[*Machine Translation Today: Translating and the Computer 15*. Papers at a conference…18-19
November 1993 (London: Aslib)]

THE SUNDIAL SPEECH UNDERSTANDING AND DIALOGUE PROJECT: RESULTS AND IMPLICATIONS FOR TRANSLATION

Norman M. Fraser
Vocalis Ltd, Chaston House, Mill Court, Great Shelford, Cambridge CB2 5LD

This paper describes the ESPRIT SUNDIAL project which ran for five years, finishing in August 1993. The objective of the project was to design and build telephone access spoken language interfaces to computer databases. After introducing the aims and objectives of the project, the problem of specifying an interactive system are outlined and the method of Wizard-of-Oz simulation described. The architecture of the resulting system is introduced, and system transaction success results of up to 96.6% are reported. In the final section, some implications for machine translation and, particularly, interpretive telephony are identified.

INTRODUCTION

This paper describes the ESPRIT SUNDIAL (Speech UNderstanding and DIALogue) project.[1] The project, which ran for five years, finished in August 1993. The objective of the project was to design and build spoken language interfaces to computer databases, capable of supporting telephone access by members of the public. After introducing the aims and objectives of the project, the problem of specifying an interactive system are outlined and the solution adopted in SUNDIAL described. The architecture of the resulting system is introduced, and performance results reported. In the final section, some implications for machine translation and, particularly, interpretive telephony are identified.

AIMS AND OBJECTIVES

The SUNDIAL project aimed to make significant advances in the state-of-the-art in spoken language processing (Peckham 1991). It did so by setting a very ambitious target, namely to produce computer systems capable of participating in natural spoken language task-oriented dialogues over the telephone, for each of English, French, German and Italian. The systems should be speaker-independent and should support large vocabulary (around 1000 words) speech recognition. The tasks chosen were flight information and reservations (English and French) and Train timetable information (German and Italian). In order to achieve this goal it would be necessary

---

[1] This research was supported by the Commission of the European Communities as part of ESPRIT project P2218, SUNDIAL.

to deliver respectable performance in each of the component technologies: signal processing, speech recognition, parsing, dialogue management, message generation and speech synthesis. Not only would these technologies have to function well in their own right, they would also have to be fully integrated in a single system with all the other speech and language components and with an application database.

This kind of approach is not without controversy. The very large DARPA ATIS program in speech understanding has placed primary emphasis on improving the component technologies and, in particular, on optimizing speech recognition (Pallet et al 1990). By contrast, SUNDIAL's focus on integration reflects the belief that component technologies do not have to be optimal, so long as they are good enough to contribute positively to overall system usability. After all, humans are not capable of perfect speech recognition, but high level interpretive competences and effective use of heuristics (including asking a speaker to repeat an utterance) allows communication to proceed even in noisy environments.

The objective of allowing telephone access to spoken dialogue systems has two principal motivations. First, the telephone has a very promising future as an interface. Almost every home and office already has a telephone installed so it need not be necessary to buy any new equipment in order to access remote data and services. There is a large number of services in existence which allow touch-tone telephone access and a small but growing number of services which make use of limited speech recognition (e.g. home banking services). SUNDIAL's choice to target telephone quality speech is a clear endorsement of the view that speech recognition over the telephone will be one of the major technology growth areas towards the end of this century.

A second reason for concentrating on telephone quality speech is that recognizing speech over the telephone is somewhat harder than recognizing speech directly using a microphone. The human speech signal occupies a frequency range of approximately 0-10000 Hz, but telephone lines are limited to a range of 400-3400Hz. Thus, a significant amount of potentially discriminative information is lost in telephone quality speech. The added difficulty of working in this environment provides strong motivation for improving the quality and robustness of dialogue management and overall system integration.

SUNDIAL ran from late 1988 to August 1993 and involved 170 person years of effort. At that time, it was the largest speech and language project in Europe. The project involved partners from five European countries: Logica (latterly Vocalis) and the University of Surrey in the U.K.; CNET, CAP Gemini Innovation and IRISA/University of Rennes in France; CSELT, Saritel and Politecnico di Torino in Italy; Daimler Benz, Siemens and the University of Erlangen in Germany; and Infovox in Sweden. Though the objective was to produce distinct demonstrators for the four languages, English, French, German, and Italian, there was a strong commitment to converge on as many aspects of the technology as possible. In this way it was hoped that general lessons would be learned about dialogue management which rose above the fine detail of any one of the actual languages investigated.

SPECIFYING THE PROBLEM

The fact that dialogue involves two parties, each of whose behaviour conditions the behaviour of the other, has serious consequences for the design of a spoken language system. There is no point in designing a computer dialogue system which takes no account of how users will behave when presented with it. Unfortunately, it is not possible reliably to predict how users will behave with such a system until it, or something very like it, exists.

The approach adopted in SUNDIAL was to collect and analyse corpora of dialogues in which real users called existing (human) telephone services. For example, for the English system a corpus of telephone calls to British Airways' (BA) flight information service was examined. This was used to bootstrap a series of so-called Wizard-of-Oz simulations (Fraser and Gilbert 1991a) in which experimental subjects believed they were talking to a computer system; in reality, they were talking to a person (the 'wizard') whose voice had been filtered through a device to make it sound synthetic.

In the first simulation, subjects were asked to carry out a set of tasks derived from the BA corpus; the wizard was required to use the actual words uttered by the BA agent whenever possible. Thus, the main factor being investigated was the effect of the user's belief about the identity of the dialogue partner on the user's utterances. It was discovered that when users believed they were talking with a computer they constrained their language significantly by comparison with the human-human language found in the BA corpus (Fraser and Gilbert 1991b). The constraints were broadly those found in 'speech to foreigners': fewer words were uttered, a smaller vocabulary was used, there was less reliance on complex grammatical constructions (such as relative clauses), and the incidence of talk-in-overlap (i.e. both parties talking at the same time), common in the human-human condition, virtually disappeared.

A series of simulations was carried out, with the lessons of each one feeding into the next. In the later simulations, some subcomponents of the real system were combined in so-called 'bionic Wizard-of-Oz' simulations (MacDermid 1993). This 'iterative design' methodology (Kelley 1984) made it possible to converge on a practical specification, sensitive to the needs of both system and users.

SYSTEM ARCHITECTURE

The overall architecture developed is shown in Figure 1. One of the objectives of the architecture was to provide a close coupling between each module with the aim of using appropriate knowledge and constraints in the process of understanding users' utterances and recovering from errors. An example of the interaction between the modules is the application of predictions derived from the dialogue context to the

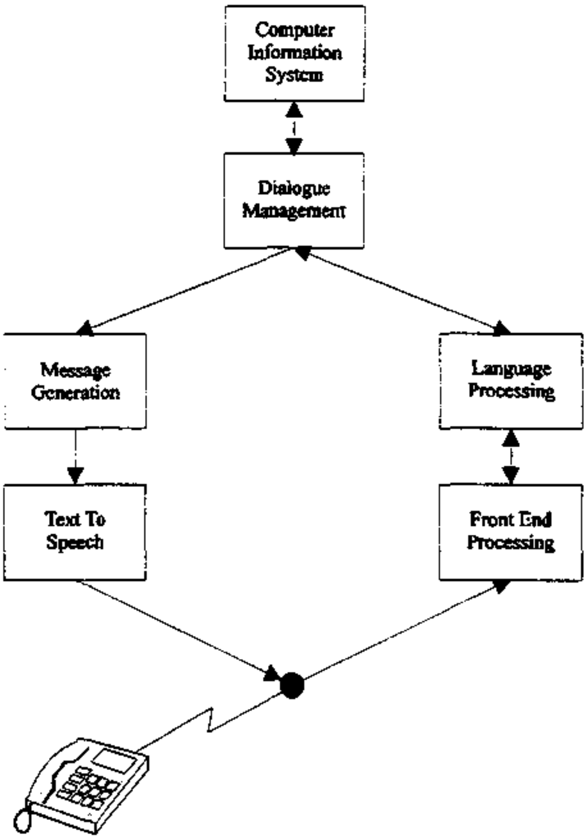


Figure 1: SUNDIAL system architecture

recognition process (Andry 1992; Niedermair 1992). As the dialogue progresses, the Dialogue Manager is able to select appropriate modes of interaction (such as spelling when a place name is consistently mis-recognised), and prompt the user accordingly.

Front End Processing

The Front End Processing module carries out acoustic-phonetic decoding of the incoming speech signal and produces a lattice or graph of word hypotheses. For all four languages, front end processing is based on Hidden Markov Modelling of sub-word units. A small number of keywords (such as 'yes' and 'no') are modelled separately to improve performance.

Speaker independent sub-word models have been constructed on the basis of a number of large multi-speaker corpora of speech recorded over the telephone. For further information on front-end processing in SUNDIAL see Charpentier et al (1993).

Linguistic Processing

The formalisms used in SUNDIAL are Unification Categorial Grammar for English and French (Andry et al 1992), Augmented Phrase Structure Grammar for German (Tropf 1989), and Dependency Grammar for Italian (Poesio and Rullent 1987). Two different parsing strategies have been used: left-to-right bottom-up parsing (Andry and Thornton 1991) and island parsing which selects starting points for parsing on the basis of the best acoustic scoring hypotheses (Baggia et al 1992).

Regardless of the parser type and the underlying formalism, all the linguistic processing systems deliver their results in a common format, using a semantic knowledge representation language called SIL (Heisterkamp et al 1992). This has the effect of making the various different configurations of front end processing and linguistic processing indistinguishable to subsequent processes in the chain of interpretation and generation.

Dialogue Management

The input to the Dialogue Manager is a context-independent interpretation of an utterance, expressed in SIL. The Dialogue Manager takes this interpretation and particularises it to the current dialogue context. This includes finding precise referents for expressions such as 'the flight', and resolving pronominal references such as 'it' and deictic expressions such as 'then'. Even the best of speech recognizers is not fool-proof, so the Dialogue Manager must be capable of establishing a reasonable degree of certainty about what has been said by initiating confirmation sub-dialogues where appropriate. It must be capable of sorting out confusing or inconsistent information, and it must do all this in a manner which is both painless and reasonable to the user. Speech understanding is a canonical example of reasoning in the face of extreme uncertainty; it is the task of the dialogue manager to co-ordinate the whole complex task of interpretation, and to establish a reasonable degree of mutual confidence between the dialogue participants (Heisterkamp 1993).

Whereas each of the four language systems adopted a separate combination of technologies for SIL production, it was decided early in the project to converge on a single 'Dialogue Manager'. This generic system would encode universal knowledge about dialogue (e.g. 'a question begs an answer') and about major categories of task (e.g. providing information from a database on request). The generic system could then be customized to work with some particular task by supplying an appropriate customization file. The general principles of task management combined with the specific customization information, constituted the interface to the application database.

The output of the Dialogue Manager is also a SIL structure encoding the propositional content and discourse function of what the system 'wants' to say next. Thus, the only language the Dialogue Manager knows about is SIL. The system is capable of being simply customised in respect of dialogue strategy. Thus, whether the system typically behaves as in (1) or (2) can be determined when the system is being customized.

(1) (Explicit confirmation)
System: Where does the flight leave from?
User: Berlin.
System: Was that from Berlin?
User: Yes.
System: What time of day does it leave?

(2) (Implicit confirmation)
System: Where does the flight leave from?
User: Berlin.
System: From Berlin. What time of day does it leave?

The generic Dialogue Manager was a very useful tool for investigating a wide variety of phenomena in a language-neutral, task-independent way. It also raised a number of valuable issues relevant to spoken language translation, as we shall see below. However, the very flexibility of the generic Dialogue Manager, coupled with the software engineering challenge of managing a piece of software written at sites in four countries, sometimes made it a difficult tool to work with. For this reason, a local variant of the generic Dialogue Manager was developed for each of the languages in the project. These local variants embodied the principal insights from the generic system in a framework specially tuned for the needs of the target application.

Message Generation

The output from the Dialogue Manager was a SIL structure containing information about both the content and function of the next system utterance. The task of the Message Generator was to take this information and turn it into a string of words in a specified natural language. Message generation in the context of a dialogue must be sensitive to what has been said previously by both system and user. The system maintained a detailed Linguistic History for this purpose (Youd and McGlashan 1992). As well as choosing an appropriate string of words, the Message Generator was required to select an appropriate intonational contour for its utterance. This was signalled by means of special markers in the string output by the Message Generator (House and Youd 1991; Youd and House 1991).

Speech Synthesis

Text-to-speech synthesis technology for each of the four languages in SUNDIAL was based on existing technology. Two different varieties of synthesiser were used. Diphone synthesizers splice together a sequence of very short stretches of human speech. Synthesizers of this type can display a very natural voice quality. Formant synthesizers generate the speech sounds artificially. This typically results in poorer voice quality but greater control over the intonation contour. A number of acceptability trials of the different synthesizers were conducted with non-expert subjects (House and Youd 1992).

EVALUATION RESULTS

The whole question of how to evaluate the performance of an interactive dialogue system has not yet been answered satisfactorily. It is simplistic to suppose that a single metric (such as a score between 0 and 100) could be used. It seems much more promising to suppose that a dialogue system could be characterized by an array of quantitative results coupled with a set of qualitative judgements on such aspects as usability and pleasantness. This is the approach taken in SUNDIAL (Simpson and Fraser 1993).

Evaluation metrics can be divided into two varieties. Black box metrics consider the performance of the whole system without reference to any of its internal details. Glass box metrics look inside the system and monitor the performance of the component technologies. Broadly speaking, glass box metrics are useful diagnostics during system development, while black box metrics are suitable for characterizing the 'goodness' of the system at achieving its ultimate objectives.

Glass box metrics

Glass box metrics which have played an important role in SUNDIAL are word accuracy, sentence accuracy and information content. Word accuracy is a measure of the speech recognizer's ability to recognize the words spoken in each utterance. As well as measuring correct word recognitions, word insertions, deletions and substitutions are recorded. Sentence accuracy is a straightforward measure of the percentage of sentences which were recognized perfectly, i.e. without insertions, deletions or substitutions. It is a rather more demanding measure than word accuracy, since it is possible to obtain a high word accuracy but a low sentence accuracy by recognizing most, but not all, words in most utterances. The problem with sentence accuracy is that it does not measure recognition at a level of granularity which is relevant to the task in hand.

Most people are familiar with the experience of missing part of what someone has said but, nonetheless, grasping the 'gist' of the utterance. Consider the following utterance:

(3) I was wondering whether you might be able to tell me the arrival time of BA 123, please.

Suppose that this was misheard as:

(4) I wonder whether you might be able to tell me the arrival time of BA 123, please.

From almost every point of view, the differences between these two sentences are trivial. However, it is enough to drag down the word accuracy score and to assign sentence (4) a sentence accuracy score of 0. It is helpful here to identify how much of the sentence actually contributes to the achievement of the task the user is attempting to perform. From this point of view, relatively few words in the utterance are important. Much of (3) could be ignored without loss of relevant information. Utterance (5) is sufficient to progress the dialogue satisfactorily.

(5) Tell me the arrival time of BA 123

The information content measure is used to assess how effective the recognizer and parser are together in identifying the task parameters mentioned in the utterance. Once again, the methodology is to compare what was actually recognized against a reference answer, and to count correct recognitions, insertions, deletions and substitutions.

Black box metrics

The most important black box metric is transaction success. This is a measure of whether or not the system succeeds in carrying out some appointed task and delivers a solution which accords with the facts. Possible values for this metric are (i) succeed, (ii) succeed with constraint relaxation, (iii) succeed in spotting that no answer exists, and (iv) fail. The fact that a system achieves a high degree of transaction success does not guarantee that it will be usable. It may be rendered useless by virtue of the slowness of its performance, for example.

Average dialogue duration is another measure which is relevant for characterizing dialogue systems. However, this misses the fact that some tasks genuinely require much more time than others. One way around this is to calculate the system's average response time. This conveys a sense of how long the user has to wait for the system to respond. This metric has to be applied with caution, however, since some tasks necessarily involve lengthy delays. For example, callers to existing flight information and reservation services often have to wait a minute or more for a

A more subtle metric is the turn correction ratio. This gives the percentage of users' turns in a corpus which are devoted to correcting some failure on the system's part. For example, the system may mishear or misunderstand the user. The insight behind this metric is that it expresses the percentage of the dialogue which is not devoted to progressing through the task.

User confidence is a key factor in spoken dialogue systems. One of the principal threats to user confidence is the production of system utterances which appear to bear no relation to the rest of the dialogue or, worse still, contradict what has gone before. The contextual appropriateness measure records the percentage of utterances which are judged to be appropriate in context.

Results

Taken together, these metrics characterize the performance of spoken language dialogue systems. It would take more space than is available here to present the full results for all the SUNDIAL systems; these can be found in (Ciaramella 1993). Here we present just the transaction success results for the Italian and English systems.

Four trials of the Italian system were carried out (Clementino and Fissore 1993). The variables which were investigated were user expertise (naive/expert) and dialogue manager (generic/Italian). All of the trials were carried out over private branch exchange (PBX) lines, i.e. lines controlled by a private switchboard. The results are reported in Table 1.

TABLE 1: Italian transaction success results over the PBX

Italian prototype	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**
Subjects	naive (10m, 10f)	naive (10m, 10f)	expert (11m, 4f)	expert (11m, 4f)
Speech input quality	PBX	PBX	PBX	PBX
Dialogue Manager	Italian variant	generic DM	Italian variant	generic DM
Number of dialogues	86	86	63	63
Transaction success	77.6%	51%	96.6%	83.3%

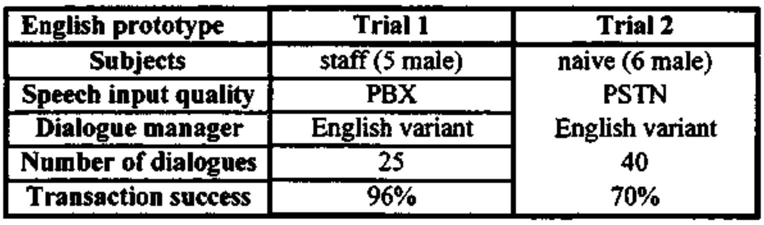
In the first and second trials, the subjects were naive users who were given tasks to perform. They were not given any special instructions about what they could or could not say. The transaction success results show significantly better performance from the Italian variant of the dialogue manager, reflecting the additional constraints it embodied. Trials 3 and 4 also contrast the different versions of the dialogue manager, this time using members of the Italian SUNDIAL project team. Once again, the Italian variant dialogue manager produces better results than the generic dialogue manager. It is noteworthy that the system performs significantly better with expert users than with naive users for both versions of the dialogue manager.

How should these results be interpreted? Trial 1 shows that more than three out of every four tasks attempted by naive users completes successfully. When the system does not succeed, it does not fail drastically. Rather, it recognizes that it is not making reasonable progress and elects to pass the caller to a human agent for task completion. The transaction success result in Trial 3 represents a very respectable target towards which users may aspire as they become more skilled in using it (more than 19 out of every 20 dialogues succeed). The nature of the skills acquired by experienced users is difficult to pin down exactly, but is likely to include the tacit knowledge absorbed by the user about how to speak in order to get the most out of the speech recognizer, and which grammatical constructions produce the best results.

A clear lesson from these trials of the Italian prototype system is that over-the-telephone task-oriented natural language dialogue for access to data and services is technically realisable within the short to medium term.

A number of trails of the English system (including the English variant dialogue manager) were carried out. Each trial was only of modest size, and was designed to investigate particular questions. Though the results shown in Table 2 are for small samples, and represent the best results achieved so far, they are nonetheless broadly in keeping with a much larger set of results drawn from a number of other trials carried out under similar but not identical conditions.

TABLE 2: English transaction success results over the PBX



English prototype	Trial 1	Trial 2
Subjects	staff (5 male)	naive (6 male)
Speech input quality	PBX	PSTN
Dialogue manager	English variant	English variant
Number of dialogues	25	40
Transaction success	96%	70%

The subjects in Trial 1 were staff of Vocalis, all of whom were familiar with the idea of speech recognition and most of whom had experience of using a speech recognizer. However, none of the subjects had contributed to the design or implementation of the SUNDIAL systems. These subjects therefore represent an intermediate group between the naive users and expert users of the Italian trials. Trial 2 investigated the use of the English SUNDIAL system over the public switched telephone network (PSTN); to be specific, the subjects were located in Guildford and the system was located in Cambridge (a distance of more than 100 miles). The diminished performance in Trial 2 may be attributable in part to the difference between staff subjects and naive subjects. However, the weakening and distorting effects of the PSTN on the speech signal is almost certainly a contributory factor.

There is clearly room for improvement in the result recorded in Trial 2. However, SUNDIAL is one of the first projects to report results for spoken language (as opposed to isolated word or connected digit) recognition over the public telephone network. The transaction success figure of 70% is an encouraging starting point for further work. Once again, transaction failure does not mean that the user is supplied with incorrect information or that the system hangs up. Rather, it means that the system recognizes that there are difficulties and switches the caller to an agent, rather than subjecting the user to the pain of a long drawn-out error correction process. Further information on the English SUNDIAL system trial results can be found in

(Ciaramella 1993). Results for the French system are reported in (Magadur et al 1993) and for the German system in (Eckert et al 1993).

Example dialogue

Perhaps the best way to get a feel for the capabilities of a dialogue system is to analyse the dialogues it produces. The following is an example of a real dialogue involving a naive user calling the system on a long distance telephone line.

```
1   System: Welcome to British Airways flight information service.
             How can I help you?
2   User:   Which terminal does BA five eight zero leave from?
3   System: Was that British Airways flight five eight zero?
4   User:   Yes.
5   System: That flight departs from Heathrow Airport Terminal 1. Do
            you have another inquiry?
6   User:   Yes.
7   System: What would you like to know?
8   User:   Can you tell me the arrival time of today's Madrid
            flight?
9   System:  What is the number of the flight?
10  User:    I don't know.
11  System:  Where does the flight leave from?
12  User:    Madrid.
13  System:  Was that from Paris?
14  User:    No, Madrid.
15  System:  Did you say from Copenhagen?
16  User:    Madrid.
17  System:  Please answer 'yes' or 'no'. Did you say from Madrid?
18  User:    Yes.
19  System:  Please tell me the destination.
20  User:    Heathrow.
21  System:  Did you say to Heathrow Airport?
22  User:    Yes.
23  System:  Please tell me the departure time.
24  User:    One o'clock in the afternoon.
25  System:  Was that thirteen hundred hours?
26  User:    Yes.
27  System:  British Airways flight four six seven leaves from Madrid
            Barajas Airport at thirteen hundred hours. It arrives at
            London Heathrow Airport, Terminal 1, at fourteen fifteen.
            Do you have another inquiry?
28  User:   No thanks.
29  System: Thank you for calling. Goodbye.
```

IMPLICATIONS FOR TRANSLATION

The SUNDIAL project set out to investigate spoken natural language dialogue by designing and implementing a prototype system for each of the four project languages. It was successful, both in producing working large-vocabulary speaker-independent systems and in bringing to light some important lessons about where effort can most effectively be directed to produce the best results.

Though the work was carried out in a multilingual environment, it was never an intention to investigate interpretation, that is, dialogue between partners speaking different languages. However, the results of the SUNDIAL project bear on the whole question of inter-language communication in a number of ways.

First, the project has shown that the enabling technology for interpretive telephony (i.e. spoken language translation over the telephone) is maturing to the point of being a practical possibility for limited domains. This observation is in keeping with recent results coming from the leading groups working on interpretive telephony around the world (Morimoto et al 1993; Woszczyna et al 1993). However, encouraging though this result is, it is important to stress that it only holds good for strictly limited domains. The idea of a general purpose interpreting machine which will cause mass unemployment amongst professional interpreters, is still a very long way from being realised.

Second, though SUNDIAL did not set out to produce an interpretive telephony system, it ended up producing one by accident! The same Dialogue Manager was the core of each of the language prototypes and the interfaces were constant across all systems. The net result was that system components could be mixed and matched. A first trial might connect an English input subsystem (Front End Processor plus Linguistic Processor) and an English output subsystem (Message Generator plus speech synthesizer) to the generic Dialogue Manager. A second trial might, with equal ease, include an English input subsystem and a German output subsystem with the Dialogue Manager. This would result in the system understanding English but speaking German. So easy was this to do, that project workers occasionally produced this effect by mistake. Once the clean modular structure to support this kind of behaviour was in place, it was a relatively simple matter to modify the system so that, rather than responding to the user's questions, it simply echoed them in a different language. Though it was never designed as such, SIL turned out to be a very effective interlingua.

This leads to a third observation, which is little more than a generalization of the previous one. It is widely acknowledged that machine translation is a very difficult problem. There has been much, often heated, debate between those who advocate surface-oriented transfer and those who support translation based on a deeper (and consequently more language-neutral) representation of the meaning of a text. There are strong arguments on both sides, but perhaps the increasing amount of work being carried out on language understanding for its own sake (as in SUNDIAL) will begin to tip the balance in favour of interlingua approaches. The experience of the SUNDIAL project was that machine translation was *relatively* straightforward in limited domains, once the problem of language understanding had been solved.

A final observation is that task-oriented translation is much simpler than general purpose translation. This is, of course, platitudinous. However, the observation does not just apply at the level of tasks such as 'making a train timetable enquiry'; it also holds for the tasks which make up the micro-structures of dialogue. Thus, if speaker A has just asked speaker B which London railway station trains from York arrive at, speaker B's range of reasonable responses is strictly limited to identifying a London Railway station, admitting to not knowing the answer, or asking for the question to be repeated.

If translation of unrestricted discourse is difficult, then translation of task-oriented discourse is easier and translation of task-oriented interactive dialogue is easiest of all, given present technology limitations.

REFERENCES

Andry, F. (1992) Static and dynamic predictions: a method to improve speech understanding in co-operative dialogues, *ICSLP'92,* Banff, October, 639-42.

Andry, F. and S. Thornton (1991) A parser for speech lattices using a UCG grammar, *Proceedings of the 2nd European Conference on Speech Communication and Technology,* Genoa, September, 219-22.

Andry, François, Norman M. Fraser, Scott McGlashan, Simon Thornton and Nick Youd (1992) Making DATR work for speech: lexicon compilation in SUNDIAL. *Computational Linguistics* 18:245-67.

Baggia, Paolo, Elisabetta Gerbino, Egidio Giachin and Claudio Rullent (1992) Real-time linguistic analysis for continuous speech understanding, *Proceedings of the 3rd Conference on Applied Natural Language Processing,* Trento, April, 33-9.

Ciaramella, Alberto, ed. (1993) Prototype performance evaluation report. SUNDIAL Technical Report WP8000-D3.

Charpentier, Francis, Giorgio Micca, Ernst-Gunter Schukat-Talamazzini and Trevor Thomas (1993) The recognition component of the SUNDIAL project. *Proceedings of the NATO Workshop,* Biebiou, Spain, July.

Clementino, D. and L. Fissore (1993) A man-machine dialogue system for speech access to train timetable information. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* Berlin, September, 1863-6.

Eckert, Wieland, Thomas Khun, H. Niemann, S. Rieck, A. Scheuer and Ernst-Gunter Schukat-Talamazzini (1993) A spoken dialogue system for German intercity train timetable enquiries. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* Berlin, September, 1871-4.

Fraser, Norman M. and G. Nigel Gilbert (199la) Simulating speech systems. *Computer Speech and Language* 5,81-99.

Fraser, Norman M. and G.Nigel Gilbert (1991b) Effects of system voice quality on user utterances in speech dialogue systems. *Proceedings of the 2nd European Conference on Speech Communication and Technology,* Genoa, September. 57-60.

Heisterkamp, P. (1993) Ambiguity and uncertainty in spoken dialogue. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* Berlin, September, 1657-60.

Heisterkamp, Paul, Scott McGlashan and Nick J. Youd (1992) Dialogue semantics for an oral dialogue system, *ICLSP'92, Banff* October, 643-6.

House, Jill and Nick J. Youd (1991) Synthesising intonation in a dialogue context, *Speech, Hearing and Language* 5, 75-90.

House, Jill and Nick J. Youd (1992) Evaluating the prosody of synthesized utterances within a dialogue system, *Proceedings of the International Conference on Spoken Language Processing,* Banff, October.

Kelley, J.F. (1984) An interactive design methodology for user-friendly natural language office information applications. *Association for Computing Machinery Transactions on Office Information Systems* 2, 26-41.

MacDermid, Catriona (1993) Features of naive callers' dialogues with a simulated speech understanding and dialogue system. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* Berlin, September, 955-8.

Magadur, Jean-Yves, Frederic Gavignet, François Andry and Francis Charpentier (1993) A French oral dialogue system for flight reservations over the telephone. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* Berlin, September, 1789-92.

Morimoto, Tsuyoshi, Toshiyuki Takezawa, Fumihiro Yato, Shigeki Sagayama, Toshihisa Tashiro, Masaaki Nagata and Akira Kurematsu (1993) ATR's speech translation system: ASURA. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* Berlin, September, 1291-4.

Niedermair, G.T. (1992) Linguistic Modelling in the Context of Oral Dialogue, *ICSLP'92,* Banff, October, 635-8.

Pallet, D.S., W.M. Fisher and J.S. Garofolo (1990) DARPA ATIS Results, June 1990. *Proceedings of the DARPA Workshop on Speech and Natural Language.* Hidden Valley, PA, June. 114-21.

Peckham, Jeremy (1991) Speech understanding and dialogue over the telephone: an overview of the ESPRIT SUNDIAL project, *Proceedings of the DARPA Workshop on Speech and Language,* Pacific Grove, CA, February, 14-27.

Peckham, Jeremy (1993) A new generation of spoken dialogue systems: results and lessons from the SUNDIAL project. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* 33-40.

Poesio, M. and Claudio Rullent (1987) Modified caseframe parsing for speech understanding systems. *Proceedings of the International Joint Conference on Artificial Intelligence,* Milan.

Simpson, Andrew and Norman M. Fraser (1993) Black box and glass box evaluation of the SUNDIAL system. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* 1423-7.

Tropf, H.S. (1989) Syntax in the spoken dialogue system SPICOS-II. *Proceedings of the European Conference on Speech Communication and Technology,* Paris, 30-3.

Woszczyna, M., N. Coccaro, A. Eisele, A. Lavie, A. McNair, T. Polzin, I. Rogina, C.P. Rose, T. Sloboda, M. Tomita, J. Tsutsumi, N. Aoki-Waibel, A. Waibel and W. Ward (1993) Recent advances in JANUSystem: a speech translation system. *Proceedings of the 3rd European Conference on Speech Communication and Technology,* Berlin, September, 1295-8.

Youd, Nick J. and Jill House (1991) Generating intonation in a voice dialogue system, *Proceedings of the 2nd European Conference on Speech Communication and Technology,* Genova, September, 1287-90

Youd, Nick J. and Scott McGlashan (1992) Generating utterances in dialogue systems, Robert Dale, Ed Hovy, D. Rosner and Olivero Stock (eds) *Aspects of Automated Natural Language Generation: Proceedings of the 6th International Workshop On Natural Language Generation,* (Lecture Notes in Artificial Intelligence 587), Berlin: Springer Verlag, 135-49.