RE in U[1], "the list" gives rise to an under-specified RD of type /entity-list/ (say, @L1). As @L1 holds predicative relationship with a past event, the reference interpretation algorithm evaluates existing context structures within the dialogue history to *assimilate* @L1, subject to the identification of a unique RD matching the type and predicative constraints. In this case, the system is able to locate 2 RDs (say @tL1 and @tL2) of type /tune-list/, which is subsumed by the type /entity-list/ in the conceptual hierarchy. Also, these 2 RDs match the predicative relational constraint as imposed on @L1 by U[1]. However, the possible referent is not unique, as it should be for identifying the target referent for a definite RE.

It is important to note that even though a referring action is intended to accomplish the referential communicative goal (Dale et al., 1995), i.e., to help the hearer in identifying the target referent, it might not always lead to the hearer identifying the referent as conceived by the speaker (Poesio et al., 2000). This is partly, because each agent involved in a dialogue can have potentially disparate knowledge resources and cognitive descriptions at his disposal.

Goodman (1986) characterizes various possible causes of miscommunication leading to an inappropriate or sub-optimal usage of referring expressions.

Within a multimodal setting, it is quite natural that miscommunications are frequent as it is strongly coupled to affordances (or rather, mis-affordances) of various modalities, as well as to the complexity of the multimodal context. Therefore, in order to impart robustness to any such system, it is imperative that the dialogue progress is incrementally enhanced in a non-monotonic way. In case of dialogue (1), the system retrieves the tune-lists which are in a predicative relation with any past event occurring /this morning/[3]. The RD @L1 thus obtained, is partitioned along the partition type of /event-Type/. The RD within this partition corresponding to the partTypeValue, /played/[4] is profiled as the possible referent and a list of 2 items is displayed along with an information-seeking speech response. This also brings the sub-structured partition under focus, implying that the objects within this partition are most likely to be referred by the user in the subsequent utterances, provided the dialogue continuity is maintained (Brennan, 2000).

In U[3], the user makes the referring action depending upon which attribute is in his perceptual context i.e. either indexicals such as "the first one", the domain attributes such as "the one by Madonna" or deictic such as [a haptic selection]. While in other scenario, say (2), the user after getting this response from the system, can recognize his mistake, rephrasing his actual request in U[3] as, "No, the one I downloaded".

Our reference and context model captures these dialogue intricacies in a coherent manner. In the first scenario, the system builds an under-specified RD for the RE, say "the one by Madonna", having /entity/ as type and /Madonna/ as an absolute modifier – a domain attribute. The activated partition of @L1, contains objects which match in type[5] with the under-specified domain. If there exits any tune-list by Madonna within this partition, the partition is further partitioned into a new partition, profiling the RD having Madonna as an artist, as the identified referent and bringing it under focus. In the other scenario, the RD corresponding to the partTypeValue, /downloaded/ is profiled and focussed. Besides, it is also evident that in U[5], the usage of pronominal is the most optimal one, as a pronominal RE marks monotonic dialogue continuity.

Thus structurally, the notion of reference domains allows transversal as well as horizontal access and update mechanisms. This enables the reference model to mimic the non-monotonic nature of dialogues, resulting into a unified description as provided by multiple modalities at the same time maintaining unified multimodal semantics.

5 Conclusions

We have outlined a reference resolution mechanism based on the cognitive grammar approach. The discussion is by no means exhaustive and complete owing to space limitations. Besides, our main objective here is to illustrate how this mechanism can be seamlessly integrated into a dialogue framework especially in a multimodal setting. Also, we argue that the particular choice of reference mechanism does have some important implications for dialogue management. In this light, it is agreeable that the reference model can be used towards building and updating dialogue structure (Seville 1999). Still, it remains to be seen how this model handles further complicated dialogue issues such as *conceptual entrainment* (Brennan 2000), use of absolute vs relative modifiers, mutual grounding etc. As a future activity, we plan to take up these issues by subsequent evaluation of our algorithm with respect to various reference phenomena encountered in a multimodal dialogue system framework.

6 References

Alexandersson J. and Becker T. (2001). Overlay as the Basic Operation for Discourse Processing in a Multimodal Dialogue System In: *Proceedings of the IJCAI Workshop ``Knowledge and Reasoning in Practical Dialogue Systems*, Seattle.

Allen J., Miller B., Ringger E., and Sikoski T. (1996). A robust system for natural spoken dialogue. In *Proceedings of the 14th Annual Meeting of the Association for Computational Linguistics* (ACL '96), June.

Asher N. (1993). *Reference to Abstract Objects in Discourse.* Kluwer Academic Publishers. Dordrecht, Boston, London.

Bos J., Mineur A-M., and Buitelaar P. (1995). *Bridging as Coercive Accommodation.* Technical

[3] We follow similar mechanism for temporal reference resolution
[4] user request for /listen/ corresponds to system action of /play/
[5] based on the subsumption criteria.

Report Number 52, Department of Computational Linguistics, Universität Saarbrücken.

Brennan S. (2000). Processes that Shape Conversation and their Implications for Computational Linguistics. In *38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, China, 1-8 October 2000.

Bunt H., and Romary L. (2002). Towards Multimodal Content Representation. In *International Standards of Terminology and Language Resources Management*, LREC 2002, Las Palmas (Spain).

Byron K. D. and Allen F.J. (2002). What's a Reference Resolution Module to do? Redefining the Role of Reference in Language Understanding Systems. In *the proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium*.

Cassell J. (2001). "Embodied Conversational Agents: Representation and Intelligence in User Interface" *AI Magazine*, Winter 2001, 22(3): 67-83.

Dale R. (1992). *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes.* MIT Press.

Dale R. and Reiter E.(1995). Computational Interpretation of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 19(2):233—263.

Dale R. and Reiter E. (1996). The Role of Gricean Maxims in the Generation of Referring Expressions. *Proc. of the 1996 AAAI Spring Symposium on Computational Models of Conversational Implicature,* Stanford University, California.

Deemter v. K. and Peters S. (1996). *Semantic Ambiguity and Underspecication*. Stanford: CSLI.

Goodman B.A. (1986). Reference Identification and Reference Identification Failures. *Computational Linguistics*, 12:273-305.

Grosz B.J. and Sidner C. (1986). Attention, Intention and the Structure of Discourse. *Computational Linguistics,* 12, 175-204.

Grosz B.J., Joshi A.K., and Weinstein S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics,* 12(2), 203-225.

Hahn U. and Strube M. (1997). Centered Segmentation: Scaling Up the Centering Model to Global Discourse Structure. *Proc. of EACL/ACL'97,* Madrid, 104-11.

Hobbs J. (1986). Resolving pronoun reference. In *Readings in Natural Language Processing*, pages 339-352. Morgan Kaufmann.

Kamp H. and Reyle U. (1993). *From Discourse to Logic.* Kluwer Academic Publishers. Dordrecht, Boston, London.

Kumar A. (2002). *Dialog Module Technical Specification*. Project MIAMM – Multidimensional Information Access using Multiple Modalities. EU project IST-20000-29487, Deliverable D5.1. LORIA, Nancy.

Kumar A. and Romary L. (2003). A Comprehensive Framework for Multimodal Meaning Representation. In *International Workshop on computational Semantics (IWCS-5)* , Tilburg, Netherlands.

Langacker R.W. (1986). *Foundations of Cognitive Grammar*. Stanford University Press. Stanford.

Langacker R.W. (1991). *Concept, image, and symbol : the cognitive basis of grammar*. Mouton de Gruyter.

McCoy K.F. and Strube M. (1999). Taking Time to Structure Discourse : Pronoun Generation Beyond Accessibility. *Proc. of the 21th Annual Conference of the Cognitive Science Society.* Vancouver, Canada, Aug. 19-21, 1999.

Pinkal M. (1999). On Semantic Underspecification. In: Bunt, H./Muskens, R. (Eds.). *Proceedings of the 2nd International Workshop on Computational Linguistics (IWCS 2)*.

Poesio M. and Reyle U. (2000). Underspecification in Reference: Some Evidence from Corpora, *Proc. of the KR-2000 Workshop on Semantic Approximation, Granularity, and Vagueness*, Breckenridge, April.

Reithinger N., Lauer C., and Romary L. (2002). MIAMM: Multidimensional Information Access using multiple modalities, In *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, 28-29 June, Copenhagen, Denmark, 2002.

Romary L. (2002). *MMIL technical specification*. Project MIAMM – *Multidimensional Information Access using Multiple Modalities.* EU project IST-20000-29487, Deliverable D6.3. LORIA, Nancy.

Salmon-Alt S. (2001). Reference Resolution within the Framework of Cognitive Grammar. *International Colloquium on Cognitive Science*, San Sebastian, Spain, May 2001.

Schauer H. (2000). Referential Structure and Coherence Structure in *Proceedings of the TALN 2000, 7e conférence annuelle sur le traitement automatique des langues naturelles*, 16-18 October, Lausanne, Switzerland, p.327-336.

Seville H. and Ramsay A. (1999). Reference-based Discourse Structure for Reference Resolution. *ACL'99 Workshop on Discourse Structure and Reference*. University of Maryland, June.

Tetreault J. R. (2001). A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, 27(4):507-520.

Vieira R. and Poesio M. (2000). An Empirically-Based System for Processing Definite Descriptions, *Computational Linguistics*, 26/4. 525-579.