# Comments on Sproat

**Kimmo Koskenniemi**
**Department of General Linguistics**
**FIN-00014 University of Helsinki**
kimmo.koskenniemi@helsinki.fi

Sproat's paper describes work done at Bell Laboratories concerning text to speech (TTS) synthesis. The paper and the work is particularly interesting and novel because the approach is genuinely multilingual. The treatment of several languages which are not closely related to each other seems to lead to more general solutions and models in the description of various components of the language.

## Relation to other research

The work is very closely related to a set of previous papers by the author, and other researchers such as Fernando Pereira, Michael Riley and Mehyar Mohri. Much research has been conducted at AT&T Research and at Bell Laboratories in the field of weighted finite-state transducers. This paper builds on top of the results of the earlier paper.

The approach resembles the work done at Xerox PARC and at RXRC in Grenoble. The use of weighted transducers seems to be a significant difference, and the resorting to dynamic construction of compositions (instead of precompiled transducers).

## Interactions

The approach is based on finite state transducers as the basic formalism in modeling the various components involved in the TTS. This choice lets one decompose the problem into smaller modules in a natural way and define the interactions between the modules in terms of the intermediate representations. This kind of a framework has the advantage of being easy to understand and explicit for the builders of the rules and descriptions for the system. At the same time, this approach may impose certain restrictions concerning the interactions between various representations and modules.

Neighboring representations may interact, but representations or modules further away cannot have any effect (unless the relevant information is repeated in all intervening representations). It would be nice to hear a short account whether such problems (with the interaction of more distant representations) are met at all in the building of the descriptions.

## Tokenization

The paper makes a full commitment to describe the tokenization, including punctuation and various abbreviations. How does the tokenization in the current project differ from that of the RXRC team, and especially of the paper of Jan-Pierre Chanod and Pasi Tapanainen (at this workshop)? (E.g. when the tokenization is ambiguous such as French *de même*.)

Also Finnish would provide demanding facilities for the processing of numerals and the percent sign, e.g. "to 521 pupils" would be in Finnish *521 oppilaalle*, but morphemically this is

*five + inessive + hundred + inessive +*
*two + inessive + ten + inessive + one + inessive +*
*pupil + inessive*

Thus, the various components of the compound numeral agree individually in number and case with the noun head.

## Modularization

The overall scheme used for relating the pronunciation $PR$ and the written form or spelling $SP$ is given as:

$$SP \xleftarrow{S} MMA = 20 \xleftarrow{L_{word}} MR \xleftrightarrow{D} MR \xrightarrow{L_{word}} MMA \xrightarrow{P} PR$$

where $MMA$ stands for the minimal morphologically motivated representation, and $MR$ for morphological representation.

This is a nice and elegant setup. Are there any arguments which would prefer the current approach over the following one (or vice versa)?=20

$$SP \xleftarrow{L} MR = 20 \xleftarrow{D} MR \xrightarrow{L_{word}} MMA \xrightarrow{P} PR$$

The lexicon transducer $D$ might be available from other sources, or it might motivated by other applications. In some languages, the orthographic form might be further away from the pronunciation.

## Experiences

I would be interested in hearing some information on the effort spent on writing the rules and descriptions for the various modules, and about the ease of writing them. What kind of methodology was followed, or would appear to be suitable?

## Weights

I would like to hear more about the use of weights. How are they incorporated in the description or rule formalism, and how are the values of individual weights set?


Proceedings of the ECAI 96 Workshop
*Extended Finite State Models of Language*
Edited by A. Kornai.

Does the introduction of weights in the transducers significantly increase the complexity of manipulating them? Which factors prevent the precompilation of the larger combined transducers? Is the presence of weights significant in this respect?

## Software and documentation

What documentation of lextools is available for people outside ATT Research and Bell? (I found the references no. 10, 11, and 15 to be available from the The Computation and Language E-Print Archive *http://xxx.lanl.gov/cmp-lg/*). Will the software be available for educational or research purposes?