SPEECH AND LANGUAGE IN AN ERA OF MEDIUM SCALE SCIENCE

Research in the Context of Elsnet

Niels Ole Bernsen, Centre for Cognitive Science
Roskilde University, Denmark
email: nob@cog.ruc.dk

In celebration of the transition from Elsnet-1 to Elsnet-2, I would like to make some remarks, or claims, on the perennial theme of "Speech and language - where are we now, what should be done etc.?" Some of the remarks concern developments which Elsnet might take into account in charting its future course, others concern developments that did not happen (yet), and still others are suggestions for action by Elsnet itself or by others.

Elsnet's Long-Term Research Goal

Starting with a development that did not happen yet, it is clear that Elsnet's long-term research goal is still not feasible:

> "The long-term technological goal which unites the participants in Elsnet is to build integrated multilingual speech and nl systems with unrestricted coverage of both spoken and written language." [1].

This does not mean, however, that important progress has not been made during the existence of Elsnet-1. Perhaps the single most important achievement in that period is the appearance of relatively large vocabulary, continuous, speaker independent speech recognition products. Arguably, the best current example of such products in Europe is the Philips train time table information system which is now being operated by Swiss Rail. Whereas this achievement by no means implies that all problems about speech recognition have been solved, it does put an end to the nagging question of when we would finally see advanced interactive speech systems on the market. The future of Elsnet's field of speech and language integration is now happening and European producers of speech technologies are hiring staff as never before.

Medium Scale Science

Early in Elsnet-1, in 1992, the Elsnet Research Task Group produced a research strategy document [2]. A central claim of the document was that research on speech and natural language systems should move away from the isolated three year project paradigm towards larger, coherently organised project structures. Except for landmark projects such as Verbmobil, this still has not

happened in Europe, neither nationally nor at the European level. Yet the reasons why it should happen have become even more apparent in the meanwhile. Briefly, this is because the field of integrated research on speech and language systems has become one of *medium scale science.* The central driving force in research is the development of advanced research prototypes which are nevertheless realistic from the end-users point of view. This should be fine from the point of view of industry because it is hard to think of a better meeting place between industry and academia than advanced yet realistic research prototypes. A guess is that the problems posed by the development of these systems will increasingly determine the research agendas in the fields of speech and natural language as a whole. However, building such systems is very demanding on resources. The scientific literature standardly offers little in terms of guidance, whether in the form of practical system development experience or in the form of practically applicable theory. Simulation-before-implementation is a necessity, lest one ends up having implemented a worthless system concept. Methods and tools often have to be developed from scratch. Large collections of data on user-system behaviour have to be created, annotated and analysed. Sometimes even the concepts of phenomena which need to be annotated, such as speech acts or types of non-cooperative system dialogue behaviour, are missing and have to be developed. Methods and tools for advanced data analysis are often missing. System evaluation concepts and methods are missing or inadequate etc. There is no way that a single, three year project involving some senior researchers and, say, four postdocs and/or Ph.D. students can handle this complexity appropriately, developing what they need as they go along. So they focus on a subset of issues, loosing important progress potential as a result.

During Elsnet-1 and partly at the instigation of Elsnet itself, relevant European efforts such as EAGLES on evaluation and ELRA, SPEECHDAT, BABEL and others on resources have slowly emerged, which is to be welcomed as part of the needed response to the problems pointed out in the Elsnet Research Strategy Paper. Yet there is no indication that these and other European efforts are aimed at supporting medium-scale European research efforts as opposed to the smaller-scale projects we have known for more than a decade.

The Need for Generalisation

Another point on which one might like to see innovation in the 5th Framework Programme, is a strong emphasis on the development of new theory, concepts, software tools and best practice methodologies related to the development of advanced interactive speech and language systems. Building a novel system that goes beyond the state of the art in system building is only half the exercise if efficient research progress is to be made. The other half is *generalising* the insights gained into theory, concepts, methods and tools that others may use to produce similar or better systems more efficiently. So far, the "see my beautiful new system" -paradigm appears to have been sufficient for most researchers and funding agencies, no doubt encouraged by the strong Framework Four emphasis on industrial relevance. It may be argued that this is not an optimum use of Europe's research potential.

Elsnet and The Web

Another important development which has happened since the 1992 Strategy paper is the emergence of the World Wide Web. This probably means that many advanced systems integrating speech and language in the future will be operating on the Web and combine speech and language with other modalities for the representation and exchange of information between system and user. Multimodality has arrived and is there to stay. Elsnet-2 should exploit these new opportunities with determination. At least two initiatives to that effect appear promising and are being discussed at the moment. One is to develop an Elsnet "virtual facility" on the Web, providing access to all kinds of information, resources, tools, papers and dissertations, training material, software etc. that are publicly available at the Elsnet node sites. Another is to create a Web-based electronic infrastructure which can serve as a "showroom" for speech and natural language technologies. Elsnet could take the initiative to develop research projects and gather existing systems that will help achieve the goals of the latter initiative.

In addition, the Elsnet virtual facility could serve another set of important objectives common to Elsnet-2 and the Copernicus project Elsnet Goes East. Many, if not most, of the objectives which follow from the Elsnet-2 goal of integrating research and industry from the Central and Eastern European countries with what happens in Western Europe could be furthered through the virtual facility and showroom mechanisms.

Multimodality and Research Integration

The fact that multimodality is there to stay, and hence that many speech and natural language systems will form components of multimodal systems, leads to another suggestion. It is that Elsnet might benefit from strengthening its links with relevant other European Networks of Excellence for mutual benefit and joint impact. Today, intelligent multimodal systems, sometimes also including cooperative human-human aspects, are being developed by a multitude of different research communities: parts of traditional artificial intelligence, intelligent multimedia systems, the speech processing community, the computational linguistics community building systems for text translation, information retrieval, spell and grammar checking etc., computer vision, computer graphics, acoustics research, computer supported cooperative work, enhanced and virtual reality, ergonomics etc. A likely scenario for the future is that these communities will have to come much closer together, driven by the technological potential for building - yes, why not give them a name?: truly natural interactive computer systems. If this scenario is on the right track then Elsnet-2 should consider acting on it.

Referring to the scenario of the preceding paragraph, one of Elsnet's goals remains that of integrating, to any meaningful extent, the speech and language communities. I don't think we ever consulted social scientists on the likelihood of this happening or on a plausible time scale for it to happen. In any case, it has not happened yet and clearly is a slow process:

> "Yet we are still far removed from a situation where it is natural for nl and speech people to work closely together, or even to be aware of each other's research agendas and achievements." [1].

The above scenario does not invalidate Elsnet's goal of speech and natural language integration. What it does, however, is to point out that the speech and natural language communities are but two communities among many others which share the more general goal of developing natural interactive systems. This could have far-reaching implications for Elsnet-2 and its objectives, implications which we are only now beginning to grasp but which are likely to increasingly influence the course of Elsnet's research strategies during Elsnet-2. Incidentally, it also seems to add even more weight behind the claim that we need more medium scale science in Europe at the expense of small-scale isolated research projects.

References

1. Elsnet-2 Network Programme, June 1996.

2. Bernsen, N.O., Granstroem, B., Giachin, E., Klein, E., Mariani, J., Nooteboom, S., Thompson, H. and Uszkoreit, H.: *European Strategic Research in Speech and Natural Language.* Report by the ELSNET Research Coordination Task Group. Draft 3.2, Edinburgh, 11 August, 1992.