

A TWO-LEVEL APPROACH TO THE HANDLING OF FOREIGN ITEMS IN SWEDISH SPEECH TECHNOLOGY APPLICATIONS

Anders Lindström* and Anna Kasaty**

*Telia Research AB, Farsta and Dept. of Information and Computer Science, Linköping University

**Dept. of Linguistics, Stockholm University

ABSTRACT

The functional criteria of a lexical component for Swedish speech recognition and speech synthesis systems are defined and extended with demands regarding proper treatment of foreign features on several linguistic levels. A corpus of Stockholm Swedish, judged to be representative of a variety spoken by younger people, is studied, and necessary extensions on the phonological and morphological level are suggested and implemented in a two-level framework – PC-Kimmo. This “extended” two-level description of Swedish is tested on some examples of English loan words and phrases drawn from the corpus.

1. BACKGROUND

Interactive systems, involving automatic speech recognition (ASR) and text-to-speech (TTS) conversion have attained quality levels that allow for inclusion in public services. However, one aspect that becomes increasingly important in a globalized economy and environment is the occurrence of foreign items in spoken (and, to some extent also in written) language. Spoken interfaces to services that involve proper names and other terms from a foreign language must be capable of handling varying degrees of foreign features on several linguistic levels. The number of such applications is growing with increasing mobility and also with the development of international tourism, as noted for instance by Billi [3].

In several smaller European language communities, e.g. in the Scandinavian countries and the Netherlands, one can consider large parts of the population, in particular younger people, to be almost bi-lingual, with a high degree of proficiency in some variety of English in addition to their own mother tongue. Therefore, the lexical component of ASR and TTS systems must be designed to cope with a number of foreign features on several linguistic levels, e.g. the phonetic, phonological (including prosodic), morphological and phrasal level. Those features are predominantly of English origin and are most certainly due to the massive cultural and commercial influence from Britain and North America during the past century. In this paper, we will study this problem within the context of Swedish speech technology applications.

Production studies by Eklund and Lindström [5, 6] and Lindström and Eklund [11] have shown that Swedish speakers include non-Swedish speech sounds, termed ‘xenophones’, to a quite high degree, in normal, everyday spoken language. Similar aspects of the basic problem of interference between native

and foreign language have also been examined for other language combinations by Trancoso [16] and Fitt [7].

Eklund and Lindström’s work was in part driven by the development of a Swedish recognizer for the *Spoken Language Translator* [13], where the domain chosen was predominantly related to flight bookings in the U.S. Similarly, in compiling and recording the speech material for a demi-syllable-based concatenative synthesizer that was being developed at Telia Research around the same time, special attention was paid to include such xenophones and also to make the demi-syllables follow what could be regarded as somewhat extended Swedish phonotactics, as compared with for instance Sigurd’s classical description of Swedish phonotactics [14].

In descriptions of the morphotactics and phonotactics of a given language, you typically distinguish patterns that clearly belong to the language from those that do not, but there is obviously a fuzzy zone between those extremes, in which the phenomena that we are dealing with here tend to fall. In an experiment involving Swedish subjects ability to form the plural forms of (Swedish) pseudo-words (i.e. words that are not existing in the language but adhere to its phonotactics), Linell [12] showed that Swedish subjects were to a quite large extent capable of inflecting pseudo-nouns according to Swedish existing paradigms, even if performance varied across words and was considerably lower on some of the neuters. Linell also showed how these performance differences could be explained by assuming that the subjects rely on comparison with existing Swedish rhyming words in carrying out these morphological operations, and therefore concluded that this was the case.

Recently, Kasaty and Koponen investigated Swedish nominal morphology and phonology [8], and implemented their results in a two-level system using PC-Kimmo [1, 2]. The functional demands they put on their system were: (1) to account for morphological structures of regular nouns by using productive and non-productive derivational and inflectional affixes, (2) to provide all possible analyses of input nouns permitted by morphotactical and phonological rules, and, (3) for each input noun in orthographic form, the system should return its transcription on a broad phonemic level, the lexical categories of all morphemes involved and all syntactically relevant morphological features (such as paradigm, primary stress location, word accent, number, gender, case and definiteness). They found that it was to a large extent possible to meet those demands using a two-level model, where a transcribed lexicon with a morphophonological description of Swedish nouns and the feature unification facilities of the PATR-II mechanism included on top of PC-Kimmo were used to handle morphotactics and to model different aspects of Swedish word prosody.

2. METHOD

In this paper, a corpus containing various foreign items is studied, and the results are used to define functional criteria on the lexical component of Swedish ASR and TTS systems. It is then described how the criteria can be met by extending the implementation of Kasaty and Koponen to include aspects of non-Swedish nominal morphotactics and phonotactics, and by extending the pronunciation lexicon to cover non-Swedish morphs.

Corpus. For the experimental purpose of this study a small corpus consisting of 415 strips from Martin Kellerman's comic *Rocky* [9] was analyzed. The language used in *Rocky* can be considered to be quite representative of current Stockholm Swedish, as spoken by people in their teens or twenties, and Kellerman apparently tries to resemble that sociolect using Swedish orthography. Indeed, in newspaper interviews, Kellerman has claimed only to be writing down what his friends say [10].

We singled out all the occurrences of foreign items and grouped them in different categories depending on language of origin, part-of-speech and orthographical integration in Swedish. Even if a lot of loan words such as *design*, *jeans*, *show*, *tejp* (tape) are included in a recently published Swedish lexicon of "new words" [15], their orthography, inflection and/or pronunciation is still foreign and needs to be handled in a special way. (Glosses are given in brackets in the examples.)

The main part of the foreign items we found was English, of which 60 were full English sentences or phrases, 3 were English sentences with German items, 3 were French and 2 German. Additionally the corpus included a large amount of proper names, which was not an issue in this study. We concentrated on the English loan words incorporated in Swedish sentences, some of which were written with Swedish orthography, others were inflected using a Swedish pattern or occurring as parts of Swedish compounds. Of a total of 174 such English loan-word tokens, 125 were nouns, 28 verbs, 12 adjectives and 9 adverbs. The nouns, standing for the majority of loans, are the ones tested in our implementation, together with some denominal verbs, e.g. *casha*, *warpa*, *beepa*, *buzza*, *scora* (English root + Swedish infinitive verb suffix "-a") and denominal adjectives e.g. *flashig* (English root + Swedish adjectival suffix "-ig"). An interesting observation about nominal loans is that the product of the loan process is sometimes just a root morpheme, e.g. *groupie*, which can be inflected, as in *groupies*, or take part in Swedish compounds e.g. *dansbandsgroupie* (dance-band groupie). In other cases, the product is a whole word model, already inflected, e.g. *kids*, *groupies*, *furbies* which is then inflected again using a Swedish pattern, e.g. *kidsen* (English plural indefinite suffix + Swedish plural definite suffix), *groupisar*, *furbisar* (English plural indefinite suffix + Swedish plural indefinite suffix, e-deletion).

Some variability was noticed in both inflection and orthography of individual loan-words. It seems possible to assign either an English or Swedish plural suffix to an English loan-word such as *geek*, as illustrated by the occurrence of both *geeks* (English indefinite plural) and *geekarna* (Swedish definite plural).

Some loans feature two spellings – English and accommodated Swedish, e.g. *date* vs. *dejt* and *styling* vs. *stajlad* (Swedish spelling + adjectival ending). This could be explained in terms of levels of integration in Swedish, according to Chrystal [4], who classified English loan-words in Swedish press in three categories: (1) 'Established' (occurring in Swedish dictionaries), (2) 'interim' (where orthography points on some level of integration in Swedish) and (3) 'non-integrated' loans.

The examples presented above are chosen to exhibit different kinds of foreign traits on the morphological level and were tested in our implementation.

Criteria. From previous experience and from the study of the *Rocky* corpus, we conclude that the foreign features that a Swedish speech technology lexical component must be capable of handling include

- expansion of the speech sound repertoire to include 'xenophones' (non-Swedish sounds) such as [dʒ, θ, ð, əʊ],
- combinations of English roots with Swedish inflections, e.g. *coolt* (neutral adjective), *geekar* (plural noun), *raidar* (raids), *bustar in* (busts in), *pushar upp* (pushes up),
- using English roots in Swedish derivational and compounding morphological processes, e.g. *flashigt* (flashy), *handsfremobil* (hands-free cellular telephones), *depparsoundtrack* (gloomy [rocker] sound-track), and also
- interactions on the prosodic and phrasal levels, e.g. *posters* with Swedish accent I as opposed to *postrar* with Swedish accent II, *drayvbajs* with accent I instead of compound accent since it derives from drive-by [shooting]s, *droppa en jätterolig punchline* (drop a very funny punch line).

These criteria should be taken in addition to the ones used by Kasaty and Koponen [8], previously listed.

All these phenomena can be studied from many different perspectives, including a diachronic (historical) one. The main goal of this study, however, is to be able to model the current behaviour of Swedish language users.

3. IMPLEMENTATION

Having defined the functional criteria we will now extend the implementation of Kasaty and Koponen [8] and attempt to construct a Swedish analyzer capable of recognizing and generating the English loan-words from our corpus using PC-Kimmo [1]. PC-Kimmo (version 2) can be regarded as a two-level framework, equipped with a PATR-II formalism on top [2]. The implementation in PC-Kimmo involves describing Swedish (nominal) morphology and phonology using three interdependent components: a transcribed morph-based lexicon, two-level rules to connect the orthographic level with the phonological, and a unification-based word grammar with feature structures to account for the morphotactics and, in our

case, also to deal with stress- and accent-related word prosody phenomena. For an excellent introduction to PC-Kimmo and a detailed description of two-level rules in general, see Antworth [1].

Unification-based word grammar. Drawing the line between lexicon and grammar was relevant for the description of our morphotactical structure of Swedish words. The lexicon serves to break a word into its morphemes using minimal morphotactic constraints, while the word grammar applies a more powerful mechanism that filters out any incorrect analyses allowed by the lexicon. The context-free rules with their corresponding feature constraints determine the resulting part-of-speech categories and the resulting word accents after derivational or inflectional processes have taken place, by passing the values of head features up the parse tree.

The morphotactical structure of Swedish words allowed in our model is illustrated by the excerpt from the word grammar shown below, where terminal categories are indicated by capital letters.

```
Word -> Stem (INFL)
Word -> Compound
Compound -> Stem JOINT Word
Stem -> (PREFIX)* ROOT (SUFFIX)*
```

Unification over morphological features, holding information about paradigm, gender, number etc., assures that only admissible sequences are accepted or generated. Properties of entities obtained through derivational processes are also computed from the features of the heads and affixes that form part of the process. The same principle is also used to assign the word accent to inflected words.

In order to cover foreign words and other neologisms, we intend to extend the definition of what could constitute a ROOT to encompass any sequence of permissible syllables according to the definitions used in the Swedish concatenative synthesizer as described previously. These include all the xenophones as described earlier [5, 6, 11], and both initial and final consonant clusters, drawn from Swedish as well as English phonotactics. In a full system, this is likely to be the cause of massive over-generation, but if you consider this as a fall-back system, when all else fails, one could successively loosen the constraints until reaching this level, which will make over-generation less of a problem.

For the purpose of these tests, we have added such syllables to cover the examples from the *Rocky* corpus. In doing so, the syllables were assigned paradigm codes to allow both for English and Swedish inflection. Assignment of Swedish paradigms is done according to rhyming principles, such as those suggested by Linell [12]. For example, the syllable /gi:k/ is assigned to paradigm s100 (allowing for English plural-s) and to s2 (in analogy with rhyming Swedish nouns, like *spik* (nail)).

Lexicon. Each lexical entry is structured into several fields, which specify the underlying lexical representation, what continuation class can follow, how the entry should be glossed and relevant information about the entry, coded as features. Word roots and various classes of affixes can be considered to be

grouped into a number of sublexicons. This makes it straightforward to treat alternative inflectional endings as they are normally discussed, namely in terms of families of inflections such as conjugations and declensions. In our implementation all lexemes, which have the same allomorphs of affixes, belong to the same morphosyntactic category or paradigm.

In order to handle loans of English plural nouns, the productive endings (/s/, /z/, /iz/) were added to the list of plural endings.

Two-level rules. The lexicon lists all the morphemes (roots and affixes) in their underlying form, in our case a phonemic representation. The rules component consists of two-level rules, which account for the relations between the underlying forms and the surface characters of Swedish orthography. A surface form can be given more than one analysis by the rules and the lexicon when submitted to the PC-Kimmo recognizer.

We augmented the number of possible grapheme-to-phoneme correspondences to be able to handle also English sounds. This is possible in languages that employ a phonographic writing system like Swedish and English, and where there is normally some relationship between the graphemes and the sounds of the language. One problem is that languages vary a lot in their graphemic/phonemic regularity. In the case of Swedish and English there is a marked degree of irregularity, and the lack of correspondence between graphemic and phonemic levels is reflected in the number of more or less arbitrary “spelling rules” that we have to use to be able to handle exceptions to the regular patterns.

The spelling rules were extended so as to permit some common English spelling conventions occurring in the expressions from the corpus, e.g. <ee>, <ea> corresponding to /i:/ (illustrated by the two-level rule below) as well as to include the xenophones.

```
i:e => C _ [.:e|.:] C
```

(Full stop is used to indicate length, C is the set of consonants.)

4. EVALUATION

The system was tested on a number of foreign items from the *Rocky* corpus. Those included inflected nominal roots, e.g. *blowjob*, *freestyle* (walkman), *flyers*, *geeks*, *geekarnas*, *kidsen*, *shotgun*, *sneakers*, *styling*, *trainee*, some denominal verbs as well as compounds, where the first or the last part of a Swedish compound consisted of an English morpheme, e.g. *bikerfest* (biker party), *crackförsäljarna* (crack dealers), *dansbandsgroupie* (dance-band groupie), *designerpåsar*, (designer bags), *hallonshots* (raspberry shots), *handsfreelurar* (hands-free headphones), *house-prylar* (things or attributes associated with the music style ‘house’), *jet-lag*, *smygdate* (sneak-away date), *sodastreamern* (the soda streamer).

We tested the system diagnostically with respect to morphemic segmentation, morphological features (e.g. part-of-speech, number, definiteness and case), word prosody (primary stress location and word accent assignment) and phonemic transcription.

Below is an example of the output of a test run.

```
PC-KIMMO>r geeks
g'i.k+s      geek(eng.)+s
1:(Word_1 (Stem_2 (ROOT_3 g'i.k 'geek(eng.)))(NUM +s '+s'))
Word: [ cat:Word head:[ paradigm:s100 root_pos:N root_gender:UTR
number:PL species:INDEF case:NOM root_accent:ACCI ] ]
1 parse found

g'i.k+s      geek(eng.)+s
1:(Word_1 (Stem_2 (ROOT_3 g'i.k 'geek(eng.)))(GEN_4+ +s
'+s'))
Word: [ cat:Word head:[ paradigm:s100 root_pos:N root_gender:UTR
number:SG species:INDEF case:GEN root_accent:ACCI ] ]
2:(Word_5 (Stem_6 (ROOT_7 g'i.k 'geek(eng.)))(GEN_4+ +s
'+s'))
Word: [ cat:Word head:[ paradigm:s2 root_pos:N root_gender:UTR
number:SG species:INDEF case:GEN root_accent:ACCI ] ]
2 parses found
```

As can be seen in this example, the English noun stem can combine with an English plural /s/ as well as with Swedish genitive /s/, whereas the Swedish (s2-assigned) syllable only combines with the genitive affix. The result of the evaluation is that it was possible to handle all items in this way, but that some further changes needed to be made to the “spelling rules” in the two-level rule component in order to do so. Proper derivation of word stress patterns also proved difficult.

5. CONCLUSIONS

We have shown by analyzing examples from a corpus how foreign features, mostly English, on the morphological and phonological level can be handled in a PC-Kimmo implementation based on a Swedish pronunciation lexicon. This was done by extending the nominal paradigm system to encompass parts of the English plural system and by extending the lexical component to allow root formation through syllable concatenation. The syllables introduced include “foreign” phonological segments (‘xenophones’) and follow phonotactic rules that are more admissive of patterns found in loan words. Problems encountered include adapting the two-level “spelling rules” to encompass spelling conventions of another language, something which will undoubtedly give rise to over-generation in the general case. Another question that needs to be solved in a more general perspective is that of word prosody prediction including primary stress placement as well as word accent assignment. This is an area in need of further studies of how “foreign” prosodic patterns interact with native paradigmatic patterns and where a host of other factors, e.g. syllable weight, probably play a role.

We also suggest carrying through the exercise of building a full lexicon that consists of a combination of existing morphemes and admissible syllables following somewhat extended Swedish phonotactics. This type of system could serve as a fall-back, and would not need to be continuously revised when new lexical items are encountered, an important feature for speech technology systems. The examples here were nouns and denominalizations, but obviously a full implementation must model all word-classes. Quantitative evaluation according to the criteria defined here could then be carried out. Finally, although *Rocky* offered some of the more hilarious moments of corpus linguistics we’ve experienced, undoubtedly a recorded corpus of current spoken Swedish would be even better suited for further studies.

6. ACKNOWLEDGEMENTS

We would like to thank Robert Eklund for many inspiring discussions, for suggesting the use of the *Rocky* corpus, and for proof reading the paper. We would also like to thank Lars Ahrenberg for his feed-back on earlier versions of the paper.

7. REFERENCES

1. Antworth, E. L. *PC-KIMMO: a two-level processor for morphological analysis*. Summer Institute of Linguistics, Dallas, 1990.
2. Antworth, E. L. *PC-KIMMO version 2*. Summer Institute of Linguistics, <http://www.sil.org/pckimmo/>, Dallas, 1995.
3. Billi, R. <http://www.linglink.lu/lejournal/article.asp?articleIndex=628>, Interview published in *Le Journal — The Journal of Record for Human Language Technology*, n.d.
4. Chrystal, J. *Engelskan i svensk dagspress*. Esselte Studium, Stockholm, 1988.
5. Eklund, R., and Lindström, A. Pronunciation in an internationalized society: A multi-dimensional problem considered. *TMH-QPSR 2/1996. Proc. FONETIK 96, Swedish Phonetics Conference, Nässlingen*, pp. 123-126, 1996.
6. Eklund, R., and Lindström, A. Xenophones: An Investigation of Phone Set Expansion and Implications for Speech Recognition and Speech Synthesis. *Speech Communication, Special Issue* (in press).
7. Fitt, S. E. *Processing Unfamiliar Words: A Study in the Perception and Production of Native and Foreign Place Names*. PhD Thesis, University of Edinburgh, 1998.
8. Kasaty, A., and Koponen, E. *Swedish Nominal Morphology Implemented within the Two-level Model in PC-Kimmo*. Unpublished term paper, Dept. of Linguistics, Stockholm University, 1999.
9. Kellerman, M. Rocky. Comic strips appearing in *Metro*, from March 28, 1998 to Nov. 18, 1999.
10. Kellerman, M. Rocky regerar...men det är Martin som bestämmer. Interview published in *Tidningen Södermalm*, p. 17, July 25, 1999.
11. Lindström, A., and Eklund, R. Xenophones Revisited: Linguistic and other underlying factors affecting the pronunciation of foreign items in Swedish. *Proc. ICPhS 99, San Francisco*, pp. 2227-2230, 1999.
12. Linell, P. Nilpor, mejder och snulkar – ett experiment med pluralbildning i svenskan. In Otterbjörk, R. And Sjöström, S. (Eds.): *Provinsiellt och universellt. En vänskrift till Karl-Hampus Dahlstedt. Publications from the Dept. of General Linguistics, No. 12, University of Umeå*, 1977.
13. Rayner, M., Carter, D., Bouillon, P., Digalakis, V., Wirén, M. (Eds.), *The Spoken Language Translator*. Cambridge University Press, in press.
14. Sigurd, B. *Phonotactic structures in Swedish*. Berlingska Boktryckeriet, Lund, 1965.
15. Svenska Språknämnden. *Nyord i svenskan från 40-tal till 80-tal*. ISBN 91-24-34322-6, Esselte, 1998.
16. Trancoso, I., Viana, C., Mascarenhas, I., and Teixeira, C. On deriving rules for nativised pronunciation in navigation queries. *Proc. Eurospeech 99, Budapest, ISSN 1018-4074*, pp. 195-198, 1999.