Reference: *Speech Communication*, Volume 35, Nos. 1–2, August 2001, pp. 81–102

Xenophones: An Investigation of Phone Set Expansion in Swedish and Implications for Speech Recognition and Speech Synthesis

Robert Eklund & Anders Lindström

Telia Research AB
Farsta, Sweden

and

Natural Language Processing Laboratory,
Department of Computer and Information Science,
Linköping University, Sweden

Robert.H.Eklund@telia.se
Anders.P.Lindstrom@telia.se

Nota Bene! Draft version. Do not quote!

Abstract

In recent years, both automatic speech recognition (ASR) and text-to-speech (TTS) conversion systems have attained quality levels that allow inclusion in everyday applications. One remaining problem to be solved in both these types of applications is that alleged phone inventories of specific languages are commonly expanded with phones from other languages, a problem that becomes the more acute in an increasingly internationalized world where multilingual automatic speech-based services are a desideratum. This paper investigates the nature of phone set expansion in Swedish. The status of these phones is discussed, and since such added phones do not have a phonemic (or allophonic) function, the term 'xenophones' is suggested. The analysis is based on a production study involving 491 subjects, and the observed xenophonic expansion is described in terms of three categories along the "awareness" and the "fidelity" dimensions. The results show that very few subjects resort to full rephonematization and that xenophonic expansion is the rule, although there is an uneven distribution depending on particular phones, spanning from phones produced by most subjects, to phones produced by almost no subjects. Of the possible explanatory factors analyzed—regional background, gender, age and educational level—the latter is by far the most important.

Zusammenfassung

In den letzten Jahren haben Systeme für automatische Spracherkennung als auch für textbasierte Sprachsynthese ein Qualitätsniveau erreicht, das ihren Einsatz in alltägliche Applikationen erlaubt. Ein Problem, das in beiden Applikationsbereichen zu lösen verbleibt, besteht darin, dass die Phoninventare einzelner Sprachen gewöhnlich mit Phonen aus anderen Sprachen ausgebaut weren – ein Problem, das sich in einer zunehmend internationalisierten Welt verschärft, in der mehrsprachige automatische sprachbasierte Dienste wünschenswert wären. In diesem Artikel wird die Natur des Phonemausbaus im Schwedischen untersucht. Der Status dieser Phone wird diskutiert, und weil sie keine phonematische (oder allophonische) Funktion haben, wird der Terminus “Xenophon” vorgeschlagen. Die Analyse gründet sich auf eine Produktionsstudie mit 491 Informanten, und der beobachtete Xenophonausbau wird in drei Kategorien entlang der Dimensionen “Bewusstheit” und “Getreue” beschrieben. Die Ergebnisse zeigen, dass sehr wenige Informanten auf vollständige Rephonematisierung zurückgreifen und dass Xenophonausbau die Regel ist, aber die Verbreitung der einzelnen Phone ist ungleichmäßig und erstreckt sich von solchen, die die meisten bis zu solchen, die fast keine Sprecher produzieren. Von den möglichen Erklärungsfaktoren, die analysiert wurden – regionale Herkunft, Geschlecht, Alter und Bildungsgrad, ist der letztere bei weitem der wichtigste.

Resumé

Au cours de ces dernières années, les systèmes de reconnaissance automatique de la parole ainsi que les systèmes de synthèse de la parole à partir du texte ont atteint une qualité leur permettant d'être intégrés à des applications quotidiennes. Un problème demeure cependant quant à ces applications, à savoir, l'expansion fréquente des inventaires de phones d'une langue spécifique à des phones issus d'autres langues, problème d'autant plus important dans un monde où les services automatisés de la parole multilingues sont un souhait. Cet article examine la nature de l'expansion des phones en suédois. Nous discuterons du statut de ces phones, et, puisque ces phones supplémentaires n'ont pas une fonction phonémique (ou allophonique), nous proposons le terme de 'xénophones'. L'analyse se base sur l'étude de la production de 491 informants. L'expansion xénophonique attestée est décrite en trois catégories selon les dimensions 'conscience' et 'fidélité'. Les résultats montrent que peu d'informants ont recours à une rephonématisation complète, l'expansion xénophonique étant la règle, bien que la distribution soit inégale selon les phonèmes, allant de ceux produits par la plupart des informants à ceux produits par très peu d'informants. Parmi les facteurs explicatifs potentiels analysés, tels les origines régionales, le sexe, l'âge et le niveau d'étude, ce dernier s'avère de loin être le plus important.

1 Introduction

In recent years, both automatic speech recognition (ASR) and text-to-speech (TTS) conversion systems have attained quality levels that allow their inclusion in everyday applications. This does not mean, however, that all problems with regard to ASR/TTS are solved. One such problem addressed in this paper is the fact that in a language such as Swedish, the alleged phone inventory is quite often expanded with phones from other languages when pronouncing foreign words and names. As a consequence, in order to create high-quality ASR and TTS systems for Swedish, these "foreign" phones need to be included in the underlying language description. Although this problem per se is uncontroversial in that several researchers acknowledge the existence of "non-Swedish" sounds in everyday spoken Swedish—to the best of our knowledge—there have been no formal studies to establish exactly what this type of extended phone sets look like, either in Swedish or in any other language. In this paper, we will first attempt to determine the nature of this extended phone set, including phone frequencies, when dealing with words and names of primarily English origin, and then show how this has been applied to enhance the quality of our Swedish TTS system by extending its phone set. We will also discuss implications for ASR, and suggest how some of these have been taken into account in a Swedish/English bilingual recognizer, and how the remaining could be catered for in the future.

1.1 Background

A hitherto somewhat neglected problem that constitutes an important issue in the development of multilingual applications is dealing with the fully normal inclusion of "foreign" items, e.g. speech sounds, in the pronunciation of foreign names and words. These foreign items might be derived from different linguistic areas, e.g., phonology, or morphology. Such speech sounds can be said to expand the phone inventory of the native language in question, a phenomenon observed in at least some languages, such as Swedish (Eklund et al., 2000; Eklund & Lindström, 1998, 1996; Lindström & Eklund, 2000, 1999a,b, in press). An example from Swedish would be the voiceless dental fricative [θ] (the first sound in the name "Thatcher"), which is not part of the Swedish phonemic inventory, but is nevertheless produced by approximately half of the population when pronouncing English words or names containing this sound in otherwise Swedish sentence contexts (Eklund & Lindström 1998, 1996, Lindström & Eklund, 2000, 1999a,b).

With a growing awareness of the need for multilingual automatic services (cf. Billi, n.d.), the handling of language users' less constrained pronunciation becomes something of a *sine qua non*.

1.2 Terminology

The Swede's pronunciation of names or words of foreign origin often exhibits sounds that are not part of what is traditionally considered the Swedish phoneme inventory. One can argue that such "added" sounds do not have a phonemic function in Swedish, and therefore must be attributed a specific status in the Swedish phonological system. Even though they are not Swedish phonemes—or allophones of Swedish phonemes—they are clearly part of the *phone* sets of individual Swedish speakers. Hence, we suggested the term *xenophones* (Eklund & Lindström,

1998), i.e. “foreign phones”, to denote such sounds. What we are dealing with here is a phenomenon that occurs precisely at the intersection between two or more phonological systems; therefore the notion of the phoneme and which system it occurs in is not trivial to define. Perhaps it would be more relevant to talk about a speaker-specific phonological system that arises from this language contact.

1.3 The xenophone problem

Appropriate treatment of the phenomenon of expanded phone sets is likely to influence the performance of any ASR or TTS system. For both these types of applications, expansion of the phone set are required, to a larger or lesser degree. For instance, a Swedish TTS system must be capable of producing the appropriate pronunciation of foreign names or words in running Swedish texts. Likewise, a Swedish recognizer must be able to cope with inter- and intra-speaker differences in foreign item pronunciation. This is probably particularly apparent in applications such as broadcast news, automatic email reading, and spoken language applications within the travel and tourism sector (involving names of foreign places and people), but also national, not to mention international, voice dialling applications.

What is also apparent in the results reported in Eklund & Lindström (1998, 1996) and Lindström & Eklund (2000, 1999a,b), is that the nature of this xenophonic expansion depends, among other things, on the particular sound in question. This leads into the field of phonological acquisition, and second language acquisition (SLA) research. The phonological processes involved when approaching a foreign language have been discussed in detail over many years (e.g. Flege, 1987; Hammarberg, 1990), and SLA research definitely provides valuable insight with regard to what factors might be at play. However, we would like to argue that although the phonological foundation may be the same in xenophonic expansion and SLA, xenophones present a different problem simply because we are facing a different situation. Within SLA, the assumed goal of the subject is to master an entire target language, often in a target language context, whereas in the case of xenophonic expansion, the subject simply includes words of foreign origin in native-language sentences, mostly within fully native-language contexts. Thus, the communicative goal may be considered different, and this in turn should affect the actual rendering of the linguistic items in question. Within SLA research, it is argued that when a language learner with a native language (L1) perceives a sound in a foreign, target language (L2), as sufficiently close to a sound in L1, the two sounds pass for the same sound, and are consequently put in the same ‘equivalence class’. The effect of this equivalence classification has been shown by e.g. Flege (1987). Hammarberg (1990), points out that whether or not an L2-sound is perceived to be identical or similar to an L1-sound is not a yes/no-decision, but is a gradual phenomenon that depends on several factors, such as the naturalness of the L1 sounds per se (in markedness terms), and/or the learner’s current level of competence in L2. A consequence of this equivalence classification is that L2 sounds that lack similar items in L1 will be learned faster, since they are more easily perceived as different. However, what counts as ‘different’ is not self-evident. For example, it is clearly not simply a question of number of phonetic features, since this would imply that certain phones, such as [z], which differs only in the [±voice] domain from the “nearest” Swedish equivalent [s], should

behave in the same way as other phones, such as [θ], that also differs in only one aspect (place of articulation) from a Swedish [s]. However, as we shall see, this is not the case. For example, while a large proportion of Swedish speakers do include [θ] in everyday speech, virtually none makes use of [z]. More examples to the same effect could be given here. Clearly, other factors, such as influence from the orthography, phonetic salience or ease-of-production must play a role in the production of foreign items.

1.4 Other underlying factors

Besides differences between the phone inventories involved—and the resultant effects on the production of foreign items—there are a number of other factors that can be expected to affect the pronunciation of foreign names or words, as Figure 1 attempts to illustrate.

FIG 1: The Language User

These include, but are not limited to, the following:

- The speaker's competence and performance capabilities with respect to the foreign language.
- The speaker's expectations of the listener's (perceptual and cognitive) competence.
- The relative social status of speaker and listener.
- The time the name/word first appeared.
- The socio-cultural (not necessarily geographical) distance to the foreign culture in question.
- The recency and frequency of occurence of the lexical item in question in both languages involved.
- The population frequency of name bearers.
- Similarities and dissimilarities between the phonological systems involved.

Some of these factors are tangible and lend themselves well to studies and modelling, while others are more elusive. For instance, the importance of the complex nature of the socio-cultural dimension is discussed in Lipski (1976). However, it can be expected that phenomena pertaining to this dimension are not easily identified. Since English is the source language in most of the examples in the present study, it is also worth noting the differences in attitude towards different varieties of English. Although British English is what is predominantly and compulsorily taught as a second language in Swedish schools since the early 1940's, the cultural influx from North America and associated media exposure quite naturally has had a large impact on Swedish speakers of English. As a result of this media exposure, not only do Swedes exhibit an awareness of different English varieties, but they also express clear differences with regard to attitude vis-à-vis those varieties, both concerning spoken language, as observed in Mobärg (1998) and written language, which is examined in Virtanen & Lindgrén (1998). Bayard & Sullivan (2000) show that Swedish listeners exhibit an ability to perceive the country of origin of an English speaker. In fact, Swedish listeners are as good as New Zealand listeners at pin-pointing the country of origin of non-New Zealand English speakers. Moreover, Swedish listeners also exhibit an appreciation of the social status associated with specific English accents, in particular English and American accents (Bayard & Sullivan, ibid.).

1.5 A suggested metric along the awareness and fidelity dimensions

In an attempt to simplify this admittedly complex picture, involving many factors, few of which are directly observable, one can choose to look at an individual language user along fewer variables regarding xenophone production. One such simplification which will be used throughout this paper consists of looking at the language user's *awareness* of the original pronunciation, their *desire* to produce something along that direction, and their *ability* to do so with some degree of *fidelity*. Swedish subjects' productions of words and names of foreign origin can then be assigned to different categories along these variables, ranging from productions very close to the source language, to fully nativized (rephonematised) pronunciations. The former would correspond to a high awareness of the source language pronunciation, and a high ability to achieve this production, coupled with an obvious desire to do so. The latter would correspond to either a low awareness of the source language pronunciation or no desire to accomodate to that. In the mid-range of this spectrum would be the case where the language user is aware of the problem, has a desire to produce something "non-Swedish", as it were, but is unable to do so.

1.6 Previous work

Despite the fact that the problem of how to handle the pronunciation of foreign items is crucial, and that references are found that date back to the 16th century (Royal correspondence cited in Dahlstedt et al., 1969), as well as other lexicographical works (e.g., Müller, 1976; Kirkness, 1976), very little actual work on the phenomenon has been reported.

Maddieson (1984) discusses how this issue was considered/addressed when determining the phonological inventories for the several hundred languages in the *UCLA Phonological Segment Inventory Database (UPSID).* Segments which only occur in interjections or foreign words, judged not to be established as loans, are considered marginal, and are simply not included in *UPSID*, while they are included in case the loans that "appear to be fully assimilated in the language concerned" (Maddieson, ibid., p. 162). There is also the possibility of including a segment, while assigning a value to a special ANOMALY variable when the segment "occurs only in foreign words or unassimilated loans but these are frequent enough to consider including the segment in the inventory" (Maddieson, ibid., p. 170). However, this option seems to have been used sparsely, and in our corner of the world examples are limited to adding the voiced sibilant fricative [ʒ] to German and its voiceless counterpart [ʃ] to Finnish. While these additions perhaps are not exhaustive, they are still, as we shall see, in line with the results in this paper. On the other hand, no such loan segments are listed for Norwegian, which is a bit surprising in relation to the results in Eklund & Lindström (1998, 1996) and Lindström & Eklund (2000, 1999a,b) for Swedish, described later, and in relation to the results in this paper.

Abelin (1985) discusses how to represent pronunciation of foreign (mainly English) words in Swedish in *Svensk Ordbok* ('Swedish Dictionary'). She concludes that the English diphthongs [eɪ] and [ɔɪ] can be approximated with the Swedish sequences [ej] and [oj], respectively, but that the English diphthongs [əʊ] and [aʊ] are harder to accommodate. The English phone [z], Abelin states, is more or less consistently pronounced as [s] in Swedish,

which lacks a voiced counterpart, and the English alveolars [r, t, d, n] are normally realized as the corresponding Swedish dentals. Carlson, Granström & Lindström (1989) describe work and related statistics on name pronunciation for a Swedish reverse directory service. Facing the task of transcription of the Greater Stockholm telephone directory, they suggest that the proper names be divided into etymologically related categories. They report on a rule-based and a neural-net approach for automatic category assignment. In Carlson, Granström & Lindström (1990) the same authors describe an iterative method for development of special letter-to-sound rules for proper names. They also investigate the use of morphological analysis, and show that this is a very productive strategy for Swedish names.

In a cross-language comparison of some lexicographic data from the ONOMASTICA project (The Onomastica Project, 1995), Gustafson (1996, 1995, 1994) lists some of the questions that need to be answered when carrying out the task of transcribing foreign names, and also suggests some possible answers. If the origin of a foreign name is not known, it is suggested that "native" pronunciation rules be used. If a name contains foreign graphemes (assuming that both the native and foreign language use Latin script, which was the case in ONOMASTICA), it is suggested that the transcriber opt for the closest native phoneme (*sic!*) if the foreign pronunciation is known, and otherwise they should be mapped to the closest graphemes in the native language, before attempting to pronounce/transcribe the name. Finally, in the event of foreign phone[me]s that do not "exist" in the native language, which is the subject of this paper, Gustafson suggests that one either opt for "native" pronunciaton rules or enlarge the native phone[me] inventory. For example, Gustafson proposes that the Swedish phone[me] inventory be enlarged by the voiced [ð] and the unvoiced [θ] dental fricative in order to cover lexical items such as "Heather" and "Keith".

Based on studies of a corpus of recorded speech, Eklund & Lindström (1998, 1996) and Lindström & Eklund (2000, 1999a,b) describe what English phones Swedes actually use in their speech, and show that a large proportion of Swedish speakers include "non-Swedish" sounds in their production system when pronouncing English words and names. They also describe the inclusion of xenophones into the Telia Research concatenative synthesizer (Eklund & Lindström, 1998).

Möbius et al. (1997) mention that the German version of the Bell Labs multilingual TTS system has been augmented with phonetic units outside the German phone inventory in order to cover English and French speech sounds.

Trancoso et al. (1999) investigate how French speakers pronounce German place names—and vice versa—in a navigation system. They conclude that even speakers with little knowledge of the foreign language can adjust their pronunciation towards either the original language of the name in question (German/French), or according to the foreign language they know best (English). To solve the problem, the authors generate two lexica, one assuming no knowledge of the foreign language—similar to the approach employed in ONOMASTICA—and another lexicon assuming very good knowledge of the foreign language, which they claim to be more in line with their data.

In her dissertation, Fitt studied the production as well as the perception of foreign place-names, both in writing and speech (Fitt, 1998), using same-media as well as cross-media experimental set-ups. From her data, she concludes that her subjects indeed have some way of assigning foreign names to different classes of origin, although the specific details of what underlies this process are as yet unknown. She also finds that lexicon-based models, involving activation of known words or set of words and/or models involving analogy with subsections of words, are better suited for explaining her results, than are sets of linguistic rules.

2 Method

In order to acquire information and knowledge about Swedish speakers' usage of xenophones, a production study was conducted. The objective was to gain knowledge in several dimensions, while controlling for a number of variables suspected to be underlying sources of variability. From preliminary studies, it was apparent that speakers' productions lie on a continuum spanning from full adaption to Swedish to full adaption to the foreign language in question. It was also apparent that there is a large degree of variability, both within and between speakers, with explanatory factors suspected to be of both linguistic and non-linguistic nature. On this continuum lie productions that are neither Swedish nor English, as it were, which was expected to provide information regarding the speakers' awareness of and attitude against/towards foreign linguistic items. The data were also expected to give some insight into what adaptations with regard to foreign items speakers may expect from human and synthetic speech. It should be pointed out that although the study was carried out from a synchronic perspective, aiming at modelling the current behaviour of Swedish language users, there is of course always a diachronic angle to this problem, which has to be taken into account. A foreign lexical item can be expected to undergo a gradual assimilation to Swedish over the years, and in a production study such as the one described here, the degree of such assimilation is difficult to control for**.** For instance, the name "Roger" is a fully normal Swedish name, and the speaker must decide whether to pronounce the "Swedish" or the "foreign" name, which would show up as a possible production difference between e.g. the names "Roger Eklund" and "Roger Moore" The same problem probably occurs for other names and words of foreign origin of long standing in Swedish.

2.1 The linguistic material

A set of twelve sentences was constructed containing seventeen English speech sounds and two phones of German/Dutch origin. The full list is given in Table 1.

TABLE 1: INCLUDED PHONES

The sounds were chosen so that they would differ phonetically from Swedish speech sounds to varying degrees. A second criterium was that the chosen sound would not be included in traditional descriptions of the Swedish phonological system. It must be pointed out that English or German/Dutch sounds that lie close to Swedish sounds were not included in the material. For instance, the phonemes /t/ and /d/, which are considered alveolars (or alveolo-dentals) in English, are pronounced as dentals in Swedish, which was considered too small a difference to be of interest. Hence, such "corresponding" sounds were omitted from the material. It could be argued that the difference between a Swedish and an English /l/ also is a very minor difference.

The chosen phones were included in commonly known names and words in twelve fully natural Swedish sentences and it was assumed that the words and names in which the xenophones appeared would be known by the majority of the subjects. The full set of sentences is shown in Table 2.

TABLE 2: SENTENCES

It should be pointed out that our focus has been on whatever sounds non-Swedish, rather than what sounds like native English, irrespective of what English variety is implied. This means that we have not concerned ourselves with finer English-specific variation, like the realisation of e.g. the [r] sound in American English as compared to British English, or (its realisation) in intervocalic as compared to postvocalic position, but instead concentrated on realisations that are clearly not Swedish renderings, and most likely derived from an underlying English variety. As will be seen, in most cases, some kind of American English realisation was encountered, even in typically British names, like "Roger Moore", but the odd post-vocal British [r] deletion was also encountered. This also corroborates Mobärg's (1998) observations that American English is preferred over British English.

2.2 Recordings

The sentences were included in a much larger set of linguistic material that was collected for the development of the Telia/SRI Swedish speech recognizer as a part of the *Spoken Language Translator* (SLT) project (Rayner et al., 2000). The subjects were equipped with a headset with a hi-fi, close-talking, microphone and a hand-held telephone, and read the material from a computer screen. They were instructed to read the material in a "natural way", i.e., neither too carefully articulated nor too sloppily, and were told that the objective of the recording session was to collect training data for a Swedish speech recognizer. The material consisted of phonologically balanced sentences, ATIS (Hemphill et al., 1990.) sentences as well as newspaper and prose text. The xenophone material was presented under the heading 'Kändisar' (Celebrities), and it can be assumed that the subjects were unaware of the fact that their pronunciation of a specific set of phones was the object of study, given the instructions provided by the session leaders (Eklund et al., 2000).

2.3 Subjects

The subjects were all Telia employees or relatives of Telia employees (Telia is Sweden's largest telecom operator). They were chosen so as to cover the parameters age, gender, and region. The subjects also filled in forms, providing biographic information such as educational level. The age span was 15 to 75. Hi-fi recordings were obtained from 491 subjects at 40 different locations, covering almost all major dialect areas in Sweden. (One intended recording site in Western Sweden had to be omitted for practical reasons, and Swedish varieties spoken outside Sweden, e.g. in Finland, were not covered either.)

2.4 Data

A total of more than 5,400 sentences containing approximately 23,750 xenophone tokens were collected and transcribed. The relative proportion of xenophone tokens in the recorded data is given in Table 3.

TABLE 3: RELATIVE FREQUENCY OF XENOPHONES

2.5 Evaluation

Four phonetically trained native speakers of Swedish, with an above-average knowledge of English (three of the transcribers have lived several years in English-speaking countries), transcribed the target phones, using a fairly narrow transcription scheme. It was a deliberate decision not to use native speakers of any English variety for the transcription task, since, as is already mentioned, the focus was not on productions that might pass for English to native speakers of English, but instead on what sounds non-Swedish to native speakers of Swedish. As a consequence, this implied that finer English-specific phonetic details were beyond the scope of the listening tasks, while finer Swedish-specific phonetic details were of interest. The transcribers also made note of sentences where the subjects applied total adjustment, using exclusively Swedish allophones in their pronunciation of the foreign items.

All phonetic transcription is to some degree subjective, and thus prone to variability due to fluctuations in the judgments of different individuals. In order to deal with this problem, the transcribers agreed upon a set of common transcription guidelines that were used. In cases deemed problematic, the transcribers consulted each other, and the phone in question was listened to by more than one transcriber before it was labelled.

In order to further control the reliability of the transcription method, a consensus test was carried out at a later stage. Five sentence tokens (recordings) from each elicitation sentence (cf. Table 2) were randomly excerpted from the entire material. These 60 sentences contained 310 xenophone tokens. Three of the transcribers (T1, T2 and T3) then transcribed the sentences, and a consensus analysis was carried out. The results showed a range between 100% agreement, concerning the diphthongs, down to 55% agreement, concerning the [r] sounds. Pooling the results (in order to normalise for the difference between the more common xenophones in the material and the rare ones), the mean figures of agreement were, given as percentages:

	T1	T2	T3
T1	—	63.7	60.5
T2		—	88.9

A few comments are called for here. First, the lower figures mainly show up for xenophones with only a few instances in the test material, i.e., five tokens, and thus are more prone to a larger degree of disagreement. The xenophones with more instances exhibit a higher degree of agreement between the three labellers. Second, certain phones, like [r], are more prone to disagreement, since there were more alternative ways to transcribe them than for phones like [w], where only two alternative pronunciations showed up. Third, at the time of the consensus test, T2 and T3 had just finished extended sessions of transcription, while T1 had not carried out any transcription work for more than a year, which might explain the higher degree of disagreement between T1 and the other labellers, than between T2 and T3. Finally, the figures for each individual transcriber/xenophone pair were virtually the same across the material. This indicates that while borderline cases might show up as disagreements between labellers, the transcription as a whole seems to exhibit a fairly high degree of consistency.

3 Results

3.1 General

The production study shows that very few of the subjects (less than a dozen) resorted to total re-phonematization. Instead, the majority of the subjects expanded their allophonic repertoire considerably, despite the fact that the foreign items were embedded in a Swedish context in a fully plausible way. The results are shown in Table 4 (vowels) and Table 5 (consonants).

TABLE 4: RESULTS VOWELS

TABLE 5: RESULTS CONSONANTS

As shown in Tables 4 and 5, there is considerable distributional variation both between segments, and also, for "the same" foreign segment, between individual words. All the target vowels (except [æ] in the name "Jackson") and diphthongs are very well approximated in 90% of the cases or more, remarkably enough also for the diphthong [əʊ], which is quite dissimilar from any Swedish vowel. The results for the consonants indicate that the subjects almost without exception produced the voiceless affricate [tʃ], while the figures for the voiced counterpart [dʒ] ranged from 21% in "James" to 48% in "Jackson", which is also quite remarkable, since there is normally no voiced affricate in the Swedish phonological system. The retroflex fricative [ʂ] that 60% of the subjects produced in "Sharon" could be regarded as a sufficient approximation of the postalveolar [ʃ], as could perhaps also the alveolo-palatal fricative [ɕ], produced by 32% of the subjects. More detailed analysis for [ʃ] shows that the alveolo-palatal fricative [ɕ] was produced by 90% of the subjects from the Southern province Skåne (Scania), which is not surprising, since Southern Swedish lacks the retroflex [ʂ]. Both the voiced [ð] and the unvoiced [θ] dental fricative were produced to a surprisingly high degree, considering the lack of similar speech sounds in Swedish. On the contrary, virtually no subjects succeeded in producing the voiced alveolar [z] and postalveolar [ʒ] fricative. Subjects also chose to opt for almost full adjustment to Swedish in the case of [l], where "Swedish" [l] differs quite audibly from the [ɬ] found in most varieties of English. Full adjustment to Swedish was also found in the case of [w], where a vast majority of the subjects produced [v].

While the above at first sight suggests the possibility of some sort of hierarchy of sound accommodation, the fact that subjects were not at all consistent across lexemes complicates the picture further. In pronouncing "Roger Moore" and "James Bond", the same subject could perfectly well produce the dental voiced affricate [dʒ] on the first target phone (in the name "Roger"), but not on the second (in the name "James") or vice versa. Another source of variability was the frequently observed within-speaker type of inconsistency, where a phrase like "Diana and Charles" (as uttered by an individual speaker) may be pronounced with xenophones on "Diana" but not on "Charles", or vice versa. This inconsistency phenomenon needs further study before any conclusions can be drawn as to possible underlying explanations.

What is also apparent from Tables 4 and 5 is that there are three major distributional outcomes: One group consisting of sounds where almost all the subjects employed xenophonic expansion (e.g. the voiceless affricate [tʃ]), another group where the majority of the subjects resorted to rephonematization (e.g. the voiced fricative [z]), and a third group with a more even distribution between the three categories mentioned above. In the following examination of possible underlying factors, we have chosen mainly to focus our discussion on results from the third group. When calculating statistical significance, we have compared category 1 productions with the categories 2 and 3 conflated.

3.2 Gender differences

The results in Table 6 indicate that gender does not play an important role in governing the production distribution. The only statistically significant difference that was found between the genders showed up in the production of the voiced affricate [dʒ] in the name "Jackson", $p < 0.029$ (Pearson chi-square, two-sided). Seen in the light of the other groups, and the entire material pooled, this difference could probably be seen as incidental. Consequently, we have not found any significant differences between the genders.

TABLE 6: GENDER DIFF

3.3 Age differences

The results in Table 7 indicate that age is an important factor across the entire material. When pooling the results in Table 7, the differences between all groups, pair-wise, were stastically significant (in most cases $p < 0.001$, Pearson chi-square, two-sided), except between the younger groups, i.e., 36–45 vs. <16, 26–35 vs. 16–25, 26–35 vs. <16, and 16–25 vs. 16, where the differences were not statistically significant. These results imply that the socio-cultural dimension is of primary importance, since one might assume that different age groups are subject to foreign cultural influx in varying degress. Across most of Table 7, the share of productions in category 3 tends to increase with the age of the subjects. In addition to this tendency, Tables 7b, 7d, 7e and 7f also show a tendency where category 3 productions exhibit a minimum around ages 26–35. One particularly interesting result is apparent from Table 7a, where the data is drawn from the example sentence number 2 in Table 2, which contains the name of the actor Roger Moore. For ages above 35, around 70% of the productions fall into category 3, with a steep fall to around 40% for younger subjects. This distribution is quite distinct from the others in Table 7, and could be attributed to at least two possible explanations. One is simply that "Roger" is also a fairly common Swedish name, at least compared with "James" etc. The other explanation would be centered around the fact that Roger Moore's career as James Bond during the 70's would have influenced those age groups frequenting the cinemas during that period. Yet another, almost anecdotal, explanation, is related to a nationally very famous Swedish TV sketch, where two comedians ("Hasse å Tage") mention "Roger Moore" several times with marked, almost exaggerated, Swedish pronunciation. This points to the fact that the specific items chosen for use in a study such as this can seriously affect and skew the results.

TABLE 7: AGE DIFFS

3.4 Regional differences

Table 8 shows how the data were divided into nine different regional (dialect) groups, based on the dialect groups suggested by Elert (1994). It goes without saying that dialect regions are not as clear-cut and easily defined as are gender and age, and the dialect groups given by Elert are defined partly for prosodic reasons, something we have yet to look at. Thus, the groups given in Table 8 differ slightly from those of Elert in order to correspond better to the task at hand.

TABLE 8: REGIONS

The results shown in Table 8 were pooled and the nine dialect groups were controlled for statistical significance, pair-wise, with the other groups. Although the differences were significant between a few of the groups (A vs. C/D and I vs. D/G, using Pearson chi-square, two-sided), these differences are difficult to account for. Clearly, more in-depth studies are required to examine what potential rôle regional background might play with regard to xenophone use.

TABLE 9: REGIONAL DIFFS

3.5 Educational differences

The rôle played by educational level is shown in Table 10.

TABLE 10: EDUCATIONAL DIFFS

The results shown in Table 10 were pooled and the four groups were compared, pair-wise, with the other groups. The differences between all groups proved to be statistically significant ($p < 0.001$, Pearson chi-square, two-sided) except the pair CLASS 1 vs. CLASS 2, i.e., the two lowest levels of education, where there was no statistically significant difference. Thus, we can safely state that there is a clear correspondence between increasing educational level and closer approximation of the foreign language pronunciation.

4 Text-To-Speech Implementation

The results above have been drawn upon in the development of Telia Research's concatenative synthesizer. The synthesizer uses different-sized units (polyphones), with the demisyllable as the basic unit. The polyphones include a set of diphones, which also serve the purpose of being used in fall-back situations. Derivational endings, words from the closed word classes and nasal and /h/ triphones are also included for speech quality reasons. During the development of a female voice in 1996, polyphones containing xenophones were also added. The total number of units in the resulting concatenative synthesizer was approximately 15,000.

For evaluation purposes, the set of sentences that was used in the data collection was synthesized in two ways: first, using non-xenophone polyphones only, and second, making use of the added xenophone polyphones. The synthesis procedure was described in Eklund & Lindström (1998), which also contains examples of the resulting synthetic speech. As could be expected, adding xenophone polyphones produced more natural-sounding speech output, as indicated by informal listening tests. More rigorous testing would of course be required to confirm this formally, but it would indeed be surprising if a synthesizer using another phone inventory than the majority of the population would be judged better than a synthesizer that is more similar to human beings in this respect.

5 Discussion

Xenophones lie at the crossroads of a number of disciplines that make human language such a fascinating topic. Any serious study of xenophones will have to take into account phonetics, phonology and probably also morphology and syntax, alongside psychology, sociology, cross-cultural influences and a host of other factors. While this study has not been able to address more than a few of the issues raised, the results still shed some light both on actual speaker behaviour and on some of the factors governing that behaviour, in particular age and educational level.

From a purely linguistic perspective, a number of theoretical questions are raised. To begin with, the "foreignness" of this category of sounds can be discussed. If most Swedes use certain sounds in everyday conversation, and/or expect them to be used, then how "foreign" are they in the language community? Moreover, in a world that is characterized by increasing international communication—economic, cultural, social—such cross-breeding between languages can be expected to become more and more frequent, and is in fact the normal situation in many parts of the world where multilingualism is the norm. Phonetically, it is not obvious how one should approach the problem, especially since the acoustic features of both the native and the foreign language are not easily specified. Comparing phone sets is complicated since the mapping necessarily involves many-to-many relations/couplings. There are many different regionally and socialectally related allophone sets both within Swedish and within the foreign language (in this case different varieties of English, and to some extent German/Dutch), and the question is which variety in the native language should be compared with which variety in the foreign language. Again, from a more theoretical phonological side, one obvious goal of this research would be to set up a model with the capability of explaining the differences in xenophone distribution across categories that have been observed. It would be surprising if such a model did not involve some elements from theoretical/phonological models within the SLA field.

As indicated in Figure 1, the human language users are obviously able to handle the multitude of factors at play in some way, and from a speech technology perspective, it will become increasingly necessary for automatic spoken language systems to handle this type of situations as well. In the following subsections, some of the consequences that these observations are bound to have for automatic speech recognition and speech synthesis will be discussed, and directions for future research will be indicated

5.1 Implications for speech recognition

A large-vocabulary automatic recognizer faces the entire variety of speech sounds within a given speech community, and the modelling of what it can expect to hear boils down to a few crucial issues.

First, the standard view on what the Swedish phone set looks like must be reconsidered, since it obviously contains a number of sounds normally not considered "Swedish", despite the fact that a large number of Swedish speakers do use them in normal conversation.

To complicate matters further, a word or name of foreign origin and containing foreign—or foreign-sounding—segments can appear in an otherwise Swedish sentence, which means that the recognizer needs to handle phones

from (at least) two languages at once. Within the SLT project (Rayner et al., 2000), an HMM-based recognizer that is able to handle English and Swedish was developed (Digalakis & Neumeyer, 2000; Weng, 2000; Weng et al., 1997). The recognizer runs both a Swedish and an English recognizer, and is capable of recognizing the odd Swedish word inside an otherwise English sentence, and vice versa. Technically, this was done by inter-connecting the so-called "back-off nodes" of the two language models, thus allowing the recognizer to switch language, as it were, between words. A problem with this method is that the lexicon used by the English recognizer will contain something like the phone string [dʒeɪmz] for the name "James" and the recognizer's acoustic models will be trained on those speech sounds (in context). Similarly, the lexicon employed by the Swedish recognizer will permit the string [jejms] and the acoustic models of that recognizer will be trained on those speech sounds (among others). However, neither recognizer will be capable of accepting the most likely phone sequence produced by a Swedish speaker, according to our findings, which would be [dʒejms]. This implies that running parallel "native" recognizers probably cannot solve the xenophone problem, and that xenophones need to be included in the training material for a single recognizer instead. On the other hand, there is a severe training problem associated with this second approach, since the amount of training data normally required to build, say, a context-dependent HMM phone model, would be quite difficult to obtain for xenophones, especially if data from several contexts is required. One technical, third, solution, drawing on advantages from the other two approaches described here, would be to train the acoustic English-related xenophone models in the Swedish recognizer by bootstrapping from English speech data.

Another issue is exactly how acute a problem xenophones present to a recognizer. This, of course, depends heavily on the context and discourse. An application like automatic handling of film ticket purchasing would surely need to cope with a large number of xenophones, since most English film titles are not translated into Swedish. Within other domains, such as bookings of summer houses in the Stockholm archipelago, xenophones are not likely to occur at all. Thus, xenophone inclusion for a given application is also an empirical and task-specific issue.

5.2 Implications for speech synthesis

As opposed to recognition, where the entire variety of a language needs be considered and catered for, a synthesizer generally only needs to cover *one* acceptable variety. The operative word here, of course, is "acceptable". Although it is our belief that a production study provides information in the acceptability domain insofar as it can be assumed that users of speech synthesis systems will be less prone to accept a synthesizer with a lower level of competence than themselves, the only safe method to gain insight in the acceptability domain would be to conduct a perception study, and ask subjects to rank the pronunciations along a few dimesions, such as intelligibility, "intelligence", pleasantness and so on. It is our belief that a low inclusion level of xenophones might not primarily show up in the intelligibility dimension. Rather, according to our findings, such a level of xenophone inclusions might give the listener an impression of a speaker with a particular age or educational level.

Another problem to consider is that "maximizing" in the xenophone dimension might leave certain listeners behind, especially concerning languages that are not so widely known as English (e.g. French, German or Russian) and that

an appropriate, intermediate, level must be found. It can be assumed that choosing too "high" a level will signal an attitude which would be perceived as high-browed and obnoxious, leading the user of the system to feel excluded. This, too, needs to be studied in further detail.

To the best of our knowledge, few attempts to include xenophones in synthesizers have so far been made. As stated previously, Möbius et al. (1997) mention the inclusion of a few English and French sounds in the German version of the Bell Labs multilingual TTS system and Eklund & Lindström (1998) report on the inclusion of English xenophones in the Telia Research Swedish synthesizer.

5.3 Future research

In the development of speech technology systems there is an increasing interest in issues such as dialectal variation, cross-language applications, handling of foreign accents etc. As mentioned earlier, this problem is becoming more acute in an increasingly internationalized world, where people tend to speak more than one language, and also tend to ask for services that pay little or no attention to national or language borders.

In our investigation a production study was carried out, which apart from obvious production results also gave some insight in the expectation/attitude dimension. However, to gain more knowledge concerning these issues, perceptually oriented studies should be carried out.

Apart from the perception studies mentioned above, a deeper look into the phonological-regional dimension is needed. The rationale for doing this is that one would want an intelligent recognizer to possess a certain level of predictive power, so that it could "tune in" to a particular speaker's use of xenophones (and idiosyncratic speech behaviour in general). However, our observations so far do not provide much hope in that dimension, since the speakers generally do not exhibit a high degree of consistency in their use of xenophones. This was exemplified by xenophone usage within a single speaker in virtually lexicalized phrases, such as "Diana and Charles". Thus, our studies so far indicate that xenophone inclusion may appear spot-wise, rather than consistently. However, this question needs more research.

Another topic that awaits studies is to what extent prosodic signalling is employed. Some subjects signalled awareness of the foreignness of names and words by using a prosodic realization that is influenced by the source-language, in this case English, either in addition to, or independently of, the use of xenophones. So far, we have not conducted any formal studies of this phenomenon, and the benefits from such knowledge of course require that recognizers make use of prosody, something which currently is not done, at least not to any large degree.

Another factor to be studied further is the role of orthography, something we have tried to normalize for by including the same sounds with different spellings (e.g., the voiced affricate [dʒ] was presented both in the name "James" and in the name "Roger"). It proved to have some effect (Eklund & Lindström, 1998), but more data are needed before any far-reaching conclusions may be drawn concerning the rôle of orthography.

Moreover, as pointed out in Fitt (1998), the line between fully foreign names and words on the one hand, and fully native names and words on the other hand, is arbitrary. In this study, we have used names and words that were

expected to be known to most (preferably all) subjects, and some of these names are also fairly common names in Swedish. Whether or not subjects did perceive the name or word as “foreign”, or to what degree, is something which is very hard to measure, since perception is clearly dependent on socio-cultural factors that are not controlled for in our study, and are indeed very hard to control for.

In particular, the vocabulary of younger people can be expected to be more coloured by the current cultural influx. As previously mentioned, a morphological approach has been proven productive for (at least) Swedish (Carlson et al. 1990, 1989), and an initial attempt at implementing a lexical component capable of handling what might be termed *xenomorphs* in a two-level formalism was recently reported on by Lindström and Kasaty (2000).

The inclusion of xenophones beyond doubt varies across languages, and it goes without saying that these observations on Swedish should not be “translated” into other languages. The typical speaker of Swedish (in Sweden) is exposed to English on an everyday basis, mainly since Swedish television never dubs foreign programs (except for programs aimed at pre-literate children), which also explains the observation that an American English rendering of the English name “Roger Moore” is more common than a British English rendering, given the fact that American TV shows vastly outnumber British TV shows on Swedish TV. This type of exposure to foreign languages does not exist in countries where foreign TV shows are always, or frequently, dubbed—or where voice-over is employed—and it is assumed that it is just one factor that influences the inclusion of xenophones in any given language.

This study is necessarily limited in scope, and clearly can not provide an exhaustive description of ‘xenophonology’. The apparent existence of phones that seem to lack underlying phonemes is per se an intriguing issue which has so far received little attention. It is our hope that this pilot study will be followed by other studies, addressing other aspects of this multi-faceted area.

Acknowledgements

The authors would like to thank Jessica Granberg and Per Sautermeister for help with the transcriptions. We would also like to thank Lars Ahrenberg, Jessica Granberg, Jaan Kaja, Eva Lindström, Robert McAllister, Kirk Sullivan and three anonymous reviewers for comments on draft versions of the paper. Thanks to Jan Anward, Björn Hammarberg and Åke Viberg for helpful discussions. Thanks to Nathalie Kirchmeyer and Hartmut Traunmüller for providing French and German translations, respectively, of the abstract. Finally, thanks to Juliana Handberg for the drawing in Figure 1.

References

Abelin, Å. 1985. Om uttalsmarkering och uttalsregler i svensk ordbok. Rapporter från Språkdata (21), Gothenburg University, Department of Computational Linguistics, Gothenburg.

Bayard, D., Sullivan, K.P.H. 2000. Perception of country of origin and social status of English speakers by Swedish and New Zealand listeners. Proc. Fonetik 2000, The Swedish Phonetics Conference, May 24–26 2000. University of Skövde. ISBN 91-630-9647-1, pp. 33–36.

Billi, R. n.d. Interview published in Le Journal – The Journal of Record for Human Language Technology. http://www.hltcentral.org/usr_docs/LeJournal/I-phonesPhonemesMultimodality.pdf

Carlson, R., Granström, B., Lindström, A. 1989. Predicting name pronunciation for a reverse directory service. Proc. European Conference on Speech Technology, vol. 1, pp. 113–116.

Carlson, R., Granström, B., Lindström, A. 1990. Methods to generate the pronunciation of proper names in Swedish. Proc. Fonetik 90, The Fourth Swedish Phonetics Conference, PHONUM 1, Umeå/Lövånger, pp. 74–77.

Dahlstedt, K.-H., Bergman, G., Ståhle, C.I. 1969. Främmande ord i svenskan. Verdandis skriftserie, Läromedelsförlagen.

Digalakis V., Neumeyer, L. 2000. Multiple Dialects and Languages. In: Rayner, M., Carter, D., Bouillon, P., Digalakis, V., Wirén, M. (Eds.), The Spoken Language Translator. Cambridge University Press, ISBN 0 521 77077 7, ch. 18, pp. 274–283.

Eklund, R., Kaja, J., Neumeyer, L., Weng F. & Digalakis, V. 2000. Porting a Recognizer to a New Language. In: Rayner, M., Carter, D., Bouillon, P., Digalakis, V., Wirén, M. (Eds.), The Spoken Language Translator. Cambridge University Press, ISBN 0 521 77077 7, ch. 17, pp. 265–273.

Eklund, R., Lindström, A. 1998. How To Handle “Foreign” Sounds in Swedish Text-to-Speech Conversion: Approaching the ‘Xenophone’ Problem. Proc. ICSLP 98, Sydney, November 30–December 5. Paper 514, vol. 7, pp. 2831–2834. CD-ROM available from Causal Productions Pty Ltd, PO Box 100, info@causal.on.net.

Eklund, R., Lindström, A. 1996. Pronunciation in an internationalized society: A multi-dimensional problem considered. Fonetik 96, The Swedish Phonetics Conference, Nässlingen, 29–31 May, 1996. TMH-QPSR 2/1996, pp. 123–126.

Elert, C.-C. Indelning och gränser inom området för den nu talade svenskan — En aktuell dialektografi. In: Edlund, L.-E. (Ed.), Kulturgränser — Myt eller verklighet? DIABAS 4, skrifter från den dialektgeografiska institutionen för nordiska språk vid Umeå universitet, Umeå 1994, pp. 215–228.

Fitt, S.E. 1998. Processing Unfamiliar Words: A Study in the Perception and Production of Native and Foreign Place Names. PhD Thesis, University of Edinburgh.

Flege, J.E. 1987. Effects of Equivalence Classification on the Production of Foreign Language Speech Sounds. In: James, A., Leather, J. (Eds.), Sound Patterns in Second Language Acquisition. Foris Publications, pp. 9–39.

Gustafson, J. 1996. A Swedish Name Pronunciation System. Licenciate thesis, TMH, Royal Institute of Technology (KTH), Stockholm, Sweden.

Gustafson, J. 1995. Transcribing Names with Foreign Origin in the Onomastica Project. Proc. ICPhS 95, Stockholm, Sweden, vol. 2, pp. 318–321.

Gustafson, J. 1994. ONOMASTICA—creating a multilingual dictionary of European names. In: Bruce, G., House, D., Touati, P. (Eds.), Papers from the Eighth Swedish Phonetics Conference, Working Papers, Lund University, vol. 43, pp. 66–69.

Hammarberg, B. 1990. Conditions on Transfer in Second Language Phonology Acquisition. In: Leather, J., James, A. (Eds.), New Sounds 90, Proc. of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech. University of Amsterdam, pp. 198–215.

Hemphill, C.T., Godfrey, J.J., Doddington, G.R. 1990. The ATIS Spoken Language Systems Pilot Corpus. Proc. DARPA Speech and Natural Language Workshop, Hidden Valley, PA, pp. 96–101.

Kirkness, A. 1976. Zur Lexikologie und Lexikographie des Fremdworts. In Probleme der Lexikologie und Lexikographie – Sprache der Gegenwart, Band XXXIX. Pädagogischer Verlag Schwann, Düsseldorf, pp. 226–241.

Lindström A., Eklund, R. 1999a. Xenophones Revisited: Linguistic and other underlying factors affecting the pronunciation of foreign items in Swedish. Proc. of ICPhS 99, San Francisco, August 1–7, Paper 0708, vol. 3, pp. 2227–2230.

Lindström A., Eklund, R. 1999b. [jàːmes] or [dʒɛɪmz] or Perhaps Something In-between? Recapping Three Years of Xenophone Studies. Proc. Fonetik 99, The Swedish Phonetics Conference, June 2–4 1999. Gothenburg Papers in Theoretical Linguistics, 81, pp. 109–112.

Lindström A., Eklund, R. 2000. How "foreign" are foreign speech sounds? Implications for speech recognition and synthesis. RTO Meeting Proceedings 28. Multi-lingual Interoperability in Speech Technology. Papers and reports presented at the Tutorial and Workshop co-sponsored by the Information Systems Technology Panel (IST) of RTO-NATO and the European Speech Communication Association (ESCA) held at Leusden, The Netherlands on 13–14 September 1999, pp. 15–19.

Lindström, A., Kasaty, A. 2000. A two-level approach to the handling of foreign items in Swedish speech technology applications. Proceedings of of ICSLP'00, 16–20 October 2000, Vol. 1, pp. 54-57.

Lipski, J.M. 1976. Prejudice and pronunciation. American Speech, vol. 51, pp. 109–118.

Maddieson, I. 1984. Patterns of sounds. Cambridge University Press, pp. 162–163, 170, 203 et passim.

Mobärg, M. 1998. Media exposure vs. educational prescription: The case of British and American English in Sweden. In: Lindquist, H., Klintborg, S., Levin, M. & Estling, M. (Eds.), The Major Varieties of English. Papers from MAVEN 97, Växjö 20–22 November 1997. Växjö University, pp. 249–262.

Müller, W. 1976. Fremdwortbegriff und Fremdwörterbuch. In: Probleme der Lexikologie und Lexikographie – Sprache der Gegenwart, Band XXXIX. Pädagogischer Verlag Schwann, Düsseldorf, pp. 211–225.

Möbius, B., Sproat, R., van Santen, J.P.H., Olive, J.P. 1997. The Bell Labs German Text-to-Speech System: An Overview. Proc. ESCA, Eurospeech 97, Rhodes, Greece, ISSN 1018–4071, vol 5, pp. 2443–2446.

The Onomastica Consortium. 1995. The Onomastica Interlanguage Pronunciation Lexicon. Proc. ESCA, Eurospeech 95, Madrid, Spain, ISSN 1018–4074, pp. 829–832.

Rayner, M., Carter, D., Bouillon, P., Digalakis, V., Wirén, M. (Eds.). 2000. The Spoken Language Translator. Cambridge University Press.

Trancoso, I., Viana, C., Mascarenhas, I., Teixeira, C. 1999. On deriving rules for nativised pronunciation in navigation queries. Proc. ESCA, Eurospeech 99, Budapest, Hungary, ISSN 1018–4074, vol. 1, pp. 195–198.

Virtanen, T., Lindgrén, S.-A. 1998. British or American English? Investigating what EFL students say and what they do. In: Lindquist, H., Klintborg, S., Levin, M. & Estling, M. (Eds.), The Major Varieties of English. Papers from MAVEN 97, Växjö 20–22 November 1997. Växjö University, pp. 273–281.

Weng, F., Stolcke, A., Cohen, M. 2000. Language Modeling for Multilingual Speech Translation. In: Rayner, M., Carter, D., Bouillon, P., Digalakis, V., Wirén, M. (Eds.), The Spoken Language Translator. Cambridge University Press, ISBN 0 521 77077 7, ch. 16, pp. 250–264.

Weng, F., Bratt, H., Neumeyer, L., Stolcke, A. 1997. A Study of Multilingual Speech Recognition. Proc. Eurospeech, vol. 1, Rhodes, Greece, Vol. 1, pp. 359–362.

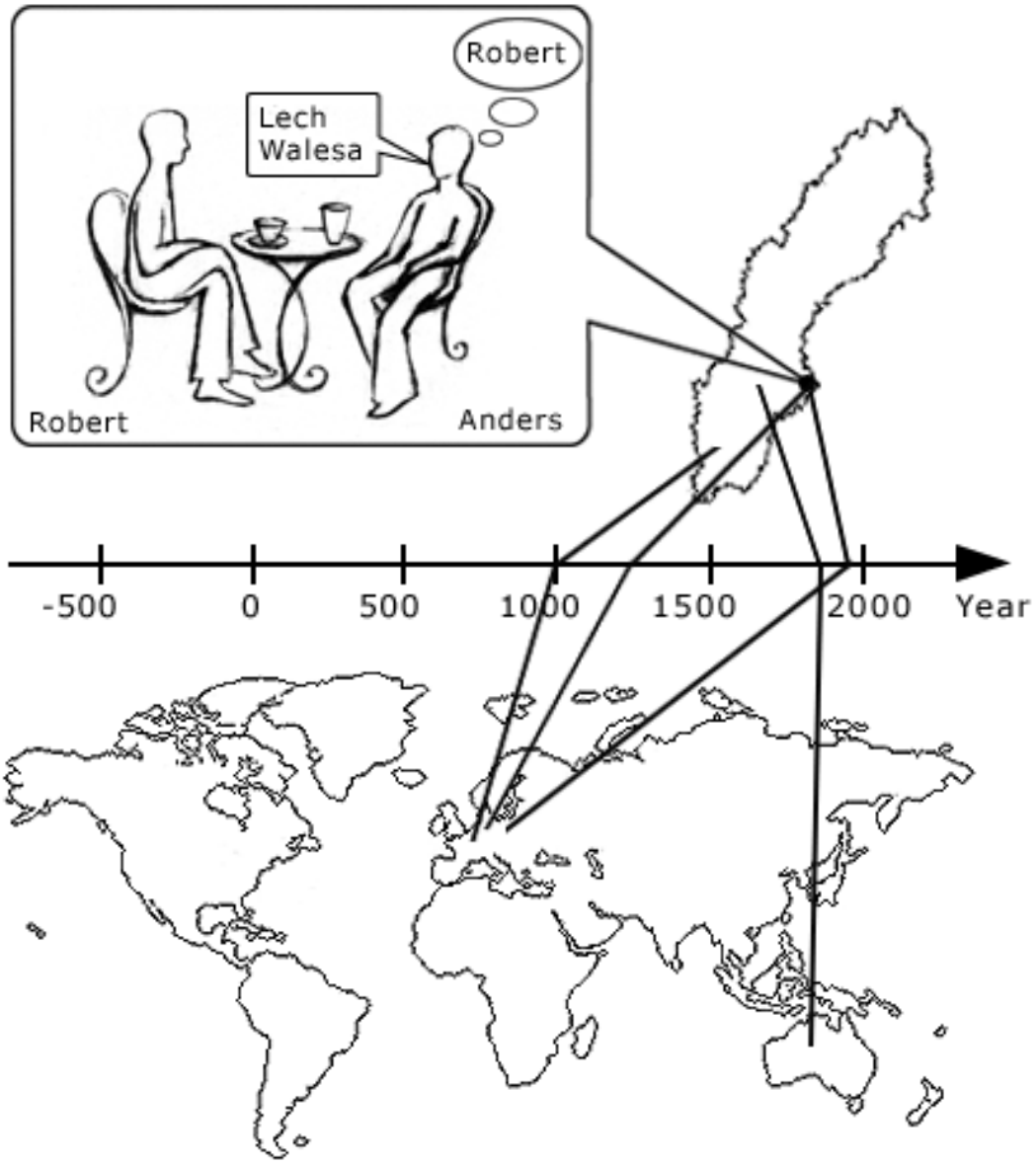

Figure 1: The language user in a typical situation. When speaker Anders is pronouncing a foreign name in speaker Robert's presence, a number of factors affect the phonetic rendering of the name, such as the name's country of origin, the time it was introduced in speaker Anders's community, what channel it passed through, Anders's knowledge of Robert's language competence and a host of other factors.

Table 1: Included target phones/xenophones and phonetic realisations observed in Swedish subjects' speech. The examples are from the material used to elicit the pronunciations.

Target Phone / Xenophone	Potential Realisation	Example Word(s)
[x]	[x, ɧ, ʂ]	Aachen
[θ]	[θ, t, ʈ, d]	thriller, Thatcher
[ð]	[ð, d, ɖ, t]	the
[ʃ]	[ʃ, ʂ, ɧ, ʂ, ɕ]	Sharon
[z]	[z, s]	music, James
[ʒ]	[ʒ, ɧ, ʂ, ɕ]	television
[tʃ]	[tʃ, ʂ, ɕ, tɕ, tʂ, ʈʂ]	Charles, Thatcher
[dʒ]	[dʒ, g, j]	Roger, James
[ɹ]	[r, ɚ, r, ʁ, ʐ]	Sharon, Roger
[Ø]	[r, ɚ, r, ʁ, ʐ, Ø]	Roger
[ɬ]	[l, ɬ, ɬ]	Lloyd, Michael, thriller, Douglas
[w]	[w, v]	we, world
[aː]	[aː, a, ɒ, ɒː]	Aachen
[æ]	[æ, a, e]	Jackson, Maggie
[ju]	[juː, ʉ̟ː, ɵ]	music
[əʊ]	[əʊ, uː, o, ɔ]	Stone
[aʊ]	[əʊ, ɔ, a]	cowboy
[eɪ]	[eɪ, ej, aj]	Baywatch
[aɪ]	[aɪ, aj, i]	Michael
[ɔɪ]	[ɔɪ, ɔj]	Cowboy, Lloyd

Table 2: The set of sentences used in the recording sessions. The sentences were presented to the subjects under the heading "Kändisar" ('Celebrities'), and contained names and words of foreign origin judged to be commonly known.

No.	Sentences with translation
1	TV-serien "Baywatch" har väckt en del uppmärksamhet på sistone. ("The TV series 'Baywatch' has attracted some attention lately.")
2	Många har Roger Moore som favorit i rollen som James Bond. ("A lot of people prefer Roger Moore's interpretation of James Bond.")
3	Den mest säljande skivan någonsin är "Thriller" av Michael Jackson. ("The best-selling record ever is 'Thriller' by Michael Jackson.")
4	Veckopressens favoriter är verkligen Diana och Charles. ("Diana and Charles are indeed the favourites of the tabloids.")
5	Intercity-tåget gick direkt från Aachen till Baden-Baden. ("The Intercity train went straight from Aachen to Baden-Baden.")
6	Jag flyger alltid business class på längre sträckor. ("I always fly business class on long trips.")
7	Det anses allmänt att John Major är en blek efterträdare till Maggie Thatcher. ("It is a widely held opinion that John Major is a rather pale successor to Maggie Thatcher.")
8	I filmen "Basic Instinct" spelade Sharon Stone mot Michael Douglas. ("Sharon Stone co-starred with Michael Douglas in the film 'Basic Instinct'.")
9	En av stumfilmens stora stjärnor var Harold Lloyd. ("Harold Lloyd was one the big stars of silent movies.")
10	Många rockstjärnor medverkade i sången "We are the World". ("Many rock stars participated in the song 'We are the World'.")
11	TV-kanalen MTV, "Music Television", har kommit att påverka hela musikbranschen. ("The TV channel MTV, 'Music Television', has affected the entire music business.")
12	En cowboy på en häst är verkligen symbolen för amerikansk kultur. ("A cowboy on a horse is really the symbol of American culture.")

Table 3: Number of xenophone tokens.

Target Phone / Xenophone	Number of transcribed tokens	Elicited from the words or names
[aɪ]	1379	Michael (Jackson), Michael (Douglas), Diana
[eɪ]	1848	Baywatch, James, Major, Basic
[ɔɪ]	812	Lloyd, cowboy
[əʊ]	460	Stone
[aʊ]	342	Cowboy
[juː]	452	Music
[æ]	1845	Jackson, Maggie, Thatcher, Diana
[aː]	1395	Aachen, Baden-Baden
[tʃ]	1388	Baywatch, Charles, Thatcher
[dʒ]	2310	Roger, James, Jackson, John, Major
[ʃ]	460	Sharon
[ʒ]	452	Television
[θ]	917	Thriller, Thatcher
[ð]	462	the
[z]	1382	James, Charles, Music
[ɹ]	3697	Roger, Moore, Thriller, John, Major, Thatcher, Sharon, are
[ɬ]	2760	Michael (J), Michael (D), Thriller, Douglas, Lloyd
[w]	924	We, World
[x]	465	Aachen
Σ	23750	

Table 4: For each target English speech sound and each occurrence in the read sentences, the resulting distribution of the Swedish subjects' productions, as obtained by manual phonetic transcription, is shown as a percentage. Based on the similarity between the produced sound and the target phone, the different productions are assigned to one of three categories along two dimensions: the AWARENESS dimension (to what extent people are aware of the difference between Swedish and English pronunciation), and the FIDELITY dimension (how well they succeed in the production of the foreign sounds). The **first category** corresponds to a high awareness among the subjects coupled with a high capability in rendering a sound close to the one in the source language. The **second category** corresponds to the case where the subjects were apparently aware that something "non-Swedish" would be appropriate, but failed to produce a good approximation. Finally, the **third category** corresponds to full adjustment to Swedish. Cases where the transcribers were unable to hear what was produced are marked by an asterisk.

		Category	***1***	***2***	***3***
		Awareness	***high***		***low***
		Fidelity	***high***	***low***	
Target	***Tokens***	***In the word***			
aɪ	454	M**i**chael (Jackson)	95.8 aj	0.4 ej	3.1 iː 0.7 ɪ
	460	M**i**chael (Douglas)	94.6 aj		3.0 iː 2.4 ɪ
	465	D**i**ana	54.6 aj	0.5 *	44.9 i
eɪ	465	J**a**mes	97.2 ej	2.4 ɛ	0.4 a
	463	M**a**jor	92.0 ej	0.9 ɛ 0.2 aj	6.9 a
	460	B**a**sic	90.7 ej	3.5 ɛ 0.4 ɛj 0.4 æː 0.2 ɔ	3.9 a 0.9 aː
	460	B**ay**watch	95.3 ej	1.4 *	3.3 aj
ɔɪ	470	Ll**oy**d	98.7 ɔɪ	1.3 *	
	342	cowb**oy**	100.0 ɔɪ		
əʊ	460	St**o**ne	89.2 əʊ	0.4 oːʊ 0.2 ɔuː 0.2 ɪuə 0.2 uːə 0.2 *	5.0 oː 4.4 uː 0.2 ɔ
aʊ	342	c**o**wboy	47.1 aʊ	2.6 ɔʊ 1.2 *	42.6 ɔ 3.2 oː
juː	452	M**u**sic	96.0 juː	0.2 jʉː 0.2 jəj	3.6 ʉː
æ	454	J**a**ckson	75.7 æ	1.1 ɛ 0.2 *	23.0 a
	463	M**a**ggie	90.2 æ	0.6 ɛ 1.5 *	7.1 a 0.6 aː
	463	Th**a**tcher	95.5 æ	1.7 i(ː) 0.2 ej 0.2 ɛ 0.2 *	2.2 a
	465	Di**a**na	55.5 æ	1.5 eː 1.0 a 1.6 *	40.4 ɑː
aː	465	**Aa**chen	12.5 aː	9.5 ɛ 7.1 ɛj 3.4 ɛː 4.5 *	38.0 a 25.0 ɑː
	465	B**a**den-Baden	15.9 aː	1.5 ɛː 1.1 *	78.9 ɑː 2.6 a
	465	Baden-B**a**den	15.7 aː	1.0 ɛ 0.7 *	80.0 ɑː 2.6 ɑː
Total	8533				

Table 5: Consonants. (See Table 4 for an explanation.)

		Category	***1***	***2***	***3***
		Awareness	***high***		***low***
		Fidelity	***high***	***low***	
Target	***Tokens***	***In the word***			
tʃ	460	Baywa**tch**	77.2 tɕ 18.9 tʂ	0.8 ʂ 3.9 *	
	465	**Ch**arles	22.5 tɕ	75.7 ɕ 1.8 *	
	463	Tha**tch**er	99.2 tɕ	0.4 ʂ 0.2 ɕ 0.2 *	
dʒ	465	Ro**g**er	32.5 dʒ	0.2 gd	67.3 g
	454	**J**ackson	47.8 dʒ	0.4 dɪ 0.2 ɕ 0.4 *	51.2 j
	463	**J**ohn	28.9 dʒ		71.1 j
	463	Ma**j**or	31.6 dʒ	0.4 *	68.0 j
	465	**J**ames	21.7 dʒ		78.3 j
ʃ	460	**Sh**aron	60.1 ʂ 31.5 ɕ 6.3 ʃ	0.4 ɧ 0.4 tɕ 0.2 sɕ 0.2 *	0.9 s
ʒ	452	Televi**s**ion	1.7 ʒ	94.6 ʂ 3.3 ɧ 0.2 ɕ 0.2 ʂɧ	
θ	454	**Th**riller	49.6 θ	0.2 ɕ	48.0 t 2.2 ʈ
	463	**Th**atcher	42.3 θ	1.3 tɕ 0.2 s 0.2 *	56.0 t
ð	462	**th**e	38.5 ð	1.1 r 0.4 h 0.4 v 0.2 j 0.2 0 0.4 *	57.5 d 1.3 t
z	465	Jame**s**	0.4 z		99.6 s
	465	Charle**s**		1.4 *	98.4 s 0.2 es
	452	Mu**s**ic			100.0 s
ɹ, Ø	465	Roge**r**	1.1 ʐ̩	2.1 *	96.8 r
	454	Thrille**r**	1.1 ʐ̩		98.9 r
	463	Majo**r**			100.0 r
	463	Thatche**r**	0.4 ʐ̩		99.6 r
	465	Moo**r**e	0.8 ʐ̩	2.6 *	96.6 r
ɹ	465	**R**oger	1.5 ʐ̩	2.4 *	96.1 r
	460	Sha**r**on	62.0 ʐ̩ 4.4 ɹ	0.4 ʁ 0.2 d 0.2 n 0.2 t 0.4 *	32.2 r
	462	a**r**e	1.5 ʐ̩	0.2 *	98.3 r
ɬ	454	Thri**ll**er	2.6 ɬ		97.4 l
	454	Michae**l** (J)	2.0 ɬ		98.0 l
	460	Michae**l** (D)	7.4 ɬ		92.6 l
	460	Doug**l**as	4.8 ɬ	0.2 0	95.0 l
	470	**Ll**oyd	17.1 ɬ	0.8 *	82.1 l
	462	wor**l**d	0.4 ɬ		99.6 l
w	462	**W**e	13.0 w	0.2 m	86.8 v
	462	**W**orld	0.9 w	0.2 b	98.9 v
x	465	Aa**ch**en	12.9 x	8.7 k 6.4 tʃ 1.8 *	33.8 ɧ 27.8 ɕ 8.6 ʃ
Total	15217				

Table 6: Distribution of Swedish subjects' productions of nine target xenophones (a–i) across **gender** (M=Male; F=Female) and category [%] (cf. Table 4). Cases where the transcribers were unable to hear what was produced are marked by an asterisk.

Gender	Tokens	Category			Tokens	Category			Tokens	Category		
		1	2	3		1	2	3		1	2	3
	a) Target dʒ in "Ro**g**er"				*b) Target* dʒ *in* "**J**ames"				*c) Target* dʒ *in* "**J**ackson"			
M	245	33 dʒ		67 g	245	20 dʒ		80 j	240	42 dʒ	1 *	57 j
F	230	33 dʒ	1 *	66 g	230	24 dʒ		76 j	226	52 dʒ	2 *	46 j
	d) Target θ *in* "**Th**riller"				*e) Target* θ *in* "**Th**atcher"				*f) Target* ð *in* "**th**e World"			
M	240	48 θ	1 t̪	51 t	243	43 θ	1 *	56 t	243	37 ð	1 r	58 d
											1 d̪	2 t
											1 v	
F	226	51 θ	3 t̪	46 t	230	42 θ	2 *	56 t	230	40 ð	1 v	56 d
											1 r	1 t
											1 *	
	g) Target əʊ *in* "St**o**ne"				*h) Target* æ *in* "M**a**ggie"				*i) Target* æ *in* "J**a**ckson"			
M	242	88 əʊ	2 *	4 uː	243	88 æ	2 *	10 a	240	74 æ		26 a
				6 oː								
F	228	91 əʊ	1 *	3 uː	230	93 æ	2 *	5 a	226	76 æ	2 ɛ	22 a
				5 oː								

Table 7: Distribution of Swedish subjects' productions of nine target xenophones (a–i) across **age** and category [%] (cf. Table 4). Cases where the transcribers were unable to hear what was produced are marked by an asterisk.

Age	Tokens	Category			Tokens	Category			Tokens	Category		
		1	2	3		1	2	3		1	2	3
	a) Target dʒ *in* "Ro**g**er"				*b) Target* dʒ *in* "**J**ames"				*c) Target* dʒ *in* "**J**ackson"			
66-75	4	50 dʒ		50 g	4			100 j	4			100 j
56-65	31	32 dʒ		68 g	31	13 dʒ		87 j	27	37 dʒ		63 j
46-55	146	27 dʒ		73 g	146	16 dʒ		84 j	143	39 dʒ	1 *	60 j
36-45	133	31 dʒ		69 g	133	20 dʒ		80 j	130	55 dʒ	1 ɕ	44 j
26-35	86	23 dʒ	3 *	74 g	86	30 dʒ		70 j	88	53 dʒ	3 dɪ	43 j
16-25	43	60 dʒ		40 g	43	35 dʒ		65 j	43	46 dʒ		54 j
<16	22	59 dʒ		41 g	22	27 dʒ		73 j	21	62 dʒ		38 j
	d) Target θ *in* "**Th**riller"				*e) Target* θ *in* "**Th**atcher"				*f) Target* ð *in* "**th**e World"			
66-75	4			100 t	4			100 t	2		50 r	50 d
56-65	27	59 θ	8 t̪	33 t	29	42 θ	3 s	52 t	30	27 ð	3 d̪	57 d
							3 tɕ				3 r	10 t
46-55	143	34 θ	4 t̪	62 t	147	36 θ	1 tɕ	63 t	146	40 ð	1 r	54 d
											4 *	1 t
36-45	130	54 θ	1 t̪	45 t	133	46 θ	1 tɕ	53 t	134	42 ð	1 d̪	54 d
											3 *	
26-35	88	62 θ		38 t	86	50 θ	1 tɕ	50 t	86	59 ð		41 d
16-25	43	56 θ	2 t̪	42 t	43	44 θ		56 t	42	38 ð	2 h	60 d
<16	21	57 θ		43 t	21	43 θ	5 *	52 t	21	18 ð		82 d
	g) Target əʊ *in* "St**o**ne"				*h) Target* æ *in* "M**a**ggie"				*i) Target* æ *in* "J**a**ckson"			
66-75	4	50 əʊ		50 uː	4	50 æ		50 a	4	25 æ		75 a
56-65	30	83 əʊ		10 oː	29	72 æ	4 *	24 a	27	63 æ		37 a
				7 uː								
46-55	143	83 əʊ	4 *	9 oː	147	88 æ		12 a	143	68 æ	4 ɛ	28 a
				4 uː								
36-45	133	90 əʊ	1 *	3 oː	133	95 æ	1 aː	4 a	130	82 æ		18 a
				6 uː								
26-35	84	92 əʊ		4 oː	86	94 æ	2 aː	2 a	88	78 æ		22 a
				4 uː			2 *					
16-25	43	100 əʊ			43	95 æ	5 *		43	91 æ		9 a
<16	23	100 əʊ			21	95 æ	5 *		21	71 æ		29 a

Table 8: Regional Groups.

Region	Recordings made in
A	Malmö, Trelleborg, Helsingborg, Kristianstad
B	Varberg, Halmstad, Falkenberg
C	Göteborg, Lidköping, Uddevalla, Trollhättan, Skövde
D	Karlskrona, Jönköping, Växjö, Kalmar
E	Linköping, Norrköping, Västervik
F	Visby
G	Eskilstuna, Stockholm, Västerås, Örebro
H	Borlänge, Särna, Mora, Falun, Idre, Orsa, Ludvika, Saxdalen, Nyhammar
I	Hudiksvall, Sundsvall, Örnsköldsvik, Östersund, Skellefteå, Umeå, Boden, Luleå, Piteå

Table 9: Distribution of Swedish subjects' productions of eleven target xenophones (a–k) across **region** and category [%] (cf. Table 4). Cases where the transcribers were unable to hear what was produced are marked by an asterisk.

Region	Tokens	Category			Tokens	Category			Tokens	Category		
		1	2	3		1	2	3		1	2	3
	a) Target dʒ *in* "Ro**g**er"				*b) Target* dʒ *in* "**J**ames"				*c) Target* dʒ *in* "**J**ackson"			
A	36	67 dʒ		33 g	36	22 dʒ		78 j	27	30 dʒ		70 j
B	40	45 dʒ		55 g	40	25 dʒ		75 j	40	45 dʒ		55 j
C	77	23 dʒ		77 g	77	13 dʒ		87 j	74	45 dʒ	1 *	54 j
D	45	41 dʒ		59 g	45	24 dʒ		76 j	45	49 dʒ	2 dɪ	47 j
											2 ɕ	
E	34	41 dʒ		59 g	34	12 dʒ		88 j	34	50 dʒ	3 dɪ	47 j
F	31	29 dʒ		71 g	31	23 dʒ		77 j	32	44 dʒ		56 j
G	75	35 dʒ		65 g	75	29 dʒ		71 j	76	64 dʒ		36 j
H	29	21 dʒ	3 *	76 g	29	21 dʒ		79 j	29	55 dʒ		45 j
I	104	22 dʒ	1 gd	77 g	104	22 dʒ		78 j	105	42 dʒ		58 j
	d) Target θ *in* "**Th**riller"				*e) Target* θ *in* "**Th**atcher"				*f) Target* ð *in* "**th**e World"			
A	27	37 θ		63 t	37	51 θ		49 t	36	44 ð		56 d
B	40	45 θ		55 t	40	55 θ	3 *	42 t	40	20 ð	2 v	78 d
C	74	53 θ	5 t̪	40 t	75	49 θ	2 s	49 t	76	41 ð	3 t	54 d
			2 *								2 *	
D	45	53 θ	9 t̪	38 t	44	52 θ		48 t	45	31 ð		69 d
E	34	47 θ	6 t̪	47 t	33	27 θ	3 ɕ	70 t	31	45 ð	3 v	45 d
											3 h	4 t
F	32	41 θ		59 t	32	41 θ		59 t	30	37 ð	3 d̪	53 d
											3 r	4 t
G	76	59 θ		41 t	73	56 θ	3 ɕ	41 t	75	40 ð	3 *	57 d
H	29	59 θ		41 t	29	31 θ		69 t	29	48 ð	4 t	48 d
I	105	44 θ		56 t	106	24 θ	2 ɕ	74 t	106	41 ð	3 r	53 d
											2 *	1 t
	g) Target əʊ *in* "St**o**ne"				*h) Target* æ *in* "M**a**ggie"				*i) Target* æ *in* "J**a**ckson"			
A	36	97 əʊ		3 uː	37	86 æ	11 *	3 a	27	74 æ		26 a
B	40	85 əʊ	2 uɵ	5 uː	40	92 æ	3 *	5 a	40	68 æ		32 a
			3 oʊ	5 oː								
C	73	82 əʊ		11 oː	75	92 æ	3 aː	5 a	74	77 æ	4 *	19 a
				7 uː								
D	44	89 əʊ	2 oːu	7 oː	44	96 æ	2 aː	2 a	45	78 æ		22 a
				2 uː								
E	34	91 əʊ		3 oː	33	94 æ	3 *	3 a	34	68 æ		32 a
				6 uː				a				
F	31	84 əʊ	3 *	10 oː	32	81 æ		19 a	32	62 æ		38 a
				3 uː								
G	75	92 əʊ		3 oː	73	88 æ		12 a	76	78 æ	1 e	21 a
				5 uː								
H	28	94 əʊ		3 oː	29	97 æ	3 *		29	86 æ		14 a
				3 uː								
I	105	90 əʊ	1 ɔ	7 oː	106	88 æ	1 aː	8 a	105	80 æ	2 ɛ	18 a
			1 oːu	1 uː			1 ɛ					
	j) Target aʊ *in* "c**o**wboy"				*k) Target* ɬ *in* "**Ll**oyd"							
A	42	57 aʊ	5 ɔʊ	38 ɔ	42	31 ɬ		69 l				
B	41	27 aʊ	5 ɔʊ	63 ɔ	41	19 ɬ		81 l				
			2.5 oɪ									
			2.5 a									
C	80	16 aʊ	4 ɔʊ	73 ɔ	81	20 ɬ	1 u	79 l				
				6 oː								
D	49	41 aʊ	6 a	51 ɔ	49	16 ɬ		84 l				
			2 ɔʊ									
E	34	35 aʊ	3 oɪ	62 ɔ	34	12 ɬ		88 l				
F	33	33 aʊ	3 ɔʊ	55 ɔ	33	15 ɬ	6 l l	79 l				
			3 oɪ	3 oː								
			3 a									
G	77	38 aʊ	1 a	61 ɔ	77	5 ɬ	1 v	94 l				
H	30	30 aʊ	7 a	60 ɔ	30	10 ɬ		90 l				
				3 oː								
I	107	32 aʊ	2 a	62 ɔ	107	22 ɬ		78 l				
			1 aw	4 oː								

Table 10: Distribution of Swedish subjects' productions of twelve target xenophones (a–l) across **level of education** and category [%] (cf. Table 4). Cases where the transcribers were unable to hear what was produced are marked by an asterisk. The subjects were asked to write down their highest level of education prior to the recording sessions. These were then tabulated into four different classes that aimed to normalize for differences in the educational system that have taken place during the decades from when the oldest subjects went to school until the youngest subjects had their education. The four different classes roughly translate as: CLASS 1: 'folkskola', roughly corresponding to 7 years of school; CLASS 2: 'realskola' or 'grundskola', roughly corresponding to 9 years of school; CLASS 3: 'gymnasium', roughly corresponding to 11–13 years of school; CLASS 4: university (or equivalent) education.

Edu-cation	Tokens	Category			Tokens	Category			Tokens	Category		
		1	2	3		1	2	3		1	2	3
	a) Target dʒ *in* "Ro**g**er"				*b) Target* dʒ *in* "**J**ames"				*c) Target* dʒ *in* "**J**ackson"			
1	18	33 dʒ		67 g	18	11 dʒ		89 j	18	11 dʒ		89 j
2	109	27 dʒ		73 g	109	9 dʒ		91 j	109	9 dʒ		91 j
3	253	36 dʒ	1 *	63 g	253	22 dʒ		78 j	253	22 dʒ		78 j
4	51	38 dʒ		62 g	53	38 dʒ		62 j	53	47 dʒ	2 dj	51 j
	d) Target θ *in* "**Th**riller"				*e) Target* θ *in* "**Th**atcher"				*f) Target* ð *in* "**th**e World"			
1	18	28 θ	5 tç	67 t	18	17 θ		83 t	19	32 ð	5 r	58 d
											5 t	
2	109	27 θ	3 ʈ	70 t	109	35 θ	3 tç	61 t	109	33 ð	2 ɖ	57 d
							1 *				3 r	4 t
											2 *	
3	253	53 θ	2 ʈ	45 t	253	42 θ	1 tç	57 t	253	38 ð	1 r	60 d
											1 *	
4	53	60 θ	2 ʈ	38 t	53	47 θ	2 s	51 t	53	55 ð	2 h	43 d
	g) Target əʊ *in* "St**o**ne"				*h) Target* aʊ *in* "c**o**wboy"				*i) Target* ɬ *in* "**Ll**oyd"			
1	18	67 əʊ	6 ɪuə	17 oː	18	5.5 aʊ		89 ɔ	18	11 ɬ		89 l
			4 *	6 uː				5.5 oː				
2	109	86 əʊ	1 ɔ	6 oː	108	22 aʊ	2 ɔʊ	68 ɔ	101	13 ɬ	1 l l	86 l
				6 uː			2 oɪ	5.5 oː			1 eɪ	1 a
							1 a				1 ø	
3	253	91 əʊ	1 *	4 oː	253	36 aʊ	3 a	56 ɔ	253	20 ɬ	0.5 l l	79 l
				3 uː			3 ɔʊ	1 oː			0.5 u	
							0.5 oɪ					
							0.5 aw					
4	53	96 əʊ		4 oː	53	53 aʊ	2 a	41 ɔ	51	25 ɬ	1 v	74 l
								4 oː				
	j) Target aː *in* "**Aa**chen"				*k) Target* aː *in* "B**a**den-Baden"				*l) Target* aː *in* "Baden-B**a**den"			
1	18		11 æ	72 a	18	6 aː		89 ɑː	18	6 aː		89 ɑː
			6 aʊ	11 ɑː				6 a				6 a
2	101	8 aː	15 æ	41 a	101	13 aː	2 æ	82 ɑː	101	12 aː	1 eɪ	86 ɑː
			6 eɪ	25 ɑː			1 eɪ	1 a				1 a
			5 æː				1 ø					
3	246	12 aː	10 eɪ	36 a	246	15 aː	2 æ	81 ɑː	246	15 aː	1.6 æ	82 ɑː
			9 æ	26 ɑː			1 æː	1 a			0.4 æː	1 a
			5 æː									
			1 oː									
			1 *									
4	51	16 aː	8 eɪ	39 a	51	20 aː	2 æː	73 ɑː	51	20 aː	2 æː	73 ɑː
			6 æ	27 ɑː				6 a				6 a
			2 oː									
			2 an									