

A new Dictionary of Swedish Pronunciation

Per Hedelin and Dieter Huber
Chalmers University of Technology

Abstract

This paper describes some aspects of a pronunciation dictionary for Swedish, "Svenskt Uttaleslexikon" (SUL), which is presently developed at our department. This dictionary provides, among other items, three kinds of information about Swedish pronunciation that are not included in standard dictionaries: information on variants, on inflected forms and compounds, and on proper names. SUL is organized as a machine-readable lexical database which in its present form contains approximately 100.000 headwords. The run-time system comprises four separate processing modules: an inflection engine, a compound engine, the transcription engine, and the dictionary search algorithm. In addition to phonetic and phonological information, SUL also aims to supply various kinds of paradigmatic, syntagmatic and statistical information, needed for the linguistic processing stages in text-to-speech synthesis and automatic speech recognition.

1. Introduction

Large-scale, high-quality, machine-readable pronunciation dictionaries are a fundamental prerequisite for progress in various areas of speech and language technology. For instance, modern speech synthesis systems require access to a dictionary in order to look up the pronunciation of individual words and compounds which are to be converted from orthographic text to speech. The other way round, automatic speech recognition systems need pronunciation information from large-vocabulary lexical databases in order to match the incoming speech signal against possible word candidates.

However, while spelling-to-sound conversion (and its reverse) admittedly constitutes one important reason for the use of specialized pronunciation dictionaries in computer speech synthesis and recognition, it is certainly not the only one. Linguistic processing modules that perform syntactic and semantic parsing are becoming more and more integrated into both synthesis and recognition systems, with the purpose to supply relevant information for e.g. prosodic processing, pause marking, disambiguation, anaphoric resolution, and the detection and correction of erroneous input. In order to be able to support these modules in an efficient way, i.e. integrating both the speech processing and natural language processing stages of the overall system, various kinds of morphological, syntactic, semantic and even context-oriented information need to be included in the dictionary.

Ordinary dictionaries do not provide these different types of information within one single, comprehensive data structure that is immediately accessible to spoken language processing (SLP) in computer speech synthesis and recognition. For instance, conventional dictionaries, in as far as

they comprise any phonological information at all, usually include only one, i.e. the most common and least stigmatized (not necessarily the same!) lexical pronunciation for each headword in some kind of broad phonemic transcription. They thus ignore the inherent variability of natural human speech, dictated by influences of sociolect, dialect, and various features of speaker idiosyncrasy. Even more importantly, they do not provide any systematic information about alternative pronunciations that typically occur in continuous, connected speech, i.e. including different kinds (and degrees!) of assimilation, reduction, neutralization, and elision of individual speech sounds, caused by influences of coarticulation, speech rate, and stress.

The intended user of a conventional dictionary is a human being. With this in mind, lexicographers have over the centuries aspired to optimize the selection, interpretation and presentation of the material to be collected in a dictionary. Machine-readable dictionaries, on the other hand, are not written primarily for intelligent human users, but have to deal with the natural language processing and speech signal processing modules of computer systems. This apparently trivial observation has a number of far-reaching consequences concerning both the selection and the organization of the lexical items. Thus, conventional dictionaries *implicitly* assume that their users have access to other sources of linguistic and factual knowledge, and are able to extract and combine information from various sources in a sophisticated and experienced manner. On the contrary, when computers access data in a machine-readable dictionary, all information must be stated *explicitly*, i.e. either by way of exhaustive listing, or retrievable by applying a set of well specified, formalized rules. As a consequence, dictionaries for applications in the fields of speech and language technology compared with traditional, human-oriented lexica, have to meet considerably higher demands in *explicitness, exhaustiveness and consistency*.

This paper discusses some aspects of a pronunciation dictionary for Swedish, "Svenskt Uttalslexikon" (SUL), which is presently developed at our department for use in computer speech synthesis and recognition. Several early versions of the work have been documented, i.e. (Hedelin 1986), (Hedelin 1987) and (Hedelin 1989). Certain additional aspects of this dictionary are also described in Huber (1990). This present article focuses mainly on the formalization and representation of *phonological* (chapter 3) and *morphological* (chapter 4) information within the SUL lexical database. The treatment of syntagmatic information in future extensions of SUL is discussed in a separate paper published within this volume (cf. Huber 1992).

2. Lexicon Structure

In standard computer terminology, SUL constitutes a lexical database which is organized in records and fields. Each record comprises a collection of various types of information pertaining to one lexical item, the lexeme. The records are again subdivided into fields storing, for instance:

- the *headword* in normal orthography;
- a *phonemic transcription*;
- information on the *word class (part-of-speech)*;
- one (or several) *paradigm codes*.

In addition to these obligatory entries, SUL also contains fields for the storage of information on alternative spellings and pronunciations, on inflectional irregularities (both in the written and in the spoken domain), on semantics and etymology, homonym indices, collocations, and different kinds of statistical data. On the whole, each record in the SUL database comprises a total of 17 data fields. Two of the more important of these fields are briefly identified below:

Headword

The headword is listed in normal orthographic spelling. Acronyms, abbreviations and blends are treated as separate records. Homonyms (i.e. both homophones and homographs) are handled by using separate headwords marked with a homonym index and a number.

Word Class (Part-of-Speech)

Each headword is classified according to its normal usage into one of the categories *noun, pronoun, verb, adverb, adjective, numeral, preposition, conjunction, interjection, and article*. Words having multiple usage, i.e. pertaining to two or more word class categories, are duplicated as headwords, i.e. one entry for each category. Subcategories (e.g. auxiliary verb, subordinating conjunction, personal pronouns) are not specified. Proper nouns, acronyms and abbreviations are marked accordingly.

The representation of information about pronunciation and morphology is described in detail in the following sections of this article. A brief authentic excerpt from the SUL database (displaying only the first 5 of the total of 17 fields) is shown for illustration purposes in table 1 below.

anammelse	/a'nammelsə/	s5	Avldng anamma+-else	
anamnes	/anam'ne:s/	s3		<i>medicin: sjukdomshistoria, sjukbikt</i>
anamnestisk	/anam'nestisk/	a3	Avldng anamnes+-tisk	<i>anamnes-</i>
anamnetisk	/anam'ne:tisk/	a3	Avldng anamnes+-isk	<i>anamnes-</i>
anamorfos	/anamor'fo:s/	s3		<i>ombildning</i>
ananas	/annanas/	s9		<i>botanik: släkte av familjen ananas-växter; en frukt</i>

Table 1. Excerpt from the SUL database

3. Pronunciation

3.1 Background

Conventional dictionaries of Swedish, in as far as they provide any phonological information at all, generally seem to treat pronunciation in a rather ad hoc and untheoretical way. That is, older dictionaries, dating back to the beginning of this century or even earlier (e.g. Dalin 1850, SAOB 1898, Lyttkens & Wulff 1911) often use notation systems that do not conform with our present standards of either broad "phonemic" or narrow "phonetic" (i.e. allophone-oriented) transcription. More recent, general-purpose dictionaries (e.g. Molde 1977, Östergren 1981, Allén et al 1986, Malmström et al 1988, Collinder 1989) usually provide pronunciation information for a subset of headwords only, and often use notation systems based on graphemic analogies rather than on established phonetic transcriptions (e.g. IPA 1949). Only few dictionaries of Swedish include systematic information on pronunciation usage for all lexical entries, all of them, however, covering only a very limited and highly specialized subset of present-day Swedish vocabulary such as for instance geographical names (e.g. SprN 1977, 1979, 1981). The only more recent, general-purpose Swedish dictionary which provides transcriptions systematically for all headwords is Lexin (1984), a 12.000-headword lexicon intended primarily for use in teaching Swedish as a second language. SFL-oriented information on the pronunciation of Swedish can also be found in a number of multi-lingual, foreign-learner dictionaries published abroad.

Until recently, the Swedish Language Council (Svenska Språknämnden) in Stockholm was preparing the publication of a major pronunciation dictionary of modern Swedish. This project has, however, been abandoned.

Clearly, there is an urgent need for the compilation of a large-scale, high-quality pronunciation dictionary of modern Swedish, which meets the standards of present-day phonetics and lexicology, and which needs to be provided both as a machine-readable electronic database for use in computer speech applications, and as a conventional hard-copy lexicon for interested human users. Similar dictionaries are already available or under development for a number of European (e.g. Wells 1990, Molbæk Hansen 1990) and non-European (e.g. EDR 1990) languages.

3.2 Notation

Principally, the transcription of pronunciation can be either narrow (allophone-oriented) or broad (phoneme oriented). For some applications (e.g. speech synthesis or speaker verification) a narrow, allophonic transcription may be desired, whereas for other applications (e.g. automatic speech recognition) a broad transcription may be more useful (cf. Huber 1990 for a more detailed discussion). Based on these requirements, SUL follows an essentially dual approach, viz:

- all transcriptions listed explicitly in the SUL database are basically *phoneme-oriented*,
- *allophone-oriented* transcriptions are generated by rule whenever required.

Phonetic (i.e. allophone-based) transcriptions of inflected word forms and compounds are thus performed first in the phoneme domain, before handling phoneme-to-allophone conversion by rule in the automatic transcription system. The resulting three-step procedure for generating both phonemic and phonetic transcriptions of inflections and compounds is illustrated below in table 2 for the inflected Swedish noun "gatorna" (Eng: the streets):

<i>database storage</i>	phoneme string	/ga.ta/
<i>word inflexion</i>	phoneme string	/ga.ta/ + /-o.rna/ = /ga.ta.rna/
<i>conversion rules</i>	allophone string	[`ga:ta.rna]

Table 2. Phoneme-to-allophone conversion in SUL.

3.2.1 Broad Transcription

The notation system used in SUL for transcriptions at the phoneme level is based on the following phoneme concepts:

Vowels: /a/ /e/ /i/ /o/ /u/ /y/ /ø/ /ɛ/ /ɞ/ /ə/

Consonants: /b/ /d/ /g/ /m/ /n/ /ŋ/ /f/ /r/ /ʌ/ /v/
/p/ /t/ /k/ /l/ /s/ /ʃ/ /ʒ/ /ç/ /h/

Prosody: /' /` / /' / /' /' / /~ / / : /

Note that in order to be able to perform the phoneme-to-allophone conversion automatically and in both directions in a consistent and unequivocal manner, we had to make some concessions to the traditional way of defining the phoneme system of Swedish in standard textbooks of phonology. These concessions refer (i) to the extensive use of prosodic markers, and (ii) to the unorthodox classification of /ə/ and /ʒ/ as phonemes during the phonemic processing stages. On the whole, our

SUL phoneme transcriptions are more detailed than what is customary for a phoneme-based transcription system in other applications (cf. for instance Elert 1970 and Garlén 1988). As stated earlier, however, all phoneme labels in SUL are written with the inherent problems of automatic allophone conversion in mind. The overriding consideration in devising our system has been to provide for simple and safe phoneme-to-allophone mapping in both directions.

The phoneme strings for each headword are stored in the transcription field of the SUL database, using normal, lowercase ASCII characters: one ASCII character per symbol.

Note also that our SUL broad transcriptions mark stress and accent with the symbols placed in the *initial* position of a syllable, as illustrated in the following examples of Swedish words: "springer" (i.e. /'sprɪŋə r/), "springor" (i.e. /'sprɪŋ(ɔ) r/) and "springbrunn" (i.e. /'sprɪŋ-²brʉn/). IPA (1949) follows the same convention, whereas Swedish phoneticians often tend to place the corresponding symbols at (or after) the stressed vowel.

The symbol /-/ as a prosodic marker is used for the dual purpose: (i) to classify a word as a prosodic compound, thereby indicating that the word is pronounced with a special version of the grave accent (accent II), and (ii) to mark the prosodic boundary. SUL thus writes for instance:

/'a:v¹ga:s-²rø:r/ and ['a:v¹ga:s-²ræ:r]

for the Swedish noun compound "avgasrör" (Eng: exhaust pipe) in the phoneme and in the allophone domain respectively. Likewise the dictionary lists:

/be'ta:lnɪŋ/,	[bɛ`tə:lnɪŋ]
/be'ta:lnɪŋs-ba'lans/,	[bɛ`tə:lnɪŋs-ba'lans]
/be'ta:lnɪŋsba'lans-prø'ble:m/	[bɛ`tə:lnɪŋsba'lans-prø'ble:m]

for compounds pronounced with the Swedish word accent II. On the other hand, whenever a morphological compound does correspond to a prosodic compound, as for instance in the Swedish words "trädgård" (i.e. /'trɛggɔrd/), "Bergman" (i.e. /'berjman/, note the prosodic distinction between the proper noun indicating the name "Bergman" and the ordinary noun "bergman" indicating an occupational group corresponding to the English "miner"), "Alingsås" (i.e. /aliŋs'o:s/) and "Dalby" (i.e. /'da:lby/), the prosodic marker /-/ is *not* used. It must be appreciated in this context that our definition of a prosodic compound is not necessarily equivalent to a compound in the morphological sense.

Finally, the distinction between long and short *syllables* is maintained by the use of prosodic stress markers, whereas the (binary) distinction between long and short *vowels* is marked in the SUL phoneme system by placing the duration sign /:/ after the vowel symbol. The distinction between long and short *consonants*, on the other hand, is not maintained in the SUL system at all, i.e. the duration sign /:/ is used for vowels only. However, for any long syllable containing a short vowel and only one consonant, the consonant symbol is repeated in all phoneme transcriptions, such as for instance in /'kall/, /'takk/ and /'takka/, but not in /'kalt/ or /'takla/.

3.2.2 Narrow Transcription

SUL uses 46 allophone labels to describe the inherent variability of the Swedish phoneme system. These labels are listed and exemplified below in the tables 3 (vowels) and 4 (consonants). The authors apologize that for technical reasons it has not been possible to correctly reproduce all the diacritic symbols in this report (cf. appendix C).

[a:]	<i>mat</i>	[a]	<i>att</i>		
[e:]	<i>vet</i>	[ɛ]	<i>vett</i>	[ə]	<i>året</i>
[i:]	<i>vit</i>	[ɪ]	<i>vitt</i>		
[u:]	<i>bo</i>	[ɔ]	<i>bott</i>		
[ɯ:]	<i>hus</i>	[ø]	<i>hund</i>		
[y:]	<i>byt</i>	[ʏ]	<i>bytt</i>		
[o:]	<i>gd</i>	[ɔ]	<i>gått</i>		
[ɛ:]	<i>säl</i>	[ɛ]	<i>vätt</i>		
[æ:]	<i>här</i>	[ɛ]	<i>kärr</i>		
[ø:]	<i>hō</i>	[ɛ]	<i>höst</i>		
[œ:]	<i>hör</i>	[œ]	<i>förr</i>		

Table 3. SUL notation for the vowel allophones

[b]	<i>bo</i>	[d]	<i>dag</i>
[d]	<i>bord</i>	[f]	<i>fisk</i>
[g]	<i>gata</i>	[h]	<i>hel</i>
[j]	<i>ljuv</i>	[k]	<i>kunna</i>
[l]	<i>leta</i>	[j]	<i>kärl</i>
[m]	<i>meta</i>	[n]	<i>natt</i>
[n]	<i>barn</i>	[ŋ]	<i>ting</i>
[p]	<i>peta</i>	[r]	<i>reta</i>
[s]	<i>sol</i>	[ʂ]	<i>fors</i>
[ʃ]	<i>sjö</i>		
[t]	<i>tak</i>	[t]	<i>fart</i>
[ɕ]	<i>tjuv</i>	[v]	<i>veta</i>

Table 4. SUL notation for the consonant allophones

In addition to these symbols, SUL also provides a special notation to mark the pronunciation of diphthongs, which, although they do not form part of the Swedish vowel system, frequently occur in loan words of foreign origin, such as for instance *Audi*. These occurrences are transcribed on the allophone level by using (i) the two vowel allophones that most accurately describe the standard Swedish pronunciation of these sounds, combined with (ii) a subscript ligature, resulting in for instance [a_ɔ] in the previous example.

Note also that the transcriptions in our dictionary frequently use the allophone [ɔ] in foreign loan words where the orthographic spelling uses "o" (e.g. in "abolition"). Some transcribers might prefer to use the allophone [ɔ] in these contexts. Another fine distinction in Swedish concerns the difference between the short allophones [ɛ] and [ɛ], which are often confused due to the dialect of Stockholm and its preference for [ɛ] in both /e/ and /e/ contexts. In both cases, i.e. for the /ɔ/ - /o/ as well as for the /e/ - /e/ distinction, SUL attempts to describe Standard Swedish (Rikssvenska) and follows the conventions established in Lyttkens & Wulff (1911).

4. Morphology in SUL

The inclusion and adequate representation of morphological information in a pronunciation dictionary, both for man and machine, serves several important purposes. Word inflection, both in the orthographic and in the phonetic domain is the main motivation. The generation of derived forms and compounds, again both in a written and a spoken version, is almost equally important. Computer systems without access to morphological data would need to resort to an exhaustive listing of all the potentially possible inflected forms for each lexeme in order to correctly associate incoming word candidates with their lexical heads as they appear in the dictionary. This kind of explicit listing is not only highly uneconomical (and unintelligent!) in practical applications, it would also be outright impossible with regard to the practically unlimited number of compounds and derivations contained in languages like Swedish.

Swedish is both an inflectional language and a language with a strong potential to form new words by compounding and by derivation. All three of these morphological processes may (and often do!) change the phonological realization of a word in a variety of ways, including for instance stress shifts, reductions, assimilations, elisions, syncope and epenthesis. These changes are governed by a large set of rules and, unfortunately, an even larger list of exceptions.

In analogy with the approach presented by Hellberg (1978), SUL attempts to capture the morphophonemic variability of Swedish in a representational format which integrates the changes of the words induced by compounding and derivation with their inflection. It must be appreciated in this context that the inflectional and derivational behaviour of words in Swedish, like in many other languages, is considerably more complex in the phonological than in the orthographical domain. For instance, stress shifts and changes in word accent patterns are manifested in spoken language only. The same applies to phenomena like vowel shortening or the realization of the mute final "e" in words like "garage". Morphological systems for use in spoken language processing therefore require a significantly larger number of paradigms than systems which operate solely in the domain of written language such as that of Hellberg.

SUL needs to operate in both domains, i.e. it needs to handle inflection, derivation and compounding both in written *and* in spoken language, in order to be useful in applications such as text-to-speech synthesis and speech-to-text conversion (i.e. automatic speech recognition). The morphological system implemented in our dictionary therefore comprises in its present form a total of 500 paradigms, as compared for instance with Hellberg's text-based system of 235 paradigms (Hellberg 1978). In order to facilitate the rule-based selection of the correct paradigm during morphological analysis and generation, these 500 paradigms are grouped into 24 *paradigm families* (the paradigm family codes used in SUL are given in the brackets):

- 13 for the nouns (s0, s1...s11, s20);
- 6 for the adjectives (a1...a6);
- 5 for the verbs (v1...v5).

A summary of these 23 paradigm families, each exemplified by a word following the main paradigm within the family, is listed below in table 5.

A complete list of all 500 paradigms grouped into the 24 paradigm families is attached to this paper as appendix A.

Observant readers will notice that the division into paradigm families implemented in SUL follows, to a certain degree, the traditional concepts of *declension* (for nouns and adjectives) and *conjugation* (for verbs), established in standard textbooks of Swedish grammar. Thus, the paradigms included in the paradigm family s1, for instance, emulate the inflection patterns of the *or*-declension, s2 the *ar*-declension, v1 the first conjugation, etc. However, irregular inflection

paradigms, which are often simply listed as exceptions in traditional descriptions of Swedish morphology, are here collected and classified into separate paradigm families, in order to permit their strictly rule-based generation and analysis within the SUL inflection and compounding system.

s0	<i>vådjan</i>	v1	<i>bada</i>	a1	<i>stark</i>
s1	<i>lampa</i>	v2	<i>våga</i>	a2	<i>röd</i>
s2	<i>burk</i>	v3	<i>blänka</i>	a3	<i>defekt</i>
s3	<i>bild</i>	v4	<i>tro</i>	a4	<i>död</i>
s4	<i>fett</i>	v5	<i>bedja</i>	a5	<i>framstående</i>
s5	<i>sko</i>			a6	<i>bra</i>
s6	<i>dike</i>				
s7	<i>trääd</i>				
s8	<i>plock</i>				
s9	<i>tro</i>				
s10	<i>jägare</i>				
s11	<i>bok</i>				
s20	<i>pengar</i>				

Table 5. Paradigm families used in SUL

Table 6 below shows a short, authentic excerpt which illustrates the main paradigm of paradigm family **s1**, generating all possible inflections of the Swedish noun "flicka" (Eng: girl) in the graphemic, phonemic and phonetic domains:

s1	a	/*/				
			<i>flicka</i>	/flicka/	-	[flik:a]
			<i>flickan</i>	/flickan/	-	[flik:an]
			<i>flickor</i>	/flickor/	-	[flik:or]
			<i>flickorna</i>	/flickorna/	-	[flik:ona]
genitiv:			<i>flickas</i>	/flickas/	-	[flik:a]
			<i>flickans</i>	/flickans/	-	[flik:an]
			<i>flickors</i>	/flickors/	-	[flik:os]
			<i>flickornas</i>	/flickornas/	-	[flik:onas]

Table 6. Inflection paradigm example.

Right hand column giving an allophone translation is included for clarity only.

A complete listing of the four paradigms belonging to the paradigm family **s1** is shown in appendix B.

During operation in the run-time system, the paradigms are automatically converted into executable rules for stem generation and for suffix merge. A dual set of inflections is thus generated, one for the orthographic and one for the phonetic domain.

5. Status and Goals

The Swedish pronunciation dictionary (SUL) described in this paper comprises in its present form a total of approximately 100.000 headwords, each with a full grapheme representation (including relevant alternative spellings), a homograph index (if required), a word class specification, the paradigm family code, and a broad "phonemic" transcription (including accentuation and word level stress) representing standard pronunciation usage. Proper names are further classified as names of persons, places, products, etc. Some etymological information is included, mainly for words of foreign origin and for "outdated" Swedish words still in use (e.g. in literary language).

The dictionary incorporates an inflection engine, a compound engine and a transcription engine, which together are capable of generating all possible inflected forms (i.e. 800.000) and a theoretically infinite number of compounds, in the orthographical as well as in the transcription domain.

Evaluation of the dictionary in practical text-to-speech synthesis showed that, depending on the material, on the average between 95 and 99 % of the orthographic words commonly found in Swedish newspaper texts are correctly handled. Word inputs missed by SUL comprise mainly proper nouns and foreign names and expressions.

These figures are based on the evaluation with text input. We have not yet evaluated the performance of SUL in the reverse task, viz. for spoken language input in applications such as automatic speech recognition.

Our goals for the future include (i) to extend the number of headwords from the present 100.000 entries to approximately 150.000 entries, (ii) to increase the number of variant transcriptions representing alternative pronunciations, and (iii) to incorporate various kinds of syntagmatic information (cf. Huber 1992).

Finally, we would like to mention that earlier versions of SUL have already been acquired by other groups of researchers working in the field of spoken language processing, both inside and outside academia. We are presently working with compiling a Macintosh based version of SUL which we hope to be able to make available to other researchers within the nearest future. We have also been approached with the suggestion to provide a hard-copy edition of SUL for the general public, based on a selection from the larger file of lexical data compiled at our department (cf. appendix C).

Acknowledgements

Our work with compiling SUL has been considerably facilitated by our cooperation with the Department of Computational Linguistics (Språkdata) at Gothenburg University. We want to thank Professor Sture Allén and Christian Sjögreen for their continuous support and for making available to us several of their lexical databases. We are also greatly indebted to Margareta Westman and Claes Garlén of the Swedish Language Council (Svenska Språknämnden) who helped us with matters of phonological representation and transcription.

References

- [Allén 1986] S. Allén et al, "Svensk Ordbok", Stockholm: Esselte Studium, 1986.
- [Collinder 1989] B. Collinder, "Stora Ordboken" (nyutgåva), Lund: Liber Förlag/Läsförlaget, 1989.
- [Dalin 1850] A.F. Dalin, "Ordbok öfver svenska språket", 1850-1852.

- [EDR 1990] EDR Technical Report, "An Overview of the EDR Electronic Dictionaries", TR-024, Japan Electronic Dictionary Research Institute, Tokyo, 1990
- [Elert 1979] C.-Ch. Elert, "Ljud och Ord i Svenskan", Almqvist & Wiksell, 1970.
- [Garlén 1979] C. Garlén, "Svenskans Fonologi", Lund: Studentlitteratur, 1988.
- [LEXIN 1984] "Svenska Ord med Uttal och Förklaringar", LEXIN-projektet, Stockholm: Esselte Studium, 1984.
- [Hellberg 1978] S. Hellberg, "The Morphology of Present-day Swedish", *Data linguistica*, Stockholm: Almqvist & Wiksell, 1978.
- [Hedelin 1986] P. Hedelin & A. Jonsson, "Svenskt uttalslexikon; 1:st edition", Technical Report, Chalmers University of Technology, 1986.
- [Hedelin 1987] P. Hedelin & A. Jonsson, "Svenskt uttalslexikon; 2:nd edition", Technical Report, Chalmers University of Technology, 1987.
- [Hedelin 1989] P. Hedelin, A. Jonsson & P. Lindblad, "Svenskt uttalslexikon; 3:rd edition", Technical Report, Chalmers University of Technology, April 1989.
- [Huber 1990] D. Huber, "An Electronic Dictionary for Computer Speech Applications", *Proceedings of the International Workshop on Electronic Dictionaries*, pp.130-140, Oiso, Japan, 1990.
- [Huber 1992] D. Huber, "Integrating Syntagmatic Information in a Dictionary for Computer Speech Applications", 1992 (this volume).
- [IPA 1949] IPA, "The Principles of the International Phonetic Association, London, 1949
- [Lytkens and Wulf, 1911] Lytkens & Wulf, "Svensk Ordlista", Gleerup, 1911.
- [Malmström, 1988] Malmström, Györki och Sjögren, "Bonniers Svenska Ordbok" (nyutgåva), Stockholm: Bonniers, 1988.
- [Molbæk Hansen 1990] P. Molbæk Hansen, "Udtaleordbog", Copenhagen: Gyldendal, 1990.
- [Molde, 1977] B. Molde, "Illustrerad Svensk Ordbok", Natur och Kultur, 1977.
- [SAOB, 1898] "Svenska akademien ordbok (SAOB)", Gleerup, 1898ff.
- [SprN, 1979] J. Sahlgren & G. Bergman, "Svenska Ortnamn med Uttalsuppgifter", Stockholm: Esselte Studium, 1979.
- [SprN, 1977] G. Bergman, "Språknämndens Uttalsordlista", Esselte Studium, (1965) 3:e uppl. 1977.
- [SprN, 1991] C. Garlén & A-C Mattisson, "Svenska Ortnamn: Uttal och Stavning", Norstedts, 1991.
- [Wells 1990] J.C. Wells, "Longman Pronunciation Dictionary", Longman, 1990
- [Östergren, 1981] O. Östergren, "Nusvensk Ordbok" (nyutgåva), Stockholm: Esselte Studium, 1981.

Per Hedelin and Dieter Huber
 Department of Information Theory
 Chalmers University of Technology
 S-412 96 Gothenburg, Sweden

Appendix A

Excerpt of the paradigm data-base

s 1	a , /*/			-
		<i>flicka</i>	/flikka/	-
		<i>flickan</i>	/flikkan/	-
		<i>flickor</i>	/flikkor/	-
		<i>flickorna</i>	/flikkorɔma/	-
genitiv:		<i>flickas</i>	/flikkas/	-
		<i>flickans</i>	/flikkans/	-
		<i>flickors</i>	/flikkorɔs/	-
		<i>flickornas</i>	/flikkorɔmas/	-
s 1	el , /ə/			-
		<i>toffel</i>	/toffel/	-
		<i>toffeln</i>	/toffelɩn/	-
		<i>tofflor</i>	/tofflor/	-
		<i>tofflorna</i>	/tofflorɔma/	-
genitiv:		<i>toffels</i>	/toffels/	-
		<i>toffelns</i>	/toffelɩns/	-
		<i>tofflors</i>	/tofflorɔs/	-
		<i>tofflornas</i>	/tofflorɔmas/	-
s 1	s , /s/			-
		<i>ros</i>	/rɔ:s/	-
		<i>rosen</i>	/rɔ:sɛn/	-
		<i>rosor</i>	/rɔ:sɔr/	-
		<i>rosorna</i>	/rɔ:sɔrna/	-
genitiv:		<i>rosens</i>	/rɔ:sɛns/	-
		<i>rosors</i>	/rɔ:sɔs/	-
		<i>rosornas</i>	/rɔ:sɔrnas/	-
s 1	* , /*/			-
		<i>våg</i>	/vo:g/	-
		<i>vågen</i>	/vo:gɛn/	-
		<i>vågor</i>	/vo:gɔr/	-
		<i>vågorna</i>	/vo:gɔrna/	-
genitiv:		<i>vågs</i>	/vo:gs/ -	-
		<i>vågens</i>	/vo:gɛns/	-
		<i>vågors</i>	/vo:gɔs/	-
		<i>vågornas</i>	/vo:gɔrnas/	-

Appendix B

The list below illustrates all paradigms by examples of words. Each example represents one paradigm.

- s0 *Alex, Alice, vädjan*
s1 *flicka, toffel, ros, våg*
s2 *choke, båge, hummer, lämmel, nagel, öken, dager, serve, sky, gran, mun, sjukdom, gam, kam, sommar, caterpillar, lapp*
s3 *aveny, dekor, professor, doktor, modul, konsul, byrå, balsam, muskel, kansler, fiber, kollega, ven, vän, dam, parallelogram, geranium, akademi, kö, analys, bild*
s4 *bageri, gage, fängelse, sekel, studium, drama, fett*
s5 *ensemble, creme, apache, damejanne, pavane, cape, grace, manege, champagne, offside, pose, bagette, sko*
s6 *altare, huvud, tabu, knä, spö, bi, dike*
s7 *garage, bekymmer, fönster, faner, kummel, kapitel, album, system, gram, ton, bäcken, vatten, träd*
s8 *natrium, minimum, beröm, flimmer, bladder, tummel, babbel, vatten, kaffe*
s9 *agio, lappri, sperma, bagage, college, avantgarde, plock, insyn, service, grape, massage, ampere, chamotte, chartreuse, tro, judendom, rom, gödsel, fräken, liter, purpur, radar, peppar, dollar, terror, fosfor, humor, standard*
s10 *kärande, bagare, kritiker, rododendron*
s11 *farao, cesar, bok, tång, fot, öra, bekant, fröken, botten, braxen, öga, hinder, finger, man, middag, söndag, måndag, tisdag, onsdag, torsdag, fredag, lördag, vardag, namnsdag, middag, land, and, baryton, morgon, son, farbror, bror, mor, mormor, farfar, far, broder, moder, fader, afton, officer, toker, hammare, dotter, spektrum, center, tema, dekanus, eforus, amaryllis, emeritus, deputerad, lus, gås, get, gnet, nöt, ledamot, rot, natt, bokstav, historia, stad, verkstad, bonde, jävul, jävel, predikan, duo, albino, lexikon, examen, spann, snö, orden, sägen, sjö, pilgrim, altare, genre, himmel, papper, lager, pansar*
s20 *pengar*
v1 *bada, andas*
v2 *tända, föda, skilja, följa, skämmas, stämma, kännas, känna, böja, sörja, föra, mala, tåla, kräva, trivas, blygas, våga*
v3 *mäta, hyfta, minna, begynna, synas, blåsa, blänka*
v4 *brås, tro*
v5 *begrava, drita, be, bedja, binda, bita, bjuda, bli, bliva, bringa, brinna, brista, bryta, byta, bära, böra, dimpa, dra, draga, dricka, driva, drypa, duga, dväljas, dyka, dö, dölja, falla, fara, förfaras, finna, fisa, flyga, flyta, fnysa, frysa, få, förgäta, förnimma, gala, ge, gifta, giva, gjuta, glida, gnida, glädja, gripa, gråta, gå, gås, göra, ha, hava, heta, hinna, hugga, hålla, idas, kliva, klyva, knipa, kryta, koka, komma, krypa, kunna, kvida, kväda, le, lida, ligga, ljuda, ljuga, ljuta, lyda, lysa, lägga, låta, löpa, minna, minnas, mysa, må, måste, niga, njuta, nypa, nysa, pipa, pysa, rinna, rysa, rädas, se, simma, sitta, sjuda, sjunga, sjunka, skina, skita, skjuta, skola, skrida, skrika, sprida, spörja, strida, vrida, rida, skriva, riva, trivas, skryta, tryta, ryta, skälva, skära, slinka, slinta, slippa, slita, sluta, slå, slåss, smita, smyga, smälla, smälta, smörja, snika, snyta, sova, spinna, spricka, springa, spritta, sticka, stiga, stinka, stjäla, stryka, ryka, strypa, stå, stöda, stödja, suga, suppa, svida, svika, svälja, välja, svälta, svära, svärja, säga, säja, sälja, sätta, taga, tiga, tjuta, töras, tvinga, två, tälja, tämja, tör, töras, vara, varda, veta, vika, vilja, vina, vinna, vänja, växa, äta, ta*
a1 *vacker, ädel, ljummen, ilsken, kavat, tunn, frigid, brydd, blond, sakta, tam, dum, ledsam, allmän, vig, avig, trygg, svensk, dov, stark*
a2 *fri, blöt, rädd, god*
a3 *beige, disträ, olovlig, alternativ, komisk, ansedd, hybrid, aktad, anlagd, brunnen, kommen, svullen, avlång, fjäderlätt, jättevacker, femdubbel, halvgammal, perifer*
a4 *arbetsfri, benvit, död, skyhö*
a5 *framstående*
a6 *bra, dålig, gammal, grov, hög, liten, låg, lång, stor, trång, tung, ung*

Appendix C

<p>a subst. s6 <i>bet</i> ['a:æ] pl <i>an</i> ['a:n] <i>ana</i> ['a:nə] • <i>alfabetets första bokstav</i>. • <i>Böjningar skrivs ofta a:et, a:n osv.</i></p>	['a:]	<p><i>övert på bl a kolonnpapird.</i> Jmfr: SAOB ['a:bakes]. SiOB ['a:b]. SAOL'' ['a:b]. (< av Latin <i>abacus</i>, av Grek. <i>abax</i>, något hopfogat)</p>
<p>å prep • <i>till, för</i>. (< av Fra: å, till, av Latin: <i>ad</i>)</p>	['a]	<p>abandon subst. s9 <i>abandonen</i> [abaŋ'dɔŋən] • <i>ledighet, otvungenhet i uppträdandet; uppsluppenhet</i>. Jmfr: L&W (1911) [abaŋ'dɔŋ]. SAOB [abaŋ'dɔŋ]. SiOB [abaŋ'dɔŋ]. SAOL'' [-aŋ'dɔŋ] el [-aŋ'dɔŋ]. (< av Fra: <i>abandon</i>, med samma betydelse, av <i>être</i> å <i>bandon</i>)</p>
<p>capella adverb alt: [aka'pɛ:l:a] • <i>utan ackompanjemang: med endast sångstämmor</i>. (< av Ital: <i>capella</i>, som i kapellet, av Latin: <i>capella</i>, liten get, diminutiv av <i>capra</i>, get)</p>	[aka'pɛ:l:a]	<p>abasi subst. s9 <i>abasin</i> [aba'si:n] • <i>medicin: oförmåga att gå</i>. (< av Grek: <i>α-, icke- + basis</i>, gång)</p>
<p>conto adverb alt: [a'konto]; el [-to]. • <i>Eller a konto</i>. • <i>i räkning</i>. Jmfr: MSEOB [a'konto]. (< av Ital: <i>o conto</i>)</p>	[a'konto]	<p>abbé subst. s3 <i>abbén</i> [a'bɛ:n] pl <i>abbéer</i> [a'bɛ:ər] <i>abbéerna</i> [a'bɛ:ərən] • <i>titel för abbot (också använt för katolsk präst i allmänhet)</i>. Jmfr: L&W (1911) [a'bɛ]. SAOB [a'bɛ]. DMTOB [a'bɛ]. (< av Fra: <i>abbé</i>, <i>abbot</i>, av Latin: <i>abbas</i>)</p>
<p>dato adverb alt: [-to]. • <i>från i dag: från utgivningsdagen</i>. Jmfr: SAOB [-'dato]. MSEOB [a'dato]. (< av Latin: <i>dato</i>, av <i>datum</i>)</p>	[a'dato]	<p>abbodissa subst. s1; evid <i>abbodissan</i> [aba'dɔ:sən] pl <i>abbodissor</i> [aba'dɔ:sər] <i>abbodissorerna</i> [aba'dɔ:sərən] • <i>föreståndarinnan för nunnekloster</i>. Jmfr: L&W (1911) [aba'dɔ:sən]. SAOB [aba'dɔ:sən]. SiOB [-a's- / -a's- /]. (< av Fra: <i>abbodissa</i>, av <i>antika latin</i> <i>abbodissa</i>)</p>
<p>à jour adjekt, s ob • <i>uppdaterad; underrättad om: från denna dag</i>. (< Fra: <i>à jour</i>)</p>	[a'ʒur]	<p>Abbekås subst. s0 alt: ['abɛ:~kɔ:s]. • <i>egennamn: ort i Malmöhus län</i>.</p>
<p>à la adverb • <i>Nästan alltid obetonat [a].</i> • <i>enligt</i>. • <i>I förbindelser som <i>biff à la Lindström, ris à la Malta</i></i>. (< Fra: <i>à la</i>)</p>	['a:lə]	<p>abbortart subst. s3; ssg alt: ['abɔ:r~a:t]. • <i>egennamn: ort i Malmöhus län</i>.</p>
<p>à la carte subst, s ob • <i>rätt på meny som tillagas på beställning</i>. (< Fra: <i>à la carte</i>)</p>	[a'lakart:]	<p>abborre subst. s2; s1 • <i>zool: en sövansens fisk (Perca fluviatilis)</i>. Jmfr: L&W (1911) [a'bɔ:rɔ]. SAOB [a'bɔ:rɔ]. SiOB [a'bɔ:rɔ]. DMTOB [a'bɔ:rɔ]. (< av Fra: <i>abborre</i>, av <i>ogh</i>, spegigt <i>brämli</i>)</p>
<p>posteriori adverb alt: [-,postɛr-]. • <i>i efterhand</i>. (< Latin <i>a posteriori</i>, eg från det efterföljande)</p>	[a,postɛr'i:ɔr]	<p>abborrgrund subst. s7; ssg alt: ['abɔ:r~grʉnd]. • <i>zool: en släkte av fiskar</i>. • <i>egennamn: ort i Blekinge län</i>.</p>
<p>priori adverb • <i>på förhand</i>. Jmfr: SpN /uprãv/. (< Latin <i>priori</i>, eg från det föregående)</p>	[apri'ɔr]	<p>Abborrtorsk subst. s0 • <i>egennamn: ort i Västerbottnens län; i Norrbottens län</i>.</p>
<p>Aage subst. s0 • <i>egennamn: mansnamn</i>.</p>	['a:ge]	<p>abbot subst. s2 alt: ['abɔ:t]. <i>abboten</i> ['abɔ:tən], pl <i>abbotar</i> ['abɔ:tər] <i>abbotarna</i> ['abɔ:tərən] • <i>föreståndare för kloster; munk</i>. Jmfr: L&W (1911) [abɔ:t]. SAOB [abɔ:t] el [a'bu:t]. Även ['abɔ:t] el [a'bu:t], <i>målan</i> [abɔ:t~bu:t]. SiOB [abɔ:t]. SpN /mãw/. DMTOB [abɔ:t]. (< av Fra: <i>abote</i>, över <i>latin</i>, av arameiska <i>abba</i>, <i>fader</i>)</p>
<p>Aalborg subst. s0 • <i>egennamn: ort på Jylland</i>.</p>	['a:il,bɔ:r]	<p>abbotsdöme subst. s6; evid alt: ['abɔ:t~dø:mə]. • <i>Eller abbotdöme</i>. • <i>Sundom med kort ö-vokal, [-,dø:mə]</i>. • <i>egennamn: ort på Jylland</i>.</p>
<p>Aarhus subst. s0 • <i>egennamn: ort på Jylland</i>.</p>	['a:ər,hʉ:s]	<p>abbotsdöme subst. s6; evid alt: ['abɔ:t~dø:mə]. • <i>Eller abbotdöme</i>. • <i>Sundom med kort ö-vokal, [-,dø:mə]</i>. • <i>egennamn: kvinnonamn</i>.</p>
<p>Aasa subst. s0 • <i>egennamn: kvinnonamn</i>.</p>	['a:sə]	<p>abbotskap subst. s7; evid alt: ['abɔ:t~skɑ:p]. • <i>egennamn: ort i Blekinge län</i>.</p>
<p>Aasarum subst. s0 • <i>Eller Asorum</i>. • <i>egennamn: ort i Blekinge län</i>.</p>	['a:sə~rʉm:]	<p>abbotsstift subst. s7; ssg alt: ['abɔ:t~stift]. • <i>tydligt</i> ['abɔ:t~stift]. • <i>Eller abbotstift</i>. • <i>egennamn: ort i Blekinge län</i>.</p>
<p>Aase subst. s0 • <i>egennamn: kvinnonamn</i>.</p>	['a:sə]	<p>abbotsstift subst. s7; ssg alt: ['abɔ:t~stift]. • <i>tydligt</i> ['abɔ:t~stift]. • <i>Eller abbotstift</i>. • <i>egennamn: mansnamn</i>.</p>
<p>Aatos subst. s0 • <i>egennamn: mansnamn</i>.</p>	['a:tɔ:s]	<p>ab förkort • <i>Skrivs ofta med versaler AB</i>. • <i>aktiebolag</i>.</p>
<p>ab förkort • <i>Skrivs ofta med versaler AB</i>. • <i>aktiebolag</i>.</p>	['a:~be:]	<p>ab förkort • <i>Skrivs ofta med versaler AB</i>. • <i>aktiebolag</i>.</p>
<p>ab förkort • <i>Skrivs ofta med versaler AB</i>. • <i>aktiebolag</i>.</p>	['ab]	<p>ab förkort • <i>Skrivs ofta med versaler AB</i>. • <i>aktiebolag</i>.</p>
<p>ABAB förkort • <i>allmänna bevaknings AB</i>.</p>	['a:bab]	<p>ab förkort • <i>Skrivs ofta med versaler AB</i>. • <i>aktiebolag</i>.</p>
<p>abakus subst. s3 alt: ['a:bakes]. • <i>Skrivs även abocus</i>. <i>abakusen</i> ['abakəsən] pl <i>abakuser</i> ['abakəsər] <i>abakuserna</i> ['abakəsərən] • <i>1. en kulram använd som räknehjälpmedel. 2. arkit: platta</i></p>	['a:bakes]	<p>ab förkort • <i>Skrivs ofta med versaler AB</i>. • <i>aktiebolag</i>.</p>