A Comparison of English Vocabulary in the Spoken Production of One Non-native Long-residency Group, One Non-native University Group and One Native Group in Three Tasks

Britt Erman and Margareta Lewis, Stockholm University

Abstract
This study investigates the rate of high- and low-frequency words, types and tokens, in L2 spoken English by two advanced Swedish groups, one studying English at a Swedish university and one living and working in London. They were compared to an English native group and to one another. The material had a multitask design: a role play, an interview and an online-narration. The results for the two non-native groups showed vulnerability on both frequency levels depending on task. Only the London group was nativelike across frequencies in the role play confirming our expectation. The results also show that the London group performed the tasks with a higher degree of lexical diversity and fluency than the university group.

Keywords: advanced English vocabulary; L2 vocabulary acquisition; fluency in L2 spoken production; high-/low-frequency words; lexical richness; nativelike performance

1. Introduction
1.1. Productive Vocabulary

A large productive vocabulary is "essential for effective communication and a large receptive vocabulary, probably made of thousands of words, needs to be in place." (Pignot-Shahov 2012: 43). Along the same lines, Schmitt advocates: "Learners need large vocabularies to successfully use a second language and so high vocabulary targets need to be set and pursued." (2008: 353).

Vocabulary is an area of L2 acquisition that has received increasing attention. In a research project involving several L2 languages of which this study is a part there has been a focus on nativelike proficiency[1] (Lindqvist 2010, Bardel et al. 2012, Forsberg Lundell and Lindqvist

---

2012, Lindqvist et al. 2013, Bardel and Gudmundson 2018, Erman and Lewis 2018). The present study compared high- and low-frequency vocabulary from different frequency bands in the oral production of two groups of advanced non-native Swedish speakers of English (NNS) and one native English-speaking group (NS). After accomplishing secondary education including 9 years of English as a foreign language, the two NNS groups had different experiences using English. One group had lived and worked in London for an average of seven years (the London Swedes; henceforth LS); the other had completed an academic year at an English department in Sweden to become teachers of L2 English (the University students; henceforth US). The NNS groups were first compared to the NS group, and then to one another. The aim of the study was to investigate the vocabulary of these groups in different situations. To this end, three tasks were included: an interview, a role play, and an online narration of a clip from a sequence of a silent film.

This investigation of advanced L2 users is one of several focusing the spoken production of the same groups as in this study. These involved the use of voiced pauses (Erman and Lewis 2013) and formulaic language (Erman and Lewis 2018) in two of the tasks. Analyzing different aspects of the vocabulary of the same participants doing the same tasks makes this study different from earlier research of advanced vocabulary in spoken production and is a particular strength of this research. Like earlier work our study includes results from frequencies of tokens (cf. Lindqvist 2010, Lindqvist et al. 2011, Bardel et al. 2012, Lindqvist et al. 2013, Forsberg Lundell and Lindqvist 2012, Bardel and Gudmundson 2018).

Frequencies of tokens in this study are related to fluency. Fluency is a complex concept which is rarely discussed or defined in a unified manner (see Freed 1995: 123). Attempts have nevertheless been made, notably, Fillmore (1979). Among features characteristic of a fluent speaker Fillmore mentions someone who easily fills time with talk (1979: 93). Fillmore also suggests that fluency is related to the appropriateness of the words used in the context. In our study tokens are real words, since all non-words, hesitations etc. have been removed (3.4). In other words, the words produced have meaning, are coherent and related to the topic and situation of each task.

Besides tokens frequencies of types are included in our study. Types are considered to relate to lexical diversity referring to the variation of

words in a text (Malvern et al. 2004). It is defined as the number of different words that a speaker uses (Jarvis 2013). According to Jarvis the term lexical diversity is synonymous with the term lexical variation. Lexical variation is frequently measured in "the proportion of words in a language sample that are not repetitions of words already encountered" (2013: 44). It is in this sense that types are applied in this study.

1.2. Background

In the Swedish compulsory school up to the age of 15, the first 2000 English words were advised to be actively learnt (see Thorén 1967, Mobärg 1997, both based on Thorndyke and Lorge 1944, and West 1953). Students at the end of upper secondary education, aged 18, were supposed to have actively learnt at least 2400 more words, totaling around 4500 words. Since his frequency-based recommendation was meant as advice for teachers Thorén, as a former Director of Education, discussed reasons, such as usefulness, for modifying and deviating from the prevailing frequency-based selection of words. In today's curricula in Sweden there are no numbers specified, but the focus has shifted towards a goal-oriented curriculum without mentioning methods or materials. A follow-up of active vocabulary knowledge in schools made as early as the 1950s and 60s testified to a declining trend, which might still be going on and in part be an explanation of our results.

1.3. Plan of the Article

The article starts by an account of vocabulary research with a focus on advanced L2 speakers' oral production (Section 2). After a presentation of participants and tasks (3.1 and 3.2), and research questions, hypotheses (3.3), method and limitations (3.4), we show the results from the two main frequency ranges (4). They are divided into the high-frequency range, 1-2000 words (4.1) and the low-frequency range, beyond 2000 words (henceforth 2000+) (4.2). In these sections the results from comparisons with the native group are presented. Section 4.3 shows results from comparisons between the two NNS groups. The main results are discussed in section 5.

2. Previous Research
2.1. Earlier Studies of Productive Vocabulary Knowledge

Although receptive vocabulary is indeed a pre-requisite for productive vocabulary, in this paper the focus is on productive vocabulary knowledge (Nation 2001, Corson 1995). Vocabulary knowledge related to different proficiency levels in L2 spoken production has been of interest the last few decades in vocabulary research (Daller et al. 2003, Milton 2007, Tidball and Treffers-Daller 2007, Lindqvist 2010, Lindqvist et al. 2011, Bardel et al. 2012, Lindqvist et al. 2013, Erman and Lewis 2013, 2015, Bardel and Gudmundson 2018.). The Lexical Frequency Profile used in some of these studies was developed by Laufer and Nation (1995) and displays the quantity of high- and low-frequency vocabulary in texts.

An assumption in most studies of vocabulary in relation to frequency is that frequency of input will affect output, so that the more frequent a word is the more likely it is to appear in an L2 speaker's production (Cobb and Horst 2004, Vermeer 2004). There is evidence that frequency plays an important role in L2 acquisition, implying that high-frequency words are shared by more L2 users than low-frequency words (Ellis 2002, Tidball and Treffers-Daller 2007). Although not further dealt with here, there are other factors to consider, such as personality, for example open-mindedness and aptitude (Forsberg Lundell and Sandgren 2013 regarding collocations). Also, similarity between L1 and L2, for example Swedish–English, facilitates the acquisition of L2 vocabulary (cf. Ringbom 1998, 2007, Jarvis 2000).

Segalowitz and Freed (2004) investigated fluency in the speech of native speakers of English studying L2 Spanish during one semester in the target language country. This group was compared to a group of learners studying L2 Spanish in their home country. The results showed that the 'study abroad' group at the end of the semester had made significant gains in fluency in that longer stretches without pauses were produced.

Apart from fluency, diversity in the sense of using different words, is of interest when evaluating proficiency. In the retelling of video films Lindqvist (2012: 269–270) investigated Swedish speakers of L3 French and compared them to native speakers. She found that the native speakers had a greater variation in their choice of words, using words with more precise as opposed to general meanings. As shown in

Lindquist more specialized vocabulary may not be as readily accessed by the L3 user, an observation that could be applicable to the online narration task in our study. In a case study of a very advanced L3 user of Italian Bardel and Gudmundson, despite a nativelike proportion of low-frequency words (types), found low-frequency words not present in the native data (2018).

According to Laufer (1995) the higher the percentage of words beyond the 2000 most frequent words in an L2 user's production, the more advanced is this person's vocabulary. Indeed, results from vocabulary studies have shown that quantitative results of frequency bands are able to distinguish not only native and non-native speakers but also speakers at different proficiency levels in L2 French (Forsberg Lundell and Lindqvist 2012), and L2 Italian (Bardel et al. 2012), both involving Swedish speakers of these L2 languages.

In view of the online narration task given its time constraints, two earlier works are of particular interest, notably Shaw and McMillion's study of reading comprehension (2008), and Hincks's study of spoken production (2010). Both involved Swedish speakers of advanced L2 English. The results showed that the Swedish participants reached nativelike levels of reading comprehension, but they needed more time than the native speakers. Hincks showed that the Swedish participants produced fewer information units in L2 English presentations than in the corresponding L1 Swedish ones. In other words, they needed more time using their L2.

Also relevant for the Retelling task, on a more general level, is the following observation made by Long: "it seems that lexical voids and collocation errors will be less easy to conceal in longer spontaneous speech samples […], especially under speeded conditions, when the NNS is less adept at planned discourse and avoidance strategies." (1990: 273).

2.2. Previous Studies of the Same Material

Focusing the same three groups as in the present study in two of the tasks (the role play and the narration task; see 3.2), Erman and Lewis (2013) found that the university group used significantly more voiced pauses ('vocalizations') in the role play than the London group and the native group, while both NNS groups significantly overused voiced pausing in the narration task compared to the NS group.

Using the same material, with a focus on the LS and NS groups, Erman and Lewis (2015) compared L2 English multiword structures and the production of low-frequency words. They found that the LS group was nativelike on both multiword structures and low-frequency words in the role play task, while they significantly underused them in comparison with the NS group in the narration task. In a more recent study of multiword structures of the same two groups on an individual level it was shown that the results of all the LS participants matched those of the native speakers in the role play, and three participants were within the native range also in the narration task (Erman and Lewis 2018).

3. The Study
3.1. Participants
The material consists of the spoken production of 20 speakers from two non-native groups and 10 from a native group (Table 1) performing three tasks.

Table 1. Participants

Informants	Time with English	Average age
10 Native speakers	All their lives	32 (28 – 38)
10 London Swedes	9 years of basic and secondary education, and an average of 7 years' residency in London	32 (28 – 38)
10 Swedish university students	9 years of basic and secondary education, and one year at an English department in Sweden	26 (20 – 32)

The selection of the London Swedes met the following three criteria: they should 1) have completed upper secondary studies in Sweden, entailing at least 9 years of English as a foreign language at school, 2) have lived and worked for at least five years in the target language country, and 3) at the time of the recording be resident in the target language country, using the L2 as a principal means of communication. Most of the London Swedes had experience of academic studies and some had received some formal instruction of English in England. Even though people growing up in Sweden have had a great deal of exposure

to English from childhood through input from films, TV, music, computers including the Internet, English has the status of a foreign language in Sweden.

The university students had studied English at an English department in Sweden and were selected from the same teacher training program. The following criteria were met: age between 20 and 35; L1 Swedish; L2 English learnt in the Swedish school system (9 years of instruction) and one year full-time English language studies at the university; a maximum stay of three months in an English-speaking country and no English-speaking parent or partner.

3.2. Tasks

The first task was an interview (hence *I*nterview). The questions included biographical data, such as their experience with English, and their current work and family situation, in other words questions pertaining to everyday life. Some of the questions focused on cultural differences between Sweden and the UK.

The second task was a role play (hence *R*ole play) in which the participants were instructed to find an acceptable solution to a problem. This task involved two speakers, a native-speaking manager and an employee (native or non-native). The employee was a legal expert phoning her/his manager to ask for two days' leave to attend a close relative's wedding at a time which clashed with an important company meeting. The participant was given five minutes to read and contemplate the instructions before making the call. The mean duration of the task was five minutes, but given its open-ended format there was individual variation.

The third task involved an online retelling of the first 14½ minutes from Charlie Chaplin's silent film *Modern Times* (hence *R*etelling task). In this task, involving time pressure, the participants were told to describe what they saw on the screen to someone who could not see it.

All three tasks were recorded and subsequently transcribed. The three tasks not only put different demands on the participants but also involved different situations and topics. The Interview and the Role play were both interactive but in different ways. The Interview gave the interviewee the freedom to expand on any topic associated to the questions asked. In this task the interviewee was in focus while the

interviewer asked questions and gave feedback. In the Role play the interactants had been given instructions for how the role was to be enacted. The employee making the request was in an inferior position since the employer had the right to grant or reject the request, which may have had an impact on the production. The Retelling task differed from the other two tasks in that the participant was requested to describe a film sequence that s/he, presumably, was not familiar with.

The material comprises approximately 115,000 words, which in view of existing specialized spoken corpora is an appreciable size (cf. Lindqvist et al. 2011, Bardel et al. 2012). The present study like several earlier studies (see section 2) has used the Lexical Frequency Profile (hence LFP). The LFP, accessible via www.lextutor.ca (Cobb), sorts the words of the transcribed texts into frequency bands. By feeding the texts into this program, we also get the total number of words over the three groups and tasks (Table 2).

Table 2. Number of words in the three tasks for the native speakers (NS), London Swedes (LS), University students (US)

Tasks/Participants	NS	LS	US	Total
Interview	21547	23753	10986	56286
Role play	3264	3138	4014	10416
Retelling task	16773	15513	15951	48237
Total	41584	42404	30951	114939

The NS and LS groups had the same native speaker acting as the interviewer, whereas for the US group a new native speaker was recruited, which resulted in differences in the number of words in this task.

3.3. Research Questions and Hypotheses
In this study two research questions were asked. Using the native English-speaking group (NS) as benchmark, the first research question was: How does the vocabulary of the two NNS groups compare with the NS group? The second research question was: How does the vocabulary of the two NNS groups compare to one another. As mentioned, two main frequency bands were examined: the first two thousand words (1-2000 frequency range; here referred to as high-frequency), and those beyond

the first two thousand words (the 2000+ frequency range; here referred to as low-frequency).

With regard to the first research question our hypothesis was that both NNS groups would be nativelike regarding types and tokens in the high-frequency range in all three tasks. This hypothesis was based on Thorén's estimate that the first 2000 words were to be learnt by the age of 15. Regarding low-frequency words (types and tokens) our expectations were different for the two NNS groups. Both NNS groups were expected to be nativelike on low-frequency words in the Interview, which mainly concerned familiar topics. Only the London Swedes having had more exposure to English and having used the language more regularly were expected to be nativelike in the Role play in this frequency range. Furthermore, this task, involving a request over the phone, was expected to be a fairly common situation for someone living in the target language country, which was the situation for the London Swedes. The Retelling task, involving an unusual situation, and performed under time pressure, was the cognitively most challenging task. In this task both NNS groups were expected to underuse low-frequency words compared to the NS group, types and tokens.

3.4. Method and Limitations

In LFP all the words are registered in terms of type and token frequency and listed alphabetically. The words have not been lemmatized, which means that type frequencies equal 'word forms'; for example, *museum, museums,* and *call, calls, called, calling* are all registered as six separate types, while representing two lemmas. The LFP sorts the word forms into four categories (or lists): the first most frequent 1000 words (K1 in Lextutor), the second most frequent 1000 words (K2 in Lextutor), the Academic Word List (AWL; Coxhead 2000), and the Off-list. The AWL list was compiled from a corpus of 3.5 million words of written academic texts from four areas: Arts, Science, Law and Commerce (Nation 2001), not included in the 2000 most frequent words. The Off-list comprises any word beyond the 2000 most frequent words and the words in the AWL. A close study of Lextutor types in the present material showed that the majority (approximately 85%) belong to different lemmas, although some high-frequency types are inflections of one and the same lemma as in the examples above.

In the literature calculations sometimes include only number of tokens (Bardel et al. 2012, Lindqvist et al. 2013, Forsberg Lundell and Lindqvist 2012), but, as mentioned, in this study the number of types was also included. For instance, on some measurements the NNS groups were nativelike on the number of tokens, but not on the number of types, which is an indication that the NNS group recycled their types more often, implying less diversity in these groups (see Lindqvist 210: 415 for the importance of including types).

The 2000+ frequency range in the present study needs some clarification. It is composed of a pruned version of the words in the Off-list combined with the words in the AWL list. The Lextutor Off-list is a heterogeneous group of items, which includes both informal words (see below) and voiced pauses (transcribed as e.g. *uhm*). To avoid a situation where words, because they are beyond the frequency bands of the first 2000 words, would unduly be considered advanced, certain items were removed (cf. Lindqvist 2010, Lindqvist et al. 2013). Therefore, all the words in the Lextutor Off-list that were of an informal character or deemed as not being part of a language vocabulary, such as voiced pauses and word fragments, were removed. Indeed, equating the Off-list words with lexical richness can be misleading (Lindqvist 2010). The words in AWL, comprising 570 word families, make up the smallest proportion of the words in all three tasks and for all three groups, covering between 1% and 2% of the texts.

The following types of items in the Off-lists have been removed: names (of people, regions, places, continents, countries including languages and nationalities, many of which are similar in Swedish and English, therefore presumably more readily accessible; cf. Horst and Collins 2006, Milton 2007, Lindqvist et al. 2013), feedback words (*yea, yeah, ok, huh, mm*), foreign words (*cher*), contractions (*wanna, gonna, gotta, coz*), swear words (*fucking*), slang words (*kids, guys, crap, ass*), voiced pauses (*eh*, *uh/uhm/um(m))*, and, finally, fragments of words (*Thur*, *archi,* and so on).

A limitation of the LFP is that it is based on written texts and the present material is spoken production. It may be the case that certain words belong to different frequency bands in oral and written production (cf. McCarthy 1998). Therefore, a selection of words from the first and second thousand most frequent words in our material were matched with those marked as spoken 1 and 2 ('S1, S2') in 'Longman Communication

3000'. 'Longman Communication 3000' lists the 3000 most frequent words in spoken and written language and can be accessed as a pdf from the Internet. This pdf is based on a corpus of 390 million words (the Longman Corpus Network) of authentic English language. The matching procedure involved selecting words from the first and second frequency bands (the LexTutor lists) with the initial letters *a, f, p,* and *s*, which were the letters with most entries in our material. Not surprisingly, quite a few high-frequency words overlapped in the two modes, especially the first thousand words containing many grammatical and other basic words. There was an agreement of between 75% and 100% in our material with Longman's list, which did not include months and weekdays. It is worth observing that the majority of K1 and K2 words, outside S1 and S2 in the list, belong to S3, the next frequency level, and not to the words marked as written in Longman. Obviously, a corpus, written or spoken, is only as representative as that on which it is based. Our results should be viewed against this background.

Another limitation is that different interviewers carried out the Interview, which resulted in different text lengths for this task. With its focus on words the LFP is independent of syntax. However, the sound files and the transcribed texts have been checked for transcription errors, and the words used have been deemed as appropriate in the context.

4. Results and Analysis
The first 1000-word span constitutes the major part of any text and covers a good 90% in our spoken material, while the second 1000-word span covers about 5%. We start by accounting for types and tokens per hundred words pertaining to the 2000 most frequent words over the three tasks (section 4.1) followed by a corresponding account of the results from the 2000+ frequency range (section 4.2) The NNS results are here compared to those of the NS group, which was our first research question. The results from comparisons between the NNS groups, our second research question, are presented in section 4.3. The threshold for significance in Tables 3–9 is set at *p* < .05.[2]

---

[2] http://www.quantpsy.org/chisq/chisq.htm

4. 1. The High-frequency Range

Words belonging to the first 2000 words, here called high-frequency range, in particular the first thousand words, are indeed common in all types of texts, as in the present material. In Tables 3, 4, and 5 the number of types and tokens per 100 words are shown for this frequency range over the three tasks, starting with the Interview.

4.1.1. The Interview

Regarding tokens per 100 words both NNS groups reached nativelike results in the Interview, indicating fluency, while they differed significantly from the NS group on types per 100 words (Table 3).

Table 3. 1-2000 words; *Group comparison of the number and mean proportion of high-frequency types and tokens <u>in the Interview</u>*; ns=non-significant.

Types	N	*/100*	Types	N	*/100*	p-value
University students	911	8.3	Native speakers	1416	6.6	$p \leq 000$
London Swedes	1366	5.8	Native speakers	1416	6.6	$p \leq 001$
Tokens	N	*/100*	Tokens	N	*/100*	p-value
University students	10521	95.8	Native speakers	20470	95.0	ns
London Swedes	22712	95.6	Native speakers	20470	95.0	ns

However, they did so in different ways. The US group used significantly more and the LS group significantly fewer types than the NS group. The lower number of types means that the LS group recycled more words than the NS group while at the same time producing a nativelike number of tokens. The higher rate of types for the US group is presumably due to their lower number of words in this task (cf. Table 2) and is therefore not entirely comparable. According to McCarthy and Jarvis "the more tokens (words) a text has the less likely it is that new words (types) will occur" (2007: 460; see also Daller et al. 2003).

A substudy was carried out to establish if the difference in number of types between the LS and NS groups was found in the first or the second

1000 most frequent words. It showed that the underrepresentation of types in the LS group compared to the NS group was found in K1 (1-1000; 776 and 799 respectively, < p .01), while they were nativelike in K2 (1001–2000), a considerably smaller category in our material.

Our hypothesis that both NNS groups would be nativelike on high-frequency words in this task given its everyday character was thus only met in the number of tokens, an indication of fluency.

4.1.2. The Role Play

In the high-frequency range in the Role play, as in the Interview, both NNS groups performed like the NS group on tokens, indicating fluency in these interactive tasks (Table 4). The LS group was also nativelike on the number of types in this task.

Table 4. 1-2000 words; *Group comparison of the number and mean proportion of high-frequency types and tokens in the Role play*; ns=non-significant.

Types	N	*/100*	Types	N	*/100*	p-value
University students	406	10.1	Native speakers	391	12.0	$p \leq 02$
London Swedes	382	12.2	Native speakers	391	12.0	ns
Tokens	N	*/100*	Tokens	N	*/100*	p-value
University students	3935	98.0	Native speakers	3155	96.7	ns
London Swedes	3015	96.1	Native speakers	3155	96.7	ns

The US group used significantly fewer types per 100 words than the NS group, while having the same number of tokens, which indicates that they recycled more words. Separating K1 and K2 showed that it was in the number of K1 types that the US group deviated significantly from the NS group, whereas they performed similarly on K2 types.

Our hypothesis that both NNS groups would be nativelike was not supported, as only the LS group was nativelike on the number of types and tokens in the high-frequency range.

4.1.3. The Retelling Task

As in the Role play the LS group was nativelike on both types and tokens per 100 words in the Retelling task. The US group, however, used significantly fewer types, and significantly more tokens than the NS group (Table 5). This result shows that the US group recycled more words than the NS group.

Table 5. 1-2000 words; *Group comparison of the number and mean proportion of high-frequency types and tokens <u>in the Retelling task</u>*; ns=non-significant.

Types	N	/100	Types	N	/100	p-value
University students	936	5.9	Native speakers	1080	6.4	$p \leq 04$
London Swedes	1030	6.6	Native speakers	1080	6.4	ns
Tokens	N	/100	Tokens	N	/100	p-value
University students	15427	96.7	Native speakers	15699	93.6	$p \leq 04$
London Swedes	14899	96.0	Native speakers	15699	93.6	ns

After separating K1 and K2 for the US and the NS groups the result showed that there was no significant difference between the two groups in either category. As is clear from Table 5, however, K1 and K2 when merged yielded a significant difference, although close to the threshold for significance (p <.04). This illustrates the fact that higher numbers are more likely to give significant results.

Our expectation that both NNS groups would perform like the natives in this task was only met for the LS group.

4.2. The Low-frequency Range
The low-frequency range includes words beyond 2000 words, here 2000+. Tables 6, 7, and 8 show types and tokens /100 words in this frequency range for the three tasks across the groups, starting with the Interview.

4.2.1. The Interview
Both NNS groups significantly underused low-frequency types and tokens in the Interview compared to the NS group (Table 6).

Table 6. 2000+ words; *Group comparison of the number and mean proportion of low-frequency types and tokens <u>in the Interview</u>*; ns=non-significant.

Types	N	*I100*	Types	N	*I100*	p-value
University students	226	2.1	Native speakers	627	2.9	*p* ≤ 000
London Swedes	537	2.3	Native speakers	627	2.9	*p* ≤ 000
Tokens	N	*I100*	Tokens	N	*I100*	p-value
University students	465	4.2	Native speakers	1077	5.0	*p* ≤ 003
London Swedes	1041	4.4	Native speakers	1077	5.0	*p* ≤ 003

Our hypothesis that both NNS groups would reach a nativelike level of low-frequency words in view of the character of this task was not supported.

4.2.2. The Role Play
Only the LS group was nativelike on low-frequency types and tokens per 100 words in the Role play, thus confirming our expectation that making a request, even over the phone to your boss, would be a fairly common experience for someone living in the target language country.

Table 7. 2000+ words; *Group comparison of the number and mean proportion of low-frequency types and tokens <u>in the Role play</u>*; ns=non-significant.

Types	N	/100	Types	N	/100	p-value
University students	42	1.0	Native speakers	56	1.7	$p \leq 02$
London Swedes	59	1.9	Native speakers	56	1.7	ns
Tokens	N	/100	Tokens	N	/100	p-value
University students	79	2.0	Native speakers	109	3.3	$p \leq 000$
London Swedes	123	3.9	Native speakers	109	3.3	ns

The US group, however, underused types and tokens compared to the NS group, which met our expectation.

4.2.3. The Retelling Task

The results in Table 8 show that both NNS groups significantly underused both types and tokens of low-frequency words compared to the NS group in the Retelling task. This result confirmed our hypothesis that neither of the NNS groups would match the NS group, indicating that this task was cognitively demanding.

Table 8. 2000+ words; *Group comparison of the number and mean proportion of low-frequency types and tokens <u>in the Retelling task</u>*; ns=non-significant.

Types	N	/100	Types	N	/100	p-value
University students	260	1.7	Native speakers	440	2.6	$p \leq 000$
London Swedes	282	1.9	Native speakers	440	2.6	$p \leq 000$
Tokens	N	/100	Tokens	N	/100	p-value
University students	524	3.3	Native speakers	1074	6.4	$p \leq 000$
London Swedes	614	4.0	Native speakers	1074	6.4	$p \leq 000$

4.3. Comparisons Between the NNS Groups

As mentioned, the difference in the number of high-frequency types between the US and LS groups in the Interview may be due to the different text sizes, a short text generating more types (McCarthy and Jarvis 2007). This unbalance in text length in the Interview therefore prevents a straight comparison. In the Role play the only significant difference between the NNS groups was that the LS group used more high-frequency types than the US group. Since the text sizes are matched the results are comparable. This is also true for the results of the high-frequency types in the Retelling task, in which the LS group used significantly more types than the US group. In sum, in the Role play and Retelling task, the LS group used more types than the US group in the high-frequency range, while at the same time being similar in the number of tokens (cf. Table 9). This implies less repetition of words and more lexical diversity for the LS group. The statistical difference in the number of types between the US and LS groups was found among the K1 words, while there was no difference in K2. This result was expected in view of the considerably larger size of K1 in our material. An overview of a comparison between the two NNS groups is given in Table 9. For numbers see Tables 3-8.

Table 9. *Comparison of the NNS groups; significant and non-significant (ns) results*

High-freq	Interview		Role play		Retelling	
	Types	Tokens	types	Tokens	Types	Tokens
US-LS	.000	ns	.01	ns	.008	ns

Low-freq	Interview		Role play		Retelling	
	Types	Tokens	types	Tokens	Types	Tokens
US-LS	ns	ns	.003	.000	ns	.002

Regarding low-frequency words there was no significant difference between the NNS groups in either types or tokens in the Interview. In the Role play, however, the LS group produced significantly more types and tokens than the US group. In the Retelling task both NNS groups

behaved similarly on types, whereas the LS group used significantly more tokens than the US group, indicating fluency.

The LS group's higher number of high-frequency types in two tasks (the Role play and the Retelling task) and low-frequency types in one task (the Role play) indicates more lexical diversity. Similarly, the LS group's higher number of low-frequency tokens in two tasks (the Role play and the Retelling task) indicates more fluency than the US group.

5. Discussion and Summing up

The aim of the present study was to compare the L2 English vocabulary of two Swedish NNS groups with a native English group. The first research question involved establishing the rate of high-frequency (1-2000) and low-frequency (2000+) words (types and tokens) in the spoken production of two NNS groups (L1 Swedish), one group studying English at a Swedish university and one living and working in London, both being compared to an English native group. The second research question involved comparing the two NNS groups with each other. The material consisted of three tasks, two interactive tasks, a Role play and an Interview, and one Retelling task from the silent film *Modern times*.

Our hypothesis that the two NNS groups would be nativelike on high-frequency words was viewed against Thorén's recommendation (1967) that Swedish students were expected to actively know the two thousand most frequent words (i.e. types) after having completed compulsory school at the age of 15. It was therefore surprising to find that it was in the first 1000 most frequent words that the NNS groups significantly underused types compared to the NS group, the LS group in the Interview, and the US group in the Role play and the Retelling task. This result contradicts the claim by Cobb and Horst (2004) and Vermeer (2004) that input affects output so that the more frequent a word is the more likely it is to appear in an L2 speaker's production. The LS result for high-frequency types in the Interview was thus unexpected since they had been living and working in the target language country for several years at the time of the recording and were asked to talk about themselves. Consequently, in spoken production, it is in this frequency range that there is room for development for both non-native groups. The larger number of high-frequency types suggests that the NS group used a

wider vocabulary when being encouraged to talk freely about familiar topics, such as in the Interview.

The two NNS groups were nativelike on the number of tokens in the Interview and in the Role play indicating fluency in the two interactive tasks. Only the LS group was nativelike on both high-frequency types and tokens in the Role play and the Retelling task, supporting earlier results that vocabulary distinguishes proficiency levels (cf. Laufer 1995, Forsberg Lundell and Lindqvist 2012, Bardel et al. 2012). In the high-frequency range in these two tasks the LS group used significantly more types than the US group, while using a similar number of tokens, which implies less repetition and possibly a wider vocabulary for the LS group.

Low-frequency types as well as tokens were significantly underrepresented in the speech of both NNS groups compared to the NS group as apparent in the Interview and the Retelling task, and for the US group also in the Role play. The NNS results, especially for the Retelling task, corroborate Long's (1990) claim that it would be more difficult to hide lexical voids in long spontaneous speech sequences under time pressure.

Comparisons between the two NNS groups, which was our second research question, showed that the LS group used significantly more high-frequency types than the US group in the Role play and the Retelling task, while behaving similarly on tokens, which implies less repetition of words. Furthermore, in the Role play the LS group used significantly more low-frequency types compared to the US group, indicating more diversity. In the Role play and the Retelling task the LS group produced significantly more low-frequency tokens than the US group, which was suggested to indicate more fluency. For the Role play the US low-frequency result converges with Erman and Lewis 2013, who showed that this group had significantly higher rates of voiced pauses in this task. This was suggested to indicate more hesitation. Both NNS groups were thus equally fluent in the high-frequency range in all three tasks, while the LS group showed more lexical diversity in the high- and low-frequency ranges in the Role play, as well as in the high-frequency range in the Retelling task. In conclusion, the LS group showed a higher degree of lexical diversity and fluency. Again, as is clear from these results frequency bands distinguish proficiency levels not only between the NNS and NS groups but also between the NNS groups.

Time may have been a crucial factor in the present study to judge from the results for the Retelling task. In Hinck's 2010 study of L2 spoken presentations the Swedish L2 speakers of English produced less information within a specified time unit than in their corresponding L1 presentations. A similar result was found in Shaw and McMillion's study of reading comprehension where non-native speakers required more time (2008). An L2 speaker may thus need more time to perform in a nativelike way. Indeed, under different conditions, the advanced L2 speaker may well be able to produce words on a nativelike level.

The results from this study agree with results from measurements of formulaic language and voiced pauses based on partly the same material (Erman and Lewis 2013, Erman and Lewis 2018), showing that the LS group was nativelike in the Role play. Measurements of different aspects of L2 vocabulary contribute to making the results more robust. Although voiced pauses (*ehm, uhm, um*) may have different functions (cf. Denke 2009) the results in the Role play showed that the US group used significantly more than the NS and LS groups in this task, suggesting that the US group searched for words to a greater extent, or simply, were more uncomfortable talking on the phone.

The multi-task design involving three tasks with different characteristics performed by the same participants distinguishes this study from earlier work. Furthermore, both types and tokens have been included, thereby shedding light on two aspects of spoken L2 production, which in this study have been called lexical diversity (types) and fluency (tokens). In agreement with Jarvis (2003) types are here indicative of lexical diversity, while our notion of fluency, inspired by Fillmore (1979), is related to the number of tokens, more specifically the number of words appropriate in the context.

In view of the results for the NNS participants in the present study regarding the most frequent English words it would be interesting to know if there is indeed an ongoing decline in active English vocabulary knowledge as suggested by Thorén. This would indicate that more emphasis on vocabulary knowledge in education is called for. In a world of increasing migration and an increasing demand for L2 competence, future studies focusing on accessibility of vocabulary, mental processes involved in production, comparisons between the speech rate and role of L1s and other background languages are welcome.

References

Bardel, Camilla, Anna Gudmundson and Christina Lindqvist. 2012. "Aspects of lexical richness in advanced learners' oral production: vocabulary acquisition and use in L2 French and Italian." *High-level L2 Acquisition, Learning and Use.* Eds. Niclas Abrahamsson and Kenneth Hyltenstam. *Studies in Second Language Acquisition* 34(2): 269–290.

Bardel, Camilla and Anna Gudmundson. 2018. "Developing lexical complexity in spoken production: Limitations and possibilities of the advanced L2 learner." *High-Level Language Proficiency in Second Language and Multilingual Contexts.* Eds. Kenneth Hyltenstam, Inge Bartning and Lars Fant. Cambridge University Press. 120–145.

Cobb, Tom and Marlise Horst. 2004. "Is there room for an academic word list in French?" *Vocabulary in a Second Language: Selection, Acquisition, and Testing.* Eds. Paul Bogaards and Batia Laufer. Amsterdam: John Benjamins. 15–38.

Corson, David J. 1995. *Using English Words.* Dordrecht: Kluwer Academic Publishers.

Coxhead, Avril. 2000. "A new academic word list." *TESOL Quarterly* 34(2): 213–238.

Daller, Michael Helmut, Roeland van Hout and Janine Treffers-Daller. 2003. "Lexical richness in spontaneous speech of bilinguals." *Applied Linguistics* 24 (2): 197–222.

Daller, Michael Helmut, James Milton and J. Treffers-Daller. 2007. *Modelling and Assessing Vocabulary Knowledge.* Cambridge: Cambridge University Press.

Denke, Annika. 2009. *Nativelike Performance: A Corpus Study of Pragmatic Markers, Repairs and Repetition in Native and Non-native English Speech.* Saarbrücken: VDM Verlag.

Ellis, Nick. 2002. "Frequency effects in language processing". *Studies in Second Language Acquisition* 24(2): 143–188.

Erman, Britt and Margareta Lewis. 2013. "Vocabulary in advanced L2 English speech". *Of Butterflies and Birds, Dialects and Genres.* Eds. N.-L. Johannesson, G. Melchers and B. Björkman. Stockholm Studies in English 104. Acta Universitatis Stockholmiensis. 93–108.

Erman, Britt and Margareta Lewis. 2015. "L2 English vocabulary in a long-residency Swedish group compared to a group of English native speakers." *Cultural migrants and optimal language acquisition.* Eds.

Fanny Forsberg Lundell and Inge Bartning. Bristol/Buffalo/Toronto: Multilingal Matters.115–134.

Erman, Britt and Margareta Lewis. 2018. "Formulaic language advanced long-residency L2 speakers." *High-Level Language Proficiency in Second Language and Multilingual Contexts.* Eds. Kenneth Hyltenstam, Inge Bartning and Lars Fant. Cambridge University Press. 96–119.

Fillmore, Charles, J. 1979. "On fluency". *Individual Differences in Language Ability and Language Behavior.* Eds. C.J. Fillmore, D. Kempler and W. Wong. Academic Press. 85–101.

Forsberg Lundell, Fanny and Christina Lindqvist. 2012. "Vocabulary development in advanced L2 French: do formulaic sequences and lexical richness develop at the same rate?" *Language, Interaction, Acquisition (LIA)* 3(1): 73–92.

Forsberg Lundell, Fanny and Maria Sandgren. 2013. "High-level proficiency in late L2 acquisition – Relationships between collocational production, language aptitude and personality." *Sensitive Periods, Aptitudes and Ultimate Attainment in L2.* Eds. Gisela Granena and Michael Long. Amsterdam: John Benjamins. 231– 258.

Freed, Barbara. 1995. "What makes us think that students who study abroad become fluent?" *Second Language Acqusition in the Study of Abroad Context.* Ed. Barbara Freed. Amsterdam: John Benjamins. 123–148.

Hincks, Rebecka. 2010. "Speaking rate and information content in English lingua franca oral presentations." *English for Specific Purposes* 29(1): 4–18.

Horst, Marlise and Laura Collins. 2006. "From 'faible' to strong: How does their vocabulary grow?" *The Canadian Modern Language Review* 63(1): 83–106.

Jarvis, Scott. 2000. "Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon." *Language Learning* 50(2): 245–309.

Jarvis, Scott. 2013. "Defining and measuring lexical diversity". *Vocabulary Knowledge: Human Ratings and Automated Measures.* Eds. Scott Jarvis and Michael Daller. Amsterdam: John Benjamins. 13–44.

Laufer, Batia. 1995. "Beyond 2000: A measure of productive lexicon in a second language." *The Current State of Interlanguage: Studies in Honor of William E. Rutherford.* Eds. Eubank, L., L. Selinker and M. Sharwood Smith. Amsterdam: John Benjamins. 265–272.

Laufer, Batia and Paul Nation. 1995. "Vocabulary size and use: Lexical richness in L2 written production." *Applied Linguistics* 16: 307–322.

Lextutor. (n.d.). *Compleat Lexical Tutor v. 8.3.* Accessed November 21, 2017. https://www.lextutor.ca

Lindqvist, Christina. 2010. "La richesse lexicale dans la production orale de l'apprenant avancé de français." *La Revue Canadienne des Langues Vivantes/The Canadian Modern Language Review* 66(3): 393–420.

Lindqvist, Christina, Camilla Bardel and Anna Gudmundson. 2011. "Lexical richness in the advanced learner's oral production of French and Italian L2." *International Review of Applied Linguistics (IRAL)* 49: 221–240.

Lindqvist, Christina. 2012. "Advanced learners' word choices in French L3". *Third Language Acquisition in Adulthood.* Eds. Jennifer Cabrelli Amaro, Susanne Flynn and Jason Rothman. Amsterdam: John Benjamins. 255–280.

Lindqvist, Christina, Anna Gudmundson and Camilla Bardel. 2013. "A new approach to measuring lexical richness in L2 oral production." *Eurosla Monographs Series* 2: 109–126.

Long, Michael. 1990. "Maturational constraints on language development." *SSLA,* 12, 251–285.

Longman Communication 3000. 2007. Accessed April 10, 2018. https://www.lextutor.ca/freq/lists_download/longman_3000_list.pdf

Malvern, David, Brian Richards, Ngoni Chipere and Pilar Durán. 2004. *Lexical Diversity and Language Development: Quantification and Assessment*. Basingstoke: Palgrave Macmillan.

McCarthy, Philip, M. and Scott Jarvis. 2007. "vocd: A theoretical and empirical evaluation." *Language Testing* 24(4): 459–488.

McCarthy, Michael, H. 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.

Milton, James. 2007. "Lexical profiles, learning styles and the construct validity of lexical size tests." *Modelling and Assessing Vocabulary Knowledge.* Eds. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge: Cambridge University Press. 133–149.

Mobärg, Mats. 1997. "Acquiring, teaching and testing vocabulary". *International Journal of Applied Linguistics* 7(2):201–222.

Nation, Paul. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Pignot-Shahov, Virginie. 2012. "Measuring L2 receptive and productive vocabulary knowledge". *Language Studies Working Papers*. Eds. C. Ciarlo and D.S. Giannoni. University of Reading. 4: 37–45.

Preacher, Kristopher, J. 2001. "Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence." Accessed June 7, 2018. https://www.quantpsy.org/chisq/chisq.htm

Ringbom, Håkan. 1998. "Vocabulary frequencies in advanced learner English: a cross-linguistic approach." *Learner English on Computer*. Ed. S. Granger. London & New York: Longman. 41–52.

Ringbom, Håkan. 2007. *Cross-linguistic Similarity in Foreign Language Learning*. Bristol/Buffalo/Toronto: Multilingal Matters

Segalowitz, Norman and Barbara, F. Freed. 2004. "Context, contact and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts." *Studies in Second Language Acquisition* 26 (2): 173–199.

Shaw, Philip and Alan McMillion. 2008. "Proficiency Effects and Compensation in Advanced Second-Language Reading." *Nordic Journal of English Studies* 7(3): 123–143.

Schmitt, Norbert. 2008. "Review article: Instructed second language vocabulary learning". *Language Teaching Research* 12(3): 329–363.

Thorén, Birger. 1967. *10 000 ord för tio års engelska*. Lund: Gleerups.

Thorndyke, E. L. and I. Lorge. 1944. *The Teacher's Word Book of 30,000 Words*. Columbia University: Teachers College.

Tidball, Françoise and Jeanine Treffers-Daller. 2007. "Exploring measures of vocabulary richness in semi-spontaneous French speech." *Modelling and assessing vocabulary knowledge*. Eds. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge: Cambridge University Press. 133–149.

Vermeer, Anne. 2004. "The relation between lexical richness and vocabulary size in Dutch L1 and L2 children." *Vocabulary in a Second Language: Selection, Acquisition, and Testing*. Eds. Paul Bogaards and Batia Laufer. Amsterdam: John Benjamins. 15–38.

West, Michael. 1953. *A General Service List of English Words*. London: Longman, Green & Company.