

English vocabulary in L1 Danish and L1 Finnish learners: Vocabulary sizes, word frequency effect, and cognate facilitation

Camilla Søballe Horslund
University of Amsterdam¹

Abstract

The study presents vocabulary sizes in native (L1) Danish and L1 Finnish learners of English differing in second language (L2) immersion. The estimated vocabulary sizes suggest that some L2 learners have vocabulary sizes within the L1 English range, and that all participants should be lexically equipped to understand spoken English. The article moreover examines the effect of word frequency and cognateness on L2 lexical knowledge and how these two effects are mediated by L2 immersion. Word frequency was found to significantly affect word definition. Contrary to the prediction, this effect was larger for L1 English speakers than for L2 learners and for immersion learners than for non-immersion learners. Significant cognate facilitation was also observed and was found to be larger for non-immersion learners than for immersion learners, as predicted.

1. Introduction²

Native speakers vary in their vocabulary size, but a conservative estimate is that adult native speakers know approximately 16,000-20,000 word families. A word family consists of a lexical root along with its derivations and inflections (Schmitt 2010: 8). The word family unit is considered more appropriate for vocabulary estimates than the lemma unit (the lexical root and its inflections), because learners beyond a minimal proficiency

¹ This article presents part of my PhD project, which was conducted at Aarhus University.

² Many thanks to Johanna Wood for help and advice in the choice of test and for comments and suggestions concerning the analysis.

level show some knowledge of word formation and are able to use this knowledge in deducing word meaning within a word family (Bertram, Laine & Virkkala 2000). Variation in L1 vocabulary sizes are primarily related to age and education (Zechmeister, Chronis, Cull, D'Anna & Healy 1995; Diack 1975: 12). It is estimated that L2 learners need a vocabulary size of 6,000 to 7,000 word families to understand spoken English and a vocabulary size of 8,000 to 9,000 word families to be able to read a novel or a newspaper unaided in English (Nation 2006).

The finding that highly frequent words are recognised faster than low frequency words is commonly referred to as the *Word Frequency Effect* in psycholinguistics. Word frequency effects are typically explained in terms of implicit learning. Repeated exposure to high frequency words is believed to strengthen the lexical representation of these words, thereby making them more readily accessible (Whitford & Tytone 2012). Word frequency effects have been found in a number of lexical tasks in the L1 (Schilling, Rayner & Chumbley 1998) and in speakers from different L1 backgrounds in their L2 (Lemhöfer, Dijkstra, Schriefers & Baayen 2008). Interestingly, research suggests that the effect of word frequency is stronger for L2 learners than for L1 speakers (Van Wijnendaele & Brysbaert 2002; de Groot, Borgwaldt, Bos & van den Eijnden 2002). Moreover, Whitford & Tytone (2012) found that the effect of word frequency on L2 reading is larger in relatively inexperienced L2 learners compared to relatively experienced L2 learners, suggesting that the effect of frequency in L2 lexical tasks is mediated by L2 experience.

Another word characteristic that affects the learnability of a word is *Cognateness*. Comensaña, Sánchez-Casas, Soares, Pinhero, Rauber, Frade & Fraga (2012: 75) define cognates as 'equivalent translations that share both form and meaning (e.g. *papel* in European Portuguese and *paper* in English)'.³ Such word similarities across languages may stem from common origins, from borrowings, or from sheer chance. A large body of research (e.g. Lemhöfer et al. 2008; Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen 2010; Balling 2012) suggests that the presence of cognates facilitates lexical tasks in the L2. Moreover, Casaponsa, Antón, Pérez, & Duñabeitia (2015) found that cognate facilitation in a lexical decision task was larger for relatively inexperienced L2 learners than

³ This is a psycholinguistic definition of cognateness, since the focus is on the psycholinguistic status of a word for L2 learners and not on the historical origin of the word as in the use of the term in historical linguistics (e.g. van Gelderen 2006: 34).

for relatively experienced L2 learners, suggesting that reliance on L1-L2 similarities plays a bigger role at lower proficiency levels. Although the value of such transparent words for L2 learners is priceless, it comes with a number of pitfalls, since not all words that look cognate indeed are so. Such pitfalls are commonly known as *False Friends* (Davidsen-Nielsen, Færch & Harder 1982: 69). The word *actually*, for instance, is a false friend for L1 Danish learners of English, as it resembles the Danish word *aktuelt* (English: ‘currently’), but the correct Danish translation is *faktisk*, which resembles the English word *factually* (Danish: ‘faktuelt’).

This study examines vocabulary sizes in L1 Danish and L1 Finnish learners of English. These two L1 backgrounds offer an interesting comparison due to the fact that Danish and Finnish differ considerably in their linguistic similarity with English (see Section 2), while Denmark and Finland present comparable learning environments for English as a foreign language. All participants had received English instruction from 3rd to 9th grade of elementary school and were exposed to a fair amount of English on a daily basis through Anglophone TV series and films, since both countries make use of interlingual subtitling of foreign TV programmes instead of dubbing (Preisler 1999; Leppänen & Nikula 2007). A comparison of the vocabulary sizes in L1 Danish and L1 Finnish learners is thus a good way to examine the effect of linguistic similarity while keeping the influence of learning environment as constant as possible.

The study also aims to investigate how amount of L2 experience, operationalised as plus/minus L2 immersion, mitigates the above-mentioned effects of word frequency and cognateness. Specifically, the study investigates 2 hypotheses:

The word frequency hypothesis: *A positive relationship between word frequency and correct word definition is expected. The word frequency effect is expected to be larger for L2 learners compared to native speakers, and for non-immersion learners compared to immersion learners.*

The cognate facilitation hypothesis: *Cognates are expected to be defined correctly more often than non-cognates by L2 learners. The cognate facilitation effect is expected to be larger for non-immersion learners compared to immersion learners.*

2. The lexicons of English, Danish, and Finnish

As any other lexicon, the lexicon of English can be divided into a native part and a borrowed part. The English lexicon is unusual, however, with respect to the large size of its borrowed part. Nevertheless, while borrowed lexemes are vast in quantity, native Anglo-Saxon lexemes dominate everyday speech, as they are the most frequent, and as most function words and affixes are original Anglo-Saxon lexemes. Throughout its history, English has borrowed words from over 350 languages, of which Latin, Old Norse, and French are the languages that have contributed the most to the English lexicon. Borrowings from Latin and Greek are especially extensive in the domain of science and academia (Crystal 2003: 24-26).

The core of the Danish vocabulary consists of lexemes from Old Norse. Since Old Norse and Old English both descend from Common Germanic, Danish and English share a substantial number of common Germanic words, most of which are still alike in both meaning and form, though semantic change has also led to false friends among the cognates. Danish is similar to English in being quite open to borrowings, though the number of loanwords in Danish is markedly lower than in English. The language from which Danish has borrowed the most is undoubtedly Low German, but French and Latin have also contributed considerably to the Danish lexicon. Similar to the situation in English, Danish words of Graeco-Latinate origin play an important part in the domain of science and academia. Importantly, borrowings into Danish since the 1950s have primarily been from English (Katlev 2013; Haberland 1994).

As a Finno-Ugric language, Finnish shares no historic cognates with English, yet lexical similarities between English and Finnish do exist due to direct and indirect borrowings. Most borrowings into Finnish come from neighbouring Germanic languages, especially from Swedish (Karlsson 1999: 1-3). Like in Danish, borrowings from English have recently increased in Finnish (Pulkkinen 1989). However, unlike in English and Danish, loanwords are usually not absorbed directly but adjusted in form, so borrowings may not be easily recognisable. The Finnish word *ranta* ('beach'), for instance, stems from the Germanic *strand*, but is rather different in form due to the phonotactics of Finnish (Sulkana & Karjalainen 1992: 369-370).

3. Methods

3.1 Participants

41 L1 Finnish learners of English (6 M, 35 F, mean age = 25.17 years) participated. They lived in and around Jyväskylä, Central Finland. The L1 Finnish participants were divided into two groups: 1) 21 immersion learners: participants who had lived in an English-speaking country (range: 2.5 months to 3 years, mean = 10.02 months), and 2) non-immersion learners: participants who had not lived in an English-speaking country.

41 L1 Danish learners of English (8 M, 33 F, mean age = 24.71 years) participated. They lived in and around Aarhus, East Jutland, Denmark. The L1 Danish participants were also divided into two groups: 1) 20 immersion learners: participants who had lived in an English-speaking country (range: 4 months to 2.17 years, mean = 10.73 months), and 2) 21 non-immersion learners: participants who had not lived in an English-speaking country.

14 L1 English speakers (2 M, 12 F, mean age = 20.65 years), participated. They lived in and around Bangor, Wales.

3.2 The Vocabulary Size Test

The Vocabulary Size Test (Nation & Beglar 2007; Nation 2012), which is based on word family frequency counts from the British National Corpus (BNC), is a multiple-choice definition test of English vocabulary. The tested word is presented in a simple, non-defining context, and four different but semantically related definitions are supplied, of which one is correct. As far as possible, all words used in the definitions are of a higher frequency than the tested word. The simple, non-defining context reflects the most frequent use of the word. The participant's task is to choose the right definition among the four options. An example is presented here:

soldier: He is a **soldier**.

- a. person in a business
- b. person who studies
- c. person who uses metal
- d. person in the army

Correct answer: d

The Vocabulary Size Test comes in a 14,000 word version and two 20,000 word versions (A and B). The 20,000 word version A was chosen in order to be able to measure the full range of vocabulary sizes. Recall that according to prior estimates, native speakers know 16,000-20,000 word families. Frequency bands 1 and 2 were left out in order to reduce fatigue (the vocabulary test was part of a larger test battery), since it was assumed that all participants knew these extremely frequent words. Since each frequency band, from 3 to 20, was tested with 5 words, the test had a total of 90 items. The test items were presented in randomised order.

All test words were judged for cognateness and false friendship with Finnish by a linguist who is an L1 speaker of Finnish⁴ and for cognateness and false friendship with Danish by the author, who is an L1 speaker of Danish. 32 cognates with Finnish and 37 cognates with Danish were found among the 90 test words. This corresponds to a cognate proportion of 41% for Danish and 36% for Finnish. Such a small difference in cognate proportion was unexpected and does not seem plausible given language history. The test words included one false friend for L1 Danish learners, namely *panzer*, which is similar to the Danish word *panser* (slang for ‘police’), but the correct Danish translation is *tank*. For this item, the response option *policewomen* was chosen by three L1 Danish participants and by no L1 Finnish or L1 English participants, thus showing a small effect of false friendship. No false friends for L1 Finnish learners were observed among the test words.

3.3 Statistics

The data was analysed by means of mixed effects models in the software program R (R Core Team 2015). The R packages used were *lme4* (Bates, Maechler, Bolker, Walker, Christensen, Singmann, Dai & Grothendieck 2015) and *optimx* (Nash 2014) for the construction of mixed effects models. Graphs were also constructed in R, by means of the package *ggplot2* (Wickham 2009).

⁴ Many thanks to Hanna Kivistö de Sousa for these judgements.

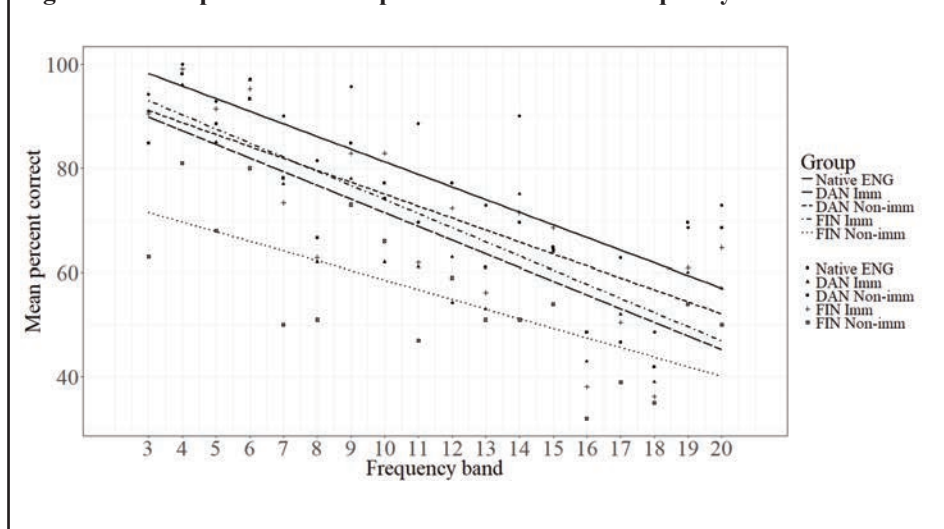
4. Results

A test-taker's vocabulary size can be estimated by multiplying the test-taker's total number of correct answers in the 20,000 version of the Vocabulary Size Test by 200. Since the first two frequency bands were left out in the present study, based on the assumption that all participants would define these first 10 items correctly, 10 points should be added to each participant's vocabulary score before multiplying by 200. The results of these calculations are presented in Table 1.

Table 1. Vocabulary sizes

Group	Mean	Range	Standard Deviation
Native English speakers			
All	15,971	14,600-18,200	1,247
L1 Danish learners			
All	14,527	9,200-17,000	1,726
Immersion	14,150	9,200-16,200	1,601
Non-immersion	14,886	10,600-17,000	1,801
L1 Finnish learners			
All	13,341	8,600-16,800	1,985
Immersion	14,590	12,200-16,800	1,251
Non-immersion	12,030	8,600-14,600	1,763

Figure 1. shows mean percent correct word definition at each frequency band (higher frequency bands corresponds to lower word frequency) for all the groups. The figure suggests a general tendency of decrease in percent correct definition as a function of increase in frequency band (decrease in word frequency) for all the groups.

Figure 1. Mean percent correct performance at each frequency band

The data was subjected to a logistic mixed effects model (Model 1) with *L1* and *Frequency Band* as fixed effects, random intercepts for *Subject* and *Word*, random subject-slopes for *Frequency Band* and random word-slopes for *L1*. The model formula was:

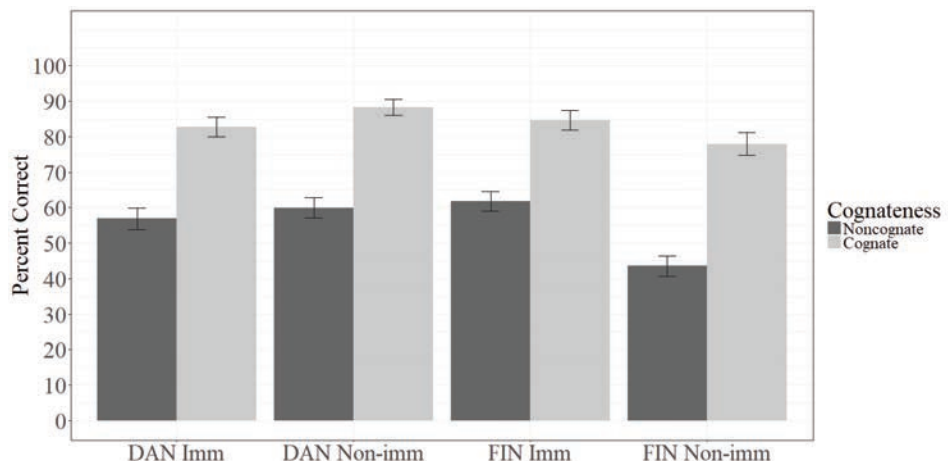
```
Model 1 <- glmer (Performance ~ L1 * FrqBand + (L1|Word) +
(FrqBand|Subject), family = "binomial")
```

Model 1 showed that the native speakers were significantly more accurate than the L2 learners ($p < 0.0001$) and that the L1 Danish learners were significantly more accurate than the L1 Finnish learners ($p = 0.0030$). Moreover, the model revealed a significant main effect of *Frequency Band* ($p < 0.0001$). Words from the lowest frequency band were 66.55 times more likely to be identified correctly than words from the highest frequency band. The significant interaction between *L1* and *Frequency Band* showed that the frequency effect was larger for the L1 English speakers than for the L2 learners ($p = 0.0069$) and for the L1 Danish learners than for the L1 Finnish learners ($p = 0.0264$). Table 2. Presents the statistics of this model.

Table 2. Statistics for Model 1. Significant effects (at the 0.05 level) are highlighted in light grey.

	Estimate (log odds ratio)	Odds ratio	Std. Error	z	p
L1 (English vs. Danish and Finnish) Coded as $+2/3$ English $-1/3$ Danish $-1/3$ Finnish	1.5315	4.625109	0.3610	4.243	2.21e-05
L1 (Danish vs. Finnish) Coded as $+1/2$ Danish $-1/2$ Finnish	0.6521	1.919568	0.2199	2.966	0.00302
Frequency Band	-4.19798	66.55176	0.4174	-5.917	3.29e-09
Frequency Band * L1 (English vs. Danish and Finnish)	-1.1883	3.281498	0.4398	-2.702	0.00690
Frequency Band * L1 (Danish vs. Finnish)	-0.5157	1.67481	0.2322	-2.221	0.02637

Figure 2. shows mean percent correct word definition for cognates and non-cognates for the immersion and non-immersion L1 Danish and L1 Finnish learners. The figure suggests that cognates are correctly defined more often than non-cognates by all the L2 groups.

Figure 2. Mean percent correct performance for cognates and non-cognates

In order to test the effect of cognate facilitation and its interactions with the other predictors, the L2 learner data was subjected to a logistic mixed effects model (Model 2) with *Frequency Band*, *L1*, *Immersion*, and *Cognateness* as fixed effects, random intercepts for *Subject* and *Word* and random subject-slopes for *Frequency Band*, *Cognateness*, and their interaction and random word-slopes for *L1*, *Immersion*, and their interaction. The model formula was:

```
Model 2 <- glmer (Performance ~ L1 * Immersion * FrqBand
* Cognateness + (L1 * Immersion|Word) + (FrqBand *
Cognateness|Subject), family = "binomial")
```

Model 2 also showed a significant frequency effect ($p < 0.0001$) and a significant difference between L1 Danish and L1 Finnish learners ($p = 0.0495$). The model moreover revealed a significant main effect of *Immersion* ($p = 0.033365$) and of *Cognateness* ($p < 0.0001$). Cognate words were 3.55 times more likely to be identified correctly than non-cognate words. The significant interaction between *Frequency Band* and *Cognateness* shows that the frequency effect is stronger for non-cognate words ($p = 0.0005$). The significant interaction between *L1* and *Immersion* shows that *Immersion* had a larger effect for L1 Finnish learners than for L1 Danish learners ($p < 0.0001$). The significant interactions between *Immersion* and *Frequency Band* ($p = 0.0018$) and *Immersion* and *Cognateness* ($p = 0.0337$) show that while the frequency effect was stronger for immersion learners than for non-immersion learners, the cognate effect showed the reverse pattern. Finally, Model 2 revealed a significant 3-way interaction between *L1*, *Immersion*, and *Frequency Band* ($p = 0.0009$). The remaining interactions did not reach significance. Table 3. presents the statistics of this model.

Table 3. Statistics for Model 2. Significant effects (at the 0.05 level) are highlighted in light grey.					
	Estimate (log odds ratio)	Odds ratio	Std. Error	z	p
L1 (Danish vs. Finnish) Coded as + ½ Danish – ½ Finnish	0.4337	1.542956	0.2208	1.964	0.049487
Immersion Coded as + ½ Imm. – ½ Non- imm.	0.4478	1.564866	0.2105	2.128	0.033365
Frequency Band	-2.2421	9.413078	0.3746	-5.986	2.15e-09
Cognateness Coded as + ½ Cognate – ½ Non-cognate	1.2680	3.553738	0.2914	4.351	1.36e-05
L1 * Immersion	-1.8163	6.149065	0.4181	-4.344	1.40e-05
L1 * Frequency Band	-0.3890	1.475505	0.2586	-1.504	0.132604
Immersion * Frequency Band	-0.6997	2.013149	0.2237	-3.128	0.001759
L1 * Cognateness	-0.3009	1.351074	0.2859	-1.053	0.292495
Immersion * Cognateness	-0.5211	1.683879	0.2454	-2.123	0.033734
Frequency Band * Cognateness	-1.4670	4.336207	0.4230	-3.468	0.000524
L1 * Immersion * Frequency Band	1.4455	4.243974	0.4371	3.307	0.000942
L1 * Cognateness * Frequency Band	-0.4962	1.642468	0.4772	-1.040	0.298430
Immersion * Cognateness * Frequency Band	-0.0490	1.05022	0.3649	-0.134	0.893170
L1 * Immersion * Cognateness	-0.4146	1.513765	0.4847	-0.855	0.392311
L1 * Immersion * Frequency Band * Cognateness	0.8906	2.436591	0.7196	1.238	0.215835

5. Discussion

The vocabulary sizes obtained from the test ranged from 14,600 to 18,200 word families for the L1 English speakers. This is slightly lower than the 16,000-20,000 word families found in previous studies, suggesting that the adopted version of the Vocabulary Size Test slightly underestimates vocabulary size. This possible underestimation may be due to the small number of items per frequency band, which makes it possible for individual items that stick out in some respect to skew the results markedly. It should be noted, however, that the L1 English speakers were young (mean age = 20.65 years), and vocabulary size typically grows throughout the life span (Zechmeister et al. 1995). On the other hand, they were all university students (eight in their third year, five in their second year, and one in her first year), and vocabulary size typically increases with degree of education (Diack 1975: 12).

The vocabulary sizes estimated for the L2 learners ranged from 8,600 to 17,000 word families. The range is clearly larger for L2 learners than for L1 English speakers. Although statistical analyses showed that the L1 English speakers significantly outperformed the L2 learners, some L2 learners obtained vocabulary sizes within the native speaker range. The larger range for L2 vocabulary sizes seems to be related to differences in L2 immersion for the L1 Finnish participants. L1 Finnish immersion learners obtained vocabulary sizes between 12,200 and 16,800 word families, while L1 Finnish non-immersion learners obtained vocabulary sizes between 8,600 and 14,600 word families. Surprisingly, L1 Danish immersion learners obtained slightly smaller vocabulary sizes, within the range of 9,200 to 16,200 word families, than L1 Danish non-immersion learners, for whom vocabulary sizes ranged from 10,600 to 17,000 word families. This reverse and less clear pattern observed for L1 Danish learners is plausibly due to the fact that a number of participants in both L1 Danish groups were students of English, while this variable was confounded with L2 immersion in the L1 Finnish participants, so that all L1 Finnish immersion participants were students of English and no L1 Finnish non-immersion participants were students of English.

The L2 vocabulary estimates obtained suggest that all participants are lexically equipped to understand spoken English, which requires a vocabulary size of 6,000 to 7,000 word families (Nation 2006), and that most participants are also lexically equipped to read novels and newspapers

unaided in English, which requires a vocabulary size of 8,000 to 9,000 word families (Nation 2006).

Based on previous research, two hypotheses were stated and tested:

The word frequency hypothesis: *A positive relationship between word frequency and correct word definition is expected. The word frequency effect is expected to be larger for L2 learners compared to native speakers, and for non-immersion learners compared to immersion learners.*

The word frequency hypothesis was partially supported by the present data. A significant effect of word frequency showed that words from the lowest frequency band were 66.55 times more likely to be identified correctly than words from the highest frequency band. However, contrary to expectation and previous research, this frequency effect was found to be significantly larger for native speakers compared to L2 learners and for immersion learners compared to non-immersion learners. This surprising finding may be a result of the corpora used to establish frequency counts not reflecting the language that L2 learners are exposed to. The 12 most frequent bands of the 14,000 word version (Nation and Beglar 2007) are based on frequency counts from the spoken section of the BNC, since these were considered more appropriate for L2 learners than counts based on the entire BNC, due to frequency counts from the entire BNC being heavily influenced by the formal, written nature of the BNC. Nevertheless, the 12 most frequent bands of the 20,000 word versions (Nation 2012) seem to be based on frequency counts from the entire BNC.

The cognate facilitation hypothesis: *Cognates are expected to be defined correctly more often than non-cognates by L2 learners. The cognate facilitation effect is expected to be larger for non-immersion learners than for immersion learners.*

The cognate facilitation hypothesis was supported by the present data. Significant cognate facilitation showed that cognate words were 3.55 times more likely to be identified correctly than non-cognate words. As predicted, this cognate facilitation was found to be significantly larger for non-immersion learners than for immersion learners.

Despite the fact that the adopted version of the Vocabulary Size Test showed an implausibly small difference in cognate proportion between Danish (41%) and Finnish (36%), L1 Danish learners were found to outperform L1 Finnish learners. The statistical analyses showed no

evidence to suggest that cognate facilitation differed between L1 Danish and L1 Finnish participants, so the explanation is unlikely to be found in cognate facilitation. The L1 Danish advantage may nevertheless reflect the closer linguistic similarity between English and Danish vis-à-vis English and Finnish. The shared Germanic origin is not only reflected in cognates but also in a more global phonological resemblance between English and Danish words. Whereas English phonotactics does not differ markedly from Danish phonotactics (Crystal 2003; Grønnum 2001), it does differ markedly from Finnish phonotactics. Important phonotactic differences between Finnish and English include vowel harmony, which is present in Finnish (Karlsson 1999) and absent in English, and consonant clusters, which are much more restricted in Finnish than in English (Sulkana & Karjalainen 1992: 369-370). These global phonological similarities may aid L2 vocabulary learning for L1 Danish learners compared to L1 Finnish learners. In other words, L1 Danish learners be more successful than L1 Finnish learners in learning English vocabulary because even non-cognate lexemes have a more familiar phonological structure for L1 Danish learners than for L1 Finnish learners. This account is in line with Ellis and Beaton's (1995) list of psycholinguistic determinants on L2 vocabulary learning, which includes the factor *pronounceability*. A word's pronounceability depends on how similar it is to the L2 learner's L1 lexicon in terms of segments and phonotactics. Importantly, support for the influence of pronounceability on word learning has been found in word learning studies in which subjects were not asked to pronounce the words (Rodgers 1969).

Summing up, the present data provided support for the cognate facilitation hypothesis but only partial support for the word frequency hypothesis. The lack of support for the word frequency hypothesis may be related to the frequency counts that the adopted version of the Vocabulary Size Test are based on. An L1 Danish advantage unrelated to cognate facilitation was moreover observed. This advantage may be related to phonotactic similarities between English and Danish.

References

- Balling, Laura Winther. 2012. Reading authentic texts: What counts as cognate? *Bilingualism: Language and Cognition* 3. 1–17. doi:10.1017/S1366728911000733.
- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai & Gabor Grothendieck. 2015. *lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1-10*. <http://CRAN.R-project.org/package=lme4> (2 November, 2015).
- Bertram, Raymond, Matti Laine & Minna Maria Virkkala. 2000. The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. *Scandinavian Journal of Psychology* 41. 287–296. doi:10.1111/1467-9450.00201.
- Casaponsa, Aina, Eneko Antón, Alejandro Pérez & Jon A. Duñabeitia. 2015. Foreign language comprehension achievement: Insights from the cognate facilitation effect. *Frontiers in Psychology* 6. 1–12. doi:10.3389/fpsyg.2015.00588.
- Comensaña, Montserrat, Rosa Sánchez-Casas, Ana Paula Soares, Ana P. Pinheiro, Andrea Rauber, Sofia Frade & Isabel Fraga. 2012. The interplay of phonology and orthography in visual cognate recognition: An ERP study. *Neuroscience Letters* 529. 75–79. doi:10.1016/j.neulet.2012.09.010.
- Crystal, David. 2003. *The Cambridge encyclopedia of the English language*. Second Edition. Cambridge: Cambridge University Press.
- Davidson-Nielsen, Niels, Claus Færch & Peter Harder. 1982. *The Danish learner*. Tunbridge Wells: Anthony Taylor.
- Davies, Mark. (2004-) British National Corpus (from Oxford University Press). Available online at <https://www.english-corpora.org/bnc/>.
- Diack, Hunter. 1975. *Test your own wordpower*. St. Albans: Paladin.
- Dijkstra, Ton, Koji Miwa, Bianca Brummelhuis, Maya Sapelli & Harald Bayen. 2010. How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language* 62. 284–301. doi:10.1016/j.jml.2009.12.003.
- Ellis, Nick C., and Allan Beaton. 1995. Psycholinguistic Determinants of Foreign Language Vocabulary Learning. In *Lexical Issues in Language Learning*. ed. Birgit Harley, 107–165. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Gelderen, Elly van. 2006. *A history of the English language*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Groot, Annette M.B. de, Susanne Borgwaldt, Mieke Bos & Ellen van den Eijnden. 2002. Lexical decision and word naming in bilinguals. *Journal of Memory and Language* 47: 91–124. doi:10.1006/jmla.2001.2840.

- Grønnum, Nina. 2001. *Fonetik og fonologi: almen og dansk*. Copenhagen: Akademisk Forlag.
- Haberland, Hartmut. 1994. Danish. In Ekkehard König & Johan van der Auwera (eds.) *The Germanic languages*, 313–348. London and New York: Routledge.
- Karlsson, Fred. 1999. *Finnish: An essential grammar*. Translated by Andrew Chesterman. London (UK): Routledge.
- Katlev, Jan. 2013. *Ordenes oprindelse*. <http://sproget.dk/temaer/ordenes-oprindelse> (4 December, 2013).
- Lemhöfer, Kristin, Ton Dijkstra, Herbert Schiefers & R. Harald. Baayen. 2008. Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology* 34(1). 12–31. doi:10.1037/0278-7393.34.1.12.
- Leppänen, Sirpa & Tarja Nikula. 2007. Diverse uses of English in Finnish society: Discourse-pragmatic insights into media, educational and business contexts. *Multilingua* 26. 333–380. doi:10.1515/MULTI.2007.017.
- Nash, John C. 2014. On best practice optimization methods in R. *Journal of Statistical Software* 60(2). 1–14. doi:10.18637/jss.v060.i02.
- Nation, Paul. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63(1). 59–82. doi:10.3138/cmlr.63.1.59.
- Nation, Paul. 2012. *Vocabulary size test instructions and description*. <http://www.victoria.ac.nz/lals/about/staff/paul-nation> (3 September, 2013).
- Nation, Paul & David Beglar. 2007. A vocabulary size test. *The Language Teacher* 31. 9–13.
- Preisler, Bent. 1999. *Danskerne og det engelske sprog*. Frederiksberg: Roskilde Universitetsforlag.
- Pulkkinen, Paavo. 1989. Anglicismerna i finska sproget. *Språk i Norden*. 89–93.
- R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (14 December, 2011).
- Rodgers, Theodore S. 1969. On measuring vocabulary difficulty. An analysis of Item Variables in learning Russian-English vocabulary pairs. *IRAL - International Review of Applied Linguistics in Language Teaching - LA English* 7 (4). 327–343. DOI: 10.1515/iral.1969.7.4.327.
- Schilling, Hildure E., Keith Rayner & James I. Chumbley. 1998. Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition* 26(6). 1271–1281. doi:10.3758/BF03201199.
- Schmitt, Norbert. 2010. *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

- Sulkana, Helena & Merja Karjalainen. 1992. *Finnish*. London and New York: Routledge.
- Whitford, Veronica & Debra Tytone. 2012. Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychon Bull Rev* 19. 73–80. doi:10.3758/s13423-011-0179-5.
- Wickham, H. 2009. *ggplot2: Elegant graphics for data analysis*. New York: Springer.
- Wijnendaele, Ilse Van & Marc Brysbaert. 2002. Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology* 28(3): 616–627. doi:10.1037//0096-1523.28.3.616.
- Zechmeister, Eugene B., Andrea M. Chronis, William L. Cull, Catherina A. D'Anna & Noreen A. Healy. 1995. Growth of a functionally important lexicon. *Journal of Reading Behavior* 27(2). 201–212. doi:10.1080/10862969509547878.