# Filled pauses as evidence of L2 proficiency: Finnish Australians speaking English

**Timo Lauttamus**[1], **John Nerbonne**[2] **& Wybo Wiersma**[2]

[1]*University of Oulu, Finland;* [2]*University of Groningen, The Netherlands*

## 1. Introduction

The paper discusses the application of a computational technique to tag a corpus containing Finnish Australian English automatically and to analyze the frequency vectors of part-of-speech (POS) trigrams using a permutation test (Nerbonne & Wiersma 2006). Our general goal is to detect the linguistic sources of the syntactic variation between two groups, the "adults", who had received their school education in Finland, and the "juveniles", who were educated in Australia. The idea of the technique is to utilize frequency profiles of trigrams of POS categories as indicators of syntactic distance between the groups and then examine potential effects of shift-induced interference (substratum transfer) on second language acquisition (SLA). Lauttamus, Nerbonne & Wiersma (2007) show that some features we describe as "contaminating" the interlanguage of the adults can be best attributed to Finnish substratum transfer. However, there are other features in our data that may also be ascribed to more "universal" primitives or universal properties of the language faculty (cf. Chambers 2003: 265-266).

To explain differential usage by the two groups, we also draw upon the strategies, processes and developmental patterns that second language learners usually evince in their interlanguage regardless of their mother tongue. We are therefore proponents of an approach to language contact where results from SLA research are highly relevant (cf. e.g. Siemund 2008: 4).

The focus of the present paper is on the disfluent speech that the adults, rather than the juveniles, produce. We apply the technique described in detail below to examine if the two groups show a differential use of *filled pauses* ('vocalized pauses'), tagged as *interjec*(tion) in our data and including items such as *um*, *uh*, and *ah*, which are characteristic of the adults. Lauttamus et al. (2007: 290-291) suggest that features such as filled pauses (FPs), repeats, false starts, and repairs, are (statistically) significant determinants that distinguish less proficient learners having acquired an L2 later in life (the adults) from more proficient learners having acquired their L2 at early age (the juveniles).

Difficulties in controlling pause duration and placement seem to be common among all learners irrespective of the target language. Paananen-Porkka (2007) argues that pausing, including FPs, is the main source for the anomalies found in English speech rhythm by Finnish comprehensive school students. She also states that "pauses not only occurred at sentence or clause boundaries, but also at

word boundaries" in all language groups that she studied, i.e. in native-speaker English, non-native-speaker English and native-speaker Finnish (pp. 259-260).

Our goal is to investigate whether FPs can be used as an indicator of L2 proficiency. We hypothesize that the adults use more FPs than the juveniles, and that they pause at "inappropriate" syntactic positions compared to native speakers of English. Paananen-Porkka (2007: 260), for example, points out that "unlike Finnish pupils, native speakers of English did not pause between the auxiliary and the main verb". In addition, she notes that there is a higher mean percentage of pauses at word boundary for Finnish speakers of English than for both L1 speakers (p. 264). It appears, however, that the difference in the use of FPs by the adults and the juveniles may be measured only in terms of frequency of use rather than in terms of position. Our preliminary results suggest that FPs in the adults' speech can occur in any position but most frequently before noun heads, as in *I just had a **uh** cup of water*, or before pronouns, as in *because **uh** we stayed three years* (possibly signaling a clause boundary). The syntactic positions of FPs will be discussed in detail in our future work.

## 2. Syntactic theory and POS tagging

Syntactic theory uses analysis trees showing constituent structure and/or dependency structure to represent syntactic structure, so a natural tool to consider for the task of detecting syntactic differences would be a parser – a program which assigns the syntactic structure appropriate for an input sentence (given a specific grammar). We decided, however, against the use of a parser, and for the more primitive technique of POS tagging (explained below) because, even though automatic parsing is already producing fair results for the edited prose of newspapers, we suspected that it would be likely to function very poorly on the conversational transcripts of second language learners. Both the conversation style of the transcripts and the frequent errors of learners would be obstacles. We return below to the selection of corpora and its motivation.

### *2.1. Tagging*

We detect syntactic differences in two corpora in a fairly simple way (Lauttamus et al. 2007). We tag the two corpora automatically (all errors are left in), i.e. we automatically detect for each word its syntactic category, or, as it is commonly referred to, its part-of-speech (POS). Below we provide an example:

| | | | |
|---|---|---|---|
| (1) | the | cat | is |
| | ART (def) | N (com, sing) | V (cop, pres) |
| | on | the | mat |
| | PREP (ge) | ART (def) | N (com, sing) |

We tagged the corpora using the set of POS tags developed for the TOSCA-ICE, which consists of 270 POS tags (Garside et al. 1997), of which 75 were never instantiated in our material. Since we aim to contribute to the study of language contact and second language learning, we chose a linguistically sensitive set, that is, a large set designed by linguists, not computer scientists. In a sample of 1,000 words we found that the tagger was correct for 87% of words, 74% of the bigrams (sequences of two words), and 65% of the trigrams (sequences of three words). The accuracy is poor compared to newspaper texts, but we are dealing with conversation, including the conversation of non-natives. Since parsing is substantially less accurate than POS tagging, we feel that this accuracy level confirms the wisdom of not trying to use the more informative technique of full parsing.

The POS tags are then collected into ordered triples, *trigrams* such as ART(def)-N(com,sing)-V(cop,pres),..., PREP(ge)-ART(def)-N(com,sing). We use POS trigrams, rather than single tags, as indications of syntactic structure in order to obtain fuller reflection of the complete syntactic structure, much of which is determined once the syntactic categories of words are known. In making this last assumption, we follow most syntactic theory, which postulates that hierarchical structure is (mostly) predictable given the knowledge of lexical categories, in particular given the lexical 'head'. Sells (1982, sec. 2.2, 5.3, 4.1) shows how this assumption was common to theories in the 1980s, and it is still recognized as useful (if imperfect given the autonomy of "constructions", which Fillmore & Kay, 1999, demonstrate). So if syntactic heads have a privileged status in determining a "projection" of syntactic structure, then we will detect syntactic differences in two varieties by quantifying the distribution of parts-of-speech in context.

### *2.2. Comparison*

We then collect all the POS trigrams found in the corpora (13,784 different POS trigrams in the case of the Finnish Australian data), and count how frequently each occurs in both of the corpora. We then compare the two rows of this 2 X 13,784 element table, asking two questions. First, we wish to know whether the distribution in the two rows might be expected by chance, in other words, whether there is a statistically significant difference in the distributions. Second, in case the overall distributions differ significantly ($p$-values at or below 0.05), we calculate which frequent POS trigrams are responsible for the skewed distribution. We suppress the technical details in this presentation, referring the interested reader to Nerbonne & Wiersma (2006).

In connection with the second goal, we examine the top 200 significant POS trigrams that contribute the most to the skewing of the distribution between the two corpora. This list was created by sorting all significant trigrams by the relative size of the difference in their frequency between the two groups,

not by their absolute frequency. This ensures that we have the trigrams that are the most characteristic for each group at the top. We turn to an examination of the Finnish Australian data below.

By analyzing differences in the frequencies of POS trigrams, we importantly identify not only deviant syntactic uses ("errors"), but also the overuse and underuse of linguistic structures, whose importance is emphasized by researchers on second language acquisition (Ellis 1994: 304-306, who calls underuse 'underrepresentation' and overuse 'over-indulgence'; de Bot et al. 2005: A3, B3). According to these studies, it is misleading to consider only errors, as second language learners likewise tend to overuse certain possibilities and tend to avoid (and therefore underuse) others. For example, de Bot et al. (2005) suggest that non-transparent constructions (such as idioms, preferred collocations, e.g. *heavy traffic, dense text,* including verb and eventive object, e.g. *make a choice, have a bath*) are systematically avoided even by very good second language learners.

We like to emphasize that our work assumes, not that syntax consists solely of part-of-speech sequences, but only that differences in part-of-speech sequences are indicative of syntactic differences in general. It is important to emphasize that we do not claim to have developed a technique that probes all conceivable syntactic differences directly, but rather a technique that detects traces of differences in superficial syntax. Those differences might naturally have causes in deeper levels of syntactic structure.

Our proposed technique for detecting syntactic differences does indeed aggregate over many indicators of syntactic difference, in a way that makes progress toward assessing the "total impact" of a first language on a second in the way Weinreich (1953:63) sought, albeit with respect to a single linguistic level, namely syntax. We do not develop a true measure of syntactic difference here as that would require further calibration and validation, preferably cross-linguistically, but we do claim to detect differences in the frequency with which different constructions are used.

## 3. The Australian English of Finnish emigrants

We shall describe the differences between the English of those who emigrated as adults and those who emigrated as children (juveniles). After studying the transcripts, we assume that the latter's English is near native, and so we focus below on the English of those who emigrated as adults.

### *3.1. Linguistic situation of the adult emigrants*

We note that the linguistic development of the two Finnish groups in Australia is best described as language shift. We are therefore concerned with bigenerational bilingualism as a series of stages in the assimilation of the Finnish ethnic minorities into a linguistically, socially and culturally English-dominant speech

community, which inevitably entails the loss of the variety of Finnish used in the speech communities and Anglicization among these ethnic groups. We note that language shift seems to take place no later than during the 2nd generation of various ethnic groups in the US, with the exceptions of Spanish and Navajo (Karttunen 1977; Veltman 1983; Smits 1996; Klintborg 1999). The evidence from Hirvonen (2001) also supports this; American Finnish does not seem to survive as a viable means of communication beyond the second generation.

The situation is similar in Australia. Clyne & Kipp (2006: 18) note that "high-shift" groups in Australia tend to be ones who are culturally closer to Anglo-Australians in contrast with some "low-shift" groups with different "core values such as religion, historical consciousness, and family cohesion". The evidence in Lauttamus et al. (2007) suggests that also Finnish Australians represent those language groups that shift to English very rapidly in the second generation. It appears that even members of the 1st generation of immigrants may demonstrate a variety of achievements, including native-like ability (cf. Piller 2002), that members of the 2nd generation speak natively and that language attrition does not wait until the 3rd generation but begins with the 1st generation (cf. Waas 1996; Schmid 2002, 2004; Cook 2003; Jarvis 2003).

Consequently, we expect to find most of the evidence for syntactic interference (substratum transfer) in the English of first generation Finnish Australians (the adults), as the second generation (the juveniles) has already shifted to English without any interference from Finnish. In a sample of 12 first generation informants from the corpus collected by Watson (1996), only three use mostly English with English speakers in their daily communication, while three out of 12 informants use Finnish most of the time. For the remaining 6, Finnish is predominantly used at home, with relatives, and with other Finnish speaking friends, whereas English is used elsewhere (Levänen 2008: 8-10). In a sample of 12 second generation informants, all of them use both English and Finnish, and there are no significant differences in the language used in different situations between them. The dominant language at work and at home is English. Finnish is mainly used with parents and other Finnish speaking relatives and friends (Levänen 2008: 11-12).

All this strengthens the argument for the shift that second generation Finnish Australians have been undergoing. The findings in Lauttamus et al. (2007) point in the direction that second generation Finnish Australians speak (almost) natively, with very little Finnish interference in their English. This is corroborated by findings in some other studies, such as Lahti (1999) and Kemppainen (2000) on lexical features, Mannila (1999) on segmental features, Laakkonen (2000) on rhythm, and Markos (2004) on hesitation phenomena.

The language contact scholarship distinguishes situations of *shift* from *maintenance* (Thomason and Kaufmann 1988; Van Coetsem 1988). The adult emigrant group, our focus here, maintains Finnish, but, more to the point, shifts

to English, the subject of our research. Their Finnish is linguistically dominant, while English is socially dominant throughout Australia. In a situation of adult language shift, we expect interference from the native (Finnish) in the acquired (English) language, beginning with pronunciation (phonology) and morphosyntax. Lexical interference is comparatively weak.

### *3.2. Finnish Australian English Corpus (FAEC)*

Greg Watson of the University of Joensuu compiled a corpus of English conversations with Finns who had emigrated to Australia nearly thirty years earlier (Watson 1996). This corpus was kindly put at our disposal. All the respondents were Finnish native speakers. We divided them into two groups, "adults", or adult emigrants, who were over 18 upon arrival in Australia, and "juveniles", the children of the adults, who were all under 17 at the time of emigration. We distinguish between adult immigrants and immigrant children based on Lenneberg's (1967) well-known critical age hypothesis, which suggests a possible biological explanation for successful L2 acquisition between age two and puberty. Note that 'adult' vs. 'juvenile' refers only to the age at emigration: all the respondents were over 30 at the time of the interviews.

The adults were 30 years old on arrival (on average), and 58.5 at the time of the one-hour interview, and the juveniles were 6 and 36, respectively. There were 62 adult and 28 juvenile interviews, and there were roughly equal numbers of males and females. The interviews were transcribed in regular orthography by trained language students and later checked by Watson. Speakers were not tested for English proficiency, but it is clear from a quick view of the data that the juveniles' English is considerably better than that of the adults'. The juveniles had gone to school in Australia, and the adults in Finland. Our corpora contain 305,000 words in total. The text size for the adults is 221,000 and for the juveniles 84,000 words. By applying permutation statistics to the data, we can eliminate the effect of the difference in the number of words contributed by the two groups. Consequently, the imbalance has no impact on the number of possible trigrams produced for comparison (section 2.2).

## 4. Disfluent speech and filled pauses

The evidence from our syntactic analysis using the POS-tag trigrams and a permutation test like the one described in detail in Nerbonne & Wiersma (2006) and Lauttamus et al. (2007) shows that there are differences between the adults and the juveniles at a statistical significance level below 0.001. We argue that some of the significant syntactic differences found in the data can be ascribed to the lower level of language proficiency of the adults, and that the observed overuse of hesitation phenomena by the adults is a concomitant of lack of English proficiency.

The adults demonstrate typical features of disfluent speech, such as (filled) pauses, repeats, false starts, incomplete or false syntactic structures, arising from difficulties in speech processing, and particularly in lexical access (Lauttamus et al. 2007). The examples (2) to (6) show some of the attested features:

(2) skin cancer and /um and uh/ and gene general
(3) but /ah I I/ never been on
(4) clubs spades /hearts and uh/ uh cl oh
(5) he was a leading-hand um /leading-hand and ah/ last last
(6) as in /a in a/ Finland because especially

As to pausing, it should be remembered that only FPs, as in (2), (3), (4), (5), were included and tagged as *interjec*(tion) in each trigram between the slashes. Paananen-Porkka (2007) suggests that features of disfluent speech can occur at any syntactic boundary, at sentence, clause, phrase or word boundary, and this is what we would like to argue as well. They are, of course, characteristic of any kind speech, native and non-native alike, but certainly more frequent in interlanguage or, more generally, in SLA where speakers demonstrate imperfect learning as they study an L2.

We applied the computational technique described above to examine if the adults and juveniles show a differential use of FPs. Our earlier research (Lauttamus et al. 2007) suggests that pausing (FPs) is a statistically significant determinant that distinguishes less proficient learners having acquired an L2 later in life (the adults) from more proficient learners having acquired their L2 at early age (the juveniles). Our goal is now to show in detail if this is the case with the English of Finnish Australians, particularly with the adults.

## 5. Findings

It should be remembered that we first deal with the top 200 trigram types out of a total of 666, which all show statistically significant differences between the adults' and the juveniles' syntax ($p \leq 0.05$).

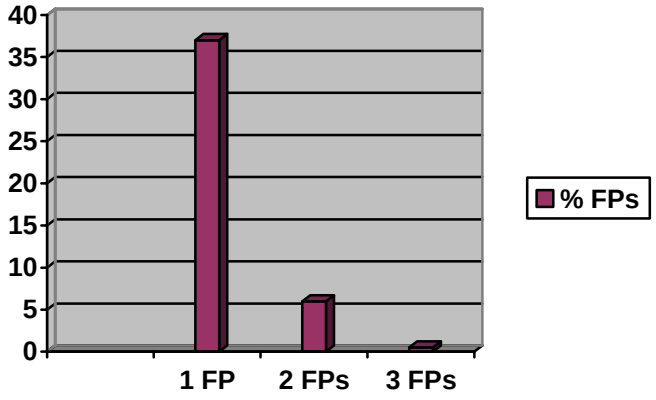

**Figure 1.** Percentage of filled pauses (FPs) in the top 200 POS-trigram types in the adults' speech.

Figure 1 shows that out of the top 200 POS-trigram types which most characteristically distinguished the adults from the juveniles, 37% (N=74/200) include at least one filled pause (1 FP), as in (7) and (8). Out of the same trigrams types, 6% (N=12/200) include at least two filled pauses (2 FPs), as in (9) and (10), and 0.5% (N=1/200) three filled pauses (i.e. a trigram type with only filled pauses). In all of these cases the adults used more filled pauses than the juveniles.

| | | | | |
|---|---|---|---|---|
| | | **Interj** | **Conj**(subord) | **Art**(def) |
| (7) | politically | /uh | when | the/ liberals were in |

| | | | | |
|---|---|---|---|---|
| | | **V**(cop,pres,encl) | **Interj** | **Adv**(inten) |
| (8) | | I'/m | ah | very/ sick |

| | | | | | |
|---|---|---|---|---|---|
| | | **Interj** | **Interj** | **Conj**(subord) | |
| (9) | and | /uh | uh | because/ | in the morning |

| | | | | | |
|---|---|---|---|---|---|
| | | **Interj** | **Pron**(pers, sing) | **Interj** | |
| (10) | and | /uh | I | uh/ | snow-skied in competitions |

For the juveniles, however, there are only 6 (six) trigram types in which they used the sequence of POS tags more frequently than the adults (using $p \leq 0.5$ as a cutoff), **none** of which **contain filled pauses**. Out of these six trigram types, four show the use of contracted forms, such as *I'll go*, *it'll be*, *we're going*, which are rarely used by the adults. The remaining two contain highly idiomatic phrases, such as *the next ten (years),* and *(I) sort of think*, not used by the adults.

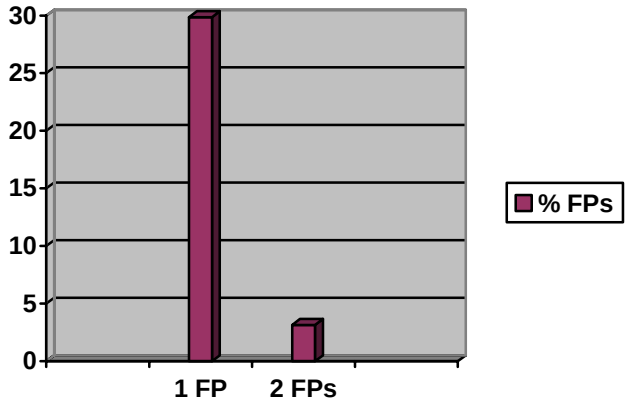

**Figure 2.** Percentage of filled pauses (FPs) in the more frequent 666 POS-trigram types in the adults' speech.

Figure 2 shows that out of all statistically significant 666 POS-trigram types produced more frequently by the adults, 29.6% (N=197/666) include at least one filled pause (1 FP). Out of the same trigrams types, 3.2% (N=21/666) include two filled pauses (2 FPs). In addition, there is one trigram type with filled pauses only.

Both figures show the same trend. We therefore argue that the highly skewed distribution of FPs across the two groups of speakers is conclusively evidenced by the data. We also argue that this anomaly can be explained in terms of the adults' lesser proficiency of English, which manifests itself as the overuse of FPs and lack of fluency.

The skewed distribution of the FPs suggested to us that the juveniles have a much more varied syntactic repertoire than the adults, who have much more limited and idiosyncratic syntactic patterns at their disposal; 'limited' in the sense that almost 40% of their top 200 trigram types include at least one FP. To corroborate the findings, we then decided to investigate what kind of impact the elimination of all FPs from the data (scripts) would have on the trigrams and their statistical significance.

The outcome of running the scripts again without the FPs shows that there are still only 6 statistically significant trigram types for the juveniles as opposed to 522 for the adults. It appears then that the elimination of the FPs has little effect on the number of POS-trigrams that differ in frequency for either the reduced tag set ($p \leq 0.0141$), e.g. ART-N-V, or the full tag set ($p \leq 0.0001$), e.g. ART(def)-N(com, sing)-V(cop, pres).

On the basis of the examination of the top 200 FP-less trigram types produced by the adults we see that 38.5% (N=77/200) of the trigram types are ungrammatical, and that some of the remaining trigram types are non-standard. Indisputably ungrammatical trigrams include, for example, omission of an obligatory article or preposition, omission of an obligatory copula or primary verb (*be* or *have*), omission of the subject, use of a redundant article with proper nouns, use of *what* as a relative pronoun (instead of *that* or *which*).

In examining the syntactic patterns, we interpreted them on the basis of our knowledge of standard (acrolectal) English, which is, we must admit, a risky undertaking. We likewise entertained interpretations based on what we know about non-standard (basilectal) varieties of English, but our knowledge is less than perfect here. Having inspected all trigram types, it became apparent that, in some, the elimination of the FPs resulted in an ungrammatical construction *per se*, such as ART(def)-PRON(poss,sing)-N(com,sing): *the my marriage*; or ART-PRON(dem)-N: *the that type* (an FP between the definite article and the pronoun

to signal correction). We will discuss the repertoire of the trigrams in more detail in future work.

## 6. Discussion

The statistically significant differential use of FPs by the adults can be explained in terms of the adults' lesser proficiency, particularly at the level of speech planning, and, consequently, lesser fluency of L2 compared to that of the juveniles. The large number of FPs found in the adults' speech as opposed to the juveniles is in agreement with the evidence attained by means of a careful phonetic analysis in Paananen-Porkka (2007:234), who argues that native speakers of Finnish show "longer pauses on average in English than in Finnish".

On the basis of the highly skewed distribution of the FPs across the two groups we hypothesize that the juveniles have a much more varied syntactic repertoire than the adults, and that the adults have much more limited and idiosyncratic (ungrammatical or non-standard) syntactic patterns at their disposal; 'limited' in the sense that almost 40% of their trigram types include at least one FP. Our findings show that the elimination of all FPs from the data has no statistically significant impact on the trigram types used by the two groups. They clearly support the argument that the adults indeed have much more *idiosyncratic* syntactic patterns at their disposal than the juveniles, who have a much more *constrained* syntactic repertoire than the adults; 'constrained' in the sense that the juveniles conform to the rules of standard (acrolectal) grammar.

In line with the evidence discussed in Oomen & Postma (2001), we would assume that filled pauses are associated with lexical search phenomena, and that both filled pauses and repetitions may signal problems in constituent construction. Christenfeld (1996: 1237) argues that filled pauses seem to be a product of speakers deliberately monitoring or stopping to correct their speech; in other words, they are attending more to what they are saying (cf. also Schourup 1982). However, Oomen & Postma (2001: 1003) suggest that "the production of filled pauses and repetitions is governed by processes that operate relatively independently of the available attentional resources", and that FPs seem to be "automatic reactions to (temporal) problems in speech planning". It is therefore reasonable to conjecture that the overuse of FPs by the adults reflects problems in the construction of syntax at the level of speech planning, and that these difficulties manifest themselves as disfluent speech (FPs).

References

Bot de, K., W. Lowie & M. Verspoor. 2005. *Second language acquisition: An advanced resource book.* London: Routledge.

Chambers, J.K. 2003. *Sociolinguistic theory. Linguistic variation and its social implications.* Oxford: Blackwell.

Christenfeld, N.S. 1996. Effects of a metronome on the filled pauses of fluent speakers. *Journal of Speech and Hearing Research* 39, 1232-1238.
Clyne, M. & S. Kipp. 2006. Australia's community languages. *International J. Soc. Lang.* 180, 7-21.
Cook, V. (ed.). 2003. *Effects of the second language on the first.* Clevedon, UK: Multilingual Matters.
Ellis, R. 1994. *The study of second language acquisition.* Oxford: Oxford University Press.
Fillmore, C. & P. Kay. 1999. Grammatical construction and linguistic generalizations. The *what's x doing y* construction. *Language* 75, 1-33.
Garside, R., G. Leech & T. McEmery. 1997. *Corpus annotation. Linguistic information from computer text corpora.* London: Longman.
Hirvonen, P. 2001. Doni finished – meni läpi – highskoulun. Borrowing, code-switching and language shift in American Finnish. In P. Sture Ureland (ed.) *Global Eurolinguistics. European languages in North America – Migration, maintenance and death*. Tübingen: Niemeyer. Pp. 297-324.
Jarvis, S. 2003. Probing the effects of the L2 on the L1. A case study. In V. Cook (ed.) *Effects of the second language on the first.* Clevedon, UK: Multilingual Matters. Pp. 81-102.
Karttunen, F. 1977. Finnish in America. A case study in monogenerational language change. In B.G. Blount & M. Sanches (eds.) *Sociocultural dimensions of language change.* New York: Academic Press. Pp. 173-184.
Kemppainen, J. 2000. Lexical features in the spoken English of Finnish Australians. Unpublished MA thesis, University of Oulu.
Klintborg, S. 1999. *The transience of American Swedish.* Lund: Lund University Press.
Laakkonen, K. 2000. A study of the realization of rhythm in the English of Finnish Australians. Unpublished MA thesis, University of Oulu.
Lahti, H. 1999. Lexical errors in the spoken English of Finnish Australians. Unpublished MA thesis, University of Oulu.
Lauttamus, T., J. Nerbonne, & W. Wiersma. 2007. Detecting syntactic contamination in emigrants. The English of Finnish Australians. *SKY Journal of Linguistics* 20, 273-307. [online version on-line at <www.ling. helsinki.fi/sky/julkaisut/SKY2007/LAUTTAMUS%20ET%20AL.pdf]
Lenneberg, E. 1967. *Biological foundations of language.* New York: John Wiley.
Levänen, H. 2008. The use of articles in the spoken English of first and second generation Finnish Australians. Unpublished MA thesis, University of Oulu.
Mannila, T. 1999. A study of phonic interference from Finnish in the English of Finnish Australians. Unpublished MA thesis, University of Oulu.

Markos, M. 2004. '*No, no swearing, no swearing allowed*'. A comparative study of hesitation phenomena in the spoken English of two generations of Finnish Australians. Unpublished MA thesis, University of Oulu.

Nerbonne, J. & W. Wiersma. 2006. A measure of aggregate syntactic distance. In J. Nerbonne & E. Hinrichs (eds.) *Linguistic distances*. Shroudsburg, PA: ACL. Pp. 82-90.

Oomen, C.C.E. & A. Postma. 2001. Effects of divided attention on the production of filled pauses and repetitions. *Journal of Speech, Language, Hearing Research* 44, 997-1004.

Paananen-Porkka, M. 2007. Speech rhythm in an interlanguage perspective: Finnish adolescents speaking English. Pragmatics, Ideology and Contact Monographs. PhD dissertation, University of Helsinki.

Piller, I. 2002. Passing for a native speaker. Identity and success in second language learning. *Journal of Sociolinguistics* 6, 179-206.

Schourup, L. 1982. Common discourse particles in English conversation. PhD dissertation, The Ohio State University, Columbus, Ohio.

Sells, P. 1982. *Lectures on contemporary syntactic theories.* Stanford: CSLI.

Smits, C. 1996. *Disintegration of inflection. The case of Iowa Dutch.* The Hague: Holland Academic Graphics.

Schmid, M. 2002. *First language attrition, use and maintenance. The case of German Jews in Anglophone countries*. Amsterdam: John Benjamins.

Schmid, M. 2004. Identity and first language attrition. A historical approach. *Estudios de Siciolingüística* 5, 41-58.

Siemund, P. 2008. Language contact: Constraints and common paths of contact-induced language change. In P. Siemund & N. Kintana (eds.) *Language contact and contact languages*. Amsterdam: John Benjamins. Pp. 3-11.

Thomason, S.G. & T. Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Van Coetsem, F. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.

Veltman, C. 1983. *Language shift in the United States.* Berlin: Mouton.

Waas, M. 1996. *Language attrition Downunder. German speakers in Australia*. Frankfurt: Peter Lang.

Watson, G. 1996. The Finnish-Australian English Corpus. *ICAME Journal: Computers in English Linguistics* 20, 41-70.

Weinreich, U. [1953] 1974. *Languages in contact.* The Hague: Mouton.