Natural Language Generation without Intentions

Arne Jönsson[1]

Abstract.

Response generation for natural language interfaces can not be seen as an isolated activity. It must be designed to facilitate cooperative user-friendly interaction. One important source of information for such a natural language generator is a model of the ongoing dialogue. There are two major approaches to dialogue modeling. One approach is based on reasoning about the goals and intentions behind a user initiative, whereas the other relies on a dialogue grammar specified from the functional role of a move. The latter approach does not allow reasoning on user intentions and goals, and will often produce less sophisticated responses. On the other hand, identifying a user's goals and intentions is not a straightforward task; furthermore, the intention-based approach relies on complex plan recognition. For many natural language applications the grammar-based approach is sufficient and it is not necessary to recognize the intentions behind a user initiative. However, if the goal is to mimic human interaction a grammar-based approach will not be accurate enough.

1 Introduction

User-friendly cooperative response generation for natural language interfaces can not rely solely on the information provided in isolated utterances. It needs to consult a model of the ongoing dialogue to utilize information on focus and dialogue structure. Research on such computational models of discourse can be motivated from two different standpoints. One is to develop general models and theories of discourse for all kinds of agents and situations. The other approach is to account for a computational model of discourse for a specific application, say a natural language interface. It is not obvious that the two approaches should present similar computational theories for discourse. Instead the different motivations should be considered when presenting theories of dialogue for natural language interfaces.

There are no studies showing that natural language interfaces should try to mimic human communication or that a generated response must resemble a human response. On the contrary, such interfaces will not only be slow, they will also provide the user with an erroneous model of its capabilities. Instead, response generation must adhere to the behaviour and capabilities of the natural language interface and produce helpful responses based on the application and the role of the agents.

Three different dialogue types can be distinguished [14]: Task dialogue, where the system guides the user's actions, examples of this is the pump assembly task, Planning dialogue, where the system assists in planning the user's actions, and Parameter dialogue, where the user's task is not known to the system, an example of this is database access. Task dialogues often need to consult a user model which may require more sophisticated dialogue models. The class Simple Service Systems [8], which can be said to incorporate both Planning and Parameter dialogues, define an important application domain for natural language interfaces. Such systems require in essence only that the user identify certain entities, parameters of the service, to the system providing the service, and once they are identified the service can be provided [8].

Essential design properties for natural language interfaces are habitability [19] and transparency. A natural language interface must clearly show the user which actions it is able to perform, which initiatives it can respond to, which it cannot respond to, and why this is the case, and also the capabilities of the underlying background system. Another important property is efficiency. The natural language interface should not slow down the interaction with the background system noticeably.

2 Computational Models of Discourse

The problem of discourse modeling can be divided into managing three structures [6]: the linguistic structure, the attentional state and the intentional structure.

The details on a component which records the objects, properties and relations that are in the focus of attention, the attentional state, need careful examination, but will not be elaborated upon in this paper. For simple service systems applications a simple model copying information from one segment to the next is often sufficient [10].

The role that is given to the intentional state, i.e. the structure of the discourse purposes, and to the linguistic structure, i.e. the structure of the sequences of utterances in the discourse, provides two orthogonal [16] approaches to dialogue management:

- One approach is what can be called the plan-based, or intention-based [9], approach. Essential to this approach is the modeling of the user's intentions and purpose of participating in the discourse. The linguistic structure is used to identify the intentional state in terms of the user's goals and intentions. These are then modeled in plans describing the actions which may possibly be carried out in different situations. The basic formalism [4] has been extended in many ways to handle various phenomena observed in human communication.
- In the other approach to dialogue management, termed the grammar-based approach, utterances are interpreted from the linguistic structure on the basis of their functional relation to the previous interaction. This approach relies on the assumption that the structure of the conversational moves can be used to model the dialogue, (cf. conversational analysis). For instance, utterances often occur in pairs, e.g. an answer follows a question. The identification of the users' goals is still an important issue. However, this can be done without reference to a speaker's underlying intent. The constraints

[1] Department of Computer and Information Science, Linköping University, S-581 83 LINKÖPING, SWEDEN, arnjo@ida.liu.se

Proceedings of the ECAI 96 Workshop *Gaps and Bridges: New Directions in Planning and Natural Language Generation.*
Edited by K. Jokinen, M. Maybury, M. Zock and I. Zukerman.

on a move can be determined from speech act information only which in turn is modeled in a dialogue grammar.

To illustrate the two approaches consider dialogue example 1. In the plan-based approach, understanding the first utterance is assumed to rely on knowledge of plans describing how to board a train, how to make a train journey, select a train, buy a ticket, etc. These plans are then used to realize that the user's goal is to make a train journey and thus, needs to select a train and that therefore the first utterance requests information about a departure time, in this case, of the train to Ottawa.

Passenger: Trains going from here to Ottawa?
Clerk: Ottawa. Next one is at four-thirty.
Passenger: How about Wednesday?
Clerk: One at nine thirty,....

Dialogue example 1 (from [13, p. 108]).

In the grammar-based approach the interpretation of an utterance stops after having identified the functional role of the linguistic structure, i.e. the speech act. In the example we can identify two speech acts. Let us call them Request and Inform. We assume that a system which acts as a clerk in a train station has the overall goal of responding to user initiatives, in this case to answer questions from the customer, i.e. to respond to a Request with an Inform. This is modeled in a rule in a dialogue grammar which is used to capture the information that if the first utterance is a Request the response is to be an Inform. Thus, the first initiative, the Request for information about trains to Ottawa will result in a response of type Inform which provides the requested information. The second utterance is treated similarly after resolving the ellipsis.

One motivation for reasoning about the user's goals and intentions is to be able to account for common phenomena from human communication, such as emotions. To illustrate this, consider the utterance *With 269 people on board?* in dialogue example 2.

Speaker 1: The Korean jet shot down by the Soviets was a spy plane.
Speaker 2: With 269 people on board?

Dialogue example 2 (from [3, p. 13]).

The problem is how to correctly communicate the doubt conveyed by that utterance. This can hardly be done if using only information on the functional role of the speech act. Such a system would probably end up answering *Yes*. On the other hand, by using plan recognition techniques in conjunction with discourse goal rules this can be accomplished by recognizing violations in the intentions and plans that the user intended to convey [3].

There are two major problems with the plan-based approach. One is the problem of identifying the primitives needed. In the plan-based approach, the user's intentions and goals must be identified to determine the relevant plans to be used in a certain application. However, it is not always clear which goals the users pursued in the interaction. In a user-advisor Wizard of Oz experiment two different coders tried to derive the dialogue structure from recognizing the users' goals and transitions in these goals and a mapping onto a task structure [7]. The inter-rater reliability between them was in some cases only 72%. This means that the coders frequently could not agree on which goal a user intended in a certain situation. If this is a difficult task for humans it is presumably even more difficult for computers. This result poses serious problems for the development of natural language systems that rely on the notion of a plan described as fulfilling a user's goal.

Efficiency is also a problem for plan recognizers. Central to the plan-based approach is the recognition by the listeners of the speaker's goals, where goals are modeled using plans. The area of plan recognition has been less rigorously studied than planning, but is considered an even more difficult task than planning [3]. General STRIPS-like planning is undecidable, but the complexity of the plan operators can be restricted to achieve tractable results. However, it is not possible to construct a polynomial-time planning algorithm for the more restricted class of problems, named the SAS-PU class [2], which probably is too restricted for practical use in natural language processing. In SAS-PU, for instance one action achieves only one effect in the world and every operator has only one effect in the world.

Removing the ability to recognize new plans by chaining together the preconditions and effects of other plans [12] also provides plans that can be recognized in polynomial time [18]. This restricts the flexibility of plan recognition, but would otherwise lead to massive increase in the size of the search space [12].

For a dialogue grammar, provided it can be written using a context-free grammar (or a grammar of less complexity), there are well-known polynomial-time algorithms that can be used for parsing. This argument also accounts for grammar formalisms utilizing feature structures, as long as the grammar is small and the number of categories limited.

3 Generating without plans

Utilizing only a dialogue model is not enough to generate user-friendly cooperative natural language responses. There are a variety of other means, not requiring an intention based model, to support the generation process, some of which will briefly be mentioned in this section.

In a series of experiments on natural language interaction for simple service systems we have utilized a principle which we term The Quantity Principle: *The system may give more information to the user than has actually been requested provided it is potentially relevant* [1, 5]. The principle can be motivated from the fact that the user reads and understands natural language at an adequate speed and is able to select information on the basis of relevance. Moreover, for information in tabular form, selection does not require excessive reading, either [1]. In our empirical investigations this was pointed out as a good feature of the system, as having all relevant information in a single table facilitated comparisons and evaluations [10].

Again consider dialogue example 1. The information provided by a computer will not necessary be the same as the one provided by a clerk at a train station. In typed interaction the response could instead be the time-table for trains going to Ottawa that particular day. That information would most probably fit on the screen and it could be provided instantly without sophisticated reasoning. In the example it might also be that the weekly time-table can be presented in one window, already as response to the first utterance, making the second request redundant.

A similar strategy can be used if a user investigates different properties about the same set of primary referents, objects, or the same property for various objects. In such cases, provided that the system utilizes tabular presentation, the information generated as response to the second utterance can be added to that of the first. In both of these cases the system provides more information to the user than she has actually asked for.

The Quantity Principle reduces the need to utilize sophisticated intention based strategies, as there is no need to further tailor the response to match the users' goals and intentions. It is, of course, most applicable when screen output is possible and especially for multi-modal generation, where a variety of modalities and multiple windows can be used [11]. For spoken interaction, for instance, it is less suitable. However, even such interaction can perform well without utilizing plan based models [20].

When presenting meta-knowledge and for applications such as argumentation systems where the organization of the knowledge base is to be explained, tabular presentation is less applicable. Systems of this kind need more advanced methods, taking into account features such as communicative goals and rhetorical structures. To some extent this can be achieved using schemata describing various aspects of the text to be generated [15].

Another aspect of generation is the cost for creating the knowledge bases required for sophisticated natural language interaction. This problem is addressed in IDAS [17], which mix a fixed set of rules with canned text generations. IDAS does not utilize plan based reasoning as the cost is too high. The cost could be reduced using control heuristics; however this also removes the systems abilities to respond appropriately in unusual situations. This strategy works well if the number of tasks to perform is small and fairly predictable. Furthermore, different applications demand different techniques [17].

4 Summary

User-friendly natural language interaction needs to consult a model of the ongoing dialogue to generate cooperative responses. Such dialogue modeling can be carried out either by utilizing only the functional role of a move, the grammar-based approach, or by also trying to convey the goals and intentions behind the move, the plan-based approach.

What is important is to carefully investigate the properties of the task and the user situation to choose the right dialogue model. If the goal is to mimic human language capabilities the plan recognition approach might be necessary, despite its complexity. There are also applications, such as Task dialogue systems, where a more sophisticated reasoning is required in order to produce meaningful responses. However, for the task of managing the dialogue in many natural language interface applications, for instance simple service systems, the less sophisticated approach of using a dialogue grammar will do just as well. This provides models that are more effective but less varied and which resembles less of human interaction, but human computer interaction need not always resemble human interaction. Instead the system must be efficient and also present the user with a model of its capabilities in order to facilitate habitabilty and transparency. This is especially important when generating responses to the user, i.e. not to produce responses which encourage the user to transgress the systems capabilities.

ACKNOWLEDGEMENTS

This work results from a research project supported by The Swedish Council of Research in the Humanities and Social Sciences (HSFR) and The Swedish National Board for Industrial and Technical Development (NUTEK) in the joint Research Program for Language Technology. The work has been carried out with the members of the Natural Language Processing Laboratory at Linköping University, Sweden, and I am especially indebted to Lars Ahrenberg, Nils Dahlbäck and Åke Thurée.

REFERENCES

[1] Lars Ahrenberg, Arne Jönsson, and Åke Thurée, 'Customizing interaction for natural language interfaces', in *Workshop on Pragmatics in Dialogue, The XIV:th Scandinavian Conference of Linguistics and the VIII:th Conference of Nordic and General Linguistics, Göteborg, Sweden*, (1993).

[2] Christer Bäckström and Inger Klein, 'Parallel non-binary planning in polynomial time', in *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence, Sydney, Australia*, pp. 268–279, (1991).

[3] Sandra Carberry, *Plan Recognition in Natural Language Dialogue*, MIT Press, Cambridge, MA, 1990.

[4] Philip. R. Cohen and C. Raymond Perrault, 'Elements of a plan-based theory of speech acts', *Cognitive Science*, **3**, 177–212, (1979).

[5] Nils Dahlbäck, 'Pronoun usage in NLI-dialogues. A wizard of Oz-study', in *Proceedings of the Third Nordic Conference on Text Comprehension in Man and machine, Linköping, Sweden*, (1993).

[6] Barbara J. Grosz and Candace L. Sidner, 'Attention, intention and the structure of discourse', *Computational Linguistics*, **12**(3), 175–204, (1986).

[7] Raymonde Guindon, 'A multidisciplinary perspective on dialogue structure in user-advisory dialogues', in *Cognitive Science and Its Applications For Human-Computer Interaction*, ed., Raymonde Guindon, Lawrence Erlbaum, (1988).

[8] Philip J. Hayes and D. Raj Reddy, 'Steps toward graceful interaction in spoken and written man-machine communication', *International Journal of Man-Machine Studies*, **19**, 231–284, (1983).

[9] Kristiina Jokinen, *Response Planning in Information-Seeking Dialogues*, Ph.D. dissertation, University of Manchester Institute of Science and Technology, 1994.

[10] Arne Jönsson, *Dialogue Management for Natural Language Interfaces – An Empirical Approach*, Ph.D. dissertation, Linköping University, 1993.

[11] Arne Jönsson, 'A dialogue manager for natural language interfaces', in *Proceedings of the Pacific Association for Computational Linguistics, Second conference, The University of Queensland, Brisbane, Australia*, (1995).

[12] Henry A. Kautz, 'A formal theory of plan recognition and its implementation', in *Reasoning About Plans*, eds., James F. Allen, Henry A. Kautz, Richard N. Pelavin, and Josh D. Tenenberg, 69–125, Morgan Kaufmann, (1991).

[13] Diane J. Litman, *Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues*, Ph.D. dissertation, University of Rochester, 1985.

[14] W. Van Loo and H. Bego, 'Agent tasks and dialogue management', in *Workshop on Pragmatics in Dialogue, The XIV:th Scandinavian Conference of Linguistics and the VIII:th Conference of Nordic and General Linguistics, Göteborg, Sweden*, (1993).

[15] Kethleen R. McKeown, 'Discourse strategies for generating natural-language text', *Artificial Intelligence*, **27**, 1–41, (1985).

[16] Rachel Reichman, 'Convention versus intention', in *The Structure of Multimodal Dialogue*, eds., M. M. Taylor, F. Néel, and D. G. Bouwhuis, 121–134, Elsevier Science Publishers B.V. (North-Holland), (1989).

[17] Ehud Reiter and Chris Mellish, 'Optimizing the costs and benefits of natural language generation', in *Proceedings of the International Joint Conference of Artificial Intelligence, Chambery, France*, pp. 1164– 1169, (1993).

[18] Marc Vilain, 'Getting serious about parsing plans: a grammatical analysis of plan recognition', in *Proceedings of the AAAI-90, Boston*, pp. 190–197, (1990).

[19] W. C Watt, 'Habitability', *American Documentation*, **July**, 338–351, (1968).

[20] Victor W. Zue, 'Toward systems that understand spoken language', *IEEE Expert*, **9**, 51–59, (1994).