

Mutual intelligibility of Finnish and Estonian vocabulary

HANNA-ILONA HÄRMÄVAARA¹,

CHARLOTTE GOOSKENS²

University of Washington¹, University of Groningen²

Abstract. This paper reports on an investigation on Finnish-Estonian mutual intelligibility carried out by means of a word translation task among 307 Finnish and 118 Estonian participants. The results confirm previous findings by Kaivapalu (2015) that both in the spoken and in the written mode Estonians understand isolated words in Finnish better than Finns understand isolated words in Estonian. Older participants performed better than younger participants.

Many of the participants had at least some previous exposure to the test language and were therefore familiar with some words and sound correspondences. However, we were also interested in testing how well speakers of the two languages could understand the other language purely on the basis of the similarity of the languages. We therefore separately analyzed the data produced by a subset of participants with no or little exposure to the test language. The results were still asymmetric, but only in the spoken mode. We looked at various linguistic and non-linguistic factors that could explain why some cognates are more easily recognized than others.

To put our results into perspective, we compared them to the results of the similar experiments involving 70 Germanic, Romance and Slavic language combinations.

Keywords: mutual intelligibility; asymmetric intelligibility; Estonian; Finnish; cognancy

1. Introduction

When an Estonian speaker encounters words in Finnish or a Finnish speaker in Estonian, one can assume her or him to understand at least some of them. This assumption is based on the similarity of these cognate languages. However, as the previous research shows (see section 1.2), Finnish and Estonian speakers do not effortlessly understand the cognate language, and there seems to be an asymmetry in how and how well the Finnish and Estonian speakers understand the other language.

This article contributes to the body of research articles that have during the past decade discussed mutual intelligibility of Finnish and Estonian, and that have mostly been written in Finnish or Estonian. The aim of this article is to summarize the existing research in English and to present a new study on mutual intelligibility of Finnish and Estonian vocabulary. This study uses the same methodology as a recent investigation on mutual intelligibility of Slavic, Germanic and Romance languages, which makes it possible to compare the Estonian-Finnish mutual intelligibility to the mutual intelligibility of other closely related languages. We will also explore the possible explanations for the asymmetry in understanding.

1.1. Finnish and Estonian

Finnish and Estonian are fairly closely related Finnic languages that belong to the Finno-Ugric language family. Finnic languages form a dialect chain around the Baltic Sea, and the varieties can be divided into northern and southern Finnic languages. Finnish belongs to the northern and Estonian to the southern branch (see Laakso 2001: 206–207). As a result of the shared genetic background and later contacts between Finnish and Estonian, the languages share a great amount of structural, lexical and functional similarities.

In addition to the original and derived Uralic vocabulary, the languages contain different historical layers of loanwords, mainly from

Indo-European and especially from Baltic and Germanic languages, and in Estonian there is a significant amount of loanwords from Finnish (see Laakso 2001 for an overview of Finnic languages; see also Häkkinen 1990; Erelt et al. 2000; Grünthal 2009). Still, there are notable differences in the vocabulary, resulting from a series of changes the languages have undergone during their independent histories.

Finnish and Estonian are spoken in neighbouring countries with a rich history of contacts. During the Soviet times Finnish had high prestige in Estonia. Even today 21% of Estonians claim to have a command of Finnish on the communicative level (European Commission 2012: 21). Estonian does not have the same status in Finland. It is not generally taught in Finnish schools, whereas in many Estonian schools Finnish is one of the foreign languages offered. The EU and the free mobility of labor have intensified the contacts between Estonia and Finland, and there is a large Estonian minority population in Finland (see Koreinik & Praakli 2017).

Due to the linguistic similarities of Finnish and Estonian and the close contacts between Estonia and Finland, speakers from the two countries are reported to sometimes communicate in the receptive multilingualism (RM) mode, each participant using their native language (see Härmävaara 2014, 2013; Härmävaara & Frick 2016; Verschik 2012; and especially Härmävaara 2017 for historical and ideological reasons of employing RM instead of or alongside with English as *lingua franca*).

When a language pair is so closely related that the speakers are able to understand each other without prior language instruction in the non-native language, it is called inherent receptive multilingualism. Acquired receptive multilingualism refers to a situation in which the participants have acquired knowledge of the non-native language and are thus able to understand each other (Bahtina & ten Thije 2012). We use the same terms for differentiating between intelligibility based on the linguistic similarity of the languages (inherent intelligibility) and intelligibility based on non-linguistic factors, such as previous experience with the non-native language (acquired intelligibility).

1.2. Previous research on mutual intelligibility of Finnish and Estonian

Mutual intelligibility of Finnish and Estonian is the topic of an ongoing Finnish-Estonian project called REMU. The project focuses on how Finnish and Estonian speakers understand the cognate language based on their native language. Different sub-studies have explored how Finns and Estonians understand texts, sentences and isolated words in the cognate language. The special emphasis of the research conducted within this project is on the relationship between actual and perceived similarity of the languages (see Ringbom 2007). This has guided the methodological choices for combining comprehension tests with self-reflection of understanding and understanding strategies.

The findings of the studies show that the languages are so similar that text level comprehension is possible. A text comprehension test, given to university students without prior knowledge of the cognate language, consisted of a text in Finnish / Estonian. The test takers were asked to write down what they understood of it, and also to reflect on why and how they thought they understood the text and its individual components. All the participants could grasp the basic idea of a text in the cognate language (Kaivapalu & Muikku-Werner 2010; Kaivapalu 2015). According to the participants' introspection, all linguistic levels were used for finding indications of similarity. Especially morpho-syntactical similarities were well recognized and used strategically while trying to comprehend the cognate language¹. However, the Finnish participants scored significantly higher in the text comprehension test than the Estonian participants (Kaivapalu 2015: 61).

To further understand if the mutual intelligibility of the languages really is asymmetric, Kaivapalu (2015) compared the results of the text comprehension tests with the results of earlier isolated word

¹ In Finnish and Estonian, grammatical functions are expressed by using inflectional formatives that are attached to stem allomorphs. Most of the inflectional formatives are similar in Finnish and Estonian (see Remes 2009: 47).

comprehension tests². The results of the isolated word comprehension tests taken by university students with no prior knowledge of the cognate language, showed that the asymmetry in comprehension was this time the other way around: in general, the Estonian participants performed better in the word comprehension test.

The Estonian participants significantly better recognized the words that were both semantically and orthographically similar in Finnish and Estonian (Est. 64.8% vs. Fin. 43.6%). Similarly, the recent loanwords and international words³ were better comprehended by the Estonian participants (56.2% vs. 35.3%). However, Finnish speakers were better at comprehending words that were in oblique meaning relationship through synonymity or partially shared semantic fields (8.0% vs. 24.2%). As an example of such relationship Kaivapalu (2015: 59) gives the Estonian word *vihmavari* ‘umbrella’ (Fin. *sateenvarjo*). The compound word can be intelligible to a Finnish speaker due to the similar latter component, *vari* – *varjo* ‘shade’, and because the first component of the Estonian word, *vihma* (genitive form of ‘rain’) is identical with the Finnish hyponym *vihma* ‘drizzle’.

Kaivapalu (2015) is careful with giving explanations for the asymmetries in understanding, but she brings up two linguistic factors that seem to have contributed to the comprehension, also according to the participants’ reflexive comments. Firstly, the proportion of loanwords compared to inherited words is larger in Estonian than in Finnish, and the Estonian participants of the isolated word test reported to have used the knowledge of other languages they knew more often than the Finnish participants did (ibid. 64–65). However, the larger proportion of loanwords worked in favour of the Finnish participants in the text comprehension test: the Estonian text contained more loanwords,

² The Estonian comprehension test was done by Muikku-Werner & Heinonen (2012) (also Heinonen 2015) and the Finnish comprehension test by Mark (2014).

³ By “recent loanwords and international words”, later referred to as “loanwords” we mean words that are borrowed recently and that have transparent cognates in many Indo-European languages the test takers report knowing.

and they were most often transparent to the Finnish participants (ibid. 65).

Secondly, Finnish has a wider dialectal variation than Estonian, and many features of standard Estonian are similar to some varieties of spoken Finnish. According to Kaivapalu (2015: 64) and in accordance with Berthele (2008) and Gooskens & Heeringa (2014), the greater variation in a native language provides better tools for recognizing sound correspondences between different variants, which could explain why the Finnish participants were better at comprehending words that are in oblique meaning relationship with each other. The Finnish participants reported having used knowledge of the variation of Finnish especially in the text comprehension test (ibid. 70).

The most difficult group of words to translate in the tests reported above, were so-called false friends, words that sound or look similar but have different meanings in Finnish and Estonian. Only 2.4% / 4.7% of the false friends were translated correctly in the studies reported by Kaivapalu (2015). Being aware of the existence of false friends and knowing some of them is a part of the folk linguistic knowledge Finns and Estonians in general have about the cognate language (see Laalo 1992)⁴, but the awareness does not necessarily help in understanding unfamiliar words. Kaivapalu & Muikku-Werner's (2010: 81) participants even reported that they often could not trust the assumed meaning of the similar looking words, since they may *possibly* mean something else.

In their study on how Finnish speakers with no prior knowledge of Estonian understand isolated sentences in Estonian, Paajanen and Muikku-Werner (2012: 249–250) discovered that the amount of correctly translated sentences with a false friend was about 30 percent – significantly more than correctly translated false friends in isolated word tests. The analysis of the introspection of the informants (ibid. 228) showed that the sentence internal context often steered the translation of

⁴ This is also reflected on in all Finnish-Estonian teaching materials (e.g. Pajusalu et al. 1999; Kuusk 1999), and there even are dictionaries on false friends between Finnish and Estonian (e.g. Alvre & Vodja 1993; Wirén 2008a, 2008b).

the false friends and non-cognates. The steering mostly led in the correct direction, whereas the general lack of context beyond the sentence level made it more difficult to avoid a misleading interpretation.

In receptive multilingual interaction, the context and the multimodal elements of interaction play a significant role in understanding. Härmävaara (forthcoming) studied receptively multilingual interaction among Finns and Estonians with language skills varying from a high competence in both languages to no active knowledge of the cognate language. By analyzing multilingual multiparty interaction, Härmävaara (2017, forthcoming) found out that the linguistic closeness of Finnish and Estonian serves as a basis for mutual understanding, and receptive multilingual interaction is in principle possible, although not an easy way to interact for the participants with no active skills in the cognate language.

In the present article we report on a study on the mutual intelligibility of Finnish and Estonian, carried out by the means of a word translation task. The task was developed within a larger project set up to investigate the mutual intelligibility of 16 closely related Germanic, Romance and Slavic languages, titled the Micrela project (Gooskens & van Heuven 2017; Gooskens et al. 2018). By using the same task for the investigation of Finnish-Estonian mutual intelligibility we are able to put the results into perspective. We tested both written and spoken words to be able to compare intelligibility in the two modes; this is the first time spoken stimulus words are used in investigating both Finnish and Estonian word comprehension⁵.

When looking at the results we are especially interested to see whether we find the same asymmetry in the Finnish-Estonian word intelligibility that has previously been reported by Kaivapalu (2015). If so, we will have a closer look at individual words to reach a greater understanding of the linguistic and non-linguistic factors that can explain this asymmetry. We will also compare the results of young and old participants, since the historical developments in the countries have changed the importance and

⁵ Heinonen (2015) used spoken Estonian words in her comprehension test.

visibility of Finnish in Estonia and Estonian in Finland (see Grünthal 2009). We formulate our research questions as follows:

1. How well do Finnish and Estonian speakers understand each other's languages?
 - a. Is there a difference between spoken and written word intelligibility?
 - b. Is there a difference between the younger and the older generation?
 - c. What is the level of mutual intelligibility compared to other European language combinations?
2. Is the Finnish-Estonian mutual intelligibility asymmetric and if so, how can it be explained?

2. Method

2.1. Material

Intelligibility can be measured at several levels of the linguistic hierarchy. When testing overall intelligibility, preference may be given to the text level since text comprehension is closer to reality, where subjects are mostly confronted with whole messages. By testing whole texts, all linguistic levels can be included as factors influencing intelligibility. However, the word level is key to speech understanding – as long as the subject correctly recognizes words, s/he will be able to piece the message together. Also, by focusing on the word level, we can get to the details of comprehension, such as to the phonetic factors, and the effect of context is excluded.

Previous research (Gooskens & van Heuven 2017) has shown that word translation scores correlate highly with intelligibility scores at sentence and text level, as measured by cloze tests and multiple choice questions (correlations between .73 and .79). The correlations with intelligibility as estimated by the participants were high too ($r = .72$ for spoken and .82 for written word intelligibility). We therefore assume that our results can be generalized to reflect the level of mutual intelligibility of Finnish and Estonian languages as a whole.

We tested intelligibility by means of a spoken and a written word translation task. To be able to compare our intelligibility results to the level of mutual intelligibility of other closely related language pairs, we used the same testing method and translations of the same testing words as in a large-scale investigation of the mutual intelligibility of 16 closely related languages from the Germanic, Romance and Slavic language families (the Micrela investigation, see Gooskens & van Heuven 2017). To test word recognition, a list was compiled of the 100 most frequently used nouns in the British National Corpus (BNC Consortium 2007). The list was slightly adapted to exclude pairs of words with similar meanings. If a word was excluded, the next one on the frequency list was added.

For the present investigation, the words were translated into Finnish and Estonian and recorded by two Finnish and two Estonian speakers. All the speakers were female, between 30 and 50 years old and the words were recorded in standard Finnish / Estonian. Each speaker contributed a half (i.e., 50 words) of the stimulus words.

In Appendix 1 the Finnish list is presented with the corresponding Estonian cognates, if any. In the same way the Estonian list with the Finnish cognates is presented. As becomes clear from Appendix 1, each word does not always have a cognate in both directions. For example, the stimulus word for ‘office’ is *kontor* in Estonian and *toimisto* in Finnish. Finnish has a cognate for *kontor*, *konttori*, but Estonian does not have one for *toimisto*. In total there were 20 such cases, 11 in the advantage of the Finnish participants and 9 in the advantage of the Estonian participants. Some words are partial cognates, i.e. translations sharing only part of the semantic field of the test word, through polysemy or homonymy. For example, as a translation of *laki* ‘law’, in addition to *seadus* ‘law’, we had to accept *lagi* ‘high point; ceiling’, since Finnish *laki* also translates to ‘high point’.

Including the partial cognates, for 70 of the 100 Estonian test words there is a corresponding Finnish cognate and for 69 Finnish words there is an Estonian cognate. Therefore only about 70% of the 100 words can be assumed to be possible for a test taker to translate based on her or his native language.

2.2. Intelligibility test

Each participant was presented with a random subset of 50 words from the larger set of 100 words in either the spoken or the written form, to keep the duration of the test within limits. The test language was the cognate language (Estonian for the Finnish participants and Finnish for the Estonian participants). Two versions of the word test were prepared, one for visual presentation and one for oral presentation. In the written version each stimulus word was presented on the computer screen and remained visible until the participant finished typing the response (by pressing the return key) with a time out after 10 seconds. In the spoken mode, the stimulus words were played twice with one second between tokens, again with a maximum time lapse of 10 seconds (trial-to-trial onset).

Participants were instructed to translate each stimulus word into their native language using the computer keyboard. The responses could be evaluated automatically, but since typing errors or synonyms might occur that were not recognized by the software we manually checked all responses that the software could not recognize. Responses were considered correct only if they were (nearly) identical to the target word or one of its synonyms. A single deletion, addition or substitution of a letter, as well as switching two letters, were considered acceptable errors, as long as the string of letters was not identical to another existing word in the participant's native language⁶. For example, Finnish *paikka* 'place, spot' can be translated to Estonian as *paik* 'place' or *koht* 'place, spot'; the response *kohe* 'at once' was judged as incorrect since it is an existing Estonian word, while the non-word *kot* was considered correct.

⁶ Some participants gave their answers in English, most likely due to the fact that the instructions for the test were in English. The responses of these participants were removed from the data.

2.3. Procedure

The participants were tested via an online application (see <http://www.micrela.nl/app>). They were first asked to complete a questionnaire about their native language, age, sex and level of education. They were also asked how much exposure they had had to the test language on six 5-point scales representing different written and spoken situations and for how many years they had learned the language. The answers to the questions were used to select participants with similar backgrounds in order to be able to compare the results of various groups of participants.

In addition to functional intelligibility established by means of an intelligibility test, intelligibility can also be measured by means of so-called judged intelligibility. Judged intelligibility is a measure of how well the participants think they understand a language. We were interested to see whether the ideas that the participants had about Finnish-Estonian intelligibility before taking part in the experiment corresponded to the results of actual testing. Therefore we asked the participants to indicate how well they thought they would be able to understand the test language (i.e., estimated intelligibility), prior to being exposed to any stimuli, on a 5-point scale from 1 = 'not at all' to 5 = 'very well'. It was not made explicit whether this question referred to the understanding of spoken or written language.

After the participants had finished the questionnaire, the intelligibility test started. The entire online session lasted approximately 15 min (questionnaire and test together).

2.4. Participants

Since the participants were tested online, no restrictions concerning their background were set beforehand. We selected participants for further analysis afterwards by matching the groups according to certain criteria. The selected participants all came from Finland or Estonia. They had all grown up and lived most of their lives in the relevant country and

spoke the language of the country as their native language at home. Participants who indicated that they also spoke the test language at home were excluded. All participants were enrolled or had completed a university education.

The criteria described above resulted in a selection of 307 Finnish and 118 Estonian participants. Twenty-eight percent of the Finnish and 51 percent of the Estonian participants were male. We made separate analyses of younger (18 to 33 years, mean 24.9 for Finnish and 26.2 for Estonian) and older (34 to 84 years, mean 43.7 for Finnish and 43.3 for Estonian) participants. This allowed us to compare the level of intelligibility among younger participants to the intelligibility among older participants. It also made it possible to compare our results of the young participants to the results of Micrela investigation, as all its participants were between 18 and 33 years old. An overview of the participants is provided in Table 1.

TABLE 1. *Number of young and old Estonian and Finnish participants in the spoken and the written test. Between brackets is the selection of young participants with minimal exposure to the test language*

Test	Estonian participants		Finnish participants	
	young	old	young	old
Spoken	33 (19)	25	122 (104)	40
Written	37 (15)	23	116 (95)	29

From the questionnaire we have information about the amount of exposure that the participants had had to the test language before taking the test, across the six 5-point scales from 1 (never) to 5 (every day). We also have information about the number of years they had learned the language. The mean scores are presented in Table 2. Estonian participants had had more exposure to Finnish than the other way round. The differences are significant ($t = -5.595$, $df = 306$, for young participants and $t = -3.103$, $df = 115$ for older participants, independent samples t-test,

$p < .001$). The young Estonians had learned Finnish for a significantly longer time than the other way round ($t = -5.112$), but the difference between the older participants is not significant ($p = .743$, $t = -.318$).

TABLE 2. Mean exposure scores across six 5-point scales from 1 (never) to 5 (every day) and the mean number of years that the participants had learned the test language. Between brackets values for the selection of young participants with minimal exposure to the test language

Variable	Estonian participants		Finnish participants	
	young	old	young	old
Exposure	1.8 (1.4)	1.9	1.5 (1.4)	1.6
Years of learning language	1.0 (0.0)	0.5	0.2 (0.0)	0.4

To be able to focus on inherent intelligibility we made an analysis of a selection of results from young participants with as little previous experience with the test language as possible. Ideally, we would have excluded all participants who had been exposed to the test language. Unfortunately, our data set would be too small had we excluded test takers who fulfil this criterion. Due to the close connections between the countries, it is difficult to find Finns or Estonians that have never been in any contact with the cognate language.

Therefore we filtered our data with less strict criteria. We selected a subset of participants who had indicated that their mean exposure on the six 5-point scales was below 2.0 (with '1' indicating no exposure) and who had not learned the test language at school.⁷ This selection included 199 Finnish (104 for the spoken and 95 for the written version) and 34 Estonian participants (19 for the spoken and 15 for the written version). There were not enough older participants that fulfilled these criteria. The exposure values and years of learning for these participants are presented between brackets in Table 2. As Table 2 shows, these participants

⁷ The same selection was made for the European languages that we compare our results to (see Figure 2).

had not learned the test language in a formal setting, but they had had some exposure to it (mean 1.4 on the scale from 1 (no exposure) to five (frequent exposure) for both groups).

3. Results

In Section 3.1., we will first present the results of the intelligibility as estimated by the participants, and then the inherent and acquired intelligibility of the written and spoken words as measured by the translation task. In Section 3.2., we will have a closer look at the word pairs that showed asymmetric mutual intelligibility.

3.1. How well do Finnish and Estonian speakers understand each other's languages?

Before taking part in the intelligibility test, the participants indicated on a scale from 1 (not at all) to 5 (very well) how well they thought they would be able to understand the test language (estimated intelligibility). In the top row of Table 3 we present the results for the old and the young participants in the two countries. None of the means are above the middle of the scale, and in general it seems that Finnish and Estonian speakers do not have very high expectations about their ability to understand the other language.

A two-way analysis of variance shows that the Estonian participants estimate their understanding of the cognate language higher (mean 2.7 for the young participants and 2.9 for the old participants) than the Finnish participants (mean 2.0 and 2.2). This difference is significant ($F = 47,314, p = .000$). The difference between the old and the young participants is also significant ($F = 4,141, p = .042$); the old participants estimate their understanding of the test language higher than the young participants. The interaction between the age of the participants and their language background is not significant ($F = 0,081, p = .776$).

In the two bottom rows of Table 3 we present the results of the spoken and written word translation tasks for the old and young participants in

the two countries. A three-way ANOVA shows no significant interaction between language background, age and test mode ($F = 2,700, p = .101$). An independent samples t-test shows that the Estonian participants perform better than the Finnish participants ($t = -7.237, df = 203, p = .000$ in the written mode and $t = -9.985, df = 218, p = .000$ in the spoken mode). The old participants perform better than the young participants ($t = -4.859, p = .000$ in the written mode and $t = -5.586, p = .000$ in the spoken mode). The participants translate more words correctly in the written mode than in the spoken mode ($t = 2.259, df = 423, p = .024$).

Both the estimated intelligibility results and the results of the functional intelligibility tests show that the mutual intelligibility of Estonian and Finnish is asymmetric. Estonian participants estimate their understanding higher, and this higher intelligibility is confirmed by functional tests of both written and spoken word intelligibility.

TABLE 3. *Estimated, spoken and written functional intelligibility scores. Between brackets are the values for the selection of young participants without previous learning of the test language (for a discussion of these results, see Section 3.2)*

Test	Estonian participants		Finnish participants	
	young	old	young	old
Estimated intelligibility	2.7	2.9	2.0	2.2
Functional spoken	61.3 (51.9)	74.6	44.6 (42.4)	52.1
Functional written	63.7 (49.7)	69.7	48.9 (46.6)	59.0

To gain more insight into how mutually intelligible the languages are, we can put the functional results into perspective by comparing them to the results from an investigation conducted with the same set of words in 16 European countries with 70 language combinations (the Micrela investigation). The participants in this investigation had the same background as far as age and education is concerned as the young Estonian and Finnish participants. Therefore, we only compare our results with the results of the young participants. In Figure 1 we present the results from the

written and spoken word intelligibility experiments from Gooskens & van Heuven 2017 (small, open symbols) together with the results from the Finnish and Estonian young participants (larger, black symbols).

In comparison to the Micrela results, Estonian participants understand Finnish rather well. In the written mode, only 29% of the results in the 70 Micrela language combinations are higher than the Estonian participants' results, and in the spoken mode 21%. However, the results of the Finnish participants are low in comparison to the Micrela results. 71% percent of the Micrela participants perform better in the written mode and 59% in the spoken mode.

A closer look at Figure 1 can give us information about which language combinations show the largest similarity to the Finnish-Estonian results. The spoken intelligibility results of the Estonian participants are closest to the results of Slovak participants tested in Polish and the written results are closest to Italian participants tested in Spanish. The spoken intelligibility results of the Finnish participants are most similar to the results of Romanian participants tested in Italian and the written results are closest to the results of the Bulgarian participants tested in Slovenian.

Figure 1 shows intelligibility among participants with varying degrees of previous knowledge of the test language (acquired intelligibility). It offers an overview of the cross-language intelligibility of related languages obtained for a fair cross-section of young, educated Europeans. However, we were also interested in getting an impression of how high the inherent intelligibility among the Estonian and Finnish participants is, compared to the European language combinations in the investigation of Gooskens & van Heuven (2017). We repeated our analysis with a selection of young participants who had not learned the language and who had had only little exposure to the language (inherent intelligibility). We present the results in Figure 2.

In the spoken mode, 32% of the participants tested in the Indo-European language combinations show higher results than the Estonian participants, and 63% perform better than the Finnish participants. The

MUTUAL INTELLIGIBILITY OF FINNISH AND ESTONIAN VOCABULARY

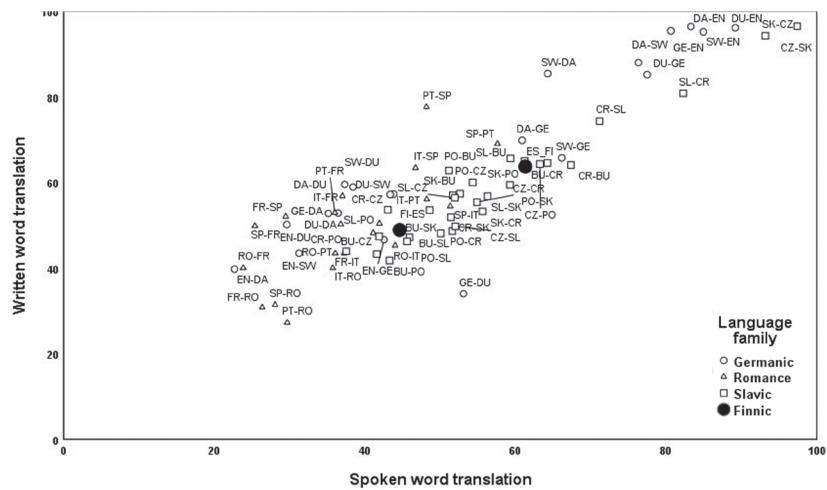


FIGURE 1. Percentages of correctly translated written and spoken words per language group for 70 language combinations from Gooskens & van Heuven (2017) and for the young Finnish and Estonian participants in the present investigation. Germanic (DA = Danish, DU = Dutch, EN = English, GE = German, SW = Swedish), Romance (FR = French, IT = Italian, PT = Portuguese, RO = Romanian, SP = Spanish), Slavic (BU = Bulgarian, CR = Croatian, CZ = Czech, PO = Polish, SK = Slovak, SL = Slovene) and Finnic (FI = Finnish, ES = Estonian). For each language combination, participant language is given first, test language second (e.g. ‘EN-DA’ means ‘English native speakers tested in Danish’)

percentages are even higher in the written mode (54% and 73%). So the rank hardly changes for the Finnish participants in comparison with the results including all participants, while the Estonian participants have a substantially lower rank, if we look at inherent intelligibility. The spoken results of the Estonian participants are closest to the results of Czech participants tested in Slovenian, and the results of the Finnish participants to the results of Dutch participants tested in Swedish. The written results of the Estonian participants are closest to the results of French

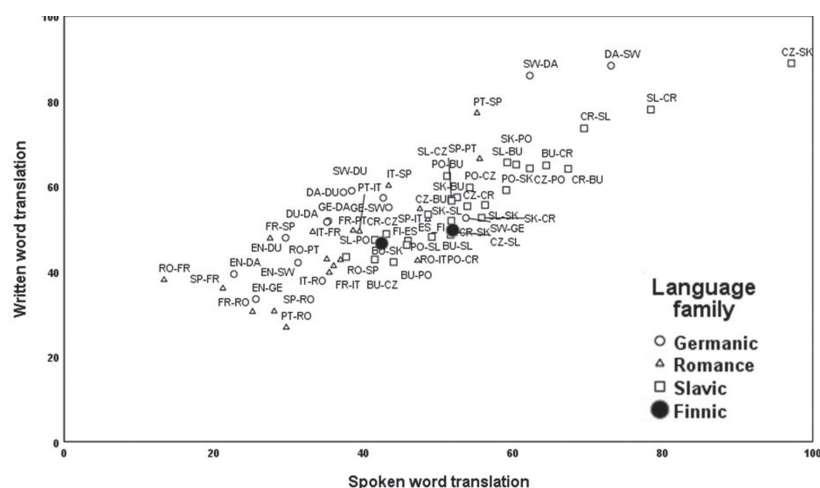


FIGURE 2. Percentages of correctly translated written and spoken words per language group for the 70 language combinations from Gooskens & van Heuven (2017) and for the young Finnish and Estonian participants in the present investigation. Only participants with a minimum of previous exposure to the test language are included (inherent intelligibility). See Figure 1 for explanation of the abbreviations

participants tested in Portuguese and the results of the Finnish participants are closest to the results of Polish participants tested in Slovenian.

3.2. How can the asymmetric Finnish-Estonian mutual intelligibility be explained?

3.2.1. Acquired intelligibility

Table 2 makes clear that on average the Estonian participants have learned the test language for a longer time and have been exposed to it more often than the Finnish participants. The intelligibility results show significant correlations with both exposure and number of years of learning at the .001 level ($r = .51$ and $.60$ for Finnish participants and

.60 and .39 for Estonian participants). This means that there is a relation between the amount of exposure that the participants have had to the test language and how well they can understand the language. The same was found in the Micrela investigation: exposure to the cognate language was the most important extra-linguistic factor that explained asymmetric intelligibility between the tested 70 language combinations (van Heuven & Gooskens 2017). Thus, it is likely that the differences in experience with the test language is part of the explanation for the asymmetric mutual intelligibility.

This becomes evident, if we look at how well the Finnish and Estonian participants translate non-cognates. Non-cognates are in principle unintelligible to a participant unless s/he has learned the language or been exposed to it. As Table 4 shows, Estonian participants translate non-cognates significantly better both in the spoken (42.7% correct translations) and the written (37.6%) data sets than the Finnish participants (6.8% and 8.4%). In this subset of the data there are 21 spoken and 15 written non-cognates that none of the Finnish participants translate correctly. On the contrary, all Finnish words are translated correctly by at least some of the Estonian participants. The result is significantly higher for the Estonian participants than in the study reported by Kaivapalu (2015).

TABLE 4. Percentages correctly translated non-cognates in the written and spoken test by Estonian and Finnish participants

Test	Estonian participants	Finnish participants
Spoken	42.7	6.8
Written	37.6	8.4

The non-cognates that at least half of the Finnish participants translate correctly are either a recent loanword (see following section) *punkt* ‘point’ (79.6% in the written data set and 78.5% in the spoken) or two fairly commonly known false friends, *raamat* ‘book’ (50% written, only 38.3% spoken) and *maja* ‘house’ (67.4% written, 58.7% spoken). Both

false friends are historically connected with their Finnish equivalents, with still somewhat close semantic fields. The Finnish false friend for *raamat* is *Raamattu* ‘Bible’, and *maja* in Finnish means ‘hut; shack’. These both are words that Finnish and Estonian speakers often encounter in the lists of false friends or hear as anecdotes about the differences between the languages. In addition to these three words, about one tenth of the Finnish participants translate correctly the word *aasta* ‘year’, *vuosi* in Finnish (15.2% written, 9.7% spoken). It may be a word that the Finnish speakers recognize for instance from Estonian products known to them, often stating that a brand has been around from a certain year.

As the Table 4 shows, the Estonian participants translate the non-cognates significantly better than the Finnish participants. Among the 10 written and 12 spoken non-cognates that at least 50% of the Estonian participants translate correctly are commonly known false friends and words that cannot be found on the lists of false friends. Commonly known false friends are for example *hallitus* ‘government’ in Finnish and ‘mold’ in Estonian (82.4% written, 70.4% spoken) and *talo* ‘house’ in Finnish, cf. *talu* ‘farm(house)’ in Estonian (75.0% written, 79.3% spoken). The Estonian participants also translate correctly words like *äiti* ‘mother’ (*ema* in Estonian) (64.5% written, 79.3% spoken), *juhlat* ‘party’ (*pidu* in Estonian) (53.6% written, 63.6% spoken), *vuosi* ‘year’ (75% written, 70.4% spoken), *tutkimus* ‘study’ (*uurimus* in Estonian) (57.1% written, 66.7% spoken) and *henkilö* ‘person’ (*isik* in Estonian) (56.3% written, 77.4% spoken).

The result that Estonians translate correctly words that are not commonly known false friends, is most likely linked to the fact that Estonian participants have had more exposure to the cognate language than the Finnish participants. Exposure can be through conscious learning or reading and hearing Finnish, but some of it can also be through Finnish loanwords used in colloquial Estonian. Based on the data consisting of Estonian online news and Internet discussions, Grünthal (2009) shows that Finnish is an important source of loanwords in Estonian even today.

For instance, the Estonized version of the word ‘party’, *juhlad*⁸ gets Estonian hits in a Google search (as of 25.10.2018), and so does *tutkida* in the meaning of ‘do research, study’.

Sometimes also the incorrect translations exhibit knowledge about the other language. For instance, Estonian participants quite often translate *lomake* ‘form’ as *puhkus* ‘vacation’, most likely based on the Finnish word *loma* ‘vacation’. Furthermore, as reported in earlier studies (see Kaivapalu 2015: 64), especially the Finnish participants use their knowledge of Finnish dialects or other languages to reach their translations. For instance, *tüdruk* ‘girl’ is translated as *juoma* ‘drink’, most likely based on the similarity the participants find between *tüdruk* and the Swedish word *dryck* ‘drink’, and as *rekka* ‘truck’ based on the similarity they find between *tüdruk* and the English word *truck*. Even though the translations are incorrect, they demonstrate that while trying to reach their translations, the test takers have used their knowledge of languages that are actual loan giving languages to Estonian. The translations also exhibit knowledge on the similarities in the morphology, and differences like the loss of word final vowels in Estonian (see the following section).

As we did not ask the participants to reflect on how they reached their translations, we cannot know the source of the acquired knowledge or how the participants use it. However, findings discussed above are in accordance with the findings of earlier studies. According to a study by Härmävaara (forthcoming), Finns and Estonians have a fairly large amount of knowledge and assumptions related to the cognate language that they also use strategically while trying to reach mutual understanding in interaction (see also Härmävaara & Frick 2016).

3.2.2. Inherent intelligibility

We were interested to know whether the asymmetry would still be present if the participants had not learned the language before and had only

⁸ According to Björklöf (2012: 55–56) *juhlad* is an established loanword in coastal dialects of Estonian.

minimal exposure to the test language. We therefore repeated our analysis with a selection of young participants who had not learned the cognate language and who had only little exposure to it.

The results of the analysis are presented between brackets in Table 3. A two-way ANOVA shows a weak interaction between language background of participants and test mode ($F = 4,253, p = .040$). The difference between test modes is not significant ($F = .433, p = .511$). Estonian participants translate more words correctly than Finnish participants do ($F = 16.852, p = .000$). Post-hoc tests show that the difference is significant for the spoken test ($t = -4,683, p = .000$) but not for the written test ($t = -1,351, p = .179$).

Since none of the participants from the two countries had learned the test language and the two groups had the same mean amount of exposure to the test language, non-linguistic factors can only play a minor role in explaining the asymmetry of spoken word comprehension for this selection of participants. We will therefore take linguistic factors into consideration as well, and have a closer look at the intelligibility results for individual spoken words. Since the asymmetry is only found for the spoken words, these words are likely to have linguistic characteristics that play a more important role for spoken word intelligibility than for written word intelligibility.

A number of words show no asymmetry. There is one word, the word for 'hand' which is *käsi* in both languages that was translated correctly by all participants. Also, there were 14 words that were unintelligible to both groups of participants and therefore did not show asymmetric intelligibility. Eleven of these words are non-cognates in both languages, such as Estonian *huvi*, Finnish *kiinnostus* 'interest'. Others include words such as 'period' that is a compound word in both languages, with a cognate as a first part and a non-cognate as a second part: *ajanjakso* in Finnish, *ajavahe* in Estonian.

The rest of the words show some asymmetry. We present the 51 words that more Estonians than Finns translated correctly in Figure 3 and the 34 words that more Finns than Estonians translated correctly in Figure 4.

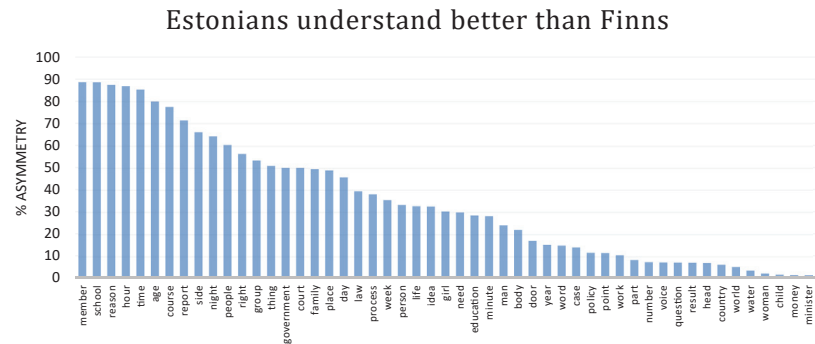


FIGURE 3. Spoken words that were better understood by the Estonian participants than the Finnish participants. The vertical axis plots the difference in percentage of correct translations between Estonian and Finnish spoken stimulus words

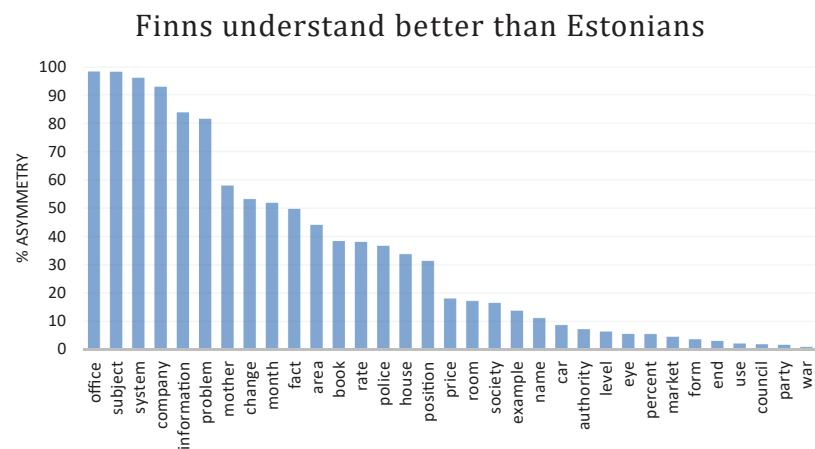


FIGURE 4. Spoken words that were better understood by the Finnish participants than the Estonian participants. The vertical axis plots the difference in percentage of correct translations between Estonian and Finnish spoken stimulus words

To get an idea why the Estonian participants translate some words better than the Finnish participants and vice versa, we will have a closer look at the 16 words that show an asymmetry of more than 30% in favor of the Finnish participants (Table 5) and the 26 words that show an asymmetry of more than 30% in favor of the Estonian participants (Table 6). We will look at a number of factors that potentially explain asymmetric word intelligibility. The tables show the test word, the corresponding cognate (if any) in the language of the participants, the test word in the other language and the phonetic transcription of the cognate pairs.

3.2.2.1. Loanwords

A large number of the words the Finnish participants understand better are recent loanwords in Estonian. When comparing Table 5 to Table 6 we notice that the list of words that the Finnish participants more often translate correctly, contains a larger proportion of loanwords (50% are loanwords) than the list of words that the Estonian participants translate correctly more often (15%). All recent loanwords in Table 5 that the Finnish participants hear when listening to Estonian have a corresponding cognate in Finnish, but most of the corresponding Finnish words do not have an Estonian cognate. For example, the Estonian word for 'system' is *süsteem*, and in Finnish there is a loanword *systeemi* that is synonymous to the test word *järjestelmä*. The Finnish test word has no cognate in Estonian. Six out of the 16 words (37.5%) listed in Table 5 have a cognate only from Estonian to Finnish but not from Finnish to Estonian.

When looking at the whole list of 100 Finnish test words (Appendix 1), we see that the Estonian list of words presented to the Finnish participants contains more loanwords (20) than the Finnish word list presented to the Estonian participants (12). Loanwords can be easy to understand not only because of the possible synonyms in the participants' native language but also because the participants can use their knowledge of the loan-giving language to understand the test word. The

TABLE 5. Words with a higher percentage of correct translations by Finnish participants than by Estonian participants. The words in parentheses indicate the (near) synonyms that give clues to the Finnish participants for translating the word, even if there is no exact cognate

Estonian test word	IPA	Finnish cognate	IPA	Finnish test word	English
kontor	kont:or	konttori	kont:tori	toimisto	office
teema	te:ma	teema	te:ma	aihe	subject
süsteem	syste:m	systeemi	syste:mi	järjestelmä	system
firma	firma	firma	firma	yritys	company
informat-sioon	informat'siio:n	informaatio	informa:tio	tieto	informa-tion
probleem	proble:m	(probleema)	proble:ma	ongelma	problem
ema	ema	(emä)	emæ	äiti	mother
muutus	mu:t:us	muutos	mu:tos	muutos	change
kuu	ku:	kuukausi	ku:kauwsi	kuukausi	month
fakt	fakt	fakta	fakta	tosiasia	fact
ala	ala	alue / (ala)	alue / ala	alue	area
raamat	ra:mat			kirja	book
määr	mæ:r	määrä	mæ:ræ	määrä	rate
politsei	politsei	poliisi	poli:si	poliisi	police
maja	maja			talo	house
asend	asend	asento	asento	asento	position

mean percentages of the correct translations of loanwords is 86.8 for the Estonian participants and 83.1 for the Finnish participants. The fact that there are more loanwords in the Estonian stimulus word set than in the Finnish stimulus word set may explain part of the overall asymmetry of the results in Table 5.

TABLE 6. Words with a higher percentage of correct translations by Estonian participants than by Finnish participants. The words in parentheses indicate the near or partial synonyms and homonyms that give clues to the Estonian participants for translating the word, even if there is no exact cognate

Finnish test word	IPA	Estonian cognate	IPA	Estonian test word	English
jäsen	jæsen	(jäse)	jæse	liige	member
koulu	kowlu	kool	ko:l	kool	school
syy	sy:	(süü)	sy:	põhjus	reason
tunti	tunti	tund	tunːd	tund	hour
aika	ɑjka	aeg	æɡ	aeg	time
ikä	ikæ	iga	iɡɑ	vanus	age
kurssi	kurs:i	kursus	kursus	kursus	course
raportti	raport:i	raport	rap-ort:	aruanne	report
puoli	puoli	pool	po:l	pool	side
yö	yø	öö	ø:	öö	night
ihmiset	ihmiset	inimesed	inimesed	inimesed	people
oikeus	ojkeus	õigus	ɔiɡus	õigus	right
ryhmä	ryhmæ	rühm	ryhm	rühm	group
asia	asia	asi	ɑsi	asi	thing
hallitus	hal:itus			valitsus	government
oikeus	ojkeus	(õigus)	ɔiɡus	kohus	court
perhe	perhe	pere	pere	pere	family
paikka	pɑjk:ɑ	paik	pɑjk	koht	place
päivä	pæjvæ	päev	pæɛv	päev	day
laki	laki	(lagi)	laɡi	seadus	law
prosessi	proses:i	protsess	protses:	protsess	process
viikko				nädal	week
henkilö	henkilø	(hing)	hiŋɡ	isik	person
elämä	elæmæ	elu	elu	elu	life
idea	idea	idee	iɔe:	idee	idea
tyttö	tyt:ø	tüdruk	tydruk	tüdruk	girl

3.2.2.2. Cognates

Similarly to the recent loanwords, some of the older test words have an asymmetric relationship, when it comes to cognancy. The effect of that is shown in Table 6, which lists the words that the Estonian participants translate better than the Finnish participants. For instance, the word for ‘member’ is more often translated correctly by the Estonians than the Finns. The Estonian test word is *liige* that does not have a Finnish cognate, but a false friend *liike* ‘movement; shop’ that the Finnish participants quite often give as a translation. The Finnish word *jäsen*, however, has a cognate *jäse* in Estonian that covers part of the semantic field of the Finnish word: Estonian *jäse* is ‘member’ in the sense of a body part, and Finnish *jäsen* can mean both a person and a body part. Seven out of the 26 words (26.9%) listed in Table 6 have a cognate only from Finnish to Estonian.

Overall, Estonian participants are good at recognizing similarity between cognates. Seventeen of the words listed in Table 6 (65.4%) have a near identical cognate in both directions. This suggests that the Estonian participants are better at perceiving the actual similarity between the cognates than the Finnish participants. We will next discuss the possible explanations for that.

3.2.2.3. Number of syllables

In 10 of the 17 cognate word pairs that the Estonian participants translate better than the Finnish participants (Table 6), the Finnish word has one more syllable than the corresponding Estonian word. Eight of the ten pairs are otherwise near identical, but the Estonian word does not show the stem vowel in the nominative form. For example the word for ‘group’ has two syllables in Finnish, *ryhmä*, and only one syllable in Estonian, *rühm*⁹. Previously it has been found that a missing syllable could

⁹ The stem vowel of *rühm* is *-a*, which makes the genitive and partitive form *rühma*, and the forms based on the genitive stem, more similar to the Finnish forms.

cause confusion in word identification (Kürschner et al. 2009; Kaivapalu & Martin 2014).

In general, the loss of a word-final vowel in Estonian is a feature well known to both Finns and Estonians. Therefore the participants most likely have a schema to work with for adding and deleting word final vowels while trying to understand the cognate language (see Härmävaara 2013). However, the Estonian participants are better at translating these words, which suggests that there is not enough information to be processed for the Finnish speakers to find a correct cognate. For instance, *koht* ‘place’ is often translated as *kohtu* ‘uterus’, instead of the correct cognate *kohta* ‘place; point’. When translating *koht* as *kohtu*, the Finnish participants add a word final vowel, but it just happens to be the wrong one. In fact, the Finnish word *kohta* ‘point, place’ includes the same stem vowel as the Estonian word *koht*.

As it is easier to remove a vowel than to guess what to add, the loss of a word-final vowels seems to be a linguistic factor that makes the cognates less transparent to the Finnish participants than to the Estonian participants (see also Kaivapalu & Martin 2014). The biggest asymmetry in comprehension is in the test words meaning ‘school’, ‘time’ and ‘hour’, which are short words, only one syllable in Estonian, *kool*, *aeg* and *tund*. In the longer word pairs (e.g. test word ‘world’ *maailma–maailm*, see Figure 3) the asymmetry is rather small; the longer words have more linguistic material to process. The Finnish participants translate these ten words that are shorter in Estonian but otherwise fairly similar cognates, more often correctly in the written data. This suggests that seeing the word makes it easier to find the correct cognate.

3.2.2.4. Morphology

Syllable structure seems to be a strong cue to the Finnish participants when finding a Finnish equivalent to an Estonian word, even when it is misleading. Particularly interesting are the Finnish translations of the Estonian stimulus words that end in *s* or *l*. Among the 26 words that

show more than 30% of asymmetry in the direction of Estonians understanding better, there are five Estonian stimulus words that end in *s* or *l*: *kursus* ‘course’, *kohus* ‘court’, *seadus* ‘law’, *protsess* ‘process’ and *nädal* ‘week’. *Protsess* is recognized by the Finnish participants as a recent loanword and is translated either as *prosessi* ‘process’ or *protesti* ‘protest’. All the other words get translations that are similar by word structure but not necessarily otherwise. For instance, Estonian *kursus* ‘course’ has a close cognate, *kurssi*, in Finnish, but the Finnish participants have given translations such as *porsas* ‘piglet’ or *puolustus* ‘defence’. Similarly, stimulus word *seadus* ‘law’ is 16 times translated as *ajatus* ‘thought’.

The participants interpret the word final *-s* not only as something that belongs to a certain word type, but also as a marker of the inessive case, and sometimes past tense¹⁰. For instance, the participants translate *kursus* ‘course’ as *korsussa* ‘in a dugout’ and *kirkossa* ‘in a church’, and *kohus* ‘court’ is translated as *kohdussa* ‘in uterus’, *kohdassa* and *paikassa* ‘on a spot, at a place’. The Finnish inessive case *-ssa* is often in spoken Finnish represented with *-s*, thus bringing the Finnish and Estonian morphology closer to each other. The Finnish participants also make a link between the word final *-l* and the adessive case *-lla*, often represented in spoken Finnish as *-l*. That can be seen in the translations to the stimulus word *nädal* ‘week’: *lähellä* ‘near’ and *täällä* ‘here’.

The Estonian participants, too, have been guided by the word final elements in their translations (see also Kaivapalu & Martin 2014). For instance, the Finnish nouns that end in *-ma* that resembles the Estonian infinitive ending *-ma*, like *elämä* ‘life’ and *järjestelmä* ‘system’, are interpreted as verbs in translations like *elama* ‘to live’ and *järjestama* ‘to rank’. Sometimes the link between the word final element and the grammatical form is not as clear. For instance, the Estonian participants have

¹⁰ The inessive ending is *-ssa* in Finnish and *-s* in Estonian. The most frequent marker of imperfect tense is in Finnish *-i*, and *-si* is used with certain word types. The other way around, the most frequent marker of imperfect tense in Estonian is *-s*, and *-i* is used only with certain verb types.

interpreted the Finnish word *tulos* ‘result’ as a form of a verb ‘to come’, *tulla* in both languages. Historically the connection is accurate. The word final *-s* is then analyzed as an inflectional ending, leading to translations like *tuli* ‘came’, *tuleb* ‘comes’, *tulema* ‘to come’ and *tulemas* ‘coming’.

All these translations indicate that the participants assume or know that there is similarity between the languages on the morphological level. That is in accordance with the study by Kaivapalu & Muikku-Werner (2010), according to which the Finnish participants recognized morpho-syntactical similarities well and used them strategically while trying to comprehend the cognate language.

3.2.2.5. Sound systems

Differences in the Finnish and Estonian sound systems may also impact recognition of the cognates. There are two cognate pairs that provoke interesting differences in translations between the written and spoken data sets: ‘hour’, *tund* in Estonian, *tunti* in Finnish, and ‘right’, *õigus* in Estonian, *oikeus* in Finnish. Both Estonian stimulus words contain sounds that do not exist in standard Finnish. Two cognate pairs is too few to make larger conclusions on the link between comprehension and differences in the sound systems, but the Finnish participants translate both words better in the written data set.

The spoken stimulus word *tund* ‘hour’ with a close cognate *tunti* provokes a type of translations that written stimulus does not. In the spoken data set 30% of the translations include a diphthong, for example *tuima* ‘grim’, *toimi* ‘deed’, *toimia* ‘to function’, *tuijottaa* ‘to stare’. Perceiving a diphthong is most likely caused by the quality of the nasal *n*. In this phonetic context, *n* is palatalized in Estonian, but standard Finnish lacks the feature of palatalization. Hence, the quality of the nasal is interpreted as a quality of the vowel, giving cues to find a cognate with a diphthong¹¹.

¹¹ The palatalization in Estonian is often realized as a “prepalatalization”, so that acoustically, something like an *i* can be perceived at the end of the preceding vowel nucleus (see Asu et al. 2016: 91).

Another feature of the Estonian sound system foreign to the Finnish sound system is the vowel marked with \ddot{o} [ɤ]. The Finnish participants do not often make the connection between the close cognates *oikeus* (Finnish) and *õigus* (Estonian) for ‘right’. Based on the written data set, making a link between these historically connected *o* and \ddot{o} is not difficult, but in the spoken data set there is a considerable variation in the way the word initial diphthong is interpreted, for example *uikutus* ‘whining’, *aikuinen* ‘adult’, *fikus* ‘rubber fig’, *ajatus* ‘thought’ and *yöuni* ‘sleep’. However, making the link between *o* and \ddot{o} may depend on the phonetic context. As a translation to *põhjus* ‘reason’ most Finnish participants give something related to *pohja* ‘base, bottom’ or *pohjoinen* ‘north’ – which are historically linked to the Estonian word *põhjus*. As a component of a diphthong \ddot{o} might be more difficult to parse.

Our data indicates that differences in the writing systems do not impact the translations as much as the differences in the sound systems. For instance, the plosives with similar sound quality, marked in Finnish with *k*, *p* and *t* and in Estonian with *g*, *b* and *d*, are often linked to each other without problems, as well as Finnish *y* and Estonian *ü*, marking the same vowel [y].

4. Conclusions and discussion

This study aimed at measuring mutual intelligibility of Finnish and Estonian by analyzing responses to a translation task of isolated words in the cognate language. The test consisted of 100 stimulus words in spoken or written Finnish and Estonian. 69 Finnish words and 70 Estonian words on the list had a corresponding cognate in the related language. This means that based only on the similarity of the languages, the participants should have been able to translate a maximum of about 70% of the words correctly.

Finnish and Estonian speakers do not have very high expectations about their ability to understand the cognate language. Estonian participants estimated their understanding a bit higher than the Finnish

participants, and this asymmetry in understanding shows in the translation data as well. The results show that the Estonian speakers understand isolated Finnish words better than the Finnish speakers understand Estonian words, which is in accordance with the results of previous research (see Kaivapalu 2015).

This was the first time a study on mutual intelligibility of Finnish and Estonian produced data directly comparable to the data on mutual intelligibility of other closely related language pairs. The comparison allowed us to put the results into perspective. Looking at the data drawn from young participants with varying degrees of previous knowledge, the Estonian participants scored well compared to the speakers of 70 combinations of closely related European languages, whereas the scores of the Finnish participants were low. If we compare to the Romance languages, Estonian participants' written results are closest to Italians tested in Spanish, and Finnish participants' spoken results are most similar to the results of Romanian participants tested in Italian.

Previous research on mutual intelligibility of Finnish and Estonian has consulted data drawn from young university students, collected in classrooms. Conducting this data collection online allowed a wider participant pool, and the test results were analyzed separately between the groups of older (34 to 84 years) and younger (18 to 33) participants. The older participant groups reported to have had more exposure to the cognate language than the younger participant groups, and they also scored higher in the functional intelligibility test. This may reflect the decreased importance of Finnish in Estonia and the increased importance of English in communication between Finns and Estonians (see also Härmävaara 2017).

The exposure to the test language turned out to be the clearest explaining factor in the overall asymmetry in understanding. The Estonian participants had on average been exposed to the test language more often than the Finnish participants, and the average number of years they had learned the language was higher. The link between exposure and intelligibility was clear. For instance, the average test score for the spoken

stimulus word test taken by the older participant group was 74.6%. That is higher than the percentage of cognates among the stimuli (71%), and in general Estonians were significantly better at translating non-cognates that are in principle only understandable to participants with previous exposure to the language. Furthermore, commonly known false friends were fairly often translated correctly by both Estonian and Finnish participants, which exhibits at least anecdotal level knowledge of the cognate language.

Narrowing the focus down to participants that did not have significant exposure to the cognate language revealed that there was statistically significant asymmetry in intelligibility only in the spoken data set. The overall mean of correct translations was higher for the Estonian participants (51.9% versus 42.4%) and the number of words that they translated better than the Finnish participants was also higher (51 versus 34 words). By analyzing the responses to the words that showed most asymmetry in understanding, we found some linguistic factors that may explain the asymmetry.

The greater number of international words and recent loanwords in the Estonian stimulus data set worked in favor of the Finnish participants, who could use their knowledge on multiple languages while translating the words. In contrast, for the Estonian participants, there were more words in the set of stimulus words that they could understand by means of an alternative cognate in their own language. Furthermore, the Estonian participants were better at recognizing Finnish cognates.

Linguistic features that seemed to make the cognates less transparent to the Finnish participants were 1) the loss of a word-final vowel in Estonian, 2) word structure that can be interpreted as morpheme + inflectional morpheme, and 3) Estonian sounds that do not exist in Finnish. Kürschner et al. (2009) got similar results when they studied how well Danish speakers understand Swedish words. They found that phonetic distance between the cognates affected the intelligibility the most. In addition, word length, different syllable numbers, foreign sounds and foreign prosodic phenomena contributed significantly to intelligibility.

Clearly some factors pertain to only a limited number of words, and also the combination of the factors plays a role. To draw stronger conclusions we would have to analyze a larger set of word pairs. The kinds of mistakes in translating discussed here may give some indications of how we may proceed to gain a better understanding of the Estonian-Finnish mutual intelligibility. The mistakes also tell something about the participants' thinking and translation process. This translation task did not ask the participants to reflect on their decisions, but the incorrect translations revealed similar phenomena as the studies that have analyzed the introspection of the participants. For instance, while translating, the participants used their knowledge of languages other than Finnish and Estonian, and they were prone to look for the hints for similarity at morphological level as well.

All in all our results show that Finnish and Estonian are mutually intelligible to a certain degree. Without significant previous exposure to the cognate language, a Finnish or an Estonian speaker can assume to understand about a half of words that s/he encounters without a context. This suggests that the lexical gap between Finnish and Estonian is significant. However, as noted in the earlier studies (e.g. Kaivapalu & Muikku-Werner 2010; Kaivapalu 2015), context helps in inferring the details that might otherwise remained blurred, and functions as a filter, either confirming or disproving the assumptions made. Therefore, even if the probability of understanding a random isolated word is low, depending on the topic and the general and immediate context, there is a good chance of there being enough similar and recognizable elements to facilitate understanding.

Acknowledgements

Charlotte Gooskens is responsible for the quantitative analysis of the data and is the primary author of Sections 2 and 3.1. Hanna-Ilona Härmävaara functioned as an expert of Finnish and Estonian languages. She is responsible for the qualitative analysis of the data and is the primary author of Sections 1 and 3.2.

We thank Jelena Golubovic for helping in the data collection and for being the technical expert on the implementation of the Micrela method. We thank Annekatriin Kaivapalu for her help in translating and recording the Estonian stimulus words, Aleksandar Mancic for programming the web application and Sophia Sansone for proofreading the text. We also thank the two anonymous reviewers for their comments. This work was supported by a grant from the Netherlands Organisation for Scientific Research (NWO) 360-70-430 and by a grant from Emil Aaltonen Foundation.

References

- Alvre, Paul, Raul Vodja 1993. *Pulma poikineen: virolais-suomalainen vertailusanakirja* [‘Comparative Finnish-Estonian dictionary’]. Porvoo: Wsoy.
- Asu, Eva Liina, Pärtel Lippus, Karl Pajusalu, Pire Teras 2016. *Eesti keele hääldus* [‘The pronunciation of the Estonian language’]. *Eesti keele varamu* 2. Tartu: Tartu Ülikooli Kirjastus.
- Bahtina, Daria, Jan D. ten Thije 2012. Receptive multilingualism. – Carol A. Chapelle (Ed.). *The Encyclopedia of Applied Linguistics*. John Wiley and Sons. <https://doi.org/10.1002/9781405198431.wbeal1001>
- Berthele, Raphael 2008. Dialekt-Standard Situationen als embryonale Mehrsprachigkeit. *Erkenntnis zum interlingualen Potenzial des Provinzlerdaseins*. – *Sociolinguistica* 22 (1), 87–107. <https://doi.org/10.1515/9783484605299.87>
- Björklöf, Sofia 2012. *Viron rantamurteen länsiryhmän sanaston alkuperä suomalaislainojen valossa* [‘The origins of the vocabulary in Eastern Coastal dialect in the light of Finnish loanwords’]. Master’s thesis. University of Helsinki. <http://urn.fi/URN:NBN:fi-fe2012121410304>
- BNC Consortium 2007. *BNC: The British National Corpus, Version 3, BNC XML Edition* Oxford University Computing Services on behalf of the BNC Consortium distributed by <http://www.natcorp.ox.ac.uk/>.
- Erelt, Mati, Tiiu Erelt, Kristiina Ross 2000. *Eesti keele käsiraamat* [‘Handbook of Estonian’]. Tallinn: Eesti Keele Sihtasutus.
- European Commission 2012. *Special Eurobarometer 386. Europeans and their Languages*.
- Gooskens, Charlotte, Wilbert Heeringa 2014. The role of dialect exposure in receptive multilingualism. – *Applied Linguistics Review* 5 (1), 247–251. <https://doi.org/10.1515/applirev-2014-0011>

- 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 J. 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>
- Grünthal, Riho 2009. Suomen kielen vaikutus viron kieleen [‘Finnish influence on Estonian language’]. – Jyrki Kalliokoski, Lari Kotilainen, Päivi Pahta (Eds.). *Kielet kohtaavat*. Tietolipas 227. Helsinki: Suomalaisen Kirjallisuuden Seura, 231–263.
- Heinonen, Maria 2015. “Koska virossa kaikki on nurinkurin”. Käännöstesti eräiden luettujen ja kuultujen vironkielisten sanojen ymmärtämisestä suomen kielen pohjalta [‘“Because everything in Estonian is backwards” – A translation test of understanding certain read and listened Estonian words on the basis of Finnish language’]. Pro gradu -tutkielma. Joensuu: Itä-Suomen yliopisto. <http://urn.fi/urn:nbn:fi:uef-20150086>
- Häkkinen, Kaisa 1990. Mistä sanat tulevat? Suomalaista etymologiaa [‘Where do words come from: Finnish etymologies’]. Tietolipas 117. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Härmävaara, Hanna-Ilona 2013. Kielten samankaltaisuus monikielisen suomalais-virolaisen vuorovaikutuksen resurssina [‘Cross-linguistic similarities as a resource of multilingual interaction between Finns and Estonians’]. – *Lähivördlusi. Lähivertailuja* 23, 60–88. <https://doi.org/10.5128/LV23.03>
- Härmävaara, Hanna-Ilona 2014. Facilitating mutual understanding in everyday interaction between Finns and Estonians. – *Applied Linguistics Review* 5 (1), 211–245. <https://doi.org/10.1515/applirev-2014-0010>
- Härmävaara, Hanna-Ilona 2017. Official language policy as a factor for using receptive multilingualism among members of an Estonian and a Finnish student organization. – Maarja Siiner, Kara D. Brown, Kadri Koreinik (Eds.). *Language Policy Beyond the State*. Springer, 201–221. https://doi.org/10.1007/978-3-319-52993-6_11
- Härmävaara, Hanna-Ilona (forthcoming). Building the bridge in interaction – receptive multilingualism among Finns and Estonians. PhD thesis. University of Helsinki.
- Härmävaara, Hanna-Ilona, Maria Frick 2016. Handling linguistic asymmetries by bilingual punning in conversations among speakers of cognate

- languages. – Sebastian Knospe, Alexander Onysko, Maik Goth (Eds.). *Crossing Languages to Play with Words: Interdisciplinary Perspectives*. Berlin: DeGruyter, 113–134. <https://doi.org/10.1515/9783110465600-008>
- Kaivapalu, Annekatrin 2015. Eesti ja soome keele vastastikune mõistmine üksik-sõna- ja tekstitasandil: lingvistilised tegurid, mõistmisprotsess ja sümmeetria [‘Mutual intelligibility of Estonian and Finnish context-free words and texts: linguistic determinants, comprehension process and symmetry’]. – *Eesti Rakenduslingvistika Ühingu aastaraamat 11*, 55–74. <https://doi.org/10.5128/ERYa11.04>
- Kaivapalu, Annekatrin, Maisa Martin 2014. Symmetry of similarity: Definition, perception, measurement: Finnish and Estonian noun morphology as a testing ground. – Heli Paulasto, Lea Meriläinen, Helka Riionheimo, Maria Kok (Eds.). *Language Contacts at the Crossroads of Disciplines*. Cambridge: Cambridge Scholars, 283–318.
- Kaivapalu, Annekatrin, Pirkko Muikku-Werner 2010. Reseptiivinen monikielisyys: miten suomenkielinen vironoppiija ymmärtää viroa äidinkielenä pohjalta [‘Receptive multilingualism: how Finnish as a first language helps learners to understand Estonian?’]. – *Lähvõrdlusi. Lähivertailuja* 20, 68–97. <https://doi.org/10.5128/LV20.03>
- Koreinik, Kadri, Kristiina Praakli 2017. Emerging language political agency among Estonian native speakers in Finland. – Maarja Siiner, Kara D. Brown, Kadri Koreinik (Eds.). *Language Policy Beyond the State*. Springer, 81–98. https://doi.org/10.1007/978-3-319-52993-6_11
- Kuusk, Margit 1999. *Suomi selväksi. Soome keele õpik* [‘Learning Finnish: textbook’]. Tallinn: Eesti Keele Sihtasutus.
- Kürschner, Sebastian, Charlotte Gooskens, Renée van Bezooijen 2009. Linguistic determinants of the intelligibility of Swedish words among Danes. – *Computing and Language Variation: International Journal of Humanities and Arts Computing* 2 (1–2), 83–100. <https://doi.org/10.1017/UPO9780748641642.006>
- Laakso, Johanna 2001. The Finnic Languages. – Östen Dahl, Maria Koptjevskaja-Tamm (Eds.). *Circum-Baltic Languages, Volume 1: Past and Present*. Studies in Language Companion Series 54. Amsterdam/Philadelphia: John Benjamins, 179–212. <https://doi.org/10.1075/slcs.54.09laa>
- Laalo, Klaus 1992. Huvitav lugu – kiinnostava juttu: suomen ja viro välisen sanaston riskiryhmä ja sen taustaa [‘Risk words between Finnish and Estonian and their background’]. *Suomi 164*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Mark, Marta 2014. Eesti õpilaste soome üksiksõnade mõistmine ['Understanding Finnish words by the Estonian students']. Bachelor's thesis. Tallinn: University of Tallinn.
- Muikku-Werner, Pirkko 2013. Vironkielisen tekstin ymmärtäminen suomen kielen pohjalta ['Understanding Estonian texts on a Finnish language base']. – Lähivõrdlusi. Lähivertailuja 23, 210–237. <https://doi.org/10.5128/LV23.09>
- 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']. – Lähivõrdlusi. Lähivertailuja 22, 157–187. <https://doi.org/10.5128/LV22.06>
- 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>
- Pajusalu, Renate, Merja Hietaharju, Viive Taro, Kai Yallop 1999. Keelesild. Viron kielen oppikirja ['A bridge between Estonian and Finnish: Textbook']. Helsinki: Otava.
- Remes, Hannu 2009. Muodot kontrastissa. Suomen ja viron vertailevaa taivutusmorfologiaa ['Forms in contrast: Contrastive inflection in Finnish and Estonian']. Oulu: Oulun yliopisto.
- Ringbom, Håkan 2007. Cross-linguistic Similarity in Foreign Language Learning. Clevedon: Multilingual Matters LTD.
- Verschik, Anna 2012. Practicing Receptive Multilingualism: Estonian-Finnish communication in Tallinn. – International Journal of Bilingualism 16 (3), 265–286. <https://doi.org/10.1177/1367006911426465>
- Wirén, Ulvi 2008a. Hääd pulmapäeva ja yli 1000 muuta viron kielen riskisanaa ['1000 Risk words between Finnish and Estonian']. Helsinki: Gummerus.
- Wirén, Ulvi 2008b. Huvittavat raamatut ja lähes 2000 muuta viron ja suomen riskisanaa ['2000 Risk words between Finnish and Estonian']. Helsinki: Gummerus.

Appendix 1

The Estonian and Finnish test words used for the translation task, the corresponding cognates (if any) in the language of the participants and the English translations. If the cognate is in brackets, it is either a partial cognate to the test word, or it is a cognate to a word homonym to the test word. All words in the brackets are also counted as correct translations. The bolded test words are the words we treated as “recent loanwords and international words”.

	Estonian test word	Finnish cognate	Finnish test word	Estonian cognate	English translation
1	aeg	aika	aika	aeg	time
2	aasta		vuosi		year
3	inimesed	ihmiset	ihmiset	inimesed	people
4	mees	mies	mies	mees	man
5	päev	päivä	päivä	päev	day
6	asi	asia	asia	asi	thing
7	laps	lapsi	lapsi	laps	child
8	valitsus		hallitus		government
9	osa	osa	osa	osa	part
10	elu	elämä	elämä	elu	life
11	juhtum		tapaus		case
12	naine	nainen	nainen	naine	woman
13	töö	työ	työ	töö	work
14	süsteem	systemi	järjestelmä		system
15	rühm	ryhmä	ryhmä	rühm	group
16	number	numero	numero	number	number
17	maailm	maailma	maailma	maailm	world
18	ala	alue	alue	ala	area
19	kursus	kurssi	kurssi	kursus	course
20	firma	firma	yritys		company

	Estonian test word	Finnish cognate	Finnish test word	Estonian cognate	English translation
21	probleem	probleema (colloq.)	ongelma		problem
22	teenistus		palvelus		service
23	käsi	käsi	käsi	käsi	hand
24	pidu	pidot	juhla		party
25	kool	koulu	koulu	kool	school
26	koht	kohta	paikka	paik	place
27	punkt		kohta	koht	point
28	maja		talo		house
29	maa	maa	maa	maa	country
30	nädal		viikko		week
31	liige		jäsen	(jäse)	member
32	lõpp	loppu	loppu	lõpp	end
33	sõna	sana	sana	sõna	word
34	näide	(näyte)	esimerkki		example
35	pere	perhe	perhe	pere	family
36	fakt	fakta	tosiasia	tõsiasi	fact
37	protsent	prosentti	prosentti	protsent	percent
38	kuu	kuukausi	kuukausi	kuu	month
39	pool	puoli	puoli	pool	side
40	öö	yö	yö	öö	night
41	silm	silmä	silmä	silm	eye
42	pea	pää	pää	pea	head
43	informat-sioon	informaatio	tieto	teave	information
44	küsimus	kysymys	kysymys	küsimus	question
45	võim	(voima)	valta		power
46	raha	raha	raha	raha	money
47	muutus	muutos	muutos	muutus	change

MUTUAL INTELLIGIBILITY OF FINNISH AND ESTONIAN VOCABULARY

	Estonian test word	Finnish cognate	Finnish test word	Estonian cognate	English translation
48	huvi		kiinnostus		interest
49	tellimus		tilaus		order
50	raamat		kirja		book
51	areng		kehitys		development
52	tuba	(tupa)	huone		room
53	vesi	vesi	vesi	vesi	water
54	ankeet		lomake		form
55	auto	auto	auto	auto	car
56	tase	taso	taso	tase	level
57	poliitika	politiikka	politiikka	poliitika	policy
58	nõukogu	neuvosto	neuvosto	nõukogu	council
59	joon		viiva		line
60	vajadus		tarve	tarve	need
61	mõju		vaikutus		effect
62	kasutus		käyttö		use
63	idee	idea	idea	idee	idea
64	uurimus		tutkimus		study
65	tüdruk	tyttö	tyttö	tüdruk	girl
66	nimi	nimi	nimi	nimi	name
67	tulemus	tulos	tulos	tulemus	result
68	keha	keho	keho	keha	body
69	sõber		ystävä		friend
70	õigus	oikeus	oikeus	õigus	right
71	autoriteet	auktoriteetti	auktoriteetti	autoriteet	authority
72	vaade		näkymä	(nägemus)	view
73	aruanne		raportti	raport	report
74	nägu		kasvot		face
75	turg	tori	tori	turg	market
76	tund	tunti	tunti	tund	hour

	Estonian test word	Finnish cognate	Finnish test word	Estonian cognate	English translation
77	määrä	määr	määrä	määr	rate
78	seadus	(säädos)	laki	(lagi)	law
79	uks	uksi	ovi		door
80	kohus		oikeus	(õigus)	court
81	kontor	konttori	toimisto		office
82	sõda	sota	sota	sõda	war
83	põhjus		syy	(süü)	reason
84	minister	ministeri	ministeri	minister	minister
85	teema	teema	aihe		subject
86	isik		henkilö	(hing)	person
87	ajavahemik	ajanjakso	ajanjakso	ajavahemik	period
88	ühiskond	yhteiskunta	yhteiskunta	ühiskond	society
89	protsess	prosessi	prosessi	protsess	process
90	ema	emä /emo	äiti		mother
91	hää	ääni	ääni	hää	voice
92	politsei	poliisi	poliisi	politsei	police
93	liik		laji		kind
94	hind	hint	hint	hind	price
95	asend	asento	asento	asend	position
96	vanus		ikä	iga	age
97	kuju		hahmo		figure
98	haridus		koulutus	koolitus	education
99	programm		ohjelma		program
100	minut	minuutti	minuutti	minut	minute

Suomen ja viron sanaston keskinäinen ymmärrettävyys

HANNA-ILONA HÄRMÄVAARA¹,

CHARLOTTE GOOSKENS²

Groningenin yliopisto², Washingtonin yliopisto¹

Tässä tutkimuksessa käsitellään suomen ja viron sanaston keskinäistä ymmärrettävyyttä. Tutkimus toteutettiin kääntämällä 100 yleisintä englannin substantiivivia (BNC Consortium 2007) kummallekin kielelle ja testaamalla Internet-sovelluksen avulla, miten hyvin äidinkieliset suomen- ja vironpuhujat kääntävät lukemansa tai kuulemansa sukukielen sanat omalle äidinkielelleen. Sanoista noin 70 prosentilla oli sukukielinen kognaatti. Testi perustui aiempaan, germaanisten, slaavilaisten ja romanisten kielten keskinäistä ymmärrettävyyttä kartoittavaan laajaan tutkimukseen (Micrela) ja käytti vertailtavuuden vuoksi samoja testisanoja ja aineistonkeruu- ja analyysimenetelmiä.

Tutkimukseen osallistui 307 suomenkielistä ja 118 vironkielistä vastaajaa. He olivat iältään 18–84-vuotiaita ja heidät jaettiin tarkempaa analyysia varten nuorten (18–33) ja vanhojen (34–84) vastaajien ryhmään. Nuorten osallistujien ryhmä vastasi iältään Micrela-tutkimuksen osallistujia. 28 prosenttia suomalaisista ja 51 prosenttia virolaisista vastaajista oli miehiä. Keräsimme osallistujilta taustatietoja siitä, miten paljon he olivat olleet sukukielen kanssa tekemisissä ja miten hyvin he uskoivat sitä ennen käännoestettä ymmärtävänsä.

Virolaiset osallistujat olivat olleet sukukielen kanssa kosketuksissa keskimäärin enemmän kuin suomalaiset osallistujat. He myös arvioivat suomen kielen ymmärryksensä paremmaksi kuin suomalaiset viron kielen ymmärryksensä, vaikkakaan kumpikaan ryhmä ei arvioinut ymmärtämistään kovin korkeaksi. Testin tulokset vahvistivat yhteyden kieleen kosketuksissa olemisen ja ymmärtämisen välillä: virolaiset ymmärsivät yksittäisiä sanoja paremmin kuin suomalaiset kummassakin ikäryhmässä ja niin puhuttuina kuin kirjoitettuina. He käänsivät oikein huomattavan määrän suomen kielen sanoja, joilla ei ole vironkielistä kognaattia. Sekä virolaisista että suomalaisista vastaajista vanhemmat osallistujat suoriutuivat testistä paremmin kuin nuoremmat.

Nuoremman ikäryhmän tuloksia verrattiin indoeurooppalaisten kieliparien puhujien sukukielen ymmärtämiseen, ja virolaiset osallistujat ymmärsivät

suomea paremmin kuin indoeurooppalaisen kielen puhuja sukukieliään keskimäärin. Suomalaiset vastaajat taas ymmärsivät keskimääräistä huonommin. Vironpuhujien tulokset olivat kuitenkin lähempänä suomenpuhujien tuloksia, kun tarkasteltiin vain niitä vastaajia, joilla oli ollut vähäisin mahdollinen kosketus sukukieleen.

Kun tutkimuksessa tarkasteltiin vastauksia vain niiltä osallistujilta, jotka eivät olleet muodollisesti opiskelleet sukukieltä ja joiden kosketus sukukieleen oli mahdollisimman vähäinen, vironkielisten vastaajien havaittiin ymmärtävän sukukieltä hieman paremmin kuin suomenkielisten. Ero oli tilastollisesti merkittävä vain puhutun kielen aineistossa. Vastauksia kvalitatiivisesti analysimalla selvitettiin, millaiset kielelliset tekijät saattaisivat selittää ymmärrettävyyden epäsymmetrisyyttä.

Epäsymmetrisyyttä aiheuttivat ensinnäkin testisanojen ja niiden sukukielisten vastineiden suhteet. Vironkielisten testisanojen joukossa oli enemmän uusia lainasanoja kuin suomenkielisten. Suomen omaperäisille sanoille taas löytyi useammin ainakin osittainen kognaatit viron kielestä. Virolaiset olivat myös parempia tunnistamaan kieltenvälistä samankaltaisuutta. Kognaattien tunnistamista hankaloitti selvimmin viron sanojen loppuheitto. Myös sanarakente, joka ohjasi tulkitsemaan sanan olevan jossakin taivutusmuodossa hankaloitti kognaatit hahmottamista. Myös viron yleissuomesta puuttuvien äännepiirteiden, *õ:n* ja palatalisaation, esiintyminen tietyissä äänneympäristöissä vaikutti hankaloittavan tunnistamista.

Avainsanat: kielten keskinäinen ymmärrettävyys; epäsymmetrinen ymmärrettävyys; suomi; viro; kognaatit

Charlotte Gooskens

University of Groningen, Department of Applied Linguistics
P.O. Box 716
NL-9700 AS Groningen, The Netherlands
c.s.gooskens@rug.nl

Hanna-Ilona Härmävaara

University of Washington, Department of Scandinavian Studies
318 Raitt hall
Box 353420
Seattle, WA 98195-3420, USA
hiharma@uw.edu