

Understanding Estonian phraseological units on the basis of Finnish: Contributing and misleading factors

PIRKKO MUIKKU-WERNER¹,

JARMO HARRI JANTUNEN²

University of Eastern Finland¹, University of Jyväskylä²

Abstract. Finns reading Estonian texts have a variety of strategies to make sense of them. One is priming, which means that language users consciously or unconsciously rely on previous knowledge of associations that occur in phraseological units. We studied how the distance between a prime and its target and theme affects the understanding of unknown words and false friends, and how the restricted semantic category affects uncovering the target.

The data of the study consist of translations of Estonian texts into Finnish, produced by native Finnish speakers with no theoretical or practical knowledge of Estonian. The participants were also asked about the process of reaching certain translation equivalents. The Finnish Language Text Collection was consulted for L1 patterning, in order to reveal whether the phraseological combinations in translation tests have counterparts in target language usage.

The analysis revealed that there are several factors that affect understanding source text lexemes: external similarity, the distance of the items in a phraseological unit, the limitedness of the semantic category and, finally, world knowledge. The participants often mentioned orthographic similarity as a starting point in revealing the meaning. Other factors that contributed to understanding a

foreign text were mentioned as well: context and semantic relations, which base on internalised metalinguistic knowledge, facilitated to associate the prime and the target. Finally, similar collocations to translation tests were found in the L1 data.

Keywords: mutual intelligibility of languages; semantic priming; prime; target; Estonian; Finnish

1. Introduction

The ability to understand but not to speak a second language has become a subject of active research, especially in the context of various cognate languages (for this kind of receptive multilingualism (RM) see e.g. Braunmüller 2007; Zeevaert & ten Thije 2007; Rehbein et al. 2012). In the case of closely related languages, a prerequisite for mutual intelligibility is essentially created by similar vocabulary and phonetic proximity of the languages in question (Gooskens et al. 2017, 2018).

Based on that observation, common lexicon has been a central starting point in studying mutual intelligibility of Finnish and languages closely related to it. Previous studies on comprehension of Estonian on the basis of Finnish (Kaivapalu 2005; Kaivapalu & Muikku-Werner 2010; Muikku-Werner 2013) have shown that the participants in translation tests utilize external resemblance of vocabulary between these two languages, and are often able to understand central parts of the text (for an overview of RM research concerning Estonian and Finnish, see Härmävaara & Gooskens 2019: 16-19).

However, similarity is not the only factor contributing to intelligibility, but various first-language (L1) skills, such as lexical inferencing, also promote comprehension. Finnish speakers' ability to understand an Estonian text is in this article approached from a cognitive perspective, focusing on lexical inferencing. It can be defined as using a variety of linguistic and non-linguistic clues to fill the gaps in understanding when encountering unfamiliar words in a text (Oxford 1990: 47). A reader of

a text can attempt to infer the meaning of a word – with no resemblance with any mother tongue unit – on the basis of linguistic clues found in the text or in the word, or on general knowledge of the topic. The activation of background knowledge helps the reader not only to make hypotheses about the meaning of an unknown word but also to confirm them (Vaurio 1998: 60).

This article offers an alternative approach for studying cognate language text comprehension, by making use of the notion of semantic priming. The concept refers to the mechanism that the readers use for building meaningful connections between lexical items: the meaning of one word (a target) is identified effectively if it is preceded or followed by another, semantically related word (a prime). The prime activates associated words (words belonging to the same semantic field, synonyms, antonyms, hypernyms, hyponyms, etc.) in the reader's mind and thus facilitates recognition of the target. (For lexical priming, see e.g. Hoey 2005, 2007; Lutjeharms 2007: 272–273.)

The present study is a follow-up to Muikku-Werner's (2017) study on how native speakers of Finnish understand Estonian, based on a translation test. The results of the translation test were contrasted to the data gathered from a control group with a cloze test in their L1, Finnish. These two sets of data were compared in order to discuss the roles of lexical inferencing, collocations and semantic priming in mutual intelligibility. The study showed that while inferring both the translation equivalents and the missing words in the cloze test, the participants relied on the cohesion of the text, knowledge of the topic and world knowledge, as well as the structures of semantic fields. Some questions, however, remained unanswered, and they are focused on in the present article.

2. Theoretical background: Collocations and lexical priming

In order to understand the mechanisms of meaningful associations of lexical items, such as collocations and other phraseological units, from the point of view of second language learning and receptive multilingualism, the concept of lexical priming is useful (Jantunen 2017: 254–255, 268–269; Muikku-Werner 2017: 205). It refers to the effect caused by repeated experience of a stimulus, and emphasises the direct connection between experience and expectation. Priming may occur at several levels of linguistic structure, notably lexical and syntactic (Hoey 2007: 8). Hoey (2007: 8) states that when using a word, we use it

“with one of its typical collocations, in its usual grammatical function, in the same semantic context, in the domain we have come to associate it with, as part of the same genre, in a familiar social context, with a similar pragmatics and to similar textual ends.”

All these previous encounters with primings and collocational language use prime language users to find the preferred and typical usage of a word. In this study, we focus on collocations, which Sinclair (1991: 170) defines as habitual occurrences of two or more lexical items within a short space of each other, stating that “words enter into meaningful relations with other words around them” (Sinclair 1996: 71).

In psycholinguistics, priming refers to the effect that previously existing stimuli have on a subsequent, related stimulus. Hoey’s (2005) theory of lexical priming suggests that language is not stored as individual items but as items with habitual contextual information, that is, as prefabricated phraseological units. Multiple encounters of formulaic utterances help L1 readers and writers to unconsciously understand and produce text smoothly and predictably (see Hoey 2007: 8; Firth [1968] 1957: 179; Sinclair 1996: 81; Jantunen 2009: 356; Pace-Sigge 2013: 13–14; Pace-Sigge & Patterson 2017: XIII–XIV). Since words are primed to occur with particular semantic sets, they form semantic associations

(Hoey 2005: 13) and collocational cohesion in a text (Halliday & Hasan 1976: 284–286).

Since semantic priming can be used as a tool for word recognition (McNamara 2005: 3–4, 11, 18), we hypothesise that inferring the meaning of an unknown word depends heavily on support from associative lexical and semantic information. This process is a form of retrieval from semantic memory and world knowledge: words that go together make associations possible (see e.g. Pace-Sigge 2013). Furthermore, if there are few alternatives for collocational combination, this limitedness can facilitate inferring the meaning of the collocate.

Priming can also enhance intelligibility of a text in a non-native language. In cases where there are enough familiar elements in a context, even words that do not share a similar meaning and spelling in the mother tongue and in the target language (i.e. that are not cognates) can be recognisable on the basis of lexical priming (Hoey 2005): a certain word can be linked to another based on lexical associations, such as collocation, but also through antonymy, hyponymy and other familiar semantic relations. In reading comprehension, lexical priming (or semantic priming) functions in a way that an item (a target) is retrieved from memory with the help of its relationships to preceding words, such as primes. Thus, a single word can be expected to precede a certain word or words to follow because recurrent previous experiences have primed the language user to expect a certain item (Page-Sigge & Patterson 2017: XIII).

Therefore, certain lexical and semantic links between concepts seem to aid Finnish speakers in understanding simple Estonian texts (Mukkku-Werner 2015, 2016, 2017; for lexical priming see also Lutjeharms 2007: 272–273). It is not necessary that the prime is located in the immediate context of the target, but it may be in a different clause: the association of the prime and the target (i.e. two collocates) is purely lexical, and semantic relation between items are irrespective of syntactic boundaries. In fact, the coherence of a text is partly built on the collocational relationships of items, which may not appear in each other's immediate context (Halliday & Hasan 1976: 284–288; Hoey 2007: 8).

Furthermore, general or background knowledge and “reading between the lines” matter: if readers encounter an unfamiliar word in a text, they can try to infer its meaning based on their general knowledge of the topic and other factual cues found in the text or in the word. The activation of background knowledge (unconscious mental structures, i.e. schema, see Minsky 1975: 212–213) from long-term memory is important, since it helps readers to make hypotheses of the meaning of an unknown word (Vaurio 1998: 60). This is a strategy called lexical inferencing. It involves a deeper processing of information given in the text, and it is supposed to improve the comprehension of texts as a whole (see e.g. O’Malley & Chamot 1990; Bernhardt 1991; Vaurio 1998: 41–44). This kind of view does not see either memory or the mental lexicon as a mere storage place for knowledge. Remembering things and accessing words can be viewed as more dynamic and context-dependent phenomena (Dufva 1999: 26–27).

In the present article, we deepen previous findings on the role of priming and world knowledge in mutual intelligibility through a series of four research questions. We first ask how the number of primes and the distance between primes (stimulus) and the target, as well as the theme of the text affect priming (RQ1). Second, if some hypernyms (here parenthood, nuclear family) of a certain semantic field (kin words) allow only a limited number of co-hyponyms, does this limitedness help or hinder inference and do the number of primes influence the process of inferring meaning (RQ2)? Furthermore, do the test takers produce information that would suggest they are aware of the priming effect (RQ3)? Finally, do the co-occurrence patterns (collocations) produced by the test participants correlate with patterns in target language usage (RQ4)?

3. Data and methods

In order to answer our four research questions, three translation tests and a corpus analysis of phraseological units in Finnish were conducted. The translation tests were as follows: A translation of a short text about Christmas, a translation of a longer text on the same theme, and finally a translation of a text about birthdays. The texts were translated from Estonian into Finnish by 75 test participants. (For information on the participants, see below. See Appendix 1 for the test texts in Estonian and their translations into English.) Using both a short and a long text on the same theme (Christmas), we seek answers to RQ1. In the longer text, the distance between primes and the target is longer, meaning that there is more text (i.e. more words) between the primes and the target. The third text (the birthday text) is used to test the relevance of schematic implication and thematic context to the inference process (RQ1).

The shorter and the longer version of the Christmas text (see Appendix 1) are about buying presents and taking photos. The primes and targets were chosen on the basis of researchers' metalinguistic knowledge and their constitution of certain semantic fields. Thus, the texts do not as such represent a natural text but are instead manipulated texts for this specific study. In the present study, it is assumed that prime and target pairs exist in certain contexts. For example, in real life when we have a dog as a pet, one probably connects the word *dog* in daily discussion with other words such as *bones* and *leash*, not with *cat*, which the word *dog* could be associated with, in a priming test. The association in this kind of test is built on schemas and semantic fields (see e.g. McNamara 2012: 3–4).

In the short Christmas (SC) text, there are two targets (italics) with five primes (bold, in brackets) that were tested: *kingitusi* 'presents' (**jõulud** 'Christmas', **raamat** 'book') and *õde* 'sister' (**isa** 'father', **ema** 'mother', **vend** 'brother'). The SC text and its translation into English are provided below as an example; other texts are available in Appendix 1.

Kohe on **jõulud** käes! **Jõulu** eel ostame *kingitusi*. Käesoleval aastal ostan sugulastele **raamatuid**. Üks **raamat** võib olla kallid, teine odav. Möödunud jõulu ajal tegime palju pilte. Sellel pildil on minu **isa**, **ema**, **vend** ja *õde*. Sellel pildil on minu naine. Ta on rase. Laps sündis juulis.

[Soon it will be **Christmas**. Before **Christmas** we buy *presents*. This year I'll buy **books** for relatives. One **book** can be expensive, the other cheap.

Last Christmas we took lots of photos. In this photo are my **father**, **mother**, **brother** and my *sister*. In this photo is my wife. She is pregnant. The child was born in July.]

In the long Christmas (LC) text, a sentence between the target *kingitusi* and the prime **raamat** was added, and the primes **isa** and **vend** for the target *õde* were omitted. Muikku-Werner's research (2017) showed that incorrect translations for *kingitus* were caused by the assumed similarity between *kingitus* and the Finnish *kinkku* 'ham' instead of the correct translation *lahja* 'present'. In the present study we want to determine, if this process is influenced by the distance between the primes and the target. (For the influence of the distance between prime and target, see McNamara 2005: 91–95.) We will also test whether the 'nuclear family' schematic field and the number of primes (**ema** vs. **ema**, **isa**, **vend**) affect finding the correct translation of the target word, in this case *õde* (RQ2). The BT text, in turn, contains two prime–target pairs, in which the first prime has been changed to **sünnipäev** 'birthday' instead of **jõulud**, and the prime for *õde* is **isa** (RQ1 and 2).

For Finnish speakers, the Estonian primes of our tests are more or less recognisable because they have cognate words in Finnish. We hypothesise that to foster understanding of the targets, the following semantic relations may be helpful. For example, **jõulud** (FI **joulu** 'Christmas') may prime *kingitus* (FI *lahja* 'present') on the basis of schematic implication (or semantic associations), and the hyponym **raamat** (FI **kirja** 'book') may also assist in inferring the meaning of *kingitus* (hypernym) by naming a potential present. **Raamat** is not a cognate

word but many Finns recognise it as a false friend. Besides, the incorrect Finnish translation **Raamattu** ('bible') may also help ending up with the correct translation for *kingitus* because it is also a potential, though rare, alternative as a present.

In her earlier study, Muikku-Werner (2017) analysed the semantic field of relatives. The diversified network of kinship offers many kinds of superordinate and subordinate categories that are well-known in L1. This familiarity was presumed to support the inferencing process. In the structure **isa/ema** ja *õde*, **isa** was expected to prime *äiti* and **ema** *isä* on the basis of the 'parenthood' semantic field, those being the co-hyponyms of the concept, even though the correct translation for *õde* is 'FI *sisko, sisar*' ('sister') (Muikku-Werner 2017: 216). In the present study, the correct translation is assumed to occur in the context of nuclear family members **isa, ema, vend** ('FI **isä, äiti, veli**', 'father, mother and brother'). (RQ2)

Answers for RQ3 were sought by asking the informants to justify their proposals for translation equivalents. Do they, for example, lean on the similarity of the items (i.e. false friends) or do they rely on the structure of semantic fields, coherence of texts or world knowledge?

The relevance of collocational frequency and patterning (retrieved from corpus analyses) is also addressed (RQ4). We are interested in finding out whether the target–prime pairs, **jõulud** 'Christmas' → *kingitusi* 'presents' ← **raamatuid** 'books' and **ema** 'mother' or **isa** 'father' → *õde* 'sister' or **isa** 'father', **ema** 'mother', **vend** 'brother' → *õde* 'sister' occur in contexts other than the translation tests. In order to compare the informants' primings and real language use and determine whether the primings have counterparts in authentic text collocations, the Finnish Language Text Collection (CSC) is consulted. The research corpus contains four newspapers (*Karjalainen, Kaleva, Turun Sanomat* and *Keskisuomalainen*, each published in different dialect areas in Finland) and one tabloid (the national *Iltalehti*). The size of the dataset is approximately 33 million words.

The LC Estonian–Finnish translation test was completed by 25 informants, a heterogeneous group that consisted of people of different ages (median age 57; 17 women, 8 men). All of them had either an upper secondary school or a university degree. The SC test was completed by 25 participants (median age 43; 11 women, 14 men, 1 other). The BT was translated by 25 test subjects (median age 46; 15 women, 10 men).

Informants have had contact with Estonian only occasionally, for example during their trips to Estonia. None of them have studied linguistics or Estonian, but most of them have studied one or more foreign languages at school. Knowledge of other languages is an important factor to take into account, when analysing the translations: previous research has shown that when inferring meanings or translation equivalents, the informants utilise not only their knowledge about the cognate languages, but also other languages they know (see Muikku-Werner & Heinonen 2012: 172–174).

4. The results of various translation tests

The following sections address different translations of the false friend target word *kingitus* (section 4.1) and the non-cognate target word *õde* (see section 4.2) in various contexts. There is some consistency in the different translation processes. However, it is important to note that there are many factors which may contribute to or detract from understanding the Estonian phraseological unit. In this article, we concentrate on some limited aspects of inference, such as relying on inclusion and schematic implications.

4.1. Shopping schema and different test variants

In this section we discuss the strength of the influence of assumed similarity (*kingitusi* ‘presents’ and FI **kinkku* ‘ham’) in the translation process in different contexts, and explore the effect of distance between the target and the prime. In order to study the distance effect, we modified

the space between the false friend target and the prime **raamatuid** ‘books’ (the hyponym of the target concept): in the LC, there is an additional sentence between the first clause with *kingitusi* and the second with **raamatuid**.

Short Christmas version (SC):

Jõulu eel ostame *kingitusi*. Käesoleval aastal ostan sugulastele **raamatuid**.

FI **Joulun** edellä ostamme *lahjoja*. Tänä vuonna ostan sukulaisille **kirjoja**.

[Before **Christmas** we buy *presents*. This year I’ll buy **books** for relatives.]

Long Christmas version (LC):

Jõulu eel ostame *kingitusi*. Kaunistame kuuse ja mängime selle ümber. Käesoleval aastal ostan sugulastele **raamatuid**.

FI **Joulun** edellä ostamme *lahjoja*. Koristamme kuusen ja leikimme sen ympärillä. Tänä vuonna ostan sukulaisille **kirjoja**.

[Before **Christmas** we buy *presents*. We decorate the Christmas tree and play around it. This year I’ll buy **books** for relatives.]

In the birthday text, the primes (**sünnipäev** and **raamatuid**) are in adjacent sentences of the target *kingitusi*.

Birthday (BT)

Kohe on minu ema **sünnipäev**. Ostame temale *kingitusi*. Käesoleval aastal ostan talle **raamatuid**.

FI Pian on äitini **syntymäpäivä**. Ostamme hänelle *lahjoja*. Tänä vuonna ostan hänelle **kirjoja**.

[Soon it will be my mother’s **birthday**. We’ll buy her *presents*. This year I’ll buy her **books**.]

Table 1 shows both absolute and relative frequencies of the correct translations. The amount of data is small, which makes it difficult to arrive at general conclusions, but it is large enough to provide tentative results.

TABLE 1. *Shopping schema – correct translations*

| Target <i>kingitusi</i> 'presents' | 'Long distance' (Christmas text LC) 1. prime jõulud 'Christmas' + 2. prime raamatuid 'books' | | 'Short distance' (Christmas text SC) 1. prime jõulud 'Christmas' + 2. prime raamatuid 'books' | | 'Short distance' (birthday text BT) 1. prime sünnipäev 'birthday' + 2. prime raamatuid 'books' | |
|--|--|------|---|------|--|------|
| | abs. | % | abs. | % | abs. | % |
| Correct responses | 16/25 | 64.0 | 14/25 | 58.3 | 9/25 | 36.0 |

The schematic association (**Christmas**; first prime) seems to be a strong connecting factor in directing the inference process, since in BT the share of the correct responses is clearly smaller (36%) than in both Christmas texts (64% and 58%). The existence of the hyponym **raamatuid** (second prime) did seemingly not help to find the hypernym *lahjoja* 'presents'. The shorter distance between the prime **raamatuid** and the target *kingitusi* did not increase the number of correct answers: the share of correct responses in the SC was in fact smaller than in the LC. The informants' statements about their translation solutions support these results (for the discussion of the supportive solutions, see section 4.3).

Table 2 shows that in the LC the share of the false friend *kinkku* 'ham' is 32%. In the SC the share is smaller (20%), and finally *kinkku* is non-existent in the BT. The lower share in the SC may result from the closeness of the second prime **raamatuid** 'books', since *kinkku* 'ham' and **raamatuid** 'books' are not likely to be associated with each other. Other translation solutions do also exist: *kirjallisuus* 'literature', which is another hypernym for **raamatuid**, was mentioned by one test participant. *Kenkä*, in turn, has some resemblance with the target word *kingitusi*, and so does the Swedish (*kung / konung*) and English words for *king*, which led to translation *kuningas* (for more about the influence of other languages on the inferring process, see Kaivapalu & Muikku-Werner 2010; Paajanen & Muikku-Werner 2012). Other translations are more or less imaginable as shopping objects or are linked with shopping, each proposed by one test subject only.

TABLE 2. *Kingitusi: various translation responses*

| Long Christmas text LC 1. prime jõulud 'Christmas' + 2. prime raamatuid 'books' | Short Christmas text SC 1. prime jõulud 'Christmas' + 2. prime raamatuid 'books' | Birthday text BT 1. prime sünnipäev 'birthday' + 2. prime raamatuid 'books' |
|--|--|---|
| <i>(joulu)lahja</i> '(Christmas) present' 64% <i>(joulu)kinkku</i> '(Christmas) ham' 32% <i>kynttilä</i> 'candle' 4% | <i>(joulu)lahja</i> '(Christmas) present' 52% <i>kinkku</i> 'ham' 20% <i>kirjallisuus</i> 'literature' 4% <i>kenkä</i> 'shoe' 4% <i>kuningas</i> 'king' 4% <i>kulta-aika</i> 'golden age' 4% <i>tavara</i> 'article' 4% <i>vähän</i> 'a little' 4% <i>riemujuhla</i> 'jubilee' 4% | <i>(teema/yhteis)lahja</i> '(themed/ shared) gift' 36% <i>lahjakortti</i> 'gift voucher' 4% <i>kiikkutuoli</i> 'rocking chair' 4% <i>kakku</i> 'cake' 4% <i>kenkälanka</i> 'shoestring' 4% <i>Kingis-jäätelö</i> 'Kingis ice cream' 4% <i>sankarikertomus</i> 'hero story' 4% <i>tuomisia</i> 'homecoming gift' 4% <i>taulu</i> 'painting' 4% <i>uutta</i> '(something) new' 4% <i>tekemistä</i> '(something) to do' 4% <i>luotolla</i> 'on credit' 4% <i>malja</i> 'bowl' 4% ? (no answer) 16% |

Inclusions of expectations already hinge on both the Christmas and birthday schema: default values include gifts (for more on default values, see e.g. Brown & Yule 1983: 223–224). When we look at the BT, we find dissimilar translation solutions compared to the SC and the LC. First, *ham* (false friend translation) has not been provided as a target, and the relationship hyponym (**book**) versus the hypernym (*present*) is more rare (36%) as the basis of lexical inference. Instead of this semantic inclusion premise, the birthday theme invites plenty of different co-hyponyms (implicated hypernym *present*). This is an interesting result because in both texts the first prime (either **jõulud** 'Christmas' or **sünnipäev** 'birthday') precedes the target, and as the second prime **raamatuid** follows it,

the textual positions do not vary. Both primes were recognised almost as well: **joulud** 100%, **sünnipäev** 92%, so the incomprehensibility of neither of them did not cause the differences. For the birthday version, the share of empty answers was 16%. The assumed similarity has most likely influenced translations such as *kiikkutuoli* ‘rocking chair’, *kenkälanka* ‘shoestring’, *kakku* ‘cake’ and *Kingis-jäätelö* ‘Kingis ice cream’. Somebody also emphasised rather incoherently the manner of payment: *luotolla* ‘on credit’.

4.2. Lexical inferencing and semantic relationships between kinship words

In this section, the lexical inferencing in the case of three different prime combinations of the ‘close relatives’ semantic field (including ‘parenthood’ or the ‘standard nuclear family’) is analysed. Previous studies (Muikku-Werner 2015, 2016) have shown that if in a cloze test (in Finnish) the prime is **isä** ‘father’, the following blank is filled with *äiti* ‘mother’. In a translation test, the Estonian word **ema** ‘mother’ primes the word *isä* ‘father’, even though the correct translation of the word *õde* would be *sisar* ‘sister’. Thus, parenthood was pictured as a category that strictly defines its alternative members. The test texts for the present study are as follows:

Long Christmas version (LC)

Möödunud jõulu ajal tegime palju pilte. Sellel pildil on minu **ema** ja *õde*.

FI Viime jouluna otimme paljon kuvia. Tässä kuvassa on minun **äiti** ja *sisar*.

[Last Christmas we took lots of photos. In this photo are my **mother** and my *sister*.]

Short Christmas version (SC)

Möödunud jõulu ajal tegime palju pilte. Sellel pildil on minu **isa**, **ema**, **vend** ja *õde*.

FI Viime jouluna otimme paljon kuvia. Tässä kuvassa on minun **isä**, **äiti**, **veli** ja *sisar*.

[Last Christmas we took lots of photos. In this photo are my **father**, **mother**, **brother** and my *sister*.]

Birthday (BT)

Eelmisel sünnipäeval tegime palju pilte. Sellel pildil on minu **isa** ja *õde*.
 FI Edellisenä syntymäpäivänä otimme paljon kuvia. Tässä kuvassa on minun **isä** ja *sisar*.

[Last Birthday we took lots of photos. In this photo are my **father** and my *sister*.]

TABLE 3. *Relatives schema – correct translations*

| Target <i>õde</i> 'sister' | Long Christmas text prime ema 'mother' | | Short Christmas text primes isa , ema , vend 'father, mother, brother' | | Birthday text prime isa 'father' | |
|----------------------------------|---|-----|---|------|---|---|
| | abs. | % | abs. | % | abs. | % |
| Correct responses | 1 | 4.0 | 15 | 56.0 | - | - |

As Table 3 shows, the translation test participants faced great difficulties finding the correct translations when either **ema** or **isa** were provided as primes. Table 4 shows that when either **ema** 'mother' or **isa** 'father' functions as a prime, the target is consistently (except for one answer) the parent of the opposite gender. The share of these incorrect translations is 44%. This indicates that parenthood is a popular semantic association. However, when the list of the standard nuclear family was supplemented with **vend** ('brother'), the share of correct translations (FI *sisko*, *sisar* 'sister') increased significantly (56%). Thus, nuclear family membership and filling in the semantic field (the pattern **isa**, **ema**, **vend**) seemed to lead to better solutions.

The categories of parents or nuclear family are quantitatively rather definite, and the expectations of listing their suitable members are high.

In the case of parenthood, they lead to semantically correct proposals. However, this sort of semantic false friend (either *mother* or *father*) is an incorrect translation of the target word *õde*. In spite of the mistranslation, both examples demonstrate the same kind of lexical inference: finding co-hyponyms of a certain semantic field, in this case, fairly limited kin subcategories.

TABLE 4. *Õde* – various responses

| Long Christmas text LC prime ema ‘mother’ | Short Christmas text SC primes isa, ema, vend ‘father, mother, brother’ | Birthday text BT prime isa ‘father’ |
|---|---|---|
| <i>isä</i> ‘father’ 44% <i>täti</i> ‘aunt’ 12% <i>sisar/sisko</i> ‘sister’ 4% <i>setä</i> ‘uncle’ 4% <i>eno</i> ‘the brother of mother’ 4% <i>sukulaiset</i> ‘relatives’ 4% <i>kohtalo</i> ‘destiny’ 4% ? (no answer) 24% | <i>sisar/sisko</i> ‘sister’ 56% <i>veli</i> ‘brother’ 4% <i>veljen vaimo</i> ‘the wife of the brother’ 4% <i>vaimo</i> ‘wife’ 4% <i>anoppi</i> ‘mother-in-law’ 4% <i>nato</i> ‘sister of the husband’ 4% <i>setä</i> ‘uncle’ 4% <i>eno</i> ‘the brother of mother’ 4% <i>kummi</i> ‘godparent’ 4% <i>muita</i> ‘others’ 4% <i>kohtalo</i> ‘destiny’ 4% ? (no answer) 4% | <i>äiti</i> ‘mother’ 44% <i>äitipuoli</i> ‘stepmother’ 4% <i>veli</i> ‘brother’ 12% <i>eno</i> ‘the brother of mother’ 4% <i>mummi</i> ‘granny’ 4% <i>ystävä</i> ‘friend’ 4% <i>korva</i> ‘ear’ 4% <i>ilo</i> ‘joy’ 4% ? (no answer) 20% |

Table 4 also shows that all tests produced translations with other different relatives (e.g. *täti* ‘aunt’, *setä* ‘uncle’, *mummi* ‘granny’), which means that the informants did consider the wider concept of *kinship* as the basis of their semantic association. It would be interesting to know if the target *sister* was inferred equally on the basis of all three primes or if **vend** ‘brother’ (category *siblings*) was the crucial influence. The potential number of primes depends on the content of the category: parenthood in its most frequent realisation is a relationship between the two, one possible standard family is described including parents and a son and a daughter. To compare, for example, the prime–target relationships (in

any language) **grandfather** and *granddaughter* or **grandfather, grandmother, grandson** and *granddaughter*, one could possibly get more evidence about the guiding influence of very closed semantic categories.

In the birthday text, the cognate word **ema** ‘mother’ was mentioned in the first sentence of the text and therefore some informants noted that *õde* could not be ‘mother’ and they suggested something else, such as ‘stepmother’. The translation *kohtalo* (‘destiny’) is based on the assumed similarity between the Estonian *õde* and Swedish *öde* (‘destiny’). The participants in the birthday text test often provided ‘no answer’.

The information about the inference processes combined with these semantic relationships can provide some support when dealing with processes for other items in similar relationships or textual positions.

4.3. Reasons for translation solutions

This section presents the reasons (translated from Finnish) the test subjects gave for their decisions. The answers are organised on the basis of semantic relationships between primes and their targets that were discussed in the previous sections (4.1 and 4.2).

1) Schematic implication (**Christmas** → *present* or *ham*)

Both the correct translation ‘present’ and incorrect translation ‘ham’ are considered suitable for *kingitus*, because they are popular items to buy at Christmas time.

(1) *kingitus* – it could be ham [FI ‘joulukinkku’]..., but the text as such did not deal with food. (LC)

(2) Nothing else comes to mind. Many people buy ham before Christmas and that word is similar to the Finnish word for ham. (LC)

2) Inclusion (hyponym **book** → hypernym *present*)

The idea of a hypernym consisting of hyponyms, a kind of inclusion, also seems to contribute to comprehension.

(3) Since it deals with Christmas, Christmas presents are usually bought before Christmas. Especially when the following sentence talks about relatives getting books. (SC)

(4) *gift* (because at Christmas, at least I do buy gifts) (LC)

As we assumed, the primes (**Christmas** and **book**) can support the process of inference. In the case of the second prime (**book**), the informant (5) has relied on the false friend translation 'Bible' (**raamatuid** '*Raamattu') and seems to be critical of the suitability of the choice. However, in principle, the Bible could be as good, although rare hyponym for present as any other book:

(5) *Kingitusi*: Context, the assumption that we are talking about Christmas and buying the Bible. As a gift? (SC)

(6) From the point of view of the text the default is that something is bought before Christmas. Later e.g. books are mentioned. (SC)

3) Same semantic field (the co-hyponyms **mother** and *father* or **father** and *mother* or *father*, *mother*, *brother* and *sister*)

In the cases of the primes **mother** or **father**, the test subjects offer *father* or *mother* as target words. They prefer to include the target with a narrow hypernym category 'parents'. The correct translation of *õde* ('sister') would presuppose a wider group of family members, more general co-hyponymy. The association of **mother** and **father** is a strong principle that creates conventional lexical relations. However, some other relatives are also suggested.

(7) Comparison of the word pair mother and father, when *ema* must be mother. (LC)

(8) *mother*. I assumed the word 'isa' meant father, father and mother sounded natural (BT)

(9) *aunt* (guess based on connection, some relative) (LC)

On the other hand, a kind of nuclear family with three members (**father, mother, brother**) was considered to need a fourth complement, also the correct translation, *sister*.

(10) Most of the text was not understood, but I could conclude that the narrator is talking that relatives take part in Christmas celebration, and *mother, father* and *brother* could be concluded on the basis of Finnish, so *sister* seemed to be a natural continuation in this story (SC)

(11) Logical choice for a word list: *mother, father...* (and who else there can be in the family) well *brother* and *sister*. (SC)

4) Other: external similarity, context, world knowledge etc.

Because assumed similarity is stronger than actual similarity, adherence to the idea of resemblance can lead to incorrect translations (see also Ringbom 2007: 7, 25–26). As previous studies (Kaivapalu 2005: 271; Kaivapalu & Muikku-Werner 2010: 83) have shown, speakers of Estonian and of Finnish trust the similarity of these cognate languages. False friends placed in the translation test on purpose, misled the test subjects, as expected (12). In other cases, other languages, such as Swedish (13, 14, 15) or even English (14, 15) could be the starting point for inference.

(12) *Kingitusi*: The beginning of the word sounds like *ham*. (SC)

(13) *Õde*: fate, taken from Swedish :) [Swedish: *öde*] (SC)

(14) *Vend* recalled the Swedish word *vän* and so I translated it as *friend*. The *öde* in turn, reminded me of the English word *other* and, on the basis of pronunciation, I inferred it to mean *others*. (SC)

(15) *öde* – my uncle (I guess from the context and the word *öde* seemed a bit like *eno*. English *uncle* also came to my mind.) (LC)

The test subjects also considered the context as one reason to bring the prime and target together.

(16) *kingitusi* – contextual guess, people usually buy them then (LC)

(17) *Kingitusi*: contextuality – gifts are usually bought on birthdays (BT)

The concepts of knowledge of world and customary behaviour were also mentioned when describing the interpretation:

- (18) *kingitus* a conclusion based on world knowledge and other content of the sentence. (LC)
- (19) *Gift*. Because the following describes what you are going to buy, that is, would fit in with the Christmas customs. (SC)

The test subjects often agreed on the explanations they provided for their solutions in the translation tests. Semantic priming, the creation of relationships between primes and targets, seems to involve internalised metalinguistic knowledge. However, it is not merely separate primes that are guiding the translation process. The thematic unity of the text and default values offered by the Christmas schema also influence the decision making (see example 1). Most of the results of this study are similar to the results of previous research on understanding Finnish or Estonian texts on the basis of the cognate L1. It has been shown that test subjects base their choices of translation equivalents on a conception of coherent texts or on contextual suitability (Kaivapalu & Muikku-Werner 2010: 83–84; Muikku-Werner 2015: 205, 2016: 332).

5. Translation task data versus corpus data

The following section discusses whether the primings produced by informants in the translation tasks actually take place in naturally occurring language. In order to compare translation task results with naturally occurring Finnish, we consulted the Finnish Language Text Collection (FTC). The collocations of the abovementioned Christmas and family member themed items are analysed using two significance tests. Two tests, namely independent sample t-test and Mutual Information test (MI), are used because they measure significance differently: T-test typically picks up grammatical words that are frequent in the datasets. The MI-test, on the other hand, places more emphasis on the relationships of lexical words. (Cf. Barnbrook 1996: 100–101.)

The translation tests revealed that the prime **joulud** ‘Christmas’ induced the target *lahja* ‘present’ in 64% of the cases in the LC test and 52% in the SC test, the share of *kinkku* ‘ham’ targets being only 32% and 20%, respectively (see Table 2). When we examine the co-occurring words of **joulu** ‘Christmas’ in the corpus data, we arrive at the following figures: **joulu** occurs in the data 3868 times. The most significant collocates according to the t-test are *alla* ‘before, below’, *ennen* ‘before’, *viettää* ‘celebrate’, *aikaan* ‘during’, and *jälkeen* ‘after’, and according to the MI test *saituri* ‘scrooge’, *menoliikenne* ‘outbound traffic’, *välipäivä* (‘a day between Christmas and New Year’), *aatonaatto* (‘eve’s eve’) and *paluuliikenne* (‘inbound traffic’). Both statistical tests score targets *kinkku* and *lahja* as significant collocates, *kinkku* being slightly more significant than *lahja*. The t-score for *kinkku* is 5.94735 ($f = 37$)¹ and it is the 14th most significant collocate of **joulu**; for *lahja* the t-score is 5.84635 ($f = 39$), and it is the 15th most significant collocation. The MI test, on the other hand, indicates that *kinkku* is the 12th most significant ($f = 37$, MI = 5.48926), whereas *lahja* is the 24th most significant ($f = 39$; MI = 3.96949). Thus, both tests reveal that *kinkku* has a slightly stronger association with *joulu* than *lahja* does. The collocational priming **joulu** → *kinkku* together with the false friend relation *kingitus* vs. *kinkku* support those solutions in which *kingitus* was translated as *kinkku*, not as *lahja*. The reason for the popularity of *lahja* in translation tests may be the second stimulus **raamatuid** ‘books’, which follows target *kingitusi* ‘presents’ in the near context.

In the BT, on the other hand, there is no contextual or extra-linguistic stimulus which could lead to *kinkku* being provided as an equivalent for *kingitus*. On the contrary, the prime **sünnipäev** ‘birthday’ acts as a stimulus for the target *lahja*. This association is not, however, supported by the corpus data: *lahja* does not seem to be a significant collocate of

¹ A web-based corpus query tool Lemmie 2.0 (available in the Language Bank of Finland 2002–2016) used in the study provides both the scores of the association strength and absolute frequencies the collocates, but no other measures, such as p-values.

syntymäpäivä, with *viettää* and *juhlia* ('celebrate') being the most significant collocates according to both tests.

The other case, the target *ōde* 'sister' revealed that the primes **isa** 'father' and **ema** 'mother' worked as stimuli for their counterparts: **isä** was related to *äiti*, and vice versa, but not to *sisar* or *sisko* 'sister'. Furthermore, the tests showed that when the informants translated the SC text in which *ōde* was listed together with other close family members (**ema**, **isa**, **vend**), the schematic relationship (family members) led to the correct translation equivalent, *sisar* or *sisko* 'sister' (see chapter 4.2).

The corpus data are in line with the observations from the translation tests. **Isä** and **äiti** are collocates, and *sisar* collocates with other family members. First, in the t-test, the second most significant collocate of *äiti* 'mother' (f = 6810) is **isä** 'father' (f = 804, t = 27.78526) (*lapsi* 'child' being the most significant); in the MI test **isä** is significant as well (MI = 5.51131, seventh most significant). The picture is reversed when we focus on the collocates of **isä** (7045): *äiti* is the most significant collocate in the t-test (f = 807, t = 27.69518), and in the MI test it is the fifth-most significant (MI = 5.31642). Thus, the collocational information based on corpus data go hand in hand with the translation tests: **isä** and *äiti* work as stimuli for each other, but not for *sisar/sisko*, which was the actual target word. The informants may have also relied on the weak priming of **äiti/isä** → *sisar/sisko*: in the collocate list of **äiti**, *sisar* is 20th most significant and *sisko* 34th most significant in the t-test, and in the MI test 21st and 24th, respectively. For **isä**, *sisar* is the 45th and *sisko* the 58th most significant collocate according to the t-test, in the MI test 33rd and 34th, respectively.

However, the nesting of family members in the SC test changes the priming effect completely: **vend** 'brother' can be seen as one stimulus for the target *sisko/sisar* in addition to the nesting that requires one missing element (**mother**, **father**, **brother** and **x**). The corpus data show that for **veli** 'brother', *sisar/sisar* is a very significant collocate (*Veli* is also a proper name for men in Finnish, but this usage is omitted from the collocation lists): in the t-test the fourth (*sisar*, t = 7.46635) and sixth (*sisko*

($t = 7.20807$) most significant and in the MI test the most significant (*sisko*, MI = 6.65903) and third most significant (*sisar*, MI = 6.49868). Consequently, the actual language usage found in the corpus supports the findings in the translation tests.

6. Discussion

On the basis of previous research (see Section 2) we know that while inferring meanings of unfamiliar Estonian words, Finnish test subjects rely on contextual information. Cues that include knowledge of the world and the theme and coherence of a text are utilised. In addition, Finnish test subjects seem to benefit, though not necessarily consciously, from semantic relationships between the prime and the target. Semantic priming strategies can be used for word recognition, if the prime is familiar on the basis of similarity between the cognate languages, or on an L2. Of course, the abovementioned strategies may sometimes result in a wrong translation.

The translation tests used in both the previous research (Muikku-Werner 2017) and in the current one, were based on texts with topics (Christmas and birthday) familiar to the participants. Despite the familiarity of the topics, the participants often failed to translate the texts correctly, which led us to dig deeper into the processes of translation. We were especially interested in finding out how different primes, and the distance between the prime and the target would affect finding a correct translation.

In the present study, we first asked how the distance between the prime and the target affects understanding an unknown word, if the target is a false friend (*kingitus* *'kinkku'). The results conveyed that the distance between the prime and the target influences the inference process: the participants performed better, when the latter prime **raamatuid** 'books' (previous prime being **jõulud** 'Christmas') was brought closer to the target *kingitusi* 'presents' by omitting one sentence. The prime **raamatuid**, despite the fact that it followed the target, decreased

the effect of assumed similarity: the incorrect translation *kinkku* ‘ham’ for *kingitus* (instead of the correct translation ‘present’) became a less popular response. However, the total number of correct translations did not increase, despite the deletion of the elements that interrupted the shopping theme, such as descriptions of Christmas tree decorations. In contrast, some other false friends for *kingitus* were produced as translations, such as *kenkä* (‘shoe’) and *kuningas* (‘king’). Thus, orthographic similarity is a strong factor in the inference process.

As expected, the schematic change from Christmas to birthdays removed ‘ham’ as a translation to ‘present’. Some responses such as *Kingis-jäätelö* ‘Kingis ice-cream’ and *kenkälanka* ‘shoestring’ are at least partially the outcome of the false-friend effect. In this thematic unit, the latter prime **raamatuid** did not guide the translation. For some unknown reasons, the hypernym ‘present’ was now a less popular equivalent for *kingitus* than it was in the Christmas texts. Instead, the test subjects preferred different co-hyponyms such as *kirja* ‘book’, *taulu* ‘painting’ or *kukka* ‘flower’.

Secondly, we wanted to know, if the narrowness of a semantic category has an impact on the translations. We assumed that the higher number of primes in the close context guides the translation process of the target. The results suggest that this hypothesis was correct. “Tight” semantic priming, such as hypernym category ‘parents’ (see also below collocational preference) can contribute to producing a sort of semantic false friend while naming the members of the group: with the primes **isa** ‘father’ or **ema** ‘mother’ the target word *õde* was translated as ‘äiti’ or ‘isä’ instead of ‘sister’. On the other hand, the addition of ‘brother’ (**vend**) as a representative of siblings, in other words completing the subcategory of family, helped the inference process and made the correct translation equivalent *õde* ‘sister’ easier to find.

Our assumptions of lexical inferencing proved to be valid by looking at both the translation data, and the reported translation strategies. The answers provided by the test participants exposed common views on the significance of different semantic categories and the potential members

of them. The participants often recognised the implications of the schematic coherence.

Finally, we assumed that L1 speakers are able to transfer their lexical and priming information related to collocational preferences of a word to an L2 and that these assumptions help to find the correct translation. This was tested by analysing corpus data that at least partly mirror Finnish speakers' mental lexicon and lexical priming, since the data are drawn from national newspapers with several types of journalistic genres. The collocation analysis done using two tests of significance confirmed that collocational regularities and lexical priming can be and are used to improve intelligibility: the collocational patterns found in the corpus have counterparts in the test participants' translations. Furthermore, not even a single instance of the target translation 'present' occurred in corpus analysis. The lack of this collocate highlights the observation that the hypernyms are more significant collocations, and the hyponyms do not get that much weight. They of course are possible and available collocates, as shown in the results of the translation test using the birthday text.

Due to the limited number of participants, these results should be considered tentative, although they are somewhat consistent with previous research. By increasing both the number of test participants and test variants (lexical items), the reliability could increase. The primes as well as other thematically important lexical elements used in our tests were cognate words. Even though in some cases trusting similarity caused incorrect translations of targets, the similarity between Estonian and Finnish is an important factor when trying to comprehend the texts.

Our study suggests that the qualities of semantic relations depend on more than the affinity of Estonian and Finnish; on more universal regularities. In addition to known primes, several prerequisites for increased intelligibility were created by the familiar Christmas or birthday themes (schema) with their conventional elements. Thus, the universality of many semantic implications observed in the test results (or mentioned by the test participants) contribute to potential applicability: semantic

relationships are not language specific, so they help both the learners and the translators of Estonian, and participants in Estonian–Finnish multilingual interaction (receptive multilingualism situations). It is possible to assume that by increasing the awareness of lexical priming and potential collocations, some lexical items would emerge as intelligible. Furthermore, by retrieving semantic information from the native language, L2 text comprehension could improve, particularly for related languages.

Acknowledgements

We would like to express our gratitude to professor Annekatrin Kaivapalu for checking the Estonian texts, our informants for providing research material and the anonymous referees for their useful comments. Furthermore, we thank the Department of Language and Communication Studies at the University of Jyväskylä for financial support for writing retreats.

References

- Barnbrook, Geoff 1996. *Language and Computers. Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.
- Bernhardt, Elizabeth B. 1991. *Reading Development in a Second Language. Theoretical, Empirical & Classroom Perspectives*. Norwood, NJ.: Ablex.
- Braunmüller, Kurt 2007. Receptive multilingualism in Northern Europe in the Middle Ages: A description of a scenario. – Jan D. ten Thije, Ludger Zeevaert (eds.). *Receptive Multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts*. Hamburg Studies on Multilingualism 6. Amsterdam: John Benjamins, 25–47. <https://doi.org/10.1075/hsm.6.04bra>
- CSC = Tieteen tietotekniikan keskus (1998). *Suomen kielen tekstikokoelman Helsinki-Korp-versio*. Kielipankki. <http://urn.fi/urn:nbn:fi:lb-2016050207>
- Dufva, Hannele 1999. Kieli, mieli ja konteksti: psykolingvistiksestä kielentutkimuksesta dialogiseen kielen psykologiaan [‘Language, mind and context: From psycholinguistics to a dialogical psychology of language’]. – Kari Sajavaara, Arja Piirainen-Marsh (eds.). *Kielenoppimisen kysymyksiä. Soveltavan kielentutkimuksen teoriaa ja käytäntöä*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto, 11–44.

- Firth, John Rupert [1968] 1957. A synopsis of linguistic theory, 1930–55. – Frank Robert Palmer (ed.). *Selected Papers of J. R. Firth 1952–1959*. London: Longman, 168–205.
- FTC = The Finnish Language Text Collection. An electronic document collection of the Finnish language containing 6 million words. Gatherers: The Department of General Linguistics, University of Helsinki; The University of Eastern Finland; CSC – IT Center for Science. <http://urn.fi/urn:nbn:fi:lb-201403268>
- Gooskens, Charlotte, Vincent van Heuven 2017. Measuring cross-linguistic intelligibility in the Germanic, Romance and Slavic language groups. – *Speech Communication* 89 (May), 25–36. <https://doi.org/10.1016/j.specom.2017.02.008>
- Gooskens, Charlotte, Vincent van Heuven, Jelena Golubović, Anja Schüppert, Femke Swarte, Stefanie Voigt 2018. Mutual intelligibility between closely related languages in Europe. – *International Journal of Multilingualism* 15 (2), 169–193. <https://doi.org/10.1080/14790718.2017.1350185>
- Halliday, Michael A. K., Ruqaiya Hasan 1976. *Cohesion in English*. London: Longman.
- Hoey, Michael 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hoey, Michael 2007. Lexical priming and literary creativity. – Michael Hoey, Michaela Mahlberg, Michael Stubbs, Wolfgang Teubert (eds.). *Text, Discourse and Corpora: Theory and Analysis*. London: Continuum, 7–29.
- Härmävaara, Hanna-Ilona, Charlotte Gooskens 2019. Mutual intelligibility of Finnish and Estonian vocabulary. – *Lähivõrdlusi. Lähivertailuja* 29, 16–19. <https://doi.org/10.5128/LV29.01>
- Jantunen, Jarmo 2009. ”Minulla on aivan paljon rahaa” – Fraseologiset yksiköt suomen kielen opetuksessa [‘I have really lots of money’ – Phraseological units in the teaching of Finnish]. – *Virittäjä* 113 (3), 356–381.
- Jantunen, Jarmo 2017. Lexical and morphological priming: A holistic phraseological analysis of the Finnish time expression *kello*. – Michael Pace-Sigge, Katie J. Patterson (eds.). *Lexical Priming: Applications and Advances*. *Studies in Corpus Linguistics* 79. Amsterdam, Philadelphia: John Benjamins, 253–271. <https://doi.org/10.1075/scl.79.10jar>
- Kaivapalu, Annekatrin 2005. Lähdekieli kielenoppimisen apuna [‘Contribution L1 to Foreign Language Acquisition’]. *Jyväskylän Studies in Humanities* 44. Jyväskylä: Jyväskylän yliopisto.

- Kaivapalu, Annekatrin, Pirkko Muikku-Werner 2010. Reseptiivinen monikielisyys: miten suomenkielinen oppija ymmärtää viroa äidinkielsä pohjalta? ['Receptive multilingualism: How Finnish as a first language helps learners to understand Estonian'] - *Lähivördlusi. Lähivertailuja* 20, 68–97. <https://doi.org/10.5128/LV20.03>
- Lutjeharms, Madeline 2007. Processing levels in foreign-language reading. - Jan D. ten Thije, Ludger Zeevaert (eds.). *Receptive Multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts. Hamburg Studies on Multilingualism 6*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 265–284. <https://doi.org/10.1075/hsm.6.18lut>
- McNamara, Timothy P. 2005. *Semantic Priming: Perspectives from Memory and Word Recognition*. New York: Psychology Press. <https://doi.org/10.4324/9780203338001>
- Minsky, Marvin 1975. A framework for representing knowledge. - Patrick Winston (ed.). *The Psychology of Computer Vision*. New York: McGraw-Hill, 211–277.
- Muikku-Werner, Pirkko 2013. Vironkielisen tekstin ymmärtäminen suomen kielen pohjalta ['Understanding Estonian texts on a Finnish language base']. - *Lähivördlusi. Lähivertailuja* 23, 210–237. <https://doi.org/10.5128/LV23.09>
- Muikku-Werner, Pirkko 2015. Tekstin semanttiset sidokset ja lähisukukielten ymmärrettävyys ['Semantic cohesion and the intelligibility of closely related languages']. - *Lähivördlusi. Lähivertailuja* 25, 191–216. <https://doi.org/10.5128/LV25.08>
- Muikku-Werner, Pirkko 2016. Monitasoinen samankaltaisuus ja sukukielisen tekstin ymmärtäminen ['Multidimensional similarity and comprehension of closely related language']. - *Lähivördlusi. Lähivertailuja* 26, 311–338. <https://doi.org/10.5128/LV26.10>
- Muikku-Werner, Pirkko 2017. Lexical inferencing and the mutual intelligibility of Estonian and Finnish. - *Nordic Journal of Linguistics* 40 (2), 201–223. <https://doi.org/10.1017/S0332586517000105>
- Muikku-Werner, Pirkko, Maria Heinonen 2012. *Lumesadu*- 'tarina' vai 'lumikasa' vai ei kumpikaan? Suomalaiset lukiolaiset viron sanoja tunnistamassa ['*Lumesadu*- 'tarina' or 'lumikasa' or something completely different? How Finnish senior high school students try to recognise Estonian words']. - *Lähivördlusi. Lähivertailuja* 22, 157–187. <https://doi.org/10.5128/LV22.06>

- O'Malley, J. Michael, Anna U. Chamot 1990. *Learning Strategies in Second Language Acquisition*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524490>
- Oxford, Rebecca L. 1990. *Language Learning Strategies*. Boston, MA: Heinle & Heinle.
- Paajanen, Ilona, Pirkko Muikku-Werner 2012. *Tee on kitsas – onko 'tee kitkerää' vai oletteko 'te saita'?* Suomalaiset opiskelijat viroa ymmärtämässä [*'Tee on kitsas' – is 'tea bitter' or are you 'penny-pinching'?* Finnish students comprehending Estonian']. - *Lähivõrdlusi. Lähivertailuja* 22, 219–258. <https://doi.org/10.5128/LV22.08>
- Pace-Sigge, Michael 2013. The concept of Lexical Priming in the context of language use. - *ICAME Journal* 37, 149–173.
- Pace-Sigge, Michael, Katie J. Patterson 2017. Introduction. - Michael Pace-Sigge, Katie J. Patterson (eds.). *Lexical Priming: Applications and Advances*. *Studies in Corpus Linguistics* 79. Amsterdam & Philadelphia: John Benjamins, XII–XXIII. <https://doi.org/10.1075/scl.79.003int>
- Rehbein, Jochen, Jan D. ten Thije, Anna Verschik 2012. *Lingua receptiva (LaRa) – remarks on the quintessence of receptive multilingualism*. - *International Journal of Bilingualism* 16 (3), 248–264. <https://doi.org/10.1177/1367006911426466>
- Ringbom, Håkan 2007. *Cross-Linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters LTD. <https://doi.org/10.21832/9781853599361>
- Sinclair, John 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John 1996. The search for units of meaning. - *Textus* IX, 75–106.
- Vaurio, Leena 1998. *Lexical Inferencing in Reading in English on the Secondary Level*. *Jyväskylä Studies in Education, Psychology and Social research* 145. Jyväskylä: University of Jyväskylä.
- Zeevaert, Ludger, Jan D. ten Thije 2007. Introduction. - Jan D. ten Thije & Ludger Zeevaert (eds.). *Receptive Multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts*. *Hamburg Studies on Multilingualism* 6. Amsterdam: John Benjamins, 1–21. <https://doi.org/10.1075/hsm.6.02zee>

Appendix

SHORT CHRISTMAS TEXT (SC)

Kohe on **jõulud** käes! Jõulu eel ostame *kingitusi*. Käesoleval aastal ostan sugulastele **raamatuid**. Üks raamat võib olla kallis, teine odav. Möödunud jõulu ajal tegime palju pilte. Sellel pildil on minu **isa, ema, vend** ja *õde*. Sellel pildil on minu naine. Ta on rase. Laps sündis juulis.

[Soon it will be **Christmas**. Before Christmas we buy *presents*. This year I'll buy **books** for relatives. One can be expensive, the other cheap. Last Christmas we took lots of photos. In this photo are my **father, mother, brother** and my *sister*. In this photo is my wife. She is pregnant. The child was born in July.]

LONG CHRISTMAS TEXT (LC) (original text Muikku-Werner 2015, 2016)

Kohe on jõulud käes! **Jõulu** eel ostame *kingitusi*. Kaunistame kuuse ja mängime selle ümber. Käesoleval aastal ostan sugulastele **raamatuid**. Üks raamat võib olla kallis, teine odav. Katame laua rikkalikult. Sellel on pähkleid ning õlgedest tähed. Potis on üks hüatsint. See on ilus lill. Peale selle armastan ka roose ja tulpe. Möödunud jõulu ajal tegime palju pilte. Sellel pildil on minu **ema** ja *õde*. Sellel pildil omakorda on minu naine. Ta on rase. Laps sündis juulis. Jõulupühadel on inimestel aega olla perekonnaga.

[Soon it will be **Christmas**. Before Christmas we buy *presents*. We decorate the Christmas tree and play around it. This year I'll buy **books** for relatives. One can be expensive, the other cheap. We'll set the table abundantly. On it there are nuts and straw stars. In the pot is a hyacinth. It is a beautiful flower. Besides it I like roses and tulips. Last Christmas we took lots of photos. In this photo are my **mother** and my *sister*. In this photo is my wife. She is pregnant. The child was born in July. In Christmas time people have time to be with their family.]

BIRTHDAY TEXT (BT)

Kohe on minu ema **sünnipäev**. Ostame temale *kingitusi*. Käesoleval aastal ostan talle **raamatuid**. Üks raamat võib olla kallis, teine odav. Eelmisel sünnipäeval tegime palju pilte. Sellel pildil on minu **isa** ja *õde*. Sellel pildil on minu naine. Ta on rase. Laps sündis juulis.

[Soon it will be my mother's **birthday**. We'll buy her *presents*. This year I'll buy her **books**. One can be expensive, the other cheap. Last birthday we took lots of photos. In this photo are my **father** and my *sister*. In this photo is my wife. She is pregnant. The child was born in July.]

Vironkielisten fraseologisten yksiköiden ymmärtäminen suomen kielen pohjalta: helpottavia ja harhauttavia seikkoja

PIRKKO MUIKKU-WERNER¹,
JARMO HARRI JANTUNEN²
Itä-Suomen yliopisto¹, Jyväskylän yliopisto²

Vironkielistä tekstiä lukevien suomalaisten on havaittu hyödyntävän monenlaisia strategioita selvittääkseen sukukielen sanojen merkityksiä. Yksi niistä on pohjustaminen (*priming*). Kun kyseessä on fraseologinen yksikkö, yksi osa (*prime*) pohjustaa toisen osan eli kohdesanan (*targetin*) esiintymistä, mikäli *prime* on tunnistettavissa äidinkielen perusteella. Tässä artikkelissa kuvaamme muutamia erityistilanteita. Miten pohjustimen ja kohdesanan välinen etäisyys tai teemanvaihdos vaikuttavat 'petollisen ystävän' ymmärtämiseen? Lisäksi tutkimme sitä, miten ahdasrajainen semanttinen kategoria ohjaa siitä puuttuvan jäsenen merkityksen löytämistä ja millaisia perusteita käännösratkaisuille on löydettävissä. Lopuksi analysoimme korpusaineiston avulla, onko kyseisen kaltaisille kytkennöille löydettävissä tukea L1-tekstien kollokaatioista.

Kyetäksemme vastaamaan näihin kysymyksiin laadimme testin, johon osallistuneiden informanttien oli käännettävä lyhyt teksti suomeksi ja perusteltava tekemiään valintoja. Yhteensä 25 testatusta kukaan ei ole opiskellut viroa.

Tulokset osoittavat, että ulkoiseen samankaltaisuuteen tukeudutaan eniten. Vahva luottamus aiheuttaa käännösvirheitä, jos kohdesana on 'petollinen ystävä'. Kun toinen tarkentava pohjustin siirretään lähemmäksi kohdesanaa tai teema vaihdetaan sellaiseksi, ettei 'petollinen ystävä' sovi siihen, väärät vastaukset vähenevät. Myös ahtaan semanttisen kategorian jäsenmäärän rajallisuus ohjaa kääntämistä. Testattavat selittävät ratkaisujaan ensisijaisesti sillä, että sanat muistuttavat toisiaan, mutta he nojautuvat myös kontekstiin, tekstien teemaan ja fraseologisten yksiköiden jäsenten välisiin semanttisiin suhteisiin. Niiden voidaan katsoa edustavan universaalia samankaltaisuutta: ilmiöt hahmotetaan lähes yhdenmukaisesti kielistä riippumatta. Lisäksi L1-korpuksista oli

löydettävissä samantapaisia kollokaatioita, joten L1:n fraseologiasuhteet saattavat myötävaikuttaa valintoihin.

Avainsanat: kielten keskinäinen ymmärrettävyys; semanttinen pohjustaminen; pohjustin; kohdesana; viro; suomi

Pirkko Muikku-Werner

Itä-Suomen yliopisto, humanistinen osasto, suomen kieli ja kulttuuritieteet

PL 111

80101 Joensuu, Finland

pirkko.muikku-werner@uef.fi

Jarmo Harri Jantunen

Jyväskylän yliopisto, kieli- ja viestintätieteiden laitos

PL 35

40014 Jyväskylän yliopisto, Finland

jarmo.jantunen@jyu.fi