A National Database for Swedish Prosody?

Eva Gårding
Department of Linguistics and Phonetics
University of Lund
Helgonabacken 12, S-22362 Lund, Sweden

INTRODUCTION

Last year a group of foreign experts were invited to evaluate language technology projects supported jointly by NUTEK (The Swedish National Board for Industrial and Technical Development) and the HSFR (The Swedish Council for Research in the Humanities and Social Sciences).

In a general introduction to the project evaluations(published as Report 19, 1993:Language Technology Research Program NUTEK, HSFR) the group gave a number of arguments in favour of the use of large, digital databases in language technology research, e.g.,

•Most of the progress in language technology in the past decade has resulted from a methodology in which large databases of text and speech and inductive procedures have been used.
•A database is necessary to coordinate and organize efficiently thework of computational linguists and speech technologists.
•Collaboration between different research institutes is advocated to solve problems which should be well-defined from the beginning.

Among the examples of automatic systems in speech recognition and speech synthesis in the USA the following are mentioned (page 2):

At AT&T four billion calls a year are processed automatically.

At NYNEX, a regional Bell Operating Company, a so-called Reverse Directory Assistance, is being developed. It uses speech synthesis technology for names and addresses, and speech recognition technology for e.g. city-names.

BBN laboratories recently demonstrated a 20000-word speaker-independent continuous speech recognition system with real-time performance for passages drawn from the Wall Street Journal.

MIT has developed a spoken dialogue system which is connected to an actual on-line air travel service, and can be used to make real air travel bookings.

AMERICAN ENGLISH

In connection with the Swedish projects involved in prosody the group recommended the setting up of a national database. It may therefore be interesting to look at a database under development for American English prosody which, although not mentioned in the evaluations, has been described at various conferences, namely the TOBI base. (TOBI is an acronym for Tones and break indices). The theory behind this system is one, according to which English is a pitch accent language and intonation (the tune) the result of a sequence of discrete pitch elements (Pierrehumbert 1980). The basic units of TOBI are five tonal configurations for the pitch accents at the word and phrase level, (H(igh) and L(ow) and combinations of H and L assigned to the accented syllable) and four acoustic criteria for the breaks. The base will consist of a variety of texts markedwith the corresponding labels (Silverman et al. 1993).

The manual labelling is done by a number of transcribers who have been previously trained to recognize the basic units. In this task the transcribers, who follow the speech from left to right and make decisions locally, are exposed to the auditory signal and at the same time to its acoustic record in terms of speech wave and F0. Later there is a strict check of the listeners' consistency andreliability.

Once TOBI is completed, it will be used for statistical treatment of prosodic patterns, for instance the distribution of the tonal configurations and the boundaries which will lead to an inventory of the intonations (tunes) used in the language. The prosodic transcription will also make it possible to establish correlations with inserted syntactic and semantic labels.

The Tobi system is supposed to work for Southern English and Australian English as well but has not been applicable to Glasgow English.

Like the automatic systems mentioned in Report 19, TOBI is mainly subsidized by commercial enterprise.

SWEDISH

To set up a national prosodic database along similar lines seems even more problematic for Swedish than for English. I shall list some of the complexities here, well aware that most of them will apply to other languages as well.

To begin with there are different rules of distribution and manifestation of the word accent(s) and phrase accent in different dialects. Add to this the fact that many speakers are in a prosodic state of flux and that we know very little about the prosodic transfer processes.

Apart from the accent configurations being dialect-specific, they are always more or less deformed by their degree of accentuation and by global intonation. A fair amount of agreement among transcribers can only be hoped for if the patterns are judged in a functional, abstract way (A1, A2, etc.), not in terms of Highs and Lows.

There is in addition an annoying and inconvenient lack of knowledge of the wider contextual and pragmatic influence on the F0 contour.

Moreover it is by no means clear what the basic perceptual cues or perceptual domains are on which listeners rely when they decode intonation. The perceptual side is not well analyzed and understood.

The usefulness of adatabase depends heavily on the accuracy and completeness of the labelling.In Swedish, the labelling of tonal configurations cannot yet be done according to uncontroversial criteria.

A database may be a waste of time and money if the labels are narrowly based on a theory which turns out to be inadequate.

In contrast to TOBI, a Swedish prosodic database is probably not attractive to the speech technology market.And if it were,the market should pay for it.

A final but very serious point. There is perhaps also the danger that this project will absorb so much cost and human effort that open-minded curiosity-driven research which used to be the hallmark of our universities will suffer. After all, there are not so many phoneticians in our country and there are many interesting research objects and applications of prosodic and other phonetic research apart from a national prosodic data base.

REFERENCES

von Hahn, W., Liberman, M. Y., Zue, V. 1993. Evaluation of the NUTEK/HSFR language technology research program. Report 19. NUTEK: 1-36.
Pierrehumbert, J. 1980. The phonetics and phonology of English intonation. Doctoral dissertation. MIT.
Silverman, K., Beckman, M., Pirelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. 1993. TOBI: A standard for labeling English prosody. Proc.ICSLP-92: 867-870.