

# English Influence on Swedish Word Formation and Segmentals

ANDERS LINDSTRÖM

## *Abstract*

The object of this study is to determine the nature and degree of English influence on contemporary spoken Swedish at the phonological and morphological levels. Particular attention is paid to certain contact-linguistically critical situations of importance for speech technology applications. Several corpora of spoken and written Swedish are analyzed regarding segmental and word formation effects. The results suggest that any description of contemporary spoken Swedish (be it formal, pedagogical or technical) needs to be extended to cover both phonological and morphological material of English origin. Previous studies of socio-linguistic and other underlying factors governing the variability observed have shown that education and age play a significant role, but that individual variability is large. It is suggested that interacting phonological constraints and their relaxation may be one way of explaining this. An attempt at a diachronic comparison is also made, showing that morphological processes involving English material are frequent in contemporary Swedish, while virtually non-existing in two corpora of spoken Swedish from the mid 1960s. Requirements for speech technology applications are also discussed.

## *1. Introduction*

Contact between cultures is undoubtedly one of the strongest driving forces behind linguistic change and evolution. This has led to bi- or even polylingualism constituting the normal linguistic situation in large parts of the world (Ladefoged and Maddieson, 1996). In contrast to this, most native speakers of the Scandinavian languages belong to

largely monolingual cultures, even if this picture is changing as a result of immigration. It is therefore especially interesting to study the effects that the dominant position of the English language in the media, entertainment and certain technical domains has had on a language such as Swedish.

In the following paragraphs, a background on spoken vs. written language, and on the influence of English on other languages, is first given, together with remarks on some of the consequences this has for certain types of speech technology applications. The second section reports on a study of the segmental aspects of English influence on Swedish, and the third section investigates word formation aspects of such influence, using data from several spoken and written language corpora. Finally, the results regarding both segmental and word formation aspects are summarized and discussed.

### 1.1 Spoken vs. Written Language

Spoken language differs fundamentally from written language in several obvious ways. First of all, spoken language constitutes a primary means of communication, acquired in early childhood by virtually everybody, whereas the skills of reading and writing need to be actively learnt at a much later developmental stage. The latter are therefore not equally well-spread even within languages that actually do have a writing system (which many lack). This is partly also due to social and economic factors.

Secondly, speech is evasive by its very nature, since it is being conveyed by an acoustic signal, and therefore cannot be erased or edited by the speaker once it has been produced. Due to this fact, spoken and written language actually differ in many linguistic respects, e.g. syntactically, regarding structure, length and frequency of phrases, in terms of vocabulary, etc. Since speech occurs in real time, as it were, the planning of spoken utterances also needs to be done on-line, often while speaking, which is one reason (among several) for the hesitations, restarts, repairs and other disfluencies which are typical of spontaneous speech (Brennan, 2000; Bell, Eklund and Gustafson, 2000). Since spoken language is normally situated in a real-world setting (or in the case of telecommunications, in two or more), the types and frequency of deictic expressions are also fundamentally different from those of written language (Jungbluth, 1999). The normal style of speech is interactive and conversational, except when reading aloud, reciting monologues etc., while producers and consumers of written language are connected across both time and space but are in a sense always off-line. In fact, the need to record

and document economic and other agreements for future retrieval in ledgers, legends and laws is probably what initially drove the development of different writing systems (Sampson, 1985). In a historical perspective, it has only very recently become possible to record the speech signal *per se*, and as such, it has turned out to be remarkably difficult to analyze, primarily because of its apparent variability. Some of this acoustic variability is possible to see as redundancy, sometimes needed to overcome noise or other obstacles in the communicative chain between speaker and listener (Shannon, 1948), while this mismatch between the continuous and variable signal and the categorically perceived discrete linguistic objects (speech sounds, words, etc.) still remains a largely unsolved but challenging paradox for the phonetic sciences (Lindblom, 1990). Thirdly, and partly related to the previous point, it is quite clear that one of the primary functions of spoken language, and perhaps what biologically drove its initial development, is to act as the social “glue” required to establish and maintain within-group cohesion and common ground but also hierarchical relationships among members of the group. Possibly as a result of this, the speech signal conveys overlaid information regarding the speaker’s identity, size, sex, age, mood, emotional state etc., which is vital in human–human interaction. The same prosodic signal parameters also provide means of conveying information which makes it possible to interpret the meaning of utterances in context at several linguistic levels, from segmentals to pragmatics. It is probably no co-incidence that colonizing nations have always sought to endow their own language and cultural habits upon the societies being colonized.

This hard coupling between spoken language variety and cultural, national or group identity is of course also what lies behind the concern about questions relating to official policies regarding language, as exemplified e.g. by the current debate regarding language use in the European Community (Phillipson, 2003). It is also the driving force behind the formation of language academies and other cultural institutions at national but also other levels, with the explicit purpose of maintaining and developing a specific language or specific variety of a language, sometimes interpreted as a mission to preserve rather than develop. In practice, such academies have often directed their work towards publishing guidelines or in other ways giving advice on what should be considered “proper” use of the written mode of the language in question. While this may of course also affect spoken language, and has indeed been shown to do so to a certain extent (Teleman, 2004), it does so in an indirect and secondary way. Because of

these fundamental differences between the spoken and the written modes of language, it can be expected that the type of crosslanguage influence investigated here should show more extensive, qualitatively different and presumably earlier effects on spoken than on written language. In view of this, the aim of the investigations reported on here is to determine the nature and degree of English influence on contemporary spoken Swedish at the phonological and morphological levels. This is partly done with the purpose of gaining knowledge of importance for the design and development of speech technology applications. Hence the perspective is descriptive and the approach employed empirical.

## 1.2 English Influence in a Global Perspective

The establishment over the past century of the English language as a kind of a global *lingua franca* (to label it using a predecessor) is indisputable (Melchers and Shaw, 2003). McArthur (1996) reports that English is used as a native language (ENL) in 36 territories, as a second language (ESL) in 51, and as a foreign language (EFL) in 141 of the 228 territories he lists. The EFL category is further divided into two groups, one, comprising 121 territories, where English is “learned as the global lingua franca” and another, comprising 20 territories (among them Denmark, Norway, Sweden and the Netherlands), where it is “virtually a second language”. Improved foreign language education is therefore one of the major factors that lie behind the influence that English has had on many languages of the world. Other factors include increase in technical mobility and ease of global access e.g. through broadcast media and the Internet, sub-titling rather than dubbing foreign language films etc. One of the most important factors is probably the strong association between youth culture, music, entertainment, movies etc. on the one hand, and the English language (in its different varieties) on the other. It can be assumed that this coupling is what underpins the use of certain words and phrases with the socio-linguistic purposes of displaying group identity and establishing common ground.

The effects of language contact occur at virtually all linguistic levels, from phonology to pragmatics (Weinreich, 1953), but the degree as well as the rate of integration of English elements in different languages seems to vary as a function of exposure, linguistic structure, attitude, foreign language education etc. These issues have therefore received attention within several areas of linguistics during at least the past century (Jespersen, 1902), spanning such diverse areas as historical linguistics,

second language acquisition, generative phonology and Optimality Theory (OT). The lexicographically oriented project *The English Element in the European Languages* (Görlach, 2001) recently charted anglicisms in 20 different European languages<sup>1</sup> and published a number of dictionaries, including one for Swedish (Antunović, 1999). Filipović (1996) concludes that the role of English as a word donor to other languages has literally exploded from the middle of the 20th century and onwards, and gives examples of how these imported elements have also affected the linguistic system of the borrowing languages, e.g. by expanding the repertoire of allowable final consonant clusters in Croatian, by adding the velar nasal phoneme to French to allow for word forms ending in /ŋ/ and by extending the Russian use of non-palatalized consonants to certain phonological positions in anglicisms, where normally the corresponding palatalized consonant would be obligatory. Filipović also notes prosodically related effects, for instance how vowel reduction in unstressed syllables, which is another highly characteristic property of Russian, is inhibited in anglicisms.

Several phonological theories have sought to model and explain the patterns found in this type of language contact. For instance, earlier ideas regarding the “markedness” of universally infrequent speech sounds and their phonotactic combinations were taken further within Natural Phonology (Donegan and Stampe, 1979), which claims that certain processes are simply more natural than others, which would explain e.g. why devoicing of final obstruents occurs in loanwords even in some languages lacking such segments in that position. In a similar vein, some of the allegedly universal constraints proposed within OT (Prince and Smolensky, 1995) also originate from typological studies of interference and assimilation phenomena due to language contact. Constraints requiring fidelity towards the underlying form are generally in conflict with constraints related to criteria for “well-formedness” of the surface forms of the language, and especially so in the case where the underlying form originates in a foreign, donor, language. This may explain differences between languages in dealing with foreign loans, but also the apparent stratification of the vocabulary of certain (or even most) languages into “native vocabulary”, “assimilated loans” and “foreign vocabulary”, as claimed by Itô and Mester (1999). This reasoning also illustrates how it is in fact quite difficult to define

---

<sup>1</sup> Among the languages covered are Albanian, Bulgarian, (Serbo-)Croatian, Danish, Dutch, Finnish, French, German, Greek, Hungarian, Icelandic, Italian, Norwegian, Polish, Romanian, Russian, Spanish and Catalan.

the notions of “native” vs. “foreign” and also how that distinction, should it be possible to define, is more or less bound to change over time.

### 1.3 Cross-linguistic Issues in Spoken Language Applications

Cross-linguistic issues have recently received increasing attention within several research areas related to spoken language dialogue applications. As Billi (n.d.) points out, the development and deployment of some of the commercially most interesting services, such as automatic train time table information (Billi and Lamel, 1997), stock market information, directory services (e.g. Carlson, Granström and Lindström, 1989; Spiegel, Macchi and Gollhardt, 1989), and call routing (Gorin, Riccardi and Wright, 1997), all evoke various multi- and cross-lingual issues, e.g. the handling of non-native speech, native speakers' handling of non-native items and names, etc. However, these applications serve to put the search-light not only on the technical side of those issues: choosing the appropriate allophone set for a speech synthesizer or recognizer is not just a matter of selecting a particular set of phonetic glyphs, but also brings up the question of how to describe the phonological system of a language in contact with another, and what factors should be considered when making such a design choice. From a purely technical point of view, foreign features at different linguistic levels would not constitute much of a problem even if they were frequent, as long as they were fossilized and could be listed as exceptions in one way or the other. However, it is quite clear that foreign traits are often assimilated or integrated into the receiving language in such a way that they also attain generative properties, e.g. in word-formation, adding considerably to the complexity of analyzing new words in both spoken and written language (Lüdeling and Schmid, 2001).

Furthermore, the attitudinal and sociolinguistic factors governing the degree of assimilation when dealing with non-native linguistic elements have quite different consequences for different types of spoken language applications. In the case of Automatic Speech Recognition (ASR), they contribute to increased variability, which is a well-known problem, studied technically for several decades. This problem also has well-established practical solutions in the case of dictation or transcription to text. In many other cases, e.g. in language learning applications, this type of variability instead constitutes a major challenge and a fundamentally unsolved problem, since the focus of such applications is on how something is pronounced, and on identifying the borderline between acceptable and unacceptable pronunciations, rather than on recognizing the identity of the words uttered.

In the case of Text-to-Speech (TTS) conversion, there is a need to be able to choose and generate the proper pronunciation variant, which poses particular problems, and particularly critical, when dealing with proper names (Carlson et al., 1989; Spiegel et al., 1989). This is an increasingly relevant issue, as persona design has become an increasingly important issue in spoken language interfaces, e.g. in speech-based call routing applications, and in interaction with multimodal, embodied animated characters in new types of computer games (Gustafson, Bell, Boye, Lindström and Wirén, 2004).

#### 1.4 English Influence on Swedish

As regards spoken Swedish, Elert (1971) noted that different degrees of nativization (from “authentic” via “re-phonematized” to “spelling-oriented”) can take place at the segmental level when dealing with a foreign lexeme or morpheme.

During 1981–1985, surveys of some 2,000 informants’ attitudes towards English loans and preferences regarding wording and grammatical constructions were made in the project *Engelskan i Sverige (EIS)* (Ljung, 1986). Recordings were made of a smaller number of subjects, reading sentences including fairly frequent English loans as well as names of (at the time) well-known athletes and politicians. Demographic data was also collected, and the recordings were labelled as either adhering to the “native” (English) pronunciation, or to some sort of Swedish approximation. Results showed that socio-linguistic factors, e.g. educational level, affected both attitudes and performance, but that even well-educated subjects make use of conventionalized adaptations to Swedish, e.g. substituting [s] for [z] when the English spelling used <rs> for the latter. In general, the younger and well-educated subjects had a more positive attitude towards anglicisms, and also used them to a higher degree. Frequency of anglicisms in newspapers was also studied, and found to be in the order of .3%.

In a recent study by Sharp (2001), which also includes a good overview of the literature on English influence on Swedish, “code-switches” of English origin in two corpora of spoken Swedish were compared. One corpus was based on recordings made at business meetings in an international shipping company, while the other consisted of casual conversation of young adults drawn from a televised reality show. Sharp found differences between the two corpora in a number of respects, including frequency of occurrence, prosodic signalling and degree of integration or accommodation.

## 2. *English Influence on Swedish Segmentals*

### 2.1 The Xenophones Production Study

Eklund and Lindström's (2001) production study was based on recordings made in 1995–1996 of nearly 500 subjects, aged 15–75, who read approximately one hour of computer-prompted sentences each (Eklund and Lindström, 1996). The primary purpose of the recordings was to collect training material to improve speech recognition, which all subjects were informed about, and part of the prompted material was the same for all subjects. Included in that section were a dozen sentences with well-known foreign names and words, most of which were English. As illustrated by Example 1, a dozen sentences with wellknown foreign names and words were included in that section. These sentences took about one minute for each subject to read, and the subjects were therefore probably not aware of the specific object of this study.

(1.)

Veckopressens favoriter är verkligen Diana och Charles.  
The tabloids' favourites are indeed Diana and Charles.  
Diana and Charles are indeed the favourites of the tabloids.

The recordings were transcribed by phonetically trained native speakers of Swedish, and a common subset of the transcriptions was later crosschecked for inter-transcriber consistency. For each subject, every allophonic transcription in 33 target positions (like the ones indicated by curly brackets in Example 2) within the 12 sentences was then assigned to one of three different categories along an axis, ranging from near-source-language (CATEGORY I) via partly accommodated (but clearly not “Swedish”) (CATEGORY II) to re-phonematized (CATEGORY III). Through this process, approximately 23,750 manually transcribed and classified tokens were collected.

(2.)

Veckopressens favoriter är verkligen D{i}{a}na och {Ch}arle{s}.

Detailed tabular results have been provided by Eklund and Lindström (2001), but to summarize some of the main findings, nearly all subjects either used or made an attempt to use “foreign” speech sounds (neutrally termed



*xenophones*). For the present study, 15 target instances in the names of English origin were selected. Just as in Ljung's (1986) study, the frequency of occurrence differed considerably across different lexical items and different positions within words or phrases, even regarding the "same" sound, as can be seen in Figure 1. Almost all vowel segments featured pronunciations close to that of the source language, notably also the diphthong [əʊ], which does not resemble any native speech sound in Swedish. The consonant segments [w], [z] and [ʒ] were produced by most subjects as [v], [s] and [ʃ], respectively. Of these, [w] occurred word-initially, while the fricatives occurred in medial position (with one exception, occurring word- finally). On the other hand, virtually all subjects rendered [tʃ], and quite a few also [dʒ], in a fashion very close to the source language. A more scattered distribution along the near-native-to-re-phonematized axis was displayed by e.g. [ð] and [θ], which were often replaced by the corresponding stops. These results largely confirm those of Ljung (1986), even if no direct comparison of the production data was possible to make, since Ljung's subjects were probably aware of the object of study, and the labelling conventions and instructions also differed between the two studies.

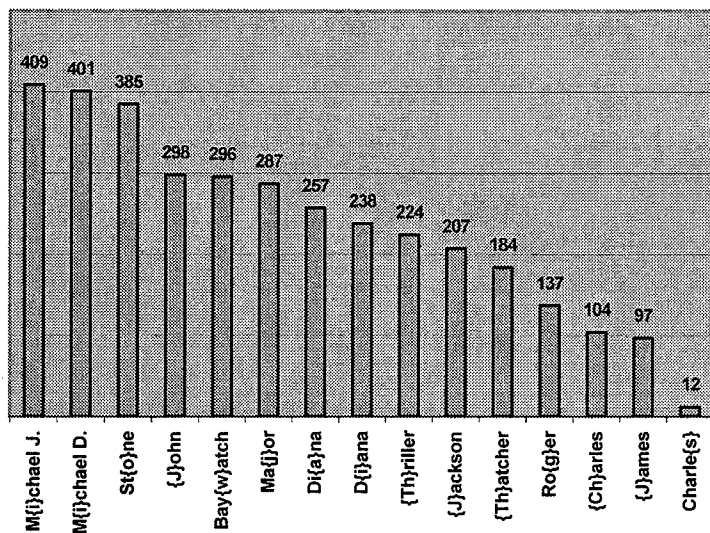


Figure 1: Number of subjects (out of 422) with non-rephonematized productions (corresponding to CATEGORIES I and II in the section on the "Xenophones" study) for 15 potential xenophone instances, positionally indicated by curly brackets in the

*orthography (e.g. counting the first segment in "Charles" as one instance, the final speech sound in the same name as another).*

As regards explanatory underlying factors, education and age showed significant effects (Pearson chi-square, two-sided), in the sense that higher education yielded a larger share of near-English pronunciations. That share was also significantly higher for subjects between 25 and 45 years of age. These effects are illustrated in Figure 2, which shows the dimensionless ratios  $r_{edu}$ ,  $r_{gen}$ , and  $r_{age} \in [0,1]$  for education, gender and age, respectively. What seems to be a slight inclination towards higher degree of xenophone production for female subjects can also be observed. The effects of educational level are shown in more detail in Figure 3, where subjects have been divided into three levels of education and results are shown for the same 15 instances. In all but two cases, subjects with *Low* education (up to 9 years of school) produced the smallest share of CATEGORIES I and II productions. It also seems to be the case that the differences between educational groups is very small for segments which were produced using xenophones (or good approximations) by a very large number of the subjects (e.g. the first vowel in "Michael" and the vowel in "Stone"). The difference seems to get accentuated as the overall number of xenophone productions goes down.

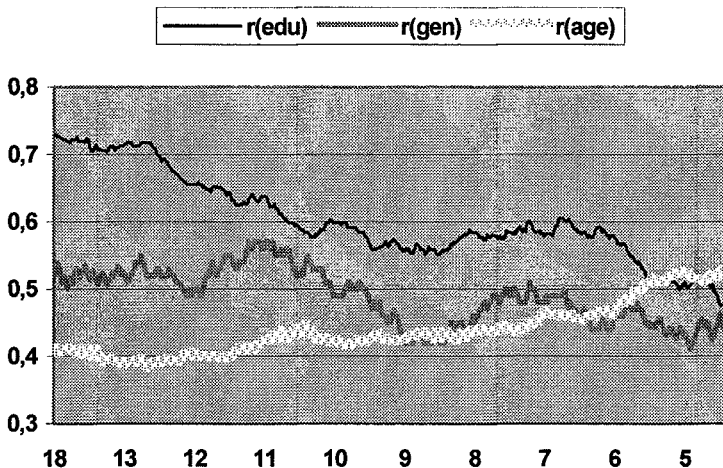
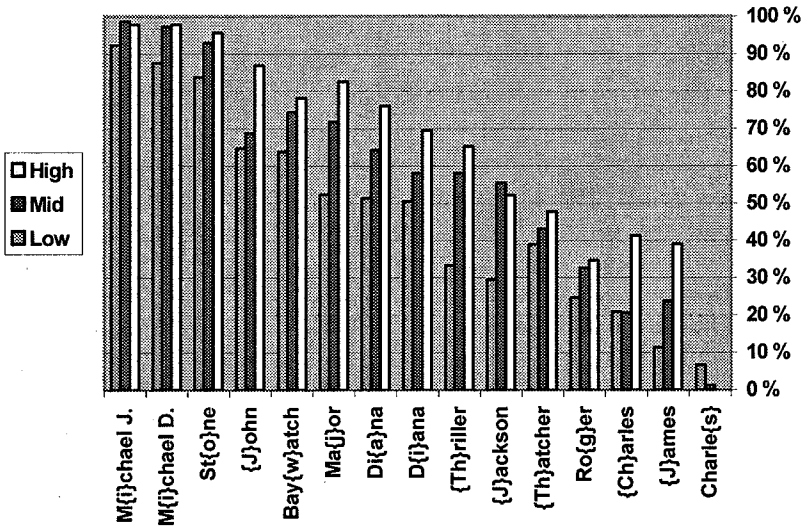


Figure 2: Education, gender and age ratios, normalized to the interval [0,1], and plotted as a smoothed average (over  $\pm 50$  data points) for 422 subjects against each subject's number of non-rephonematized productions (corresponding to CATEGORIES I and II in the section on the "Xenophones" study) out of the 15

*xenophone instances in Figure 1. The interval endpoint "1" on the dimensionless ordinate axis stands for high education, female gender and high age, respectively.*



*Figure 3: Percentage of subjects per educational level with nonrephonematized productions (corresponding to CATEGORIES I and II in the section on the "Xenophones" study) plotted for the 15 xenophone instances in Figure 1. The educational level is coded as Low ( $\leq 9$  years of school), Mid (10-13 years of school) or High (University education).*

### 3. English Influence on Swedish Word Formation

As pointed out for instance by Schmid, Lüdeling, Säuberlich, Heid and Möbius (2001), it is the productive nature of non-native elements which calls for their modelling, both from a linguistic and a computational point of view. In order to investigate such properties, e.g. using statistical methods (Baayen, 1992), extensive (and transcribed) corpora of spoken Swedish would be called for, but unfortunately such resources do not exist in abundance. In the following sections, some corpora of both spoken and written Swedish that are of some relevance to the problem of word formation are presented, analyzed and discussed in more detail. However, it should be borne in mind that, with the exception of the corpus presented in

Section 3.3, these corpora were originally collected and compiled for entirely different purposes by other researchers.

### 3.1 A Written Corpus of Elicited Slang

In her studies of the language of Swedish adolescents, Kotsinas (2000) carried out an experiment where 2,000 youngsters, aged 15–19, were given 55 Swedish keywords and asked to write down as many alternative words for each keyword as possible. Even if this corpus uses the written mode, it can be expected to reveal at least some things regarding the spoken language of the informants, due to these instructions. On the other hand, apart from the obvious restriction in domain, induced by the relatively small set of keywords, the results may be both limited by, and sometimes, as noted by Kotsinas herself, apparently also detrimentally “enhanced” by the imagination of some of the subjects.

Some 17% of the answers were judged to be of (recent) English origin, the most frequent when normalized for spelling differences being: *party, facefejs, cash, crazy, cool, gay, boring, happy, sure, strange, ugly, fatso, babe, kid, money, bull, nice, super, bitch, chicken, loser, bimbo, hip, and scared*. Kotsinas also found that elements of English origin combined quite freely with Swedish elements, at least in compounding, producing forms such as *asboring, stenugly* and *råstrange*, where the Swedish words “as” (carcass/cadaver), “sten” (stone) and “rå” (raw) are used as generic reinforcement adverbial particles in combination with the English adjectives.

### 3.2 TV show transcriptions

As part of the project *Samtalspråkets grammatik* (Grammar in conversation: A Study of Swedish) (GRIS) (Nordberg, 1999), an episode of “Tryck Till”, a music program primarily targeted at adolescents, was transcribed by Öqvist (2000a) for the purpose of discourse particle analysis (Öqvist, 2000b). Even if the recording situation is rather special, the language is not scripted, and does not seem to have been inhibited by the TV studio setting. Examples 3–6 below serve to illustrate how both simplex English words and complex nominals take part in compounding with Swedish nouns.

(3.)

girlpowersparkar i fejset  
girl power kicks to the face  
“Girl Power” kicks to the face

(4.)

precis som i wannabe-videon så  
exactly as in the wannabe video  
Exactly as in the Wannabe video

(5.)

careless whisper videon  
(the) careless whisper video  
the “Careless Whisper” video

(6.)

de gör e att nån slags (eh) roadmovie westerngrej  
they are doing, are, to, some kind of (er) road-movie western thing  
they are doing some kind of road-movie/western thing

*Transcriptions by Öqvist (2000a)*

*Glosses and translations by the present author*

Swedish prosodic conventions regarding compounding require that the primary stressed syllable of the initial constituent remains stressed (with any prior unstressed syllables regarded as extra-metrical), that the primary stressed syllable of the final constituent receives secondary stress, and that all intervening syllables are demoted in terms of stress (Bruce, 1998). Our speaker therefore needs to re-analyze the original English stress patterns and internal prosodic hierarchy of “Careless Whisper” and “Girl Power”, respectively, in order to come to the conclusion that the first syllable of “Whisper”, not of “Careless Whisper”, should become the initial primary stressed syllable of the Swedish compound. On the other hand, the original initial stress pattern of “Girl Power” is natural to retain when forming the Swedish compound. Example 7 shows unassimilated use of the English adjective “catchy”, whereas the adjectival present participle form “coachande” in Example 8 is the result of ordinary (productive) Swedish derivational processes. Despite this, a pronunciation with [əʊ] is used in this TV show (and is probably more or less the required one in most dialects/sociolects).

## English Influence on Swedish Word Formation and Segmentals

(7.)

jävlit catchy låt  
bloody catchy tune  
bloody catchy tune

(8.)

om du får säga några coachande ord till mej och till Jonna och John  
if you are allowed to say a few encouraging words to me and to Jonna and John  
if you are allowed to say a few encouraging words to me and to Jonna and John

*Transcriptions by Öqvist (2000a)*

*Glosses and translations by the present author*

Examples 9–10 serve to illustrate how “Mr. Latino Lover” receives Swedish nominal definiteness inflection (by adding the morpheme /n/) when required by a later co-reference situation.

(9.)

(mmm) dom fortsätter in i gränden där sen (eh) mister latino lover och  
(mmm) they continue into that alley there later (er) mister latino lover and

den här damen då som han fick på fall med sina magiska ögon  
this lady then whom he made fall for his magic eyes  
later, they continue into that alley, that is, Mr. Latino Lover and this lady who had fallen  
for his magic eyes

(10.)

ja som e så svårflörtad till å me ja kom på (ehhh) fall av den här latino lover  
I, who am so hard to get even I came on (er) fall by this latino lovern  
even I, being quite hard to get, fell for this Latino Lover

*Transcriptions by Öqvist (2000a)*

*Glosses and translations by the present author*

Finally, Example 11 shows almost idiomatic use of an English phrase in the midst of otherwise “Swedish” material.

(11.)

ämen keep up the good work ä våga slåss [för de ni älskar]  
 now, keep up the good work and don't hesitate to take a fight for the things you love  
 now, keep up the good work and don't hesitate to take a fight for the things you love/cherish

*Transcriptions by Öqvist (2000a)*

*Glosses and translations by the present author*

### 3.3 Written Dialogue from the Rocky Comic Strips

Lindström and Kasaty (2000) analyzed 415 strips taken from the Swedish comic *Rocky* by Martin Kellerman, who claims that the characters in the comic as well as their language are entirely based on himself and his friends, all adolescents in central Stockholm in the mid- to late 90s. Even if such an introspective statement can (and should) be questioned, it can be assumed that these comic strips at least reflect how Kellerman *wants* us to perceive himself, his friends, and their common language. A large number of examples of foreign items were found in the material, the vast majority of which was of apparent English origin (125 nouns, 28 verbs, 12 adjectives and 9 adverbs), as shown along with tentative analyses in Examples 12–15. Both inflectional (Example 12), derivational (Example 13) and compounding (Example 14) processes involving English items were found. Morphologically motivated usage of English plural /s/ in joining compounding elements (Example 14) or as an intrusive /s/ (Example 15) was also observed.

(12.)

geeks, geekarna, drajv-bajs  
 geeks, geek +PLUR+DEFSwe, drive-by [shooting]s

(13.)

casha, beepa, flashig  
 cash +DEVERBAL/a/Swe, beep +DEVERBAL/a/Swe, flash +ADJECTIVAL/g/Swe

(14.)

hallonshots, handsfreemobiler, depparsoundtrack,  
 raspberry shots, hands-freeEng cell phonesSwe, gloomy rockerSwe sound trackEng,

dansbandsgroupie, bikerfest, sodastreamern  
 (Swedish) dance band + /s/Swe groupie, biker party, soda streamer+DEFSwe





Swedish context, as also noted by Svenska Akademin (Teleman, Hellberg and Andersson, 1999).

### 3.5 The Parole Corpus

The Swedish *Parole* corpus (Parole Consortium, n.d.; Gronostaj, n.d.) is a morphosyntactically annotated text collection comprising approximately 19 million running words, compiled by Språkdata in Göteborg. Even if this corpus is based on written rather than spoken language, it was included for the sake of comparison, and also because its sheer size could make it valuable as a resource for finding general word formation patterns. The cumulative frequency distribution of this and two spoken language corpora is shown for comparison in Figure 4. An example concordance from the *Parole* corpus is shown in Figure 5, involving the obviously even in text quite established term *grunge*<sup>2</sup>. This specific example shows how English terminology (in this case the name of a genre) from the cultural scene has quickly been picked up and become a productive part of the Swedish vocabulary. Drawing on the results from the two production studies, it can be expected that this term contributes to Swedish phonology by adding pronunciations such as [gø-aɔʒ], [gø-aɪʃ], [grandʒ], grantʃ], [granç], and [grans], all of which extend beyond traditional descriptions of Swedish phonotactics.

### 3.6 Language and Music Worlds of High School Students

Within the project *Gymnasisters språk- och musikvärldar (GSM)* (Andersson, Edström, Lilliestam, Norrby and Wirdenäs, 1999; Norrby and Wirdenäs, 1998), 27 group conversations, encompassing approximately 20 hours of speech have been collected and orthographically transcribed. The cumulative frequency distribution of this corpus is shown in Figure 4, and some examples drawn from that corpus are shown in Table 1, where each word is shown along with its rank, frequency, relative frequency and cumulative coverage in the actual corpus. One thing to note is the high productivity featured by *skate[board]*. Another observation, which relates directly to the phonological level, discussed in Section 2, is that several of

---

<sup>2</sup> Thanks to Mats Wirén for pointing out that it apparently takes Swedish neutral ground for the English and French to meet, as seen here in *grunge-couture*.

## English Influence on Swedish Word Formation and Segmentals

the example words are highly likely to elicit pronunciations using xenophone extensions, e.g. by [dʒ] in *Fugees*, *energy*, and *Prodigy*.

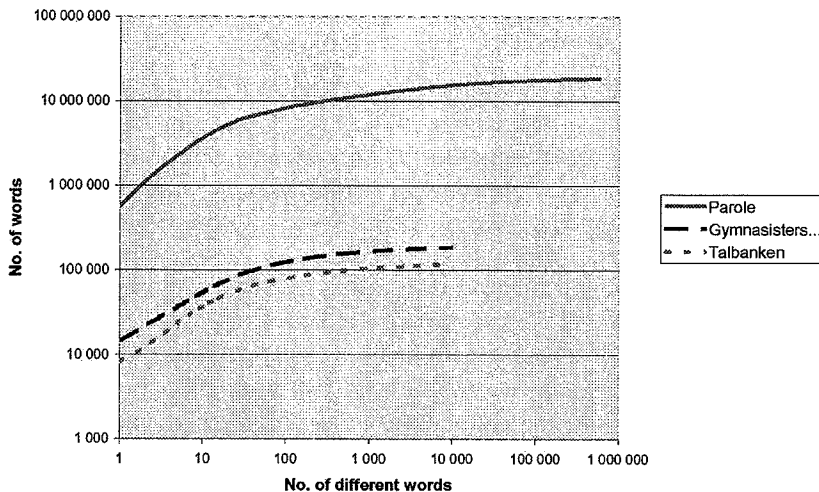


Figure 4: Cumulative frequency distributions for the text corpus *Parole*, and the spoken language corpora *Gymnasisters språk- och musikvärldar* and *Talbanken*.

igt av Staffan Astner , det är fullt rockkomp med grungeprofil  
både här och var , och syntetiska rytmer  
k blir det alltid oavsett om det är rockmusiker i grungepaltor  
eller biståndsarbetare i skjorta och slips  
cerat slag . Alltid med humor och mycket stänk av grunge-couture .  
Fyra olika klanrutor lyckades hon samla  
trutlavor , glest hakskägg och brunråpräckliga , grungesäckiga  
modepaltor . En av brädorna har varumärket  
sångskickliga band slår ut det mesta i popgenren grunge . Så långt  
allting väl . Men snälla , snälla Viktor  
jade producera dansorienterad pop , mil ifrån den grungeväg som  
svepte över världen just då . - Vi har  
göra program för dem . Sommarbio-trion känns som grunge-  
mutationerna av ville , valle och Viktor .  
t så mycket antiglamour och anti-divatrender inom grungen . - Jag  
tror folk riktigt gick och längtade efter  
gasonic skrev jag till filmen SFW som är en slags grungefilm . Jag  
tycker så oerhört illa om de där  
iensers som den sekulära . Här finns heavy-metal , grunge och rap .  
Gospel Gangstas heter en rapgrupp som  
Uchida och Ralf Gothoni . Sound Garden och andra grungeband från  
USA spelar den 30 augusti . Koreografen

Figure 5: A concordance from the Parole corpus for the morpheme grunge.

Word	Rank	f	Rel.f [%]	Cumul. f [%]
de	1	14148	7.48	7.5
e	2	7743	4.09	11.6
ja	3	5473	2.89	14.5
jag	4	4748	2.51	17.0
så	5	4470	2.36	19.3
på	6	4244	2.24	21.6
inte	7	3819	2.02	23.6
Beatles	197	106	0.06	75.0
Fugees	264	78	0.04	78.2
energy	447	35	0.02	83.1
hardcore	579	25	0.01	85.2
skateare	682	20	0.01	86.4
coolt	724	19	0.01	86.8
cool	866	15	0.01	88.0
aqua	868	15	0.01	88.0
soul	878	14	0.01	88.1
you	962	12	0.01	88.7
metal	988	12	0.01	88.9
coola	1016	12	0.01	89.1
power	1050	11	0.01	89.3
wow	1087	10	0.01	89.5
skate	1098	10	0.01	89.5
westcoast	1159	9	< 0.01	89.8
metallica	1314	8	< 0.01	90.6
house	1336	8	< 0.01	90.7
freestyle	1494	7	< 0.01	91.3
Prodigy	1522	7	< 0.01	91.4
coolio	1677	6	< 0.01	91.8
sound	1769	5	< 0.01	92.1
goatrance	4142	2	< 0.01	95.8
coolhetsstatus	4388	2	< 0.01	96.0
speedar	6360	1	< 0.01	97.2
speeda	6361	1	< 0.01	97.2
soundtracks	6377	1	< 0.01	97.2
soundtrackmusik	6378	1	< 0.01	97.2
sounds	6380	1	< 0.01	97.2
soulmusik	6381	1	< 0.01	97.2
soullåtar	6382	1	< 0.01	97.2
snowboardtävlingar	6428	1	< 0.01	97.2
snowboardare	6429	1	< 0.01	97.2
skatesvängen	6659	1	< 0.01	97.4
skatepunkarna	6660	1	< 0.01	97.4
skatepunk	6661	1	< 0.01	97.4
skatekulturen	6663	1	< 0.01	97.4
skateboardmen	6664	1	< 0.01	97.4
skateboardkulturen	6665	1	< 0.01	97.4
skateboardare	6666	1	< 0.01	97.4
skatearmusik	6667	1	< 0.01	97.4
skatearaktigt	6668	1	< 0.01	97.4
skatearkläder	6669	1	< 0.01	97.4
skateaktigt	6670	1	< 0.01	97.4
Total	11,635	189,246		100.0

Table 1: Example words, along with rank, frequency, relative frequency and cumulative coverage in the corpus *Gymnasisters språk- och musikvärldar*.

### 3.7 Two corpora from the 1960s

Within the *Talbanken* project (Einarsson, 1976), more than 115,000 words of interviews, conversation and debate were recorded and transcribed, as described by Teleman (1974) and Einarsson (1978). The cumulative frequency distribution of this corpus is also included in Figure 4. As expected, the cumulative frequency distributions of the three corpora show that the two spoken language corpora are quite similar, while the much larger text corpus *Parole* corpus (*Parole*) behaves differently. In order to cover 90% of each of the three corpora, it takes 14% of the 8,289 different words in *Talbanken*, 10% of 11,635 different words in *GSM*, but only 5% of the 573,546 different words in *Parole*. The share of *hapax legomena*, however, is roughly the same across all three corpora, namely 56% in *Talbanken*, 60% in *GSM* and 54% in *Parole*. This is also similar to the 50% reported by Bell and Gustafson (1999) for the *August* man-machine dialogue corpus.

The *Talbanken* interviews were part of a sociological study regarding attitudes towards labour immigration, which (as expected) elicited topics such as ethnicity, foreign language learning, culture etc. In spite of this, when exactly the same directed semi-automatic search procedure that was used with the *Parole* and *GSM* corpora was applied to *Talbanken*, only a couple of examples turned up:

(17.)

han skrev deeds and demonstrations  
 he wrote deeds and demonstrations  
 he wrote "Deeds and Demonstrations"

(18.)

vänta ja har ju en kursare te mej som nyligen har skrivit en  
 wait I've got a class-mate (of mine) who recently has written an  
  
 uppsats om abort men hon har väl toucha lite ve de  
 essay on abortion but she has (I think) touched a little on that

wait—I've got a class-mate who recently wrote an essay on abortion but I think she touched slightly upon that

*Transcriptions by Einarsson (1976)*

*Glosses and translations by the present author*

Of these, Example 17 should perhaps be disregarded as being more of a title quotation, while Example 18 shows Swedish verbal morphology used in conjunction with the borrowed English root "touch".

This almost complete lack of anglicisms in the *Talbanken* corpus could of course also be due to other causes than language contact, e.g. factors related to the interview situation. To eliminate the risk that any perceived distance between interviewer and interviewee inhibited cross-linguistic word formation processes, a set of highly informal conversations, recorded in a project by Bengt Nordberg and others in 1967–1968, were also studied (Pettersson and Forsberg, 1970). Transcriptions of two hours of conversational-style interviews with five subjects, aged 17–23, plus a conversation between two female subjects, aged 20 and 21, comprising a total of 16,250 running words, were analyzed. The topics of conversation were school, language education, hobbies, sports, travel, TV shows, etc. The only example of English influence was found in the unsupervised conversation between the two female subjects, who quote an English song title:

(19.)

han f̥œŋ ɪn n̥ɔntɪŋ po: mɪn banspe:lare... d̥ɛ n hæ:r ju: ɑ:r ðɪ o'nɪr 'oan  
he recorded something on my tape recorder... this one You are the only one  
he recorded some song on my tape recorder... the one called "You are the only one"

Transcriptions in "landsmålsalfabetet" by several transcribers, as described by Pettersson and Forsberg (1970)

Transliteration in IPA, glosses and translations by the present author

These results appear to support the hypothesis that English influence on spoken Swedish is much more widespread today than a couple of decades ago.

#### 4. Discussion

From the corpora studied, we can conclude that word formation processes like derivation, inflection and compounding are highly productive in contemporary spoken Swedish also when incorporating material (morphemes, simple lexemes or even complex nominals) borrowed from

English. In the spoken language corpora from the 1960s, hardly any such examples were found, despite the fact that the topic of conversation should, if anything, elicit precisely that. In the productive processes we see today, the foreign material can or has to undergo adaptations in order to fit in with morphotactic or morphonological constraints, as e.g. in the case of stress pattern and word accent restrictions in Swedish compounding. Sometimes virtually no adaptation occurs, e.g. when retaining English plural endings, instead of employing one from an appropriate productive Swedish paradigm. We have also revisited existing data from a production study of the segmental aspects, which indicates that there are plenty of cases in everyday communicative situations where the phonological system simply needs to be extended with xenophones, in order to model, produce or perceive contemporary and socially acceptable spoken Swedish.

The effects of educational level in the production study, coupled with well-known socio-linguistic grounding mechanisms, seem to suggest that selecting the appropriate level of xenophone inclusion should be of importance for the perceived persona e.g. in the generation of synthetic speech. It is also worth noting that the differences between educational groups seems to get accentuated with decreasing overall share of xenophone productions. One way of interpreting this is that in some cases, pronunciations involving xenophones of English origin are already conventionalized, and consequently produced by virtually all subjects. While this may be technically interesting, since it will require special treatment e.g. in spoken language dialogue systems, it is probably of less interest, linguistically; borrowing the terminology from OT, one might simply say that constraints requiring faithfulness towards the underlying forms of English origin completely outrank conflicting "well-formedness" constraints, requiring re-phonematization. In less conventionalized cases, however, education, age, and possibly also gender, seem to determine to what *degree* faithfulness constraints are allowed to outrank conflicting "well-formedness" constraints. This is in line with the reasoning by Davidson (2001), who claims that studying the phonotactics of a vocabulary in equilibrium gives little insight regarding the interaction between conflicting constraints, compared to what can be extracted from the type of contact-linguistic situation we are dealing with here.

One question often raised in other studies is whether the English influence on (particularly written) Swedish is large or not. This question is often associated with a debate where some regard this type of influence as a problem or even perceive it as a "threat" against

relatively small languages, e.g. those of the Nordic countries, or specific domains within those languages, e.g. computing, engineering, etc. It should be borne in mind, however, that linguistic borrowing, boosted by cultural contact of various kinds, is (and always has been) one of the most fundamental driving forces in linguistic development, with the English language itself being a very obvious result of such a process. Also, as can be seen from the many examples we have given, although terminology borrowed from English is often allowed to expand the phonotactic repertoire, spoken Swedish is still subject to seemingly stable “native” morphological, morphonological and prosodic constraints. These processes need to be further studied and descriptions of contemporary Swedish need to be revised and extended to take them into account, rather than treating them as a marginal phenomenon. The relative stability of some of these “well-formedness” constraints does not mean that Swedish speech and language technology applications will face no problems related to English and other foreign linguistic elements, quite the opposite. At first glance, the problem may seem to be marginal and of minor importance— after all, each item in Table 1 accounts for a relatively small fraction of the entire corpus. However, one needs to remember that while the most frequent items (many of which belong to the closed word classes) rapidly yield high coverage in terms of cumulative frequency, the productive nature of Swedish morphology in connection with English items, as we have seen, in fact makes that section of the vocabulary infinitely large. As we saw, the share of *hapax legomena* is 54% in *Parole*. However, that corresponds to no less than 310,973 items. Even if these unique word forms probably also include a few inevitable typos, which have escaped the meticulous annotation process, they are primarily formed through the productive morphological processes, of which we have just seen numerous examples. The two spoken language corpora in Figure 4 are not really very different from *Parole* in terms of growth, they are only a lot smaller, with approximately 5,000 *hapax legomena* each. Chances are that the next time a speech corpus of a similar size is collected, its set of *hapax legomena* will not have much overlap with any of these corpora (Good, 1953). It has therefore also proven necessary to develop and evaluate any lexical component against functional criteria, rather than using data-driven methods (Lindström, 2003). Furthermore, these items may cause a disproportionate amount of trouble for spoken language applications when mis-pronounced and/or mis-recognized, e.g. when repeating someone’s given

name in a dialogue situation. What is intended as an act of clarification may then well be perceived as an insult, since errors in pronouncing proper names are especially prone to either cause serious identification mistakes, or possibly offend the bearer of the name, or both.

## 5. Conclusions

We have shown how English influences contemporary spoken Swedish both segmentally and at the word formation level, while hardly any such examples could be found in corpora from the 1960s. The Swedish language seems to permit or even require the use of a number of xenophones of English origin in these contact situations, whereas Swedish prosodic and morphotactic restrictions are imposed and seem more difficult to violate in the word formation processes. Examples and frequency data from several corpora were analyzed, compared and discussed, and it was found that the processes of inflection, derivation and compounding are highly productive in contemporary Swedish, also in contact with English. This is of theoretical interest, has practical consequences for speech technology applications, but available spoken language data is limited and the area definitely calls for many further empirical studies.

## Acknowledgements

The work presented here was carried out at TeliaSonera R&D and at the Department of Computer and Information Science, Linköping University as part of the Industry Research School in Applied IT and Software Engineering. The project was partly funded by KK-stiftelsen (The Knowledge Foundation), and was initiated by Leif Bengtsson of TeliaSonera, and by Lars Ahrenberg, Bertil Lyberg and Sture Hägglund of Linköping University, to all of whom the author is grateful. Special thanks to Robert Eklund for inspiring collaboration throughout the whole project and for taking care of the documentation, to Anna Kasaty for collaboration on the “Rocky” study, to Anke Lüdeling for keeping my spirits high, and to Linda Bell for providing a mixture of humor, support and a special attitude. Thanks also to Jessica Granberg and Per Sautermeister for transcribing large parts of the Xenophone data. The author is grateful to the following researchers, for generously sharing their data, providing useful information, and helping with literature: Jens Allwood and Leif Grönqvist (*GSLC*), Goranka Antunović (*dictionary of*



*Anglicisms*), Jan Anward, Jenny Öqvist and Niklas Norén (*GRIS*), Jan Einarsson (*Talbanken*), Ulla-Britt Kotsinas (*slang corpus*), Magnus Ljung (*EIS*), Eva Sundgren and Bengt Nordberg (*Eskilstuna interviews*), and Karolina Wirdenäs (*GSM*). Thanks are also due to Joakim Gustafson for fruitful ideas regarding the data analysis, and to Per Näsman for helping out with the statistical analysis. Finally, the author would like to thank Lars Ahrenberg, Linda Bell, Petra Hansson and the patient editors of this issue for relevant and much needed feed-back on earlier versions of this paper.

*Voice Technologies, TeliaSonera R&D, Stockholm*  
*NLPLAB, Dept. of Computer and Information Science,*  
*Linköping University*

## References

- Allwood, J., Grönqvist, L., Ahlsén, E. and Gunnarsson, M.: 2002, Göteborgskorpusen för talspråk, *Nydanske Sprogstudier* (30), 39–58. special issue on “Korpuslingvistik”.
- Andersson, L.-G., Edström, K.-O., Lilliestam, L., Norrby, C. and Wirdenäs, K.: 1999, Gymnasisters språk- och musikvärldar, Project Web Site. <http://svenska.gu.se/~svekw/gsmidx.html>.
- Antunović, G.: 1999, *A dictionary of Anglicisms in Swedish*, University of Zagreb, Zagreb.
- Baayen, H.: 1992, Quantitative aspects of morphological productivity 1991, in G. Booij and J. van Marle (eds), *Yearbook of Morphology*, Kluwer Academic Publishers, Dordrecht, pp. 109–150.
- Bell, L. and Gustafson, J.: 1999, Utterance types in the August dialogues, *Proceedings of IDS'99, ESCA workshop on Interactive Dialogue in Multi-Modal Systems*, ESCA, pp. 81–84.
- Bell, L., Eklund, R. and Gustafson, J.: 2000, A comparison of disfluency distribution in a unimodal and a multimodal speech interface, *Proc. of the 6th ICSLP*, Vol. 3, Beijing, pp. 626–629.
- Billi, R. and Lamel, L. F.: 1997, Railtel: Railway telephone services, *Speech Communication* 23, 63–65.
- Billi, R.: n.d., Interview, [http://www.hltcentral.org/usr\\_docs/LeJournal/I-phonesPhonemesMultimodality.pdf](http://www.hltcentral.org/usr_docs/LeJournal/I-phonesPhonemesMultimodality.pdf).
- Brennan, S. E.: 2000, Processes that shape conversation and their implications for computational linguistics, *Proc. of the 38th Annual Meeting of the ACL*, ACL, Hong Kong.

## English Influence on Swedish Word Formation and Segmentals

- Bruce, G.: 1998, *Allmän och svensk prosodi*, number 16 in *Praktisk lingvistik*, Institutionen för lingvistik, Lunds universitet, Lund.
- Carlson, R., Granström, B. and Lindström, A.: 1989, Predicting name pronunciation for a reverse directory service, *Proc. of Eurospeech*, Vol. 1, Paris, pp. 113–116.
- Davidson, L.: 2001, Hidden rankings in the final state of the English grammar, in G. Horwood and S.-K. Kim (eds), *Ruling Papers*, Vol. II, Rutgers University, New Brunswick, pp. 21–48.
- Donegan, P. J. and Stampe, D.: 1979, The study of natural phonology, in D. A. Dinnsen (ed.), *Current approaches to phonological theory*, Indiana University Press, Bloomington, chapter 6, pp. 126–173.
- Einarsson, J.: 1976, Talbankens talspråkskonkordans, Transcribed corpus.
- Einarsson, J.: 1978, Talad och skriven svenska. Sociolingvistiska studier, Lund.
- Eklund, R. and Lindström, A.: 1996, Pronunciation in an internationalized society: A multidimensional problem considered, *Proc. of the Swedish Phonetics Meeting (Fonetik 96) (TMH-QPSR 2/1996)*, Nässlingen, Stockholm, pp. 123–126.
- Eklund, R. and Lindström, A.: 2001, Xenophones: An investigation of phone set expansion in Swedish and implications for speech recognition and speech synthesis, *Speech Communication* 35(1-2), 81–102.
- Elert, C.-C.: 1971, *Uttalsbeteckningar i svenska ordlistor, uppslags- och läroböcker*, number 44, Kungl. Boktryckeriet P. A. Norstedt och Söner, pp. 51–66.
- Filipović, R.: 1996, English as a word donor to other languages of Europe, in R. Hartmann (ed.), *English Language in Europe*, European Studies Series, Intellect, Oxford, pp. 37–46.  
<http://www.intellectbooks.com/europa/number5/hart.htm>.
- Good, I. J.: 1953, The population frequencies of species and the estimation of population parameters, *Biometrika* 40, 237–264.
- Gorin, A. L., Riccardi, G. and Wright, J. H.: 1997, How may I help you?, *Speech Communication* 23(1/2), 113–127.
- Görlach, M. (ed.): 2001, *A Dictionary of European Anglicisms*, Oxford University Press, Oxford.
- Gronostaj, M. T.: n.d., The Swedish PAROLE lexicon. A language engineering resource with access to morphological and syntactic information in Swedish, elaborated by Språkdata, Göteborgs Universitet, Web publication. <http://spraakbanken.gu.se/>.
- Gustafson, J., Bell, L., Boye, J., Lindström, A. and Wirén, M.: 2004, The NICE fairy-tale game system, *5th SIGdial Workshop on Discourse and Dialogue*, ACL, Boston. <http://sigdial04.eml-research.de/sigdial/>.

- Itô, J. and Mester, A.: 1999, *The Handbook of Japanese Linguistics*, Blackwell Publishers, Malden, MA, and Oxford, U.K., chapter The Structure of the Phonological Lexicon, pp. 62–100.
- Jespersen, O.: 1902, Engelsk og nordisk: En afhandling om låneord, *Nordisk tidsskrift for vetenskap, konst och industri* p. 500–514.
- Jungbluth, K.: 1999, Two- and three-dimensional deictic systems between speech and writing—Evidences from the use of demonstratives in romance languages, in E. André, M. Poesio and H. Rieser (eds), *Proc. of the Workshop on Deixis, Demonstration and Deictic Belief. Workshop held at the 11th European Summer School in Logic, Language and Information (ESSLLI), ESSLLI, Utrecht University, Utrecht*, pp. 12–19. Paper 4.
- Kotsinas, U.-B.: 2000, Engelska i svensk slang, in B.-L. Gunnarsson, S. Hellberg and K. Svartholm (eds), *Kontakt, variation och förändring — studier i Stockholmspråk. Ett urval uppsatser av Ulla-Britt Kotsinas, number 22 in Stockholm Studies in Scandinavian Philology*, Almqvist & Wiksell International, Stockholm, pp. 137–159.
- Ladefoged, P. and Maddieson, I.: 1996, *Sounds of the World's Languages*, Blackwell, Oxford.
- Lindblom, B.: 1990, Explaining phonetic variation: A sketch of the H and H theory, in W. J. Hardcastle and A. Marchal (eds), *Speech Production and Speech Modelling*, Kluwer Academic Publishers, pp. 403–439.
- Lindström, A. and Kasaty, A.: 2000, A two-level approach to the handling of foreign items in Swedish speech technology applications, *Proc. of the 6th ICSLP*, Vol. 1, Beijing, p. 54–57.
- Lindström, A.: 2003, Non-native linguistic elements in spoken Swedish, *Proc. of the 15th ICPhS*, Barcelona, pp. 2353–2356.
- Ljung, M.: 1986, *Skinheads, Hackers och Lama Ankor. Engelskan i 80-talets svenska*, Trevi, Stockholm.
- Lüdeling, A. and Schmid, T.: 2001, Does origin determine the combinatory properties of morphological elements in German?, in J. DeCesaris (ed.), *Proceedings of the third Mediterranean Meeting on Morphology*, Barcelona.
- McArthur, T.: 1996, English in the world and in Europe, in R. Hartmann (ed.), *English Language in Europe*, European Studies Series, Intellect, Oxford, pp. 3–15.
- Melchers, G. and Shaw, P.: 2003, *World Englishes. An Introduction*, Edward Arnold.
- Nordberg, B.: 1999, Samtalsspråkets grammatik. (Grammar in conversation: A study of Swedish). Available at <http://www.tema.liu.se/tema-k/gris/syfte.html>. Project Granted by the Bank of Sweden Tercentenary Foundation.

## English Influence on Swedish Word Formation and Segmentals

- Norrby, C. and Wirdenäs, K.: 1998, Language and music worlds of senior high school students, in I.-L. Pedersen and J. Scheuer (eds), *Sprog, køn og kommunikation. Rapport fra 3. Nordiske konference om sprog og køn*, C. A. Reitzels forlag, Copenhagen, pp. 155–163.
- Öqvist, J.: 2000a, Tryck till, *Unpublished transcriptions*, Linköping University, Linköping, From the project Samtalsspråkets Grammatik (Nordberg, 1999).
- Parole Consortium: n.d., Parole/simple project home page, Web publication. <http://www.ub.es/gilcub/SIMPLE/simple.html>.
- Pettersson, P. A. and Forsberg, K.: 1970, Beskrivning och register över Eskilstunainspelningar, *FUMS Rapport nr. 10 (del 1-2)*, Forskningskommittén i Uppsala för Modern Svenska, Uppsala.
- Phillipson, R.: 2003, *English-Only Europe? Challenging Language Policy*, Routledge.
- Prince, A. and Smolensky, P.: 1995, *Optimality Theory: Constraint Interaction in Generative Grammar*, Linguistic Inquiry Monograph Series, MIT Press, Cambridge, MA.
- Sampson, G.: 1985, *Writing Systems: A Linguistic Introduction*, Stanford University Press, Stanford, California.
- Schmid, T., Lüdeling, A., Säuberlich, B., Heid, U. and Möbius, B.: 2001, DeKo: Ein system zur analyse komplexer wörter, in H. Lobin (ed.), *Proc. of GLDV*, Gießen, pp. 49–57.
- Shannon, C. E.: 1948, A mathematical theory of communication, *Bell System Technical Journal* 27, 370–423, 623–656.
- Sharp, H.: 2001, *English in Spoken Swedish. A Corpus Study of Two Discourse Domains*, PhD thesis, Stockholm University, Stockholm.
- Spiegel, M. F., Macchi, M. J. and Gollhardt, K. D.: 1989, Synthesis of names by a demisyllable-based speech synthesizer (Spokesman), *Proc. of Eurospeech*, Vol. 1, Paris, pp. 117–120.
- Teleman, U.: 1974, Manual för grammatisk beskrivning av talad och skriven svenska (Mamba), Studentlitteratur, Lund.
- Teleman, U.: 2004, *Tradis och Funkis: Svensk Språkvård och Språkpolitik efter 1800*, Norstedts ordbok.
- Teleman, U., Hellberg, S. & Andersson, E.: 1999. *Svenska Akademiens Grammatik*. Stockholm: Norstedts: 1999,
- Weinreich, U.: 1953, *Languages in contact*, Mouton, The Hague.