

Pronunciation in an internationalized society: a multi-dimensional problem considered

Robert Eklund and Anders Lindström
Telia Research AB, Haninge, Sweden

Abstract

This paper deals with the treatment of foreign words and proper names in Swedish. Preliminary results from a production study are presented, and guidelines are suggested for broad, phonematic transcription, covering alternative pronunciations. Such a transcription scheme is a prerequisite for applications such as speech synthesis and multi-dialectal speaker-independent speech recognition.

Introduction

The general problem considered in this paper is the accommodation of so-called foreign words, and in particular foreign names, in a target language, in this case Swedish.

The concept of foreign (or alien) words vs. accommodated words is discussed by Dahlstedt [2], who, in turn, cites Swedish (royal) correspondence on the subject, dating as far back as to 1584. Dahlstedt defines the following neighbouring word categories:

citat *främmande* *hemtama*
citations foreign/alien accommodated

These categories are to be understood from a synchronic point of view, and they are not (necessarily) related to diachronic, historical or etymological categories. Dahlstedt cites Lundell [7] as having given the best definitional criteria to date for foreign words w.r.t. Swedish. These include:

1. Pronunciation
 - (a) The phone repertoire expands to include “foreign” phones, e.g. [x] (German *ach-laut*), [θ] (English lisp sound)
 - (b) Phone sequences that do not conform to Swedish phonotactic structure are allowed
2. Accentuation, e.g. primary stress location and choice of word tone
3. Number of syllables
4. Inflexional affixes
5. Derivational affixes
6. Spelling

Dahlstedt finds that only criteria 1 (a) and 4 can serve to reliably identify “foreign words”. In any case, the boundaries between the three word categories are hard, if at all possible, to define.

Elert looks at this problem from a production-related point of view [3], according to which the language user can choose between three pronunciation strategies when confronted with a “foreign” word:

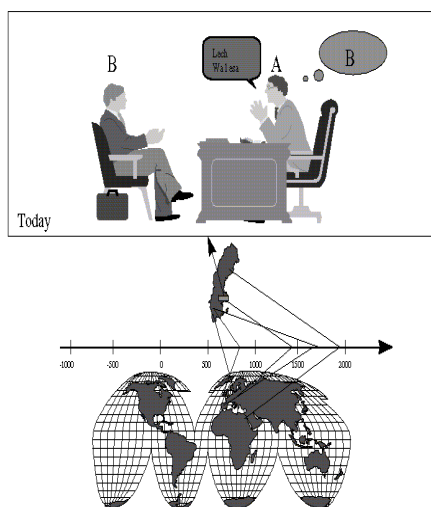
autentiskt fonemanpassat stavningsanpassat
authentic re-phonematized orthographical

The problem

The categories defined by Dahlstedt can be seen as symptoms of the type of user behaviour that Elert describes. However, Elert’s division of production strategies into three pronunciation types seems a bit coarse, especially when considering the multitude of factors that affect the speaker.

Fig. 1 illustrates a typical situation, where language user A is about to pronounce a foreign proper name. His choice of pronunciation strategy is affected by a number of factors, among which are:

1. A’s own competence regarding the original pronunciation
2. A’s performance capability when rendering the original pronunciation or an approximation thereof
3. A’s expectations on B’s (perceptual and cognitive) competence
4. The relative social status of A and B
5. The socio-cultural (not necessarily geographical) distance to the country of origin of the name in question
6. The time the name first appeared
7. The population frequency of bearers of this name
8. The frequency and recency of occurrence of this particular name in society, news etc.

Figure 1: *The language user in a difficult situation*

Some of these factors are tangible and lend themselves well to studies and modelling, others are more evasive, but may nonetheless be important. For instance, the importance and complex socio-linguistic nature of factors 4 and 5 is discussed in some detail by Lipski [6]. The importance of factor 3 (A's expectations on B's competence) must also be pointed out: There is no *universally* applicable level, and whatever level A opts for, he is likely to exclude *some* listeners, or at least make the comprehension task more difficult for them.

Phone repertoire

Due to the complexity of the situation that the language user is subjected to, we believe that it would be reasonable to think of the productions as forming a continuum, rather than only three distinct categories. The choice of phone repertoire alone must give rise to numerous combinations of phones into individual phone sets. Fig. 2 shows some of the possible mappings of “foreign” phones onto the Swedish system. In principle, the original phones should of course be qualified by their respective language. For instance, a Greek [θ] and an English [θ] may well be mapped onto different phones in the system of a Swedish individual. Note that, for most speakers, this phone set is much larger than what is assumed in conventional descriptions of Swedish [4].

Investigations

Two aspects of the general problem outlined above have been investigated. Both concern the pronun-

Original	Swedish	Example
[x]	[x, ʃ, s̥]	A <u>achen</u>
[θ]	[θ, t]	J <u>eth</u> ro, <u>thr</u> iller
[ð]	[ð, d., d]	<u>th</u> e
[ʃ]	[ʃ, s̥]	<u>Sh</u> aron
[z]	[z, s]	<u>z</u> gn, <u>Oz</u>
[ʒ]	[(ʒ), s̥]	le <u>ig</u> ure
[tʃ]	[tʃ, tʃ, tʃ, tʃ, ɛ]	<u>Th</u> atcher, <u>Ch</u> arles
[tʃ]	[tʃ, j]	<u>G</u> ianna, <u>J</u> ames
[ɹ]	[ɹ, r]	<u>P</u> rice, <u>R</u> oger
[t]	[t, l]	<u>El</u> vis
[w]	[w, v]	ply <u>w</u> ood, <u>W</u> hite
[aː]	[aː, aː, a]	<u>A</u> achen, <u>G</u> onzales
[æ]	[æ, a]	<u>S</u> haron, <u>J</u> ackson
[uə]	[uə, uː]	<u>N</u> ewman
[aʊ]	[aʊ]	<u>a</u> ula, <u>m</u> ouse
[eʊ]	[eʊ]	<u>E</u> uropa
[əʊ]	[əʊ]	<u>S</u> tone, <u>O</u> wen
[ei]	[ei, ej]	<u>B</u> aywatch
[ai]	[ai, aj]	<u>M</u> ichael
[oi]	[oi, oj]	<u>L</u> loyd

Figure 2: *Some possible mappings of foreign phones to Swedish approximations.*

ciation of “foreign” proper names. In such a situation speakers are likely to do their best to approximate the original pronunciation, and to avoid a reading too much inspired by the orthography.

First, preliminary observations regarding production competence are presented. Then the implications of the discussion above for lexicographic work within applications such as text-to-speech conversion and automatic speech recognition (ASR) are examined.

Production data/study

The main incentive to look at speakers' production of foreign items, is that, by studying the pronunciations Swedish speakers use when uttering foreign names, one might get an idea of what renderings that assumedly are considered acceptable within the Swedish language environment.

Therefore, twelve sentences were created that contained foreign words and proper names of mainly English origin. The words and names were chosen so that they were likely to be known by most Swedish speakers. The sentences were recorded by 70 subjects—35 from Skåne and 35 from Stockholm—who read the sentences one by one from a computer screen. The recordings formed part of a large recording session where the bulk of the material was recorded for entirely different reasons, so there is good reason to assume that the subjects were not aware of the purpose of this particular session. In this way, 857 sentence tokens, covering 17 non-Swedish phones and more than 2 500 phones, were collected.

Some provisos need be mentioned here concerning the problems associated with this kind of production study. First, it is very hard to know what kind of knowledge one is actually examining, that is to say, whether one is studying language, word or world knowledge, since these aspects to a large extent are interrelated. Second, approximations of foreign phones do not take discrete values. A Swedish attempt at a dental fricative in e.g. *Thatcher*, might not sound like native English, but still be very different from a Swedish [t]. Third, the material in this study is way too small to provide a sufficient base for more far-reaching or definite conclusions to be drawn. Thus, the observations below should be read as fairly 'impressionistic' descriptions of a more general kind.

The subjects generally did well on diphthongs. Almost everybody produced a good [ei] (*James*, *Baywatch*). A majority did a good [ai] (*Michael*, *Diana*) and [oi] (*Lloyd*, *cowboy*). This is probably due to the fact that the aforementioned English diphthongs may be approximated well by the appropriate vowel + [j], a pattern which conforms with Swedish phonotactics. Consequently, the worst results were encountered with [əʊ] in *Stone*, where a good half of the speakers used the Swedish [u:] instead. In all these cases, substitute phones were obtained by a Swedish interpretation of the orthography.

Two non-diphthongal vowels were included, the English [æ] (*Harold*, *Jackson*, *Sharon*) and the German [a:] (*Aachen*). Most subjects produced a high-perfect [æ] phone, whereas [a:] was replaced with either Swedish [a] or [ɑ:]. The excellent results for the [æ] phone are probably due to the fact that this sound is phonetically very close to a proper Swedish sound, namely [æ], which is an allophone of [ɛ] normally preceding a retroflex.

The English dental fricatives [θ] (*Thriller*, *Thatcher*) and [ð] (*the*) were both produced by about 50 % of the speakers. The commonest substitute for [θ] was [t], and [ð] was replaced by either [d.] or [t].

Decent affricates (*Charles*, *James*, *Thatcher*, *Roger*) were produced by around 50 % of the speakers. What was used in lieu of the English phones was often determined by the orthography. Thus, the initial affricate in *James* would be substituted by [j], while the medial affricate in *Roger* would be replaced by [g].

The [ɜ] sound (*Roger*, *Sharon*) was produced by about 50 % of the speakers. If the English sound was not aimed for, Swedish [r] was used.

Very few subjects produced the voiced consonants [z] (*James*) and [ʒ] (*television*), the first of which being approximated by but a handful of the speakers, the latter by none.

English [w] (*we*) was produced by around 20 % of the speakers, most of whom used Swedish [v]

instead.

The 'thick' English [t] (*Lloyd*) was not used by any of the subjects, Swedish dental [l] being used instead.

The German ach-laut [x] (*Aachen*) was either rendered as [fj] or [ç].

An overwhelming majority of the subjects were fairly inconsistent as to their use of English phones. Thus, they might use the voiced affricate in *Roger* but not in *James*, or vice versa. Also, concerning *Diana & Charles*, one would perhaps have expected that if *Diana* was pronounced in an English way, *Charles* would be, too. This was not the case, however. *Diana* was frequently pronounced with the English diphthong and vowel, but *Charles* without the affricate, or vice versa.

Only two speakers exhibited a consistently Swedish pronunciation of the names, such as *Roger* as [ro:ger], *James* as [ja:mes], and so on. All the other speakers did something extra-Swedish to approximate the foreign sounds.

A final observation is that no apparent differences were noted between the speakers from Skåne and the speakers from Stockholm.

Lexicography

Given the discussion so far, how does one go about producing a phonetically transcribed Swedish lexicon of names, useful for applications such as speech synthesis and speaker independent (multi-dialectal) speech recognition?

This question has received much interest recently, because of the applications that can be built using this type of information [1, 5]. In ASR, the lexicon needs to cover *all* pronunciation variants. Dialectal and regional variants must also be taken into account. For synthesis, the problem is to choose a pronunciation strategy, suitable from a pragmatic point of view. In a particular dialogue situation, the speaker, be it man or machine, needs to make a decision regarding the pronunciation of "difficult" words, e.g. proper names. The human language user has access to various linguistic and phonetic devices, such as discourse markers and voice quality changes, that can be used to signal "distance" to a particular word or uncertainty regarding its pronunciation. A current "state-of-the-art" text-to-speech system will typically not have access to such devices, but can at least resort to spelling out entries, when in doubt regarding the "correctness" of the pronunciation.

In order to be able to meet the needs of different applications, as stated above, a set of guidelines for lexicographic work were formulated. They are summarized in Fig 3.

This gives the possibility to model different speaking styles, different degrees of reduction and different dialects. In speech synthesis, it will also

1. Use a broad, phonematic level of transcription
2. Indicate *all* possible pronunciations (in a Swedish social setting)
3. Index these pronunciations by origin, stylistic value etc.

Figure 3: *Transcription guidelines.*

implicitly allow for a ranking of different error types, something which can be used to determine in which cases fall-back strategies should be called upon. For instance, the uncertainty regarding the choice between [u] and [o], as in the final vowel of *Barbro*, is probably just a reflection of the free variation between [u] and [o] that occurs in unstressed position in ordinary words. More serious problems can occur, when a particular allomorph is coupled to either a stylistic value, or a particular language/region of origin, or when there is simply a large phonetic difference between alternative pronunciations.

A transcription scheme, such as the one outlined above, has been applied to a large lexicon, drawn from an application involving proper names. The ordinary Swedish phone set was enlarged by addition of [ʈ], [ɖʒ], [w], [ɘ], [θ], [ð], [aɪ], [æ], [ʊə], [aʊ], [eʊ] and [əʊ]. The diphthongs [oi, ai, ei] were approximated using the “vowel + [j]” strategy.

Statistics of phone usage were collected based on a part of that lexicon which contained 51 003 entries. In that material, there were 665 entries with at least one occurrence of [ʈ], 293 [aʊ], 67 [w], 42 [ɘ], 33 [æ], 29 [əʊ], 25 [ɖʒ], 11 [θ], 7 [ʊə] and 5 [eʊ]. The transcribers did not use either [aɪ] or [ð]. Examples of transcriptions produced in this manner are shown in Fig 4.

One of these transcriptions, that of *Lindstein*, displays the type of pronunciation conditioned by “country of origin” that was briefly mentioned above: The morph *stein* seems to correspond generally to the four allomorphs /stejn, stajn, ʂtejn, ʂtajn/, but it seems that the three allomorphs that have retained more of a German pronunciation (/stajn, ʂtejn, ʂtajn/) do not combine very well with the Swedish compound accent, the latter indicated by ‘...~...’ in Fig 4. Only the orthographically inspired /stejn/ can form part of a true Swedish compound.

Conclusions. Future Work

We have tried to show how important it is to have the full linguistic picture before your eyes, even when performing seemingly trivial tasks, such as producing phonetically transcribed lexica for use in speech technology applications.

An enlarged Swedish phone inventory to be used in the treatment of “foreign” proper names

Barbro b'arbr(o|)
Lindstein 1'ind(s|\$)t(e|a)jn, 1'ind~st"ejn
Steinwall (s|\$)t'(e|a)jnval

Figure 4: *Verbatim example transcriptions, using a regular expression notation, and a “computer-friendly” non-IPA representation of the extended phone set discussed in this paper.*

was suggested, based on some preliminary observations regarding the production of such items. Some results from a transcription task involving such an enlarged phone inventory were also presented and the advantages of this approach were indicated.

Future work could include a more thorough categorization of the different types of alternative transcriptions, possibly coupled with acceptance tests, using synthesized stimuli. The coupling between linguistic “origin” and pronunciation could also be studied, using the data obtained through the transcription experiment.

Acknowledgement

We thank Carin Lindberg, Elisabeth Zetterholm and Johanna Etzler for carrying out the mind-bogglingly tedious task of hand-transcribing a large lexicon, but above all for their valuable contributions to the continuous discussion on the subject of this paper. Special thanks to Camilla Lindberg, who, by keeping a low profile, allowed her mother to work extensively on this project during a critical period.

References

- [1] Rolf Carlson, Björn Granström, and Anders Lindström. Predicting name pronunciation for a reverse directory service. In *Proc. of the European Conf. on Speech Technology*, volume 1, pages 113–116, Paris, 1989.
- [2] Karl-Hampus Dahlstedt, Gösta Bergman, and Carl Ivar Ståhle. *Främmande ord i nusvenskan*. Verdandis skriftserie. Läromedelsförlagen, 1969. (First ed. appeared in 1962).
- [3] Claes-Christian Elert. *Studier i dagens svenska*, chapter Uttalsbeteckningar i svenska ordlistor, uppslags- och läroböcker. Kungl. Boktryckeriet P. A. Norstedt och Söner, 1971.
- [4] Claes-Christian Elert. *Allmän och svensk fonetik*. Norstedts, 1989.
- [5] Joakim Gustafson. ONOMASTICA—creating a multilingual dictionary of European names. In Gösta Bruce, David House, and Paul Touati, editors, *Papers from the Eighth Swedish Phonetics Conference*, volume 43 of *Working Papers*, pages 66–69, Lund, Backagården, May 1994. Lund University.
- [6] John M. Lipski. Prejudice and pronunciation. *American Speech*, 51:109–118, 1976.
- [7] Johan August Lundell. *Svensk ordlista med reformstävning ock uttalsbeteckning*. Unknown publisher, 1893.