Some innovations in a Norwegian text-to-speech system

Gustafson, K.

journal: STL-QPSR
volume: 30
number: 1
year: 1989
pages: 081-085

http://www.speech.kth.se/qpsr
INTRODUCTION: A NORWEGIAN TEXT-TO-SPEECH SYSTEM

Norwegian orthography stands in a fairly close relationship to a phonemic representation of speech in the standard dialects. Therefore it is possible for a text-to-speech system to generate with a relatively small number of rules phonemic transcriptions that are on the whole quite acceptable; the majority of the mistakes such a system is likely to make will be ones related to stress placement and choice of toneme; the two 'word tones' of the standard dialects, usually referred to as toneme 1 and toneme 2, are exemplified by (1) /H'ENER/ and (2) /H"ENER/ respectively (the phonemic transcriptions in this paper follow the principles used in the Norwegian versions of the KTH/INFOVOX text-to-speech systems). Many segmental choices also depend on the position of stress within the word, cf. (3) FILOLOG /FILOL'A:G/ vs. (4) FILOLOGI /FILOLOG'I:/.

The first version of a text-to-speech system for Norwegian, based on Standard Eastern Norwegian pronunciation of the Bokmål variety of the written language, was made in 1984 through collaboration between Bjørn Granstrøm of KTH and Ivar Utne of the University of Bergen. The system was developed within the framework of the multilingual text-to-speech system at KTH (Carlson, Granstrøm, & Hunnicutt (1982); Carlson & Granstrøm (1986)) and produced by INFOVOX A/B. This version had a very acceptable quality, with respect to both phonetic realization and accuracy of the transcriptions. I have later contributed to some minor improvements of this system.

However, there are some outstanding problems that have to be overcome before one can be satisfied that the system is as good as current technology would allow.

During the last few months I have made an attempt at overcoming some of these problems; this attempt has led me to rewrite entirely the grapheme-to-phoneme rules (the GRAF rules), and to make some far-reaching reforms in the system of look-up in the lexicon (LEX). In this paper I shall deal with some of the principles and details involved in this work.

THE PROBLEM OF POSITION OF WORD STRESS AND CHOICE OF TONEME

As regards position of word stress, uncompounded Norwegian polysyllables, with relatively few exceptions, fall into two categories, those of Germanic origin (GERMANIC), and those of Greek/Latin/Romance origin (LATIN). It is possible to establish a fairly small set of rules which will assign word stress correctly provided one knows which of the two classes a particular word belongs to (these rules have as yet to be satisfactorily formulated); one major difficulty for an automatic system is to decide the class membership. In the first Norwegian version only a rudimentary attempt was made to solve this problem; the main principle regulating stress placement was to identify specific strings of graphemes that either attract the stress or avoid it, and that are associated with one or the other of the tonemes. With this solution there are very many cases where the wrong choice will be unavoidable as a result of the system's inability to discriminate between the two main word categories. The problems associated with choice of toneme are partly of a lexical nature (some lexemes have the one, others have the other), and partly to do with the particular combinations of individual lexemes and inflectional endings. These two points can be brought out with the following examples of verbs in the infinitive and the present and preterite tenses, and nouns in the definite and indefi-
nite forms of the singular and the plural (a hyphen is used to mark the end of the lexical look-up items):

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>vb. inf.</td>
<td>LOP-E /L&quot;Ö:PE/</td>
<td>HOPP-E /H&quot;ÅPE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pres.</td>
<td>LÖP-ER /L.&quot;Ö:PER/</td>
<td>HOPP-ER /H&quot;ÅPER/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pret.</td>
<td>LÖP- /L.&quot;Ö:P/</td>
<td>HOPP-ET /H&quot;ÅPET/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n. sg. indef.</td>
<td>LÖP- /L.&quot;Ö:P/</td>
<td>HOPP- /H&quot;ÅP/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>def.</td>
<td>LÖP-ET /L.&quot;Ö:PE/</td>
<td>HOPP-ET /H&quot;ÅPE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pl. indef.</td>
<td>LÖP- /L.&quot;Ö:P/</td>
<td>HOPP- /H&quot;ÅP/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>def.</td>
<td>LÖP-ENE /L.&quot;Ö:PENE/</td>
<td>HOPP-ENE /H&quot;ÅPENE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vb. inf.</td>
<td>KJÖP-E /KJ&quot;Ö:PE/</td>
<td>HEND-E /H&quot;ÅNE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pres.</td>
<td>KJÖP-ER /KJ&quot;Ö:PER/</td>
<td>HEND-ER /H&quot;ÅNER/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pret.</td>
<td>KJÖP-TE /KJ&quot;ÖPTE/</td>
<td>HEND-TE /H&quot;ÅENTE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n. sg. indef.</td>
<td>KJÖP- /KJ&quot;Ö:P/</td>
<td>HEND-ER /H&quot;ÅNER/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>def.</td>
<td>KJÖP-ET /KJ&quot;Ö:PE/</td>
<td>HEND-ENE /H&quot;ÅENE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pl. indef.</td>
<td>KJÖP-ENE /KJ&quot;Ö:PENE/</td>
<td>HEND-ENE /H&quot;ÅENE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n. sg. indef.</td>
<td>BÖK- /B&quot;Ö:K/</td>
<td>HEND-ENE /H&quot;ÅENE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>def.</td>
<td>BÖK-EN /B&quot;Ö:KEN/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pl. indef.</td>
<td>BÖK-ER /B&quot;Ö:KER/</td>
<td>BÖK-ER /B&quot;Ö:KER/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>def.</td>
<td>BÖK-ENE /B&quot;Ö:KENE/</td>
<td>BÖK-ENE /B&quot;Ö:KENE/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is plain from the above examples (which could be multiplied), it is not possible to establish rules (other than of an ad hoc nature) to regulate the choice of toneme in front of e.g. the endings -ER, -EN, -ET, -ENE. An additional complication is the fact that in normal reading style the -ET is pronounced without the final consonant when it marks the definite singular of neuter nouns, but with a final /T/ when it marks the preterite of verbs. Again, it is impossible to formulate rules based on the graphemic shape of the words that will reliably result in the right choice being made.

**THE PROBLEM OF INAPPROPRIATE SELECTION OF A STEM**

In the original version of the Norwegian system the lexical look-up procedure accepts a word from the LEX when the word being examined can be analysed as consisting of this word as a STEM + an ending that is recognized by the system. But it is clear that this brings us no nearer to a solution of the problem of choice of toneme unless the lexical entry is somehow marked for which toneme is associated with which ending. One of the innovations I have made is precisely to this effect; I have introduced a small list of toneme class markers which have the desired effect. However, the problems involving endings and other modifications of the graphemic shape of words extend further than to the choice of toneme; sometimes stress position is affected, sometimes vowel length (cf. the preterite of KJÖPE), and sometimes the phonemic shape of a word will be radically different when two or more endings recognized by the system are attached to the same word STEM; for instance, if the system has the following lexical entries:

(15) OL /"O:EL/,
(16) JEG /J"El+/,
(17) TEL /TELEF"O:N/ 

and recognizes the following ENDINGS: -A, -T, -ER, -EN (which are all common inflectional endings), then it will not assign the correct pronunciation to...
unless these particular word forms are also listed in the LEX; if they are so listed, then the system will search for and find the full, correct forms without proceeding to analyse them into STEM + ENDING. It is therefore possible to avoid mispronunciations like /J'EIER/ (with rule-governed deletion of the +), /O:ELA/, /TELEF'O:NT/ (which were all produced by one of our versions) by listing the full forms in the LEX. But the need for separate lexical entries for many inflectional forms of one and the same lexeme was precisely what the STEM + ENDING procedure was meant to reduce. Besides, it is often impractical or even impossible to list all conceivable modifications of a STEM.

THE INTRODUCTION OF ALLOMORPHS

I have introduced the possibility of listing any number of ALLOMORPHS in the lexical entries; these ALLOMORPHS need of course not be allomorphs in the conventional sense - they are simply different phonemic representations of one and the same lexical look-up item. With this technique it is possible to specify which endings attach to which ALLOMORPH, including the case of there being no ending. The following examples will illustrate the extensive gain that can be had from this; the one lexical entry

\[ \text{KJÖP /KJ'Ö:Ps4aKJ'Ö:PeaKJ'ÖPf/} \]

will correctly generate among others the following forms:


and

\[ \text{TEL /T'E:LaTELf/TELEF'O:Nua/} \]

will generate e.g.:

\[ \text{TEL-T /T'ELT/, (23) TEL-TE /T'ELTE/, TEL-E /T'E:LE/, TEL-/TELEF'O:/N.} \]

Similarly,

\[ \text{LETTVIN /L'ETVI:NaL'ETVINf/} \]

will automatically make the correct distinction between (24) /L'ETVI:N/, /L'ETVI:NEN/ on the one hand, and (25) /L'ETVINTE/, /L'ETVINTE/ on the other.

In such entries "s1" ("s2", etc.) and "v1" ("v2", etc.) mark the respective ALLOMORPHS for toneme class (and "s1" etc. cause a dropping of a final T in the -ET ending of nouns); "f", "e", "r" and others indicate different categories of endings, "u" indicates "no ending", and "a" is used as an ALLOMORPH delimiter.
THE PROBLEM OF GERMANIC VS. LATIN AND OF COMPOUNDS

The problem of deciding on LATIN or GERMANIC I have attempted to solve by devising a system of marking in the GRAF rules elements that are diagnostic of the one category or the other in different ways; it then becomes possible to make reasonably successful guesses as to class membership on the basis of this and to calculate the position of the word stress accordingly. One outstanding problem is to calculate the position of word stress in compounds; a rule which has wide application is that if a word is compounded of two or more LATIN words then the stress will fall on the rightmost stressable element of the leftmost constituent "word"; for instance, in (26) ADMINISTRASJONS-DEPARTEMENTS-SEKRETÄR the stress falls as indicated by underlining (but the very last vowel retains vowel length as a reminder that this syllable is stressed when the word stands alone. Again, it is possible to formulate rules that will predict the correct stress with reasonable reliability in such cases, although much remains to be done before mine work satisfactorily in this respect.

WORDS WITH GERMANIC PREFIXES

The system of marking elements of a word as an indication of how those elements affect stress and toneme can be used to good effect in a large class of words of the following make-up: prefix + monosyllabic verbal root or verbal noun. The following table shows typical cases (in orthography, but with toneme marked in front of stressed vowel):

| (27) FOR-ST'Å | (31) 'AN-STÅ | (35) BE-ST'Å |
| (28) FOR-ST'AND | (32) 'AN-STAND | (36) BE-ST'AND |
| (29) FOR-SL'Å | (33) 'AN-SLÅ | (37) BE-SL'Å |
| (30) FOR-SLAG | (34) 'AN-SLAG | (38) BE-SL'AG |

Simple rules can be formulated involving a limited class of prefixes and stems, which will generate correctly most of the possible combinations of these. In this way it is possible to relieve the LEX of quite a heavy burden.

Conclusion: how can we deal with the problems that look insoluble?

There will of course still be some cases where e.g. the correct toneme cannot be established within the limits of the technology at present available; for instance the two readings of BÖKER quoted above cannot be distinguished on lexical or grammatical grounds. One possibility which seems worth exploring is to operate with intermediate phonetic realizations in this and some other cases. Granström & Gustafson (1987) look at some aspects of this; the results of a perception test show listeners to be more favourably disposed towards "toneme 1 1/2" than to either of the "true" tonemes when these were used incorrectly.
Meaning of Norwegian words quoted

(1) "hands" (noun)    (2) "happens"    (3) "philologist"
(4) "philology"  (5) "run" (verb)    (6) "run" (noun)
(7) "jump" (verb)  (8) "jump" (noun)  (9) "buy" (verb)
(10) "purchase" (noun)  (11) "happen"  (12) "hands" (noun)
(13) "beech"    (14) "books" (noun) 
(15) abbreviation for 'Olympiske leker': "The Olympic Games".
(16) 1st p. sg. personal pronoun: "T"; the + marks it as a function word that loses its
word stress in certain cases
(17) abbreviation: "telephone"  (18) boy's name
(19) "hunter"  (20) "tent"
(21) "frost (in the ground)" def. sg.  (22) "buying" (participle)
(23) "counted"  (24) "light wine", indef. and def. sg.
(25) "quick and easy" indef. sg. and indef. pl./def.
(26) "administration department secretary"  (27) "understand"
(28) "wits"   (29) "suffice"  (30) "proposal"
(31) "be proper"  (32) "dignity"  (33) "estimate" (verb)
(34) "estimate" (noun)  (35) "last" (verb)  (36) "stock (of animals)"
(37) "stud" (verb)  (38) "fitting" (noun)

References
