ISCA Archive
http://www.isca-speech.org/archive

3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

AMBIGUITY AND UNCERTAINTY IN SPOKEN DIALOGUE

Paul Heisterkamp

Daimler-Benz AG, Institute for Information Technology
Wilhelm-Runge-Str. 11, D-89081 Ulm, Germany
heisterkamp@dbulml.uucp

ABSTRACT

In actual spoken dialogues, ambiguity is a special case of uncertainty. Uncertainty is the overall standard condition of a speech understanding system, not only because of possible deficiencies in the recognition, but also because people may make mistakes. The system has to have ways to deal with this condition. In SUNDIAL[1], the strategy chosen is to find the optimal interpretation with regard to context and situational setting, and to use the result of this interpretation in the following system utterance, so that mutual assurance of understanding what was meant is constantly given.

Keywords: speech dialogue, ambiguity, uncertainty, dialogue strategy, semantics, SUNDIAL.

10.21437/Eurospeech.1993-373

1. AMBIGUITY AND UNCERTAINTY

Ambiguity is a very common phenomenon in language. It arises when linguistic units can receive more than one interpretation in a given context. The context may range from some phones to whole texts, and accordingly ambiguity can be observed at every level from phonetics via semantics to pragmatics. Ambiguity has received special attention because here what leads to non-uniqueness in interpretation is inherent in the linguistic material itself, i.e. this material is taken not to contain any production or reception errors. In actual communication, however, be it between humans or between human and machine, this assumption can not be made. There are production as well as reception errors. People *do* mispronounce, use 'wrong' grammar and also use 'wrong' words. On the reception side, channel quality (noise, phone lines) or insufficiencies of the hearing apparatus also lead to errors. So, in actual spoken communication, ambiguity is a special case of uncertainty.

In human-human communication, these errors mostly go unnoticed for a number of reasons that commonly are subsumed under the label 'robustness' of human speech understanding (*cf.* /10/). The interpretation is guided by contextual expectations, huge masses of language and world knowledge and the whole range knowledge of the context and situation are fully exploited in interpretation, and finally the most plausible interpretation is achieved; still, this interpretation is not necessarily the one intended by the speaker. But as long as the reaction of the dialogue partner is appropriate, i.e. continues the communication in a non-deviative way, no explicit repair will take place. So, the robustness is only partly due to the quality of the processing and/or internal repair mechanisms. It is rather the continuation of the dialogue itself that serves as an implicit or explicit confirmation of the mutual understanding.

Any understanding system must have ways to deal with ambiguity and uncertainty. We will argue in the following that for a successful man-machine communication in spoken language, this constant mutual assurance of 'speaking of the same thing' is very helpful, if not crucial, for a dialogue system that provides easy and user friendly access to an application system.

The following example illustrates a syntactic (and consequently semantic) ambiguity that can be resolved using knowledge (or rather assumptions) of the situation and dialogue setting, and an ambiguity through underspecification that can not be resolved by internal knowledge:

(1) Tell me the trains from Ulm to Hamburg around eight o'clock tomorrow.

This sentence is syntactically ambiguous, in that the 'around eight o'clock tomorrow' can taken to modify either the 'trains to Hamburg' or the 'tell me', yielding two different readings, *viz.*:

(2) Tell me *the trains from Ulm to Hamburg around eight o'clock tomorrow.*

(3) Tell me *the trains from Ulm to Hamburg* around eight o'clock tomorrow.

Now, reading (3) obviously makes no sense in an interactive information system. Still, if syntactic processing allows that reading, at some stage it has to be decided to rule it out and go for reading (2) instead, i.e. take 'around eight o'clock tomorrow' to modify 'the trains to Hamburg'. The knowledge needed for this decision should be built into such a system.

The other ambiguity in (1) is the 'eight o'clock' time expression. It is underspecified in that it lacks an a.m./p.m. part which would make it clear whether it refers to '08.00 hours' or '20.00 hours'. Now, one can use a lot of internal inference like that the ride from Ulm to Hamburg takes seven hours

[1] The SUNDIAL system (Speech UNderstanding and DIALogue) was partly funded by the Commission of the European Communities as project 2218 in the ESPRIT-II programme (1989-1993). The partners in the project were Logica Cambridge (Vocalis), University of Surrey, CAP Gemini Innovation, CNET, Irisa, CSELT, Siemens, Universität Erlangen and Daimler-Benz.

and that it is unlikely that someone wants to arrive in the middle of the night, but still it might be the case that the caller has to work in Ulm until 19.30 hours and wants to go home to Hamburg that same evening. The internal reasoning can only give hints to make a more intelligent guess. The question is undecidable by internal means alone. The only way out of this is to ask the dialogue partner to specify exactly what is meant.

2. UNCERTAINTY IN A DIALOGUE SYSTEM

In a speech understanding computer system like SUNDIAL, every level can be a source of uncertainty. This starts from the recognition side. As the recognizer works on the basis of stochastic modelling (Hidden Markov Models and statistic language models), the words or sequence of words found in the input signal have per se only a certain probability to be the words that were actually spoken. Furthermore, the recognizer will always find the words (or non-words) it knows in any non-silence part of the signal, so that careful thresholding is applied to limit the number of word hypotheses and still make sure that a maximum of the actual words are still among the hypotheses. For the linguistic analysis, the same applies mutatis mutandis. The grammatical constructions a human will use are supposed to be known, as well as the semantics of the words used; and so on for the other stages of processing.

This technique is an example of so-called overcoded abduction (*cf.* /2, 6/): if the thresholding does not rule out the whole utterance, the system will always find *an* interpretation, namely the most plausible one, as it is assumed that the utterance to be processed is in the range of the system's knowledge. As this knowledge is limited, there never is a guarantee that this is really the case. Any interpretation thus must be treated as being a hypothesis, a working hypothesis can only be considered valid if it receives a confirmation from the dialogue partner.

In order for the dialogue to continue as smooth as possible to achieve the caller's goal in a minimum of dialogue turns and time, it is possible to interpret any non-contradiction as a confirmation: the dialogue continues without any correction from either side. The interpretation of non-contradiction as confirmation is only possible if the hypothetical interpretation is made known to the dialogue partner. The system must use the results of it's interpretation in its following utterance to give the caller knowledge of what it has understood and by this, to provoke contradiction (or correction) if it (or the caller) has got something wrong. In the following, we highlight some features of the SUNDIAL system that were built into the system to ensure an optimal interpretation on the one hand and on the other to implement the continuous confirmation of hypotheses, ensuring at the same time a maximum naturalness of the resulting dialogues.

3. CONFIRMATION IN SUNDIAL

SUNDIAL is a speech dialogue system for train and flight time table inquiries over the telephone. It is implemented for English, Italian, and German, using a common dialogue management component. A French system uses a more specified dialogue manager. The German system is configured for German train time table inquires.

As SUNDIAL is a dialogue system, it is able to perform extended continuous interaction with the user. This means, that is can use the full flexibility of its dialogue capabilities to ensure mutual coherence with the dialogue partner. Also, as it performs semantic interpretation in context (*cf.* /6/), it is able to deal with uncertainty in recognition and parsing by applying contextual knowledge to ensure optimal understanding.

In the actual implementation, the German version of the SUNDIAL system makes use of expectations to determine the choice of language models at the recognition side to find the best string of words in the utterance (*cf.* /1,8/). This string is then submitted to the parser, which first tries to parse it with a very restrictive grammar. If the parse fails, it loads a more 'robust' grammar and tries to find a parse again. Upon failure, it sends a respective message to the dialogue manager.

The SUNDIAL dialogue manager[2] is designed to react flexibly to different conditions. It first tries to incorporate the surface semantic description of the caller utterance into its contextual model. This may result into one of for types of change in the contextual model. Any sub-part of the utterance may either be new for the system, a repetition, a contradiction, or it may fail, i.e. not fit into the contextual model at all. These changes lead to an appropriate change in the goal states of the pragmatic component, as do changes coming from the application system interface, like a request for a new parameter or the delivery of a solution for the caller's request. From these goal states, the pragmatic component determines its next utterance, trying to move into the direction of reaching as may goals as possible under the constraint of not overloading the caller with information or different requests.

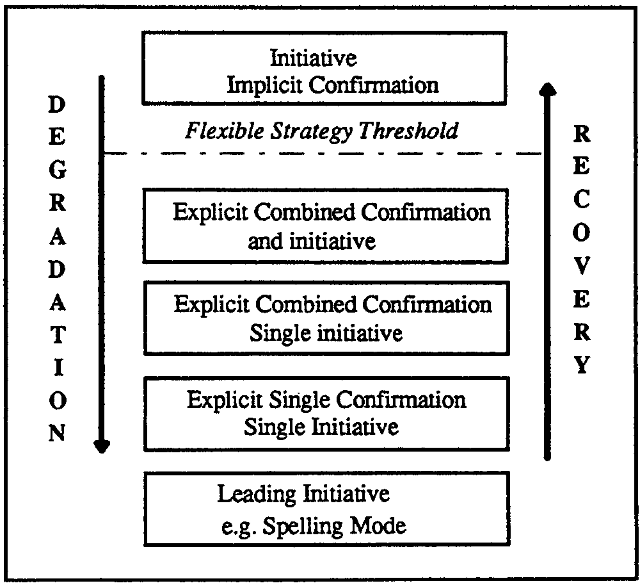

Fig. 1: Dynamic Dialogue Strategy

2 For an older implementation cf. /3/.

The type of goals and the possibility of their combination is determined by the current dialogue strategy. The dialogue strategy is one of an ordered set of possibilities that serve to determine the optimal continuation of the dialogue. It is ordered in that its setting corresponds to the minimal number of dialogue turns to be executed in order to solve the request of a caller, given that a caller mentions all the relevant parameters for the request in the first utterance. The choice of the current strategy follows a degradation and recovery meta-strategy, and the maximum strategy threshold can be set by the system developer, as shown in Figure 1. This flexible strategy enables the system to carry out a user-friendly dialogue.

The dialogue strategy's main result in any actual dialogue is the way that confirmations are handled by the system. The current strategy determines whether an item that has received a 'new_for_system' type change of the contextual model needs a confirmation or not, and whether that confirmation may be used together with the confirmation of another item or a new system initiative. The strategy is dynamic in that the system adjust its setting according to whether it encounters contradiction from the caller or not. In case the caller contradicts the system, it assumed that something in its understanding did go wrong, and so it tries to restrict the understanding task by asking more specifically, finally degrading to a 'Leading Initiative' mode, where the caller is asked not only to provide an isolated task parameter, but also to give it in some specific form that makes e.g. the recognition more reliable because a specifically restricted language model can be used.

The SUNDIAL system is perfectly able to deal with a dialogue without any confirmation turns. However, even actual human-human dialogues hardly go without confirmation requests (*cf.* /6/), and as it has been shown above, it is completely impracticable if there are ambiguities of the sort shown above in the caller utterance, especially underspecified time expressions.

So, we are using the strategy next in rank as the initial setting. Example 1 shows a SUNDIAL dialogue for the German system, starting from this setting. It is an actual dialogue from the current system implementation under laboratory conditions. There are no misunderstandings and consequent repair sequences:

system1:	This is the Automatic Intercity Inquiry System. Which information do you want?
caller2:	I need a connection from Essen to Hamburg tomorrow.
system3:	You want to travel from Essen to Hamburg. You want to travel on Wednesday, September 22nd. At what time do you want to leave?
caller4:	At 16.00 hours.
system5:	You want to leave around 16.00 hours?
caller6:	Yes.
system7:	I have found two connections for you. The first train leaves Essen at 15.45, 15.45, arriving in Hamburg main station 19.32. The second train leaves Essen at 16.18, 16.18, arriving in Hamburg main station 19.43. Do you have another inquiry?
caller8:	No thanks [Hangup]

Example 1: Combined Confirmation Strategy

The system makes it clear to the caller what it has understood by explicitly repeating the parameters provided by the user. The principle we follow here is to submit for implicit confirmation the parts of the caller's utterance that are important for the solution of the request, and to make this submission as specific as possible (*cf.* /7/). In utterance system3, it can be seen that the system internally expands the 'tomorrow' from caller2 to a full date expression, including the day of the week, day and month. The caller thus can check whether the system really means the day that was intended. This can be especially important when the caller calls around or shortly after midnight, when it is not always clear which is 'tomorrow'. In system7, the departure time is given twice, as this is the most important information for the caller.

Example 2 illustrates how the system degrades its strategy, and recovers from this.

system1:	This is the Automatic Intercity Inquiry System. Which information do you want?
caller2:	I want to go from Essen to Hamburg.
system3:	You want to travel from Essen to Hamm? On what day do you want to travel?
caller4:	To Hamburg.
system5:	You want to travel to Hannover?
caller6:	No, Hamburg.
system7:	You want to travel to Hamm?
caller8:	No.
system9:	Please spell the name of the arrival city.
caller10:	h-a-m-b-u-r-g.
system11:	You want to travel to Hamburg?
caller12:	Yes.
system13:	On what day do you want to travel?
caller14:	On wednesday.
system15:	You want to travel on wednesday, September 22nd? At what time do you want to leave?
caller16:	Around four o'clock.
system17:	You want to leave around 16.00 hours?
caller18:	Yes.
system19:	I have found two connections for you. (...) Do you have another inquiry?
caller20:	No thanks [Hangup]

Example 2: Degradation and recovery

In example 2, utterance caller4, the caller contradicts the system by correcting it. Though there has been a misrecognition and the actual name of the city is wrong, the contextual interpretation notices that there is a contradiction in the values of the arrival city, and infers from this that there was some mistake in the interpretation of utterance caller2. As the other value coming from that utterance, the departure city, was not corrected, the system assumes that this was interpreted correctly and it is therefore taken to be confirmed. The detected mistake in the interpretation of caller2 lets the system set its current strategy to a more restricted mode. In system5, it now tries to only get a confirmation for the value it has understood. It gets a contradiction in caller6, but it is also supplied with a new value, which it takes to be a value for the dialogue object that is currently in the focus of attention, *viz.* the arrival city. It tries again to get a confirmation for this in system7, but as it receives a contradiction, it is clear that there must be some severe recognition or parsing problem. It now enters the 'Leading Initiative' mode[3], informing the parser and recognizer of this change. After receiving a value in this mode, it goes back to 'Single Confirmation' in system11, and as this goes smooth, it goes back up to the original 'Combined Confirmation an Initiative' in system15.

Example 2 also illustrates the internal inference for underspecified time expressions. Following the same principle of submitting items for confirmation as specifically as possible, the system infers from the underspecified 'four o'clock' in caller16 the unambiguous 'sixteen hundred hours' in system17.

Note that at any stage in the dialogue, the caller is free to make so-called overinformative utterances, i.e. not only answer the system's question but also add new things. Starting from utterance system 13, the dialogue from example 2 thus could continue as shown in example 3.

system13:	On what day do you want to travel?
caller14:	On wednesday around four.
system15:	You want to travel on wednesday, September 22nd? You want to leave around 16.00 hours?
caller16:	Yes.
system17:	I have found two connections for you. (...) Do you have another inquiry?
caller18:	No thanks [Hangup]

Example 3: Overinformative utterance

4. FURTHER WORK

At Daimler-Benz, we currently are working to improve the system's robustness and performance. Starting from the view of interpretation as a process yielding a hypothetical result, we investigate the possibility of combining the contextual interpretation capabilities with the syntactic-semantic processing by an island parser which we developed in another project (cf. /4/)[4]. Taking predicted words in the graph of word hypotheses as the seeds for islands, this parser tries to build a maximum spanning structure by expanding these seeds left and right. It does not, however, have to find a parse for the utterance as a whole, as this may consist of more than one sentences, isolated prepositional phrases etc. As this parser operates with a chart, it is even possible to stop processing after a time-out and to continue the dialogue semantic interpretation with the best non-overlapping partial results found until then. It is also possible to use this parser in combination with word spotting techniques, using keyword candidates as seed and to look for syntacto-semantic confirmation or refutation of the keywords. This would then result in a structure spotting rather than word spotting technique.

Secondly, we are working on a generator for the system utterances, specially built to meet the requirements for interactive speech systems[5]. Currently, the system uses sentence templates and their combination for its utterances. The generator shall be able to use several sources and levels of input, *viz.* prefabricated templates or surface forms already used by the system or the caller, syntactic descriptions where e.g. change of rection is necessary, or semantic descriptions.
The output of the generator will be the input to a synthesizer rather than a simple string, making it possible to exploit the specific capabilities of speech, e.g. to mark focus by prosodic means (*cf.* /11/).The generator shall also be able to make use of negative constraints: when a formulation of the system has lead to a request for repetition from the caller's side, the generator shall make sure that is does not repeat, i.e. not use exactly the same words, syntax etc. It shall rather strive for an alternative formulation. Finally, the synthesis component has to trace exactly where it is in the utterance, and this tracing has to be co-indexed with the syntactic description of the utterance. The synthesizer will be short-circuited with the recognizer. The caller then can interrupt the system utterance and the system can find the exact spot that caused the caller's interference. If it is just a 'Pardon?', the system can reformulate the utterance. If it is a contradiction, the system can find out a which point there has been some misunderstanding, and use the appropriate measures, as described above.

[3]At this point, the system would have entered 'Leading Initiative' even if it had been supplied with a new value in caller8, because the arrival city has been negotiated twice in single confirmation mode without a result.
[4]This work is carried out in co-operation with the FORWISS, Erlangen.

[5]This work is carried out in co-operation with the DFKI (German Centre for Artificial Intelligence), Saarbrücken.

REFERENCES

/1/ *Andry, F.*: Static and dynamic predictions: A method to improve speech understanding in cooperative dialogues. In: /9/.
/2/ *Ballim, A.; Wilks, Y.*: Artificial Believers. Hillsdale, N.J.: L. Erlbaum 1991.
/3/ *Bilange, E.*: Dialogue personne-machine. Paris: Hermes 1992.
/4/ *Brietzmann, A.*: "Reif für die Insel": Syntaktische Analyse natürlich gesprochener Sprache durch bidirektionales Chart-Parsing. In: Mangold, H. (ed.): Sprachliche Mensch-Maschine-Kommunikation. München, Wien: Oldenbourg 1992.
/5/ *Heisterkamp, P.; McGlashan, S.; Youd, N.*: Dialogue semantics for an oral dialogue system. In: /9/.
/6/ *Hitzenberger, L. et al.*: FACID Fachsprachlicher Korpus informationsabfragender Dialoge DICOS. Regensburg: FG Ling. Informationswissenschaft, Univ. Regensburg 1986.
/7/ *Licher, E.* Setting up a model for the design of system responses in computerized information dialogues. Diploma Thesis, TU Eindhoven, Faculteit Wijsbegeerte en Maatschapijwetenschappen, 1990.
/8/ *Niedermair, G.*: Linguistic modelling in the context of oral dialogue. In: /9/.
/9/ Proceedings of ICSLP-92, Banff, Alberta, Canada, 1992.
/10/*Reilly, R. G. (ed.)*: Communication failure in dialogue and discourse. Amsterdam: North Holland 1987.
/11/*Youd, N.*: The production of prosodic focus and contour in dialogue. PhD Thesis, The Open University, 1992.