

# Xenophenomena: studies of foreign language influence at several linguistic levels

Anders Lindström†‡ and Robert Eklund†‡

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

† Telia Research AB, SE-123 86 Farsta, Sweden

‡ Dept. of Computer Science, NLPLab, Linköping University, Sweden

## Introduction

Languages have always been influenced by other languages in various ways, through cultural contacts, migration, trade and other channels. In an increasingly internationalized world, where contacts across national borders are commonplace, sometimes politically driven/pushed by bodies such as the EU, foreign language influences have become stronger than ever. Moreover, besides cultural influx through media such as TV and radio, multilingual automatic applications have become an important area of study for automatic speech recognition services, raising issues like how Germans pronounce French place names, and vice versa (Trancoso et al., 1999). Similarly, automatic speech synthesis also needs to cover pronunciation of foreign items, which has been observed by e.g. Eklund & Lindström (1996; 1998; 2001) and Möbius et al. (1997). While speech recognition and speech synthesis mainly are affected by “foreignness” of speech *sounds*, languages are also influenced at other linguistic levels, such as vocabulary, idioms such as ‘catch-phrases’ and ‘buzz-words’, translated or original expressions and so on, as observed by e.g. Ljung (1988).

## Xenophonology

In order to investigate the degree and character of phone set expansion in Swedish, Eklund & Lindström (ibid.) analyzed data from 491 subjects balanced over age, gender and educational level. A total of 5,400 sentences and 23,750 phone tokens were recorded on location at 40 different places in Sweden. The results indicated that Swedes to a very high degree expand their phone sets when uttering foreign words and names (in this case mainly of English origin; to a minor degree of German and Dutch). It was also found that almost no one resorted to full accommodation to “corresponding” Swedish sounds. The status of these sounds is cumbersome. They are clearly not *phonemes*, since these sounds never change the meaning of a word. Moreover, they cannot be labeled *allophones*, given that there are no underlying phonemes. Consequently, Eklund & Lindström (1998) suggested the term *xenophones* to describe this type of sounds.

## Xenomorphology

In earlier attempts aimed at proper name pronunciation in Swedish, a morphological approach has proven productive (Carlson, Granström & Lindström, 1989). More recently, Lindström & Kasaty (2000) made an initial attempt to implement a lexical component capable of handling what might be termed xenomorphs in a two-level formalism. They used a corpus of youth’s speech that featured a number of different foreign traits of English, French and German origin. Lindström & Kasaty (2000) concluded that a lexical component designed to cope with current Swedish including “foreign” language elements has to be capable of handling: *a)* Expansion of the speech sound repertoire to include ‘xenophones’ (non-Swedish sounds); *b)* Foreign roots with Swedish inflections; *c)* English roots in Swedish derivations and compounds; *d)* Interaction between foreign items and Swedish prosody. Lindström & Kasaty then used the PATR-II extensions available in PC-KIMMO (Antworth, 1990) to model Swedish morphotactics in a word grammar using unification over features such as paradigm, gender and number. In Eklund & Lindström (1998) it was demonstrated (with synthesis examples in the electronic proceedings) how this technique transfers directly into a Swedish concatenative demi-syllable synthesis. The problem of paradigm assignment in the case of foreign morphs is handled in a way which was inspired by Linell (1977), and Lindström & Kasaty (2000) therefore used a rhyming principle in assigning “Swedish” paradigm codes to the foreign morphs.

## Attitudes

The entailing issue here is of course how people react to different degrees of accommodation or “foreignness” of xeno-items. Ljung (1998) reports that negative attitudes vis-à-vis foreign expressions at all levels (pronunciation, idioms, translated idioms) correlates with a high degree of education. This observation runs counter with the data presented in Eklund & Lindström (ibid.), where high education correlated with a high degree of xenophone inclusion. One possible explanation for this discrepancy is that while Ljung (ibid.) based his observations on subjects who were aware of what they were tested for, Eklund & Lindström (ibid.) analyzed data from subjects who were unaware of the objective of the task. Thus, it could be assumed that, when explicitly asked, speakers with a high degree of education consider it politically correct to guard one’s native tongue against foreign influence, while the same subjects, given the said high education, are more prone to source language renderings of foreign items.

## Discussion

As we have seen, xenophone production seems to be strongly governed by educational level, at least in Swedish users. There are also results by Bayard and Sullivan (2000) showing that Swedish listeners are indeed able to identify social status and even recognise and distinguish different varieties and accents of English. On the perceptual side, one can therefore conclude that there is strong reason to suspect that people should be quite good at perceiving differences in xenophone usage, possibly linking that to identification along social/educational strata. In the production study presented, differences in xenophone usage were found, and the distribution differed across xenophones and seemed to depend on such factors as the identity of the lexical items in question and their particular spelling. As mentioned in Eklund & Lindström (ibid.), a number of underlying factors are surely involved in the exploitation of foreign items, including: *a*) The speaker’s competence and performance capabilities with respect to the foreign language; *b*) The speaker’s expectations of the listener’s competence; *c*) The relative social status of the speaker and listener; *d*) The time the word or name first appeared; *e*) The socio-cultural distance to the foreign country in question; *f*) The recency and frequency of occurrence of the name/word in both language concerned; *g*) Similarities and dissimilarities between the languages concerned. This makes xenophenomena a multidimensional problem in the truest sense of the word.

## References

- Antworth, E.L. 1990. *PC-KIMMO: a two-level processor for morphological analysis*. Summer Institute of Linguistics, Dallas, 1990.
- Bayard, D. & K.P.H. Sullivan. 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, 24–26 May 2000*, University of Skövde, pp. 33–36.
- Carlson, R., B. Granström & A. Lindström. Predicting name pronunciation for a reverse directory service. *Proc. Eur. Conf. Speech Technology*, vol. 1, pp. 113–116.
- Eklund, R. & A. Lindström. 2001. Xenophones: An investigation of phone set expansion in Swedish and implications for speech recognition and speech synthesis. *SPEECH COMMUNICATION* 35, pp. 81–102.
- Eklund, R. & A. Lindström. 1998. How to handle “foreign” sounds in Swedish text-to-speech conversion: Approaching the ‘xenophone’ problem. *Proc. International Conference on Spoken Language Processing*, Sydney, 30 November–5 December, 1998, vol. 7, pp. 2831–2834.
- Eklund, R. & A. Lindström. 1996. Pronunciation in an internationalized society: A multi-dimensional problem considered. *Proc. Fonetik 96, The Swedish Phonetics Conference, Nässlingen, 29–31 May 1996*, pp. 123–126. TMH-QPSR 2/1996.
- Lindström, A. & A. Kasaty. 2000. A two-level approach to the handling of foreign items in Swedish speech technology applications. *Proc. International Conference on Spoken Language Processing*, Beijing, 16–20 October, 2000, vol. 1, pp. 54–57.
- Linell, P. 1977. Knilpor, mejder och snulkar – ett experiment med pluralbildning i svenskan. In Otterbjörk, R. & Sjöström, S. (Eds.): *Provinsiellt och universellt. En vänskrift till Karl-Hampus Dahlstedt*. Publications from the Dept. of General Linguistics, No. 12, University of Umeå, 1977.
- Ljung, M. 1988. *Skinheads, hackers och lama ankor*. Bokförlaget Trevi, Stockholm 1998.
- Möbius, B., R. Sproat, J.P.H. van Santen & J.P. Olive. 1997. The Bell Labs German text-to-speech system: An overview. *Proc. ESCA, Eurospeech 97*, Rhodes, Greece, vol. 5, pp. 2443–2446.
- Trancoso, I., C. Viana, I. Mascarenhas & C. Teixeira. 1999. On deriving rules for nativised pronunciation in navigation queries. *Proc. ESCA, Eurospeech 99*, Budapest, Hungary, vol 1., pp 195–198.