

From Speech Acts to Search Acts: a Semantic Approach to Speech Acts Recognition

Marc Cavazza

University of Teesside
Borough Road, TS1 3BA, Middlesbrough, United Kingdom
m.o.cavazza@tees.ac.uk

Abstract

This paper describes the implementation of a human-computer dialogue system based on speech acts. The system has been developed as part of a conversational character for Interactive Television that assists the user in choosing TV programmes from an on-line Electronic Programme Guide (EPG). We have defined a set of specialised speech acts to account for the fact that dialogue actions correspond to the construction of a programme description. These speech acts can be recognised in the course of dialogue by comparing the semantic content of the user utterance with the search filter constructed from the previous dialogue history. The first step consists in parsing the user input to produce a semantic representation. This semantic representation is used to generate a search filter for the EPG. User replies are matched to previous search filters to determine speech acts for acceptance, rejection or refinement of the current selections. These mechanisms are illustrated with example dialogues from the system.

1. Introduction

Speech Acts theory [Austin, 1962] [Searle, 1969] [Récanati, 1979] [Berrendonner, 1981] provides a framework for the implementation of dialogue systems for Information Access applications, as it helps breaking down the information exchange between the user and the system into minimal dialogue units. Recent work in speech acts-based dialogue systems has emphasised the definition of core speech acts [Poesio and Traum, 1997], the specialisation of speech acts according to the task at hand [Busemann et al., 1997] (see also [Lee and Wilks, 1996] [Bunt, 1989]) and the importance of speech acts identification [Traum and Hinkelman, 1992]. The latter point is especially relevant in relation with acceptance and rejection of proposals [Walker, 1994; 1996].

In this paper, we describe the implementation of a dialogue system based on speech acts for Interactive Television. This system is a conversational character, which assists the user in the selection of programmes (Figure 1). The rationale behind the use of dialogue is that it enables users to concentrate on single programme features at each dialogue turn and to refine their selection according to previous results and system's suggestions. During dialogue, the system constructs and updates a filter corresponding to user preferences as these are incrementally refined through dialogue. This filter is used to search the programme database (or Electronic Programme Guide, EPG), which is a hierarchical structure of standard editorial categories defined by broadcasters. More specifically, we will discuss the relations between the semantic content of user utterances, the identification of user speech acts and the subsequent updating of the search filter. Despite the traditional definition of speech acts, Berrendonner [1981] and Reacanati [1979] have actually suggested that utterances can qualify as speech acts on the basis of their representational content.

We will first describe the specific set of core speech acts we have defined for our application. We will then show how these task-oriented speech acts can be identified by comparing the semantic contents of the latest user utterance with that of the search filter.



Figure 1. The Virtual Interactive Presenter.

2. Definition of Task-oriented Speech Acts

We have retained most of core speech acts as previously defined e.g. in [Traum and Hinkelman, 1992] such as *yes-or-no-questions*, *request*, *accept* and *wh-questions*. However, we have further specified those speech acts that are connected to the search process, such as *inform* or *reject*. Instead of *inform*, we describe *initial* and *specify* speech acts. Instead of a single rejection speech act, we distinguish between rejections and alternatives and we further refine rejection according to the rejected category. The rationale for using specialised speech acts is that the system's response as well as some dialogue control mechanisms are easier to define in terms of a larger set of speech acts.

Specifications are detected each time the new utterance provides previously non-existing information that is compatible with the information currently available. That is to say, adding a compatible subgenre to a programme genre, adding specific features such as cast or parental rating to a programme description. These do not have to follow a strict top-down refinement.

Rejections are identified each time incompatible features between the current utterance and the search filter are detected. Explicit rejections are introduced by a set of illocutionary speech acts whose surface form most often

include negation (e.g., “I don't want to watch that”). The system also allows some explicit rejections in the form of negative comments (e.g., “this is daft”). However, an important form of rejection is constituted by indirect speech acts. Such rejections, as identified by Searle [1975], “constitute a rejection of the proposal, but not in virtue of its meaning”. A typical example occurs when the user requests a different movie genre than the one currently considered. Rejections are subsequently typed according to the category or feature rejected.

Alternatives express variants of the current selection. They have to be interpreted contextually. For instance, “can I have another western?” expresses an alternative choice for the instance, which does not reject the subgenre (western). The alternative “can I have another movie?” is slightly more ambiguous, in that it can imply a rejection of the programme subgenre (i.e., the movie genre). Indefinite alternatives (“can I have something else?”) cannot be interpreted as such and require additional feedback from the user. Finally, the sentence “Do you have another movie with James Woods?” would reject the top selection, and potentially specify the search, if cast was not already a search feature.

In the remainder of the paper, we concentrate on those speech acts which have a direct impact on database search, mainly rejections, refinements and alternatives. Though traditional speech acts such as greetings, openings and wh-questions are part of the system, we will not discuss them in this paper.

3. From Semantic Analysis to Speech Act Identification

The first step consists in analysing the user utterance to produce a semantic representation. The semantic representation is a feature structure based on semantic categories that correspond to the taxonomic categories of the EPG (e.g. movie, documentary, news...). To this extent, semantic content is more important than logical form, as even the relations that structure these feature representations (e.g. “cast” or “audience”) are semantic in nature. The semantic features correspond to the set of editorial categories in the EPG, enhanced with inference mechanisms that can derive categories from connotations (e.g. “entertaining”) and infer relevant features (such as deriving the parental rating from the audience). Sentence analysis involves both a syntactic and a semantic step. The syntactic formalism used is a simplified variant of Tree-Adjoining Grammars, whose trees are enriched with semantic features [Cavazza, 1998]. However, the sole purpose of syntax is to ensure propagation of semantic features according to the tree operations carried out during parsing. The system is designed to accept partial parses: in that case the partial semantic structures are aggregated on the basis of feature compatibility. Semantic structures obtained from user input analysis are represented in the various examples illustrating this section as LISP feature structures immediately following the user utterance (see also Figure 2).

These semantic structures are used to instantiate a search filter, which takes the form of a partially instantiated EPG programme record. To this extent, the semantic features in the semantic representation are matched with the Electronic Programme Guide categories. Semantic features that correspond to EPG categories are

assembled into a search filter, which can include both positive and negative criteria. This search filter is incrementally refined and updated throughout the dialogue process, to reflect the progressive specification of the user's choice through the course of dialogue.

The next step consists in identifying the user's speech act. In the general instance, speech act identification is a complex process that involves surface forms, semantic content and dialogue pragmatics. Original work on speech acts refers to surface structure for their identification [Searle, 1975], but this is also because in the traditional conception semantic content is denied a role in speech act characterisation. More recent proposals in Computational Linguistics combine surface signals with other heuristics [Hinkelman and Allen, 1989]. Walker [1994; 1996] has specifically studied the recognition of utterance rejection, though not explicitly referring to speech acts, and bases this recognition on content comparison. We thus claim that speech acts, as we have defined them, can be identified by comparing the semantic content of the current utterance with the global filter constructed so far. In other words, the semantic difference between the new filter and the current filter can be used to identify speech acts. A total of 25 rules are used for speech act identification by comparing the semantic contents of the user utterance and the search filter. These rules define speech act recognition in terms of their semantic content differences. For instance, if the latest filter instantiates the same EPG category with a different feature, it consists in a rejection of that category (see examples below). In addition, a distinction is established between speech acts that can be immediately followed by a new EPG search and those who require additional information (characterised as `no_search`). For instance, a speech act rejecting a high-level category without providing a replacement one cannot trigger a new EPG search.

Additional heuristics are used to identify a direct query targeted at a selected programme instance, from the illocutionary nature of the user reply (e.g., “what is its rating?” “who is starring?”, etc.), using wh-questions, anaphora and definiteness. These direct speech acts are identified through a set of heuristic rules, which track the explicit mention of a programme feature (e.g., “cast”, “parental rating”, “starting time”) while a specific programme instance is considered (as signalled through pronouns, determiners and deictics). They are not used for filter comparison.

In the next sections, we illustrate different speech acts on sample dialogues (these correspond to actual system runs in its current development status). Each user utterance is followed by its corresponding semantic structure, in the form of a feature structure obtained from the LISP implementation of the dialogue system. The search filter derived from this structure is next, in the form of an attribute-value pair. Finally, the specific speech act recognised from that utterance is presented. Speech acts are represented as tuples, the first item in the list being the speech act category. For the particular `initial` speech acts, which open the dialogue, the tuple includes the initial search filter as well. System replies (labeled “S”) have sometimes been edited for the sake of clarity, as the natural language generation component is still under development.

U1: Do you have any action movies?
 ((QUESTION) (EXIST)
 (PROGRAMME ((GENRE MOVIE)
 (SUB_GENRE ACTION) (INDET))))
Filter: ((GENRE MOVIE)
 (SUB_GENRE ACTION))
SA: (INITIAL ((GENRE MOVIE)
 (SUB_GENRE ACTION)) SEARCH)

S2: *There are 5 action movies. The first one is: "Raw Deal"*

U3: I'd like a movie I can watch with my kids
 ((REQUEST)
 (OBJECT ((GENRE MOVIE) (INDET)))
 (AUDIENCE USER))
 ((QUESTION)) ((VIEW))
 (AUDIENCE CHILDREN)
 (PAR_RATING FAMILY) (POSS)))
Filter: ((GENRE MOVIE)
 (SUB_GENRE ACTION)
 (PAR_RATING FAMILY))
SA: (SPECIFY PAR_RATING FAMILY SEARCH)

S4: *Here is "Last Action Hero"*

In this first example, the user opens the dialogue by requesting a specific movie genre. In response to a first selection proposed by the system as a result of the EPG search, the user specifies an additional feature (parental rating), though through an *indirect* speech act (describing the audience). While this utterance is also a rejection of the proposal, the speech act is recognised as a specification, because parental rating had not been previously grounded during dialogue. The system recognises the specification by detecting a new feature in the incoming filter. The system actually rejects the current selection in its subsequent search as it is incompatible with the parental rating introduced.

U5: Can I see a western tonight?
 ((QUESTION)
 (SUBJECT ((AUDIENCE USER))
 (CHOICE+ ((VIEW))
 (PROGRAMME ((SUB_GENRE WESTERN)
 (INDET))))))
Filter: ((SUB_GENRE WESTERN))
SA: (INITIAL ((SUB_GENRE WESTERN)) SEARCH)

S6: *Here is the western "Unforgiven"*

U7: Can I have a thriller?
 ((QUESTION) (EXIST)
 (PROGRAMME
 ((SUB_GENRE THRILLER) (INDET))))
Filter: ((SUB_GENRE THRILLER))
SA: (REJECT SUB_GENRE THRILLER SEARCH)

In this second example, the user requests a different movie genre from the one already in the filter following his/her opening request. Because the movie genre was previously mentioned in dialogue, this is recognised as a rejection of the movie genre, through an indirect speech act. This is detected by comparing the latest filter with the

aggregated filter that was a result of the previous utterance. This comparison explicitly detects a new subgenre, incompatible with the previous one. However, because a new category is provided, the system can trigger a new EPG search that would correspond to the updated filter (this is represented by the `search` feature in the speech act representation).

U8: Do you have any thrillers?
 ((QUESTION) (EXIST)
 (PROGRAMME
 ((SUB_GENRE THRILLER) (INDET))))
Filter: ((SUB_GENRE THRILLER))
SA: (INITIAL
 (SUB_GENRE THRILLER) SEARCH)

S9: *Here is the thriller "the Specialist"*

U10: Can I have another movie with James Woods?
 ((QUESTION) (SUBJECT ((AUDIENCE USER))
 (CHOICE+ ((VIEW))
 (PROGRAMME
 ((GENRE MOVIE) (OTHER-CHOICE+))
 (INSTR ((CAST WOODS)))))
Filter: ((GENRE MOVIE) (CAST WOODS))
SA: (OTHER GENRE SEARCH)

This example illustrates the treatment of alternatives (other speech acts). Namely, that the alternative rejects the proposed movie instance, which is deleted from the selection, and only retains the most generic feature ("movie") deleting the subcategory as well ("thriller"). The alternative also refines the selection by introducing a specific cast. This is actually taken into account when updating the filter, a process that takes place after speech act recognition. In this particular case, the alternative has enough information for the EPG new search to be directly triggered.

U11: Is there a movie with John Wayne?
 ((QUESTION) (EXIST)
 (PROGRAMME
 ((GENRE MOVIE) (INDET))
 (INSTR ((CAST WAYNE)))))
Filter: ((CAST WAYNE))
SA: (INITIAL ((CAST WAYNE)) SEARCH)

S12: *Here is the western "Rio Bravo"*

U13: I don't want a western
 ((CHOICE+) (NEGATIVE)
 (AUDIENCE ((AUDIENCE USER))
 (PROGRAMME
 ((SUB_GENRE WESTERN) (INDET))))
Filter: ((SUB_GENRE WESTERN) (NEGATIVE))
SA: (REJECT SUB_GENRE NIL NO_SEARCH)

This is a simple and explicit rejection of the movie genre. In this case we find more appropriate that the system returns to the user asking him to specify the subcategory of his choice. This avoids that the system enumerates all possible choices, which can be painful for the user.

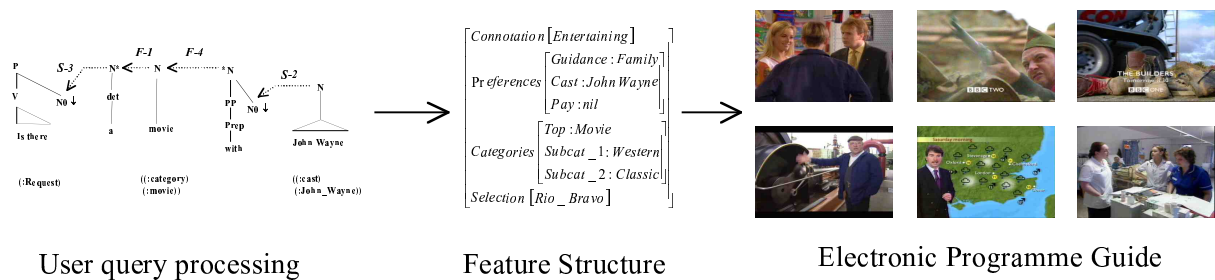


Figure 2. The Natural Language Processing step.

4. Conclusions

We have presented a semantic approach for the recognition and interpretation of speech acts in an Information Access application. This application shares similarities with previous systems [Nagao and Takeuchi, 1994], [Sadek, 1996] [Beskow and McGlashan, 1997], though it proposes a unified treatment of speech act recognition and database search, based on the semantic contents (rather than logical structure) of the dialogue units. This approach appears appropriate for the processing of indirect speech acts as well. The system is however still under development and the above results should be re-assessed in the context of the full EPG database.

Acknowledgements

The Virtual Interactive Presenter is a LINK Broadcast project funded by the DTI. Steve Francis is thanked for Figure 1. EPG contents have been provided by the BBC.

5. References

Austin, J. (1962). *How to Do Things with Words*. Oxford, Oxford University Press.

Berrendonner, A. (1981). *Elements de Pragmatique Linguistique*. Editions de Minuit, Paris (in French).

Beskow, J., and McGlashan, S. (1997). Olga: A Conversational Agent with Gestures. In: *Proceedings of the IJCAI'97 workshop on Animated Interface Agents - Making them Intelligent*, Nagoya, Japan, August 1997.

Bunt, H.C. (1989). Information dialogues as communicative action in relation to information processing and partner modelling. In: Taylor, M.M., Néel, F. and Bouwhuis, D.G. (Eds.), *The Structure of Multimodal Dialogue*, Amsterdam, North-Holland.

Busemann, S. Declerck, T., Digne, A., Dini, L., Klein, J. and Schmeier, S. (1997). Natural Language Dialogue Service for Appointment Scheduling Agents. In: *Proceedings of ANLP'97*, Washington DC.

Cavazza, M. (1998). An Integrated TFG Parser with Explicit Tree Typing. In: *Proceedings of the fourth TAG+ workshop*, IRCS, University of Pennsylvania.

Hinkelman, E.A. and Allen, J.F. (1989). Two Constraints on Speech Act Ambiguity. *Proceedings of ACL-89*, pp. 212-219.

Lee, M. and Wilks, Y. (1996). An ascription-based approach to speech acts. In: *Proceedings of the 16th*

International Conference on Computational Linguistics (COLING'96), Copenhagen.

Nagao, K. and Takeuchi, A. (1994). Speech Dialogue with Facial Displays: Multimodal Human-Computer Conversation. In: *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL'94)*, pp. 102-109.

Poesio, M. and Traum, D. (1997). Representing Conversation Acts in a Unified Semantic/Pragmatic Framework. In: *Proceedings of the AAAI Fall Symposium on Communicative Actions in Humans and Machine*, Cambridge (MA).

Récanati, F. (1979). *La transparence et l'énonciation*, Editions du Seuil, Paris (in French).

Sadek, D. (1996). Le dialogue homme-machine: de l'ergonomie des interfaces à l'agent dialoguant intelligent. In: J. Caelen (Ed.), *Nouvelles Interfaces Homme-Machine*, OFTA, Paris: Tec & Doc (in French).

Searle, J. (1969). *Speech Acts: an Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J.R. (1975). Indirect Speech Acts. In: P. Cole and J.L. Morgan (Eds.), *Syntax and Semantics*, vol. 3: *Speech Acts*, pp. 59-82, New York, Academic Press.

Traum, D. and Hinkelman, E.A. (1992). Conversation Acts in Task-Oriented Spoken Dialogue. *Computational Intelligence*, vol. 8, n. 3.

Walker, M.A. (1996). Inferring Acceptance and Rejection in Dialogue by Default Rules of Inference. *Language and Speech*, 39-2.

Walker, M.A. (1994). Rejection by Implicature. In *Proceedings of the 20th Meeting of the Berkeley Linguistics Society*,