

ORIGINAL RESEARCH

The Measurement of Mutual Intelligibility between West-Slavic Languages

Lukáš Kyjánek* and Jiří Haviger**

*Faculty of Science, University of Hradec Kralove, Czech Republic

**Faculty of Informatics and Management, University of Hradec Kralove, Czech Republic

ARTICLE HISTORY

Compiled June 6, 2019

Abstract

The West Slavic languages are close enough for their speakers to use their mother tongues when communicating with each other. Haugen (1966) referred to this type of communication as ‘semicomunication’. Research into the mutual intelligibility of the semicomunication of West Slavic languages has been done exclusively by sociolinguistic methods so far.

The present paper, therefore, uses the mathematical-linguistic methods proposed by Moberg, Gooskens, Nerbonne, and Vaillette (2007) to measure the mutual intelligibility of semicomunication. These methods are based on calculating the conditional entropy. The calculations are done on the phonetic and graphemic planes of Czech, Slovak, and Polish.

Conditional entropies were the lowest between Czech and Slovak, whereas between both of these languages and Polish they were much higher. This asymmetry was confirmed for every studied relationship. However, its extent was varied for each. There were also demonstrated the differences between content and function words, and various styles of material.

The measurements in this research proved, as had done Moberg et al. (2007) as well, that conditional entropy correlates with mutual intelligibility. The general conclusions mostly confirm the sociolinguistic conclusions. In many cases, however, these conclusions have been refined and supplemented.

KEYWORDS

Mutual intelligibility; Linguistic distance; West Slavic languages; Levenshtein distance; Conditional entropy

1. Introduction

In 1966, Haugen noticed that Scandinavians use their mother tongues (Danish, Norwegian and Swedish) to communicate with each other and they still understand each other well. He referred to this language phenomenon as *semicomunication*, described it with the example of these Scandinavian languages, and he also listed several other languages (including Czech, Slovak and Polish) between which semicomunication could be possible.

However, the term semicomunication was not fully accepted and was even slightly misinterpreted, for example by Budovičová (1987). Braunmüller (2002), therefore, sug-

CONTACT L. Kyjánek. E-mail: l.kyjaneck@gmail.com. ORCID: orcid.org/0000-0001-6354-8549.

CONTACT J. Haviger. E-mail: jiri.haviger@uhk.cz. ORCID: orcid.org/0000-0001-8353-497X.

gested using the term *receptive multilingualism* for this type of communication. Zeevaert (2007), nevertheless, argues that it is essential to distinguish between receptive multilingualism for (a) unrelated or remotely related languages, so-called *Abstandsprache*; and (b) closely related languages, so-called *Ausbausprache*.

A significant phenomenon with semicommunication is the asymmetry of the mutual intelligibility. Haugen (1966), Börestam Uhlmann (2002), Moberg et al. (2007), Schüp-pert and Gooskens (2011) confirm that, for example, in the relationship of Danish and Swedish, the Danes understand Swedish better than the Swedes understand Danish.

There are basically three factors affecting the mutual intelligibility of semicommunication: the recipient's attitude towards the language and his/her will to cooperate, the recipient's experience with the producer's language, and the linguistic distance between the producer's and the recipient's languages (Gooskens, 2006, 2007).

1.1. *The situation of the West Slavic languages*

The Slavic languages are divided into three branches. One of them is West Slavic branch, represented by Czech, Slovak, Polish, Sorbian and Kashubian. However, the materials used for the investigation of the last two languages are often limited, and the number of speakers is not as high as the number of Czech, Slovak or Polish speakers. Therefore, much research, including this, do not consider Sorbian and Kashubian.

Czech, Slovak and Polish are related both genetically and typologically (they are fleective). The Czech language is the most fleective West Slavic language. The Slovak language is similar to Czech, but it is closer to the South Slavic languages than the Czech language is. The Polish language is the least fleective West Slavic language, but at the same time it is the most analytical one (Comrie & Corbett, 2002). All the previously mentioned languages are also referred to as *Ausbausprache* (Kloss, 1967).

The language contacts of Czech and Slovak have lasted for quite a long time, and periods of their development have been both highly convergent and divergent (Lipowski, 2005). After the dissolution of Czechoslovakia in 1993, the language contacts of these languages have further weakened (Musilová, 2000). Until 1993, the language contacts had been supported in schools, at work and in the cultural environment through language planning. Since 1993, the Slovak language has been slowly disappearing from the Czech language environment (the knowledge of Slovak is no longer a part of the Czech educational program, Slovak newspapers are not being published in the Czech Republic, bilingual broadcasts have been stopped, etc.). The Czech language has been disappearing more slowly from the Slovak language environment, although a legal obligation to dub all non-Slovak audio-visual works for children under 12 years has been introduced in Slovakia. Even though semicommunication still works between Czechs and Slovaks, there are fears that both languages could become completely foreign, especially for the young generation that has not experienced a common state (Dobrotová & Musilová, 2012).

The language contacts of Czech and Polish or Slovak and Polish were not so intense in the past and they are not so in the present. Except for border areas, where the Polish language has been commonly used in various areas of life since the 19th century, the Czechs and Slovaks have not got easy access to the Polish language nor to Polish culture (Nábělková, 2008). Lotko (1998) also mentions that there are signs of xenophobia against the Poles.

Research on the mutual intelligibility of semicommunication between the West Slavic languages has so far been conducted exclusively through sociolinguistic methods and

the main focus has been on the relationship of Czech and Slovak. The results of this research (Budovičová, 1987; Dobrotová & Musilová, 2012; Golubović, 2016; Musilová, 2000; Nekvapil, Sloboda, & Wagner, 2009) assert, among other things, that Czech and Slovak are extremely close, Slovak is closer to Polish than is Czech, and that (Golubović, 2016, p. 85):

- the Slovaks understand written Czech better than the Czechs understand written Slovak,
- the Poles understand written Czech better than the Czechs understand written Polish,
- the Poles understand written Slovak better than the Slovaks understand written Polish;

- the Slovaks understand spoken Czech better than the Czechs understand spoken Slovak,
- the Czechs understand spoken Polish better than the Poles understand spoken Czech,
- the Slovaks understand spoken Polish better than the Poles understand spoken Slovak.

1.2. *Research questions*

The presented research examines the differences between West Slavic languages. It focuses solely on the linguistic factor of mutual intelligibility on both the graphemic and the phonetic planes. The methodology of Moberg et al. (2007) is followed through the research. In their article, they ‘assume a relationship between the non-linguistic factors (attitude and experience) and the intelligibility score, but correlations are low and the direct relationship is difficult to prove.’ (p. 52–53) The interpretation of the results is based on language distance as a strong indicator of potential the mutual intelligibility of semicommunication.

1.2.1. *Objectives*

For a comprehensive description and comparison of the results of mutual intelligibility on both language planes, the focus is on differences in the following areas:

- (1) overall mutual intelligibility,
- (2) asymmetry of mutual intelligibility,
- (3) mutual intelligibility of content/function words,
- (4) mutual intelligibility of various styles of material.

In the areas of overall mutual intelligibility and its asymmetry, the results are the same as the sociolinguistic research results, as expected (see Section 1.1).

In contrast, as to the mutual intelligibility of content/function words and various styles of material, different results are expected. For example, for Scandinavian languages, Moberg et al. (2007) measured significantly different values for content words than for function words. Different values could be also expected for various styles of material. There has not yet been any research dealing with the distinction between these two areas in the West Slavic languages.

Significantly different results for the graphemic and phonetic planes are not expected, at least not between Czech and Slovak. However, a difference between the two languages and Polish might occur, because Polish uses digraph orthography, which Czech and Slovak stopped using in the 15th century (except for the digraphs /ch/, /dz/, /dž/). Nevertheless, the pronunciation of all three languages is similar.

2. Methods

The methods used represent the real process of interlocutors during semicommunication. A recipient with a vocabulary of his/her mother tongue (code) receives a producer's code, and compares it with his/her own mother code. The recipient tries to understand the information based only on similarities in the compared codes (Stopyra, 2010). The producer's and recipient's vocabularies used in this research are contained in the databases of word pairs (see Section 2.1). The word pairs are transcribed for the phonetic plane research, but not for the graphemic plane research. For research purposes, the word pairs were also divided by their morphological tags into the two following categories (a) content words; and (b) function words. The stylistic classification of every word pair is determined by the source subcorpus.

All phonemes/graphemes had been aligned using Levenshtein distance before calculations of intelligibility. However, the computations were done using conditional entropy because the Levenshtein distance cannot capture the asymmetry of mutual intelligibility.

2.1. Materials

The input material for the conditional entropy calculation was a database of the most frequently used words and their translation equivalents for each studied language. Each word pair consisted of a word and its translation equivalent with the same semantic meaning. Phonemes/graphemes of the words in the word pairs were aligned by Levenshtein distance (Levenshtein, 1966). Moberg et al. (2007) used the 2000 most frequently used word pairs, although it turned out that about 800 word pairs would suffice to stabilize the value of the conditional entropy. Even a smaller sample of data brings satisfactory results.

The words in the database for this investigation of West Slavic languages came from the parallel corpus Intercorp v9 2016 (Bańczyk, Dybalska, & Vavřín, 2016; Nábělková & Vavřín, 2016; Rosen & Vavřín, 2016; Čermák & Rosen, 2012) of the Czech National Corpus organization¹. The four subcorpora (*Acquis*, *Core*, *Europarl*, *Subtitles*) of this corpus contained the Czech, Slovak and Polish vocabulary. From each subcorpus, the 2000 most frequently used words, including their morphological tags, to divide the words into content or function categories, were taken. The translation equivalents for the words were obtained from the individual subcorpora by the web application Treq v1.1, by the Czech National Corpus organization. The application offered several translation equivalents to every word, but the most frequently used equivalent was taken from the application.

- *Acquis* contains automatically aligned European Union legislation from 1950 to the present. Thus, the vocabulary of this subcorpus represents the style of official documents.
- *Core* contains manually aligned belletristic texts. Thus, the vocabulary of this subcorpus represents the belles lettres style.
- *Europarl* contains automatically aligned minutes of the 2007–2011 European Parliament meetings. They are in the form of monologues. The vocabulary of this subcorpus represents the style of official documents.
- *Subtitles* contains automatically aligned movie subtitles from the OpenSubtitles.org online database. The vocabulary of this subcorpus is therefore, on the borderline between the belles lettres style and simple speech style.

IPA was used for the phonetic transcription of the words which were used for studying

¹<http://korpus.cz/> for the organization; <http://ucnk.korpus.cz/intercorp/> for the corpus

Table 1. Final numbers of word pairs for each language relationship. The default number of words for each subcorpus was 2000.

Subcorpus	category	CS-SK	CS-PL	SK-PL
Acquis	all words	1772	1775	1770
	content words	1678	1681	1677
	function words	94	94	93
Core	all words	1986	1988	1984
	content words	1855	1857	1854
	function words	131	131	130
Europarl	all words	1960	1966	1957
	content words	1858	1864	1856
	function words	102	102	101
Subtitles	all words	1964	1968	1960
	content words	1842	1846	1839
	function words	122	122	121

the phonetic plane of the languages. For this investigation, programs for the phonetic transcriptions of Czech, Slovak and Polish were designed. The rules of phonetic transcription and orthophonies of all three languages were elaborated according to Palková (1994) for Czech, Král and Sabol (1989) for Slovak, and Balowski (1993) for Polish. If there were any disputed situations, Lipowski (2011) would be taken into consideration.

During the preparation phase of words for the Levenshtein distance alignment and the conditional entropy calculation, the final numbers of word pairs were reduced (see Table 1). The default number of words for each subcorpus was reduced due to the Treq application and phonetic transcription programs. The Treq application did not find translation equivalents for all words, and the phonetic transcription programs had not been designed for transcribing digits and foreign-language characters (for example β , etc.). Such word pairs were automatically excluded.

2.2. Levenshtein distance

The Levenshtein distance between two strings of characters is a metric specifying the minimum number of operations that would have to be made on the characters of the two strings to make the two strings identical. Three types of operations are allowed: adding, deleting, and replacing a character. Using this metric, the phonemes/graphemes were aligned, yielding the minimal Levenshtein distance between the words compared. In many cases, the two compared words were not of the same length. In such cases, a fill character ($/\#/$), not representing any phoneme/grapheme, was required for the free space. Table 2 shows this situation.

Table 2. The examples of outputs aligned by Levenshtein distance. The Czech and Slovak graphemic and phonetic transcribed word pairs mean, in English, *wooden* and *toy*. While the graphemic example does not need any fill character, the phonetic one need it for Slovak at position 5 (deleting operation), because the pronunciation of Czech $/ě/$ is realized as $[je]$ if $/ě/$ follows $/v/$. In both examples, the adding operation can be seen e.g., at position 1, and the replacing operation at position 2.

position:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(graph) CS	d	ř	e	v	ě	n	á			h	r	a	č	k	a
(graph) SK	d	r	e	v	e	n	á			h	r	a	č	k	a
(phon) CS	d	r̥	ɛ	v	j	ɛ	n	a:		f̥i	r	a	t̥ɕ	k	a
(phon) SK	d	r	ɛ	v	#	ɛ	n	a:		f̥i	r	a	t̥ɕ	k	a

In dialectometry research, different weights are often set for operations and the average differences/similarities of words or languages, and hence the levels of mutual intelligibility, are calculated (Heeringa, 2004). However, such results do not provide a quantification of the asymmetry of mutual intelligibility.

2.3. Conditional entropy

‘Conditional entropy measures entropy, or uncertainty in a random variable when another is known.’ (Moberg et al., 2007, p. 55) This means that it measures the uncertainty of the recipient’s understanding when he/she is trying to understand the received code. The higher the measured conditional entropy, the worse the recipient’s understanding is. Conditional entropy measures entropy, uncertainty from the point of view of each recipient separately, so that it can quantify the asymmetry of mutual intelligibility.

Conditional entropy is calculated using the following formula:

$$H(X|Y) = - \sum_{x \in X, y \in Y} p(x, y) \log_2 p(x|y). \quad (1)$$

Here, $p(x, y)$ is the probability of occurrence of the characters x and y at the same position (aligned pairs x, y), i.e. $p(x, y) = p(x \cap y)$. The conditional probability $p(x|y)$ means the probability of the occurrence of a character x if we know that character y is at the same position, i.e. $p(x|y) = \frac{p(x \cap y)}{p(y)}$. Precisely the denominator of this fraction provides a quantification of the asymmetry, that $H(X|Y)$ does not have to be equal to $H(Y|X)$.

2.3.1. An example of a conditional entropy calculation

To illustrate how conditional entropy on language data works, let’s consider a database of two phonetically transcribed word pairs aligned by Levenshtein distance, as in Table 2 in the previous subchapter. Thus, this database contains 14 aligned cells with phoneme pairs. Table 3 shows the calculation of the conditional and joint probabilities which are required in the formula for conditional entropy.

Table 3. Example of a database of two phonetically transcribed word pairs (aligned by Levenshtein distance) with calculated conditional and joint probabilities for each pair of phonemes.

position:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
CS	d	ɾ	ɛ	v	j	ɛ	n	a:	fi	r	a	ɛ	k	a
SK	d	r	ɛ	v	#	ɛ	n	a:	fi	r	a	ɛ	k	a
$p(SK CS)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$p(CS SK)$	1	0.5	1	1	1	1	1	1	1	0.5	1	1	1	1
$p(CS, SK)$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$
$p(SK, CS)$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{2}{14}$

Already at this calculation phase, the asymmetry between Czech and Slovak can be noticed. Czech uses two different phonemes at the second and tenth positions (the alveolar sonorant trill /r/, and the raised alveolar non-sonorant trill /ɾ/). Slovak uses only one phoneme (the alveolar sonorant trill /r/) in these positions. In practice, this means that Czechs need to choose one phoneme from /r/ or /ɾ/ for the position of the phoneme /r/ in Slovak. In contrast, Slovaks do not need to choose, because /r/ in Slovak

is always used at the positions of /r/ and /r̥/ in Czech. The conditional probabilities in Formula 1, $p(SK|CS)$ and $p(CS|SK)$, describe this situation mathematically for each aligned pair of phonemes (the results are in Table 3). In the formula for calculating the conditional entropy, these probabilities are given in logarithms: $\log_2 1 = 0$; $\log_2 0.5 = -1$.

The joint probability, $p(CS, SK) = p(SK, CS)$, is the relative frequency of occurrences of the individual aligned pairs of phonemes. Most of these pairs are found in the illustrative database just once, therefore, their joint probability is $\frac{1}{14}$. Only the aligned pairs in the third and eleventh cells are found in the database twice (also in the sixth and fourteenth cells), so their joint probability is $\frac{2}{14}$.

After fitting in Formula 1, conditional entropies really provide the quantification of the asymmetry in the relationship of Czech and Slovak. $H(SK|CS) = 0$ and $H(CS|SK) \doteq 0.143$. Based on this illustrative database of two word pairs, the Slovaks would understand Czech better than the Czechs understand Slovak, because $H(CS|SK) > H(SK|CS)$.

3. Results

In this section, the conditional entropies measured for every each pair of language relationships in both directions are presented. A low conditional entropy $H(native|foreign)$ implies that the alignment of phonemes/graphemes is frequent for both languages. A low conditional entropy, i.e. a low level of the recipient's uncertainty, means good intelligibility, whereas a high conditional entropy, i.e. a high level of the recipient's uncertainty, means bad intelligibility.

Now, it is important to stress that 'conditional entropy corresponds well with the results of intelligibility tests,' (Moberg et al., 2007, p. 64) and that the interpretations of the overall mutual intelligibility, its asymmetry, and the mutual intelligibility of various styles of material, ought to be based on categories of all words because 'the relationship is clearest when all words are included,' (Moberg et al., 2007, p. 64). This respects the process of sociolinguistic research (a set of all words is presented to the respondents).

The two following subsections contain a page with graphs of the conditional entropies for all the objectives of this research (see Section 1.2.1). Every measured language relationship (Czech–Slovak, Czech–Polish, Slovak–Polish) is plotted in its own graph for both linguistic planes (phonetic and graphemic). The vertical position of a point represents the overall mutual intelligibility, and the horizontal position of a point represents the asymmetry of the mutual intelligibility. The diagonal line in the graphs always shows the symmetric relation between languages. The distance of a point from the line of symmetry, the diagonal in the graph, shows the extent of the asymmetry.

The exact conditional entropies, relative asymmetries, and their extent for each measured language relationship, are presented in Appendix A for the phonetic plane and in Appendix B for the graphemic plane.

3.1. Phonetic plane

The results of the conditional entropies measured on the phonetic plane represent the language systems' potential for mutual intelligibility in the spoken form of semicommunication. Figures 1-3 present the conditional entropies for each language relationship, measured on the phonetic plane.

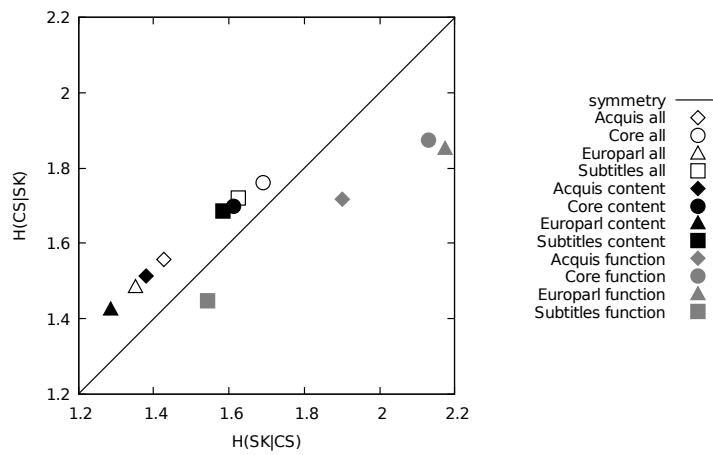


Figure 1. Conditional entropies for the relationship of Czech and Slovak on the phonetic plane.

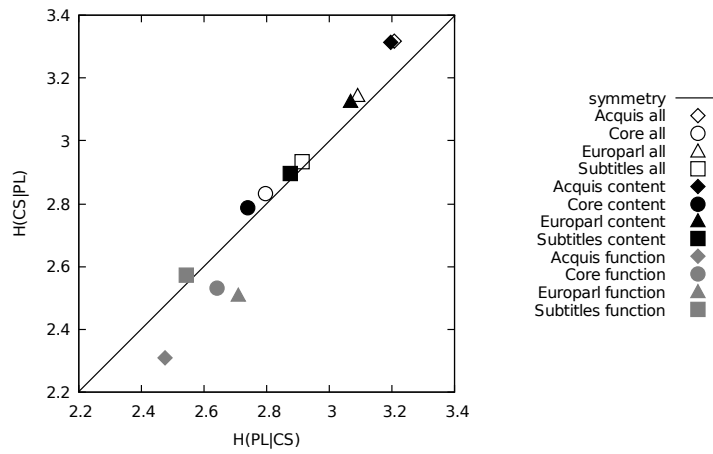


Figure 2. Conditional entropies for the relationship of Czech and Polish on the phonetic plane.

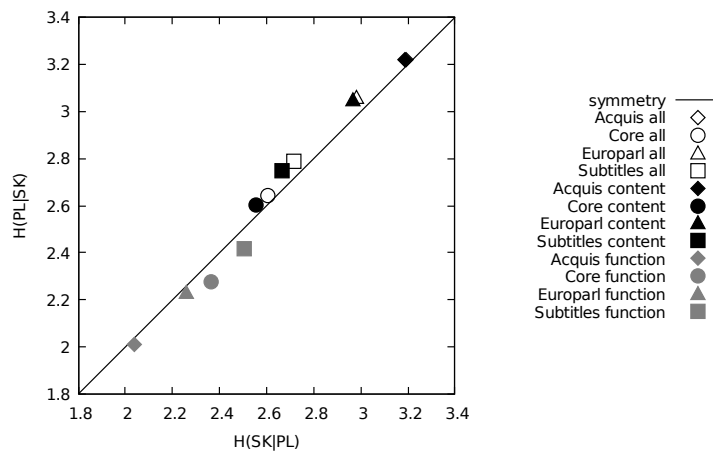


Figure 3. Conditional entropies for the relationship of Slovak and Polish on the phonetic plane.

3.1.1. Overall mutual intelligibility

Between Czech and Slovak, very low conditional entropies were measured (on the x -axis from 1.2850 to 2.1719, on the y -axis from 1.4259 to 1.8753). Such results indicate an extraordinary proximity of the phonetic plane of both languages and their very good mutual intelligibility for spoken semicommunication.

Between Czech and Polish, relatively high conditional entropies were measured (on the x -axis from 2.4756 to 3.2059, on the y -axis from 2.3098 to 3.3167). The conditional entropies in this relationship were the highest measured on the phonetic plane. This signifies not only the substantial distance between the two languages but also a much worse mutual intelligibility for spoken semicommunication between the Czechs and Poles than between the Czechs and Slovaks.

Between Slovak and Polish, slightly lower conditional entropies than for the relationship of Czech and Polish were measured (on the x -axis from 2.0398 to 3.1926, on the y -axis from 2.0102 to 3.2221). These conditional entropies signify the substantial distance between Slovak and Polish, but not as much as between Czech and Polish. Therefore, the overall mutual intelligibility of spoken semicommunication between the Slovaks and Poles is worse than between the Slovaks and Czechs, but it is not as bad as between the Czechs and Poles.

The results in this area corresponded with all the expectations based on the results of sociolinguistic research (see Section 1.1).

3.1.2. Asymmetry of mutual intelligibility

Between Czech and Slovak, $H(CS|SK) > H(SK|CS)$ applied to all the above mentioned points. This means that the Slovaks better understand spoken Czech than the Czechs understand spoken Slovak. Moreover, the extent of the asymmetries was large.

Between Czech and Polish, $H(CS|PL) > H(PL|CS)$ applied to all the above mentioned points, which means that Poles better understand spoken Czech than the Czechs understand spoken Polish. However, the extent of the asymmetries was very small.

Between Slovak and Polish, $H(PL|SK) > H(SK|PL)$ applied to all the above mentioned points. This means that the Slovaks better understand spoken Polish than the Poles understand spoken Slovak. The extent of the asymmetries was small, similar to the relationship of Czech and Polish.

The results in this area did not fulfil all the expectations based on the results of sociolinguistic researches (see Section 1.1). The results corresponded with the expectations for the relationship of Czech and Slovak, and for Slovak and Polish. Yet the opposite result was expected in the relationship of Czech and Polish. However, it should be mentioned that the extent of the asymmetries in this relationship and in the relationship of Slovak and Polish was quite close to symmetry.

3.1.3. Mutual intelligibility of content and function words

There were intriguing results for the relation between Czech and Slovak. The categories of function words (except for Subtitles) had higher conditional entropies than the categories of content words. This means that spoken content words in the relationship of Czech and Slovak are more mutually intelligible than spoken function words. This was not reflected in any other measured relationship. From the perspective of the asymmetry of mutual intelligibility, the opposite results were measured for both categories. The Slovaks better understand Czech spoken content words than the Czechs understand content words in Slovak, but the Czechs better understand Slovak spoken function words

rather than vice versa. The extent of the asymmetries was large.

There were lower conditional entropies between Czech and Polish for the category of function words than for the category of content words. Thus, the Czechs and Poles better understand spoken function words than content words. From the perspective of asymmetry, the function words were once again on the opposite side of the asymmetry than the content words were (except for Subtitles). This means that Poles better understand Czech spoken content words than the Czechs understand content words in Polish, but the Czechs better understand Polish spoken function words rather than vice versa. The extent of the asymmetries for the category of content words was small, but they were large for function words.

There were lower conditional entropies between Slovak and Polish for the category of function words than for the category of content words. According to these results, spoken function words in the relationship of Slovak and Polish are more mutually intelligible than spoken content words. In this case, all the function words (including Subtitles) were on the opposite side of the asymmetry from the content words. The Slovaks better understand Polish spoken content words than the Poles understand content words in Slovak, but the Poles better understand Slovak spoken function words rather than vice versa. The extent of the asymmetries for both categories was close to symmetry.

For the West Slavic languages, there has been no sociolinguistic research focusing solely on the understanding of content and function words, therefore the results of the present paper cannot be compared with any other results. However, the emergence of the opposite trend of asymmetry among these West Slavic languages is intriguing, since the results for the Scandinavian languages (Moberg et al., 2007) have always shown the content and function words on the same side of the asymmetry. Additionally, the category of function words always had lower conditional entropies than the category of content words. None of these assumptions have been fulfilled in the relationship of Czech and Slovak.

3.1.4. Mutual intelligibility of various styles of material

The source (subcorpus) of an individual word pair determines the style of material. The specific styles of subcorpora are described in Section 2.1.

Between Czech and Slovak, Europarl and Acquis had the lowest conditional entropies, and Subtitles and Core had higher values. The subcorpora Europarl and Acquis were not only close based on the measured entropies, but also by their styles. The vocabulary of both belongs to the style of official documents. A similar situation held between the subcorpora Subtitles and Core (stylised simple speech style, or belles lettres style). Thus, the spoken style of official documents is for Czechs and Slovaks more mutually intelligible than the spoken belles lettres style.

Between Czech and Polish, the measured results were exactly the opposite to the relationship of Czech and Slovak. Core and Subtitles had the lowest conditional entropies, and Europarl and Acquis had much higher values. Thus, the spoken belles lettres style is for Czechs and Poles more mutually intelligible than the spoken style of official documents. The distribution of points in the graph corresponded with the individual styles again. Subtitles and Core were again close to each other, just like Europarl and Acquis, although the distance between them was already greater.

Between Slovak and Polish, the results were almost identical to the results for the relationship of Czech and Polish. The lowest conditional entropies (i.e. the highest intelligibility) were measured for Core and Subtitles, much higher for Europarl and Acquis. The individual styles were distinctly separate and the extent of their asymmetries was

very small.

Again, the results in this area cannot be compared with the results of sociolinguistic research.

3.2. *The graphemic plane*

The results of the conditional entropies measured on the graphemic plane of these languages represent the potential of the language systems for mutual intelligibility in the written form of semicommunication, i.e. when the recipient reads text written in a foreign language. The following page contains graphs of the conditional entropies for each individual language relationship measured on the graphemic plane.

3.2.1. *Overall mutual intelligibility*

Very low conditional entropies between Czech and Slovak were measured (on the x -axis from 1.3472 to 1.6191, on the y -axis from 1.3136 to 1.5611). Such results indicate an extraordinary proximity and mutual intelligibility on the graphemic plane.

High conditional entropies between Czech and Polish were measured (on the x -axis from 2.7937 to 3.1353, on the y -axis from 2.8767 to 3.2278). The conditional entropies in this relationship were the highest of those measured on the graphemic plane. These results indicate the substantial distance between the two languages and a much worse mutual intelligibility of written semicommunication between the Czechs and Poles than between the Czechs and Slovaks.

Slightly lower conditional entropies between Slovak and Polish were measured than for the relationship of Czech and Polish (on the x -axis from 2.6406 to 3.0764, on the y -axis from 2.5401 to 3.0043). These conditional entropies reflect the substantial distance between Slovak and Polish, but not as much as in the relationship of Czech and Polish. Thus, the overall mutual intelligibility of written semicommunication between the Slovaks and Poles is worse than between the Slovaks and Czechs, but it is not as bad as between the Czechs and Poles.

The results in this area corresponded with all the expectations based on the results of sociolinguistic research (see Section 1.1).

3.2.2. *Asymmetry of mutual intelligibility*

For all measured relationships, a small degree of asymmetry was found.

Between Czech and Slovak, $H(SK|CS) > H(CS|SK)$ applied to all the above mentioned points. This means that the Czechs better understand written Slovak than the Slovaks understand written Czech.

Between Czech and Polish, $H(CS|PL) > H(PL|CS)$ applied to all the above mentioned points. This means that Poles better understand written Czech than the Czechs understand written Polish.

Between Slovak and Polish, $H(SK|PL) > H(PL|SK)$ applied to all the above mentioned points. This means that the Poles better understand written Slovak than the Slovaks understand written Polish.

The results in this area did not fulfil all the expectations based on the results of sociolinguistic research (see Section 1.1). The results corresponded with the expectations for the relationship of Czech and Polish, and of Slovak and Polish. The opposite result was expected in the relationship of Czech and Slovak.

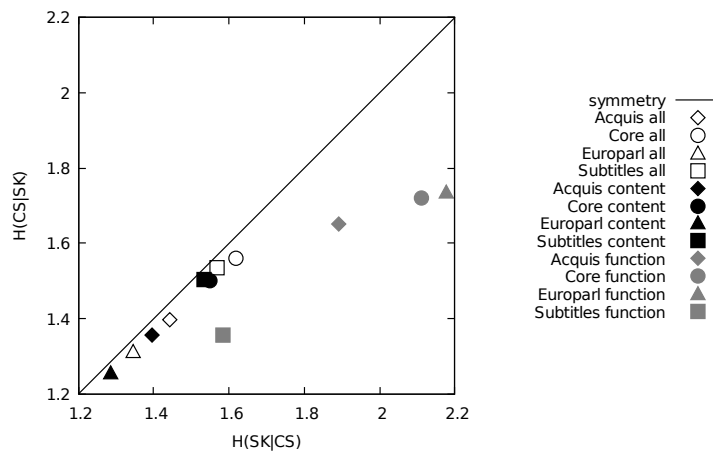


Figure 4. Conditional entropies for the relationship of Czech and Slovak on the graphemic plane.

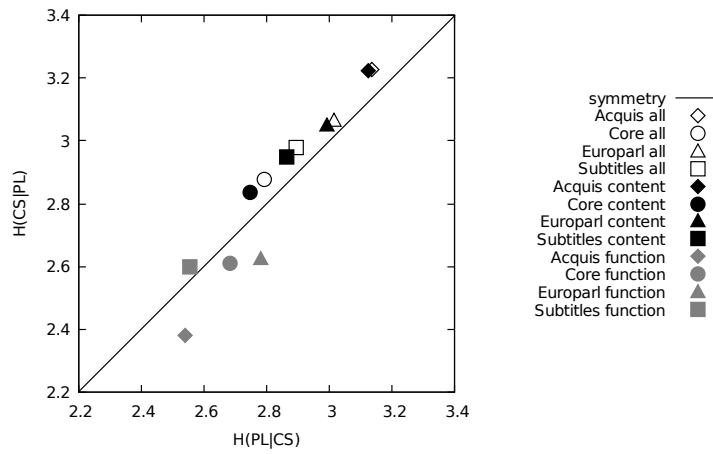


Figure 5. Conditional entropies for the relationship of Czech and Polish on the graphemic plane.

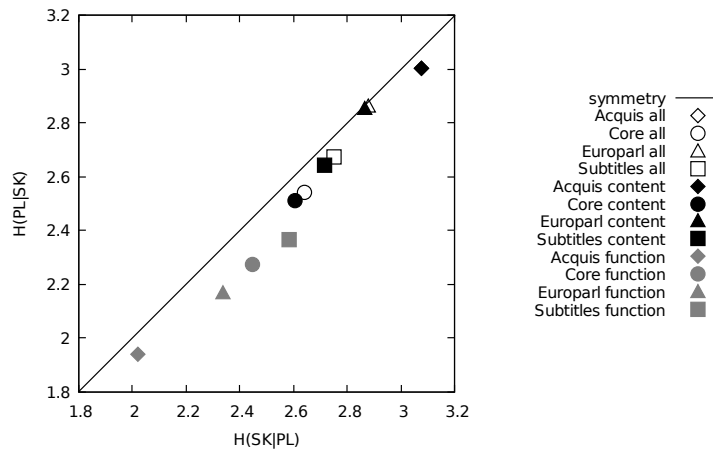


Figure 6. Conditional entropies for the relationship of Slovak and Polish on the graphemic plane.

3.2.3. *Mutual intelligibility of content and function words*

Between Czech and Slovak, the category of function words (except Subtitles) had higher conditional entropies than the category of content words, meaning that written content words are, in the relationship of Czech and Slovak, more mutually intelligible than written function words. Regarding the asymmetry of mutual intelligibility, both categories were on the same side. The Czechs better understand Slovak written content and function words than the Slovaks understand these words in Czech. The extent of the asymmetries was small for content words, and large for function words.

Lower conditional entropies between Czech and Polish were measured for function words than for content words. Thus, the Czechs and the Poles better understand written function words than content words. Regarding this asymmetry, function words were on the opposite side from content words (except Subtitles). According to these results, the Poles better understand Czech written content words than the Czechs understand content words in Polish, but the Czechs better understand Polish written function words rather than vice versa. The extent of the asymmetries was varied.

Lower conditional entropies between Slovak and Polish were measured for function words than for content words. The Slovaks and the Poles better understand written function words than content words. Both categories on the same side of the asymmetry. Thus, the Poles better understand Slovak written content and function words than the Slovaks understand these words in Polish. The extent of the asymmetries for both categories was small, but a little larger for the category of function words.

Once more, comparing the conditional entropies with the results of sociolinguistic research is not possible, because such research for the West Slavic languages has not yet been done.

3.2.4. *Mutual intelligibility of various styles of material*

Between Czech and Slovak, Europarl and Acquis had the lowest conditional entropies, and Subtitles and Core had higher values. Thus, the written style of official documents is for Czechs and Slovaks more mutually intelligible than the written belles lettres style. From the perspective of the asymmetry of mutual intelligibility, $H(SK|CS) > H(CS|SK)$ applied to all the above mentioned points, which means that the Czechs better understand all Slovak written styles rather than vice versa.

For the relation of Czech to Polish, the measured results were exactly the opposite from those for the relation of Czech to Slovak. Core and Subtitles had the lowest conditional entropies, and Europarl and Acquis had much higher values. Therefore, the written belles lettres style is for Czechs and Poles more mutually intelligible than the written style of official documents. From the point of view of the asymmetry of mutual intelligibility, $H(CS|PL) > H(PL|CS)$ applied to all the above mentioned points, meaning that the Poles better understand all Czech written styles rather than vice versa.

For the relation of Slovak to Polish, the results were almost identical to those for Czech and Polish (but the conditional entropies were lower by 0.2 bits for the relationship of Slovak and Polish). The lowest conditional entropies (i.e. the highest intelligibility) were measured for Core and Subtitles. The conditional entropies were much higher for Europarl and Acquis. From the point of view of the asymmetry of mutual intelligibility, $H(SK|PL) > H(PL|SK)$ applied to all the above mentioned points, which means that the Poles better understand all Slovak written styles rather than vice versa.

Here, too, the results in this area cannot be compared with the results of sociolinguistic research.

3.3. *Differences between the intelligibility of the phonetic and graphemic planes*

The results on the phonetic and the graphemic planes of the West Slavic languages were not very different, and in many cases corresponded to the predictions based on the results of sociolinguistic research. The following subsections describe some of the tendencies observed while comparing both planes. Interpretations and explanations are provided.

3.3.1. *Overall mutual intelligibility*

The results for the overall mutual intelligibility confirmed all predictions based on the results of sociolinguistic research. Czech and Slovak were the closest (thus, the most intelligible) among the West Slavic languages. Polish was distant from both languages, but not enough to be a barrier for semicommunication in practice. The measured conditional entropies were very close to the values measured for Scandinavian languages (see Moberg et al. (2007)). Polish was a little bit closer to Slovak than Czech.

The conditional entropies were generally lower on the graphemic plane than on the phonetic plane, which means that recipients of West Slavic languages understand better the written form of semicommunication than the spoken form. This could be explained by the general rigidity of the writing systems of all languages. Although the pronunciation of all West Slavic languages is very similar, the written form seems to be even more similar.

3.3.2. *Asymmetry of mutual intelligibility*

Some of these results about the asymmetry of the mutual intelligibility are new. Based on sociolinguistic research, it was assumed the asymmetries would differ for different planes. This prediction was confirmed by the new research presented here.

The extent of the asymmetries did not indicate any particular trends, but the asymmetry appeared on both planes of all measured relationships.

The result of the relationship of Czech and Slovak on the graphemic plane did not confirm the results of sociolinguistic research, even though the relationship between these two languages has been extensively described and has been being sociolinguistically explored by various linguists for several decades. The same sociolinguistic results on the phonetic plane was not confirmed for the relationship of Czech and Polish. It is necessary to point out that linguists are not as interested in the relationship of Czech and Polish, and hence there is not an extensive sociolinguistic literature on the mutual intelligibility between these languages.

3.3.3. *Mutual intelligibility of content and function words*

The results of the presented research in the area of the mutual intelligibility of content and function words provide unique results on the phonetic and the graphemic planes among West Slavic languages. Nevertheless, the results for function words were, to a certain extent, indefinite. There has been no sociolinguistic research in this area with which these results could be compared.

On both measured planes, it was positively confirmed that the overall mutual intelligibility and its asymmetry clearly correlate with the intelligibility and asymmetry of content words. Their conditional entropies were always close and they always stood on the same side of the asymmetry.

On the other hand, function words behaved unpredictably on both measured planes. On the phonetic plane, function words usually stood on the opposite side of the asymmetry than the others. On both planes, in the relationship of Czech and Slovak, the function words had higher conditional entropies than the other category. Both of these observations could be explained by the small number of word pairs in the category of function words. Although the results of the conditional entropies should be sufficient for fewer than 800 pairs (Moberg et al., 2007), it seems that the number of pairs in the category of function words was probably insufficient to stabilize the results. However, the extraordinary proximity of Czech and Slovak should affect the conditional entropies of function words. In such a close relationship, even fewer irregularities could cause a bigger extension.

3.3.4. Mutual intelligibility of various styles of material

The results of the presented research in the area of the mutual intelligibility of various styles of material showed a clear individual tendency for each measured language relationship on both planes. For the relationship of Czech and Slovak, the style of official documents was more intelligible than the belles lettres style. For the relationship of Czech and Polish, and Slovak and Polish, the belles lettres style was more intelligible than the style of official documents. These two opposite trends could be explained by the separate development of the individual languages.

Lipowski (2005) mentioned the convergent and divergent periods of the development of Czech and Slovak. Due to national pride, a strong divergent tendency in the belles lettres style is assumed (see Section 1.1). On the other hand, in practice Czech and Slovak translators of European documents (Acquis and Europarl subcorpora) cooperate, thanks to the proximity of these languages. This could be the reason why the style of official documents is more intelligible than belles lettres in the relationship of Czech and Slovak.

Lotko (1999) mentioned the tendency of strong Europeanization of current Polish. This is not so strong for Czech or Slovak, and therefore the understanding by Czechs and Slovaks of the Polish (contemporary) style of official documents may be worse than the (traditional) belles lettres style.

There were interesting results concerning the subcorpus Subtitles, representing a semi belles-lettres style or simple speech style (stylised common language including substandard forms of language). The conditional entropies on both studied planes in all language relationships always placed this subcorpus in the middle of the measured spectrum. It seems that Subtitles could best reflect the languages that are commonly used for communication in the Czech Republic, Slovakia, and Poland.

The conditional entropies on the phonetic plane visibly separated individual styles. The subcorpora Acquis and Europarl, representing the style of official documents, always had very similar conditional entropies. Similarly, the conditional entropies for the subcorpora Core and Subtitles were very close. On the graphemic plane, the distance between the conditional entropies of the individual subcorpora was not so obvious.

4. Discussion

The purpose of this research was to describe and quantify the mutual intelligibility of the West Slavic languages based on their language distance, and to compare the results with the results of sociolinguistic research (see Section 1.1). The measurement of this

purely linguistic factor on the Czech, Slovak and Polish phonetic and graphemic planes was the first such realized using conditional entropy. In addition to the overall mutual intelligibility, it also takes into consideration the asymmetry of mutual intelligibility. Interesting results were revealed by sorting the used material according to content and function words and according to their style (the style of official documents, belles lettres, and the simple speech style).

The following part contains a brief summary of the results.

Overall mutual intelligibility

- The Czechs and Slovaks understand each other the best of all, in terms of the measured relationships on both planes.
- Czech and Slovak have a distance from Polish on both planes, but not large enough to prevent mutual intelligibility.
- The Poles better understand the Slovaks than the Czechs on both planes.

The measured results of this research in the area of overall mutual intelligibility confirmed all the results of sociolinguistic research (Budovičová, 1987; Dobrotová & Musilová, 2012; Golubović, 2016; Musilová, 2000; Nekvapil et al., 2009) on the phonetic and the graphemic planes.

Asymmetry of mutual intelligibility

- The Slovaks better understand spoken Czech than vice versa.
- The Poles better understand spoken Czech than vice versa.
- The Slovaks better understand spoken Polish than vice versa.
- The extent of the asymmetry of the relationship of Czech and Slovak was quite large.
- The extent of the asymmetry of the relationship of Czech and Polish, as well as Slovak and Polish, was small.

- The Czechs better understand written Slovak than vice versa.
- The Poles better understand written Czech than vice versa.
- The Poles better understand written Slovak than vice versa.
- The extent of the asymmetries of all measured relationships was small.

The results of this research in the area of the asymmetry of mutual intelligibility confirmed most of the results of sociolinguistic research (Budovičová, 1987; Dobrotová & Musilová, 2012; Golubović, 2016; Musilová, 2000), except for the relationship of Czech and Polish on the phonetic plane and Czech and Slovak on the graphemic plane.

Mutual intelligibility of content and function words

- Content words contribute more than function words to the overall mutual intelligibility and its asymmetry.
- The function words tend to be on the opposite side of the asymmetry than the content words.
- Function words have greater asymmetries on both planes than do the content words.
- The Czechs and Slovaks better understand the content words of their language relationship than the function words.
- The Czechs and Poles better understand the function words of their language relationship than the content words.
- The Slovaks and Poles better understand the function words of their language relationship than the content words.

This research establishes only certain trends, as the number of function words may have been insufficient for some of the measured relationships to stabilize the conditional entropies. The conclusions in this area should be further verified.

Mutual intelligibility of various styles of material

- The Czechs and Slovaks better understand the style of official documents of their language relationship than the belles lettres style.
- The Czechs and Poles better understand the belles lettres style of their language relationship than the style of official documents.
- The Slovaks and Poles better understand the belles lettres style of their language relationship than the style of official documents.
- The conditional entropies distinctly separate the individual styles of material (especially on the phonetic plane).
- The subcorpus Subtitles could probably be the best reflection of the languages commonly used for communication in the Czech Republic, Slovakia and Poland.

The measured conditional entropies for the various styles of material brought up challenging results. More detailed research in this area would be highly appropriate.

The methods used have proven themselves well, although they have technical and conceptual limitations (technical in the area of the phoneme/grapheme alignment using the Levenshtein distance, conceptual in the area of not including extra-linguistic factors of mutual intelligibility). However, the potential for practical use is considerable, for example in learning closely related languages (intercomprehension methods). Partial results of the methods used may be very helpful in this. Based on the joint and conditional probability results for each aligned pair of phonemes/graphemes, it would be possible to suggest more appropriate approaches and textbooks to support the teaching and learning process.

Acknowledgement

The contribution were supported by a project of Students Grant Agency – FIM, University of Hradec Kralove, Czech Republic.

Appendix A. Exact values of conditional entropies (CE) and relative asymmetries (RA) on the phonetic plane

Table A1. CE and RA for relationship of Czech and Slovak.

Subcorpus	category	Conditional entropy		Relative asymmetry
		$H(CS SK)$	$H(SK CS)$	$\left \frac{H(CS SK) - H(SK CS)}{H(CS SK)} \right $
Acquis	all words	1.5569	1.4281	0.0827
	content words	1.5149	1.3808	0.0885
	function words	1.7170	1.9005	0.1069
Core	all words	1.7626	1.6901	0.0411
	content words	1.6976	1.6135	0.0495
	function words	1.8753	2.1292	0.1354
Europarl	all words	1.4853	1.3509	0.0905
	content words	1.4259	1.2850	0.0988
	function words	1.8513	2.1719	0.1732
Subtitles	all words	1.7205	1.6251	0.0554
	content words	1.6866	1.5829	0.0615
	function words	1.4478	1.5444	0.0667

Table A2. CE and RA for the relationship of Czech and Polish.

Subcorpus	category	Conditional entropy		Relative asymmetry
		$H(CS PL)$	$H(PL CS)$	$\left \frac{H(CS PL) - H(PL CS)}{H(CS PL)} \right $
Acquis	all words	3.3167	3.2059	0.0334
	content words	3.3124	3.1940	0.0357
	function words	2.3098	2.4756	0.0718
Core	all words	2.8313	2.7942	0.0131
	content words	2.7862	2.7407	0.0163
	function words	2.5305	2.6426	0.0443
Europarl	all words	3.1431	3.0904	0.0168
	content words	3.1262	3.0652	0.0195
	function words	2.5088	2.7095	0.0800
Subtitles	all words	2.9344	2.9137	0.0071
	content words	2.8970	2.8738	0.0080
	function words	2.5727	2.5451	0.0107

Table A3. CE and RA for the relationship of Slovak and Polish.

Subcorpus	category	Conditional entropy		Relative asymmetry
		$H(SK PL)$	$H(PL SK)$	$\left \frac{H(PL SK) - H(SK PL)}{H(PL SK)} \right $
Acquis	all words	3.1926	3.2221	0.0092
	content words	3.1887	3.2175	0.0090
	function words	2.0398	2.0102	0.0147
Core	all words	2.6028	2.6425	0.0150
	content words	2.5543	2.6003	0.0177
	function words	2.3639	2.2791	0.0372
Europarl	all words	2.9789	3.0599	0.0265
	content words	2.9647	3.0485	0.0275
	function words	2.2580	2.2322	0.0116
Subtitles	all words	2.7134	2.7880	0.0268
	content words	2.6667	2.7460	0.0289
	function words	2.5057	2.4151	0.0375

Appendix B. Exact values of conditional entropies (CE) and relative asymmetries (RA) on the graphemic plane

Table B1. CE and RA for the relationship of Czech and Slovak.

Subcorpus	category	Conditional entropy		Relative asymmetry
		$H(CS SK)$	$H(SK CS)$	$\left \frac{H(CS SK) - H(SK CS)}{H(CS SK)} \right $
Acquis	all words	1.3991	1.4416	0.0304
	content words	1.3566	1.3973	0.0300
	function words	1.6508	1.8919	0.1461
Core	all words	1.5611	1.6191	0.0372
	content words	1.5017	1.5499	0.0321
	function words	1.7209	2.1100	0.2261
Europarl	all words	1.3136	1.3472	0.0256
	content words	1.2563	1.2861	0.0237
	function words	1.7347	2.1776	0.2553
Subtitles	all words	1.5340	1.5689	0.0228
	content words	1.5039	1.5323	0.0189
	function words	1.3577	1.5847	0.1672

Table B2. CE and RA for the relationship of Czech and Polish.

Subcorpus	category	Conditional entropy		Relative asymmetry
		$H(CS PL)$	$H(PL CS)$	$\left \frac{H(CS PL) - H(PL CS)}{H(CS PL)} \right $
Acquis	all words	3.2278	3.1353	0.0287
	content words	3.2237	3.1246	0.0307
	function words	2.3808	2.5385	0.0662
Core	all words	2.8767	2.7937	0.0289
	content words	2.8371	2.7459	0.0321
	function words	2.6089	2.6838	0.0287
Europarl	all words	3.0659	3.0150	0.0166
	content words	3.0500	2.9918	0.0191
	function words	2.6262	2.7821	0.0594
Subtitles	all words	2.9771	2.8940	0.0279
	content words	2.9503	2.8643	0.0291
	function words	2.5972	2.5546	0.0164

Table B3. CE and RA for the relationship of Slovak and Polish.

Subcorpus	category	Conditional entropy		Relative asymmetry
		$H(SK PL)$	$H(PL SK)$	$\left \frac{H(PL SK) - H(SK PL)}{H(PL SK)} \right $
Acquis	all words	3.0764	3.0043	0.0240
	content words	3.0749	3.0031	0.0239
	function words	2.0236	1.9399	0.0431
Core	all words	2.6406	2.5401	0.0396
	content words	2.6068	2.5098	0.0386
	function words	2.4468	2.2736	0.0762
Europarl	all words	2.8784	2.8617	0.0058
	content words	2.8656	2.8536	0.0042
	function words	2.3373	2.1692	0.0775
Subtitles	all words	2.7499	2.6729	0.0288
	content words	2.7150	2.6446	0.0266
	function words	2.5848	2.3642	0.0933

References

- Balowski, M. (1993). *Title in Czech: Fonetika a fonologie současné polštiny [Title in English: Contemporary polish phonetics and phonology]*. Praha: Karolinum.
- Bańczyk, L., Dybalska, R. R., & Vavřín, M. (2016). *Title in Czech: Korpus InterCorp -- polština, verze 9 z 09.09.2016 [Title in English: The corpus InterCorp – Polish, version 9 from 09.09.2016]*. [online]. Praha: Ústav českého národního korpusu FF UK. Retrieved from <http://www.korpus.cz>
- Braunmüller, K. (2002). Semicommunication and accommodation: Observations from the linguistic situation in Scandinavia. *International Journal of Applied Linguistics*, 12(1), 1–23. Retrieved from <http://dx.doi.org/10.1111/1473-4192.00022>
- Budovičová, V. (1987). *Title in Slovak: Semikomunikácia ako lingvistický problém [Title in English: Semicommunication as a linguistic problem]*. *Studia Academica Slovaca*, 16, 49–66.
- Börestam Uhlmann, U. (2002). Interscandinavian comprehension and Scandinavian language community: Real or ideal? *Pour un modélisation de l'apprentissage simultané de plusieurs langues apparentées ou voisines. Publications de la Faculté de Lettres de Nice, CID Diffusion, Paris*, 129–136. Retrieved from <http://logatome.eu/pmasplav2001.pdf>
- Comrie, B., & Corbett, G. G. (2002). *The slavonic languages*. London and New York: Routledge.
- Dobrotová, I., & Musilová, K. (2012). *Title in Czech: Z dějin kontaktu dvou blízkých slovanských jazyků [Title in English: From the history of contact of two close Slavic languages]*. *Bohemistika*, 2012(1), 35–60. Retrieved from <http://hdl.handle.net/10593/3002>
- Golubović, J. (2016). *Mutual intelligibility in the Slavic language area*. Groningen: University of Groningen.
- Gooskens, C. (2006). Linguistic and extra-Linguistic predictors of inter-Scandinavian intelligibility. *Linguistics in the Netherlands 2006*, 23(1), 101–113. Retrieved from <http://www.tandfonline.com/doi/abs/10.2167/jmmd511.0>
- Gooskens, C. (2007). The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development*, 28(6), 445–467. Retrieved from <http://www.tandfonline.com/doi/abs/10.2167/jmmd511.0>
- Haugen, E. (1966). Semicommunication: The language gap in Scandinavia. *Sociological Inquiry*, 36(2), 280–297. Retrieved from <http://dx.doi.org/10.1111/j.1475-682X.1966.tb00630.x>
- Heeringa, W. J. (2004). *Measuring dialect pronunciation differences using Levenshtein distance* (Doctoral dissertation, University of Groningen). Retrieved from <http://www.rug.nl/research/portal/files/9800656/thesis.pdf>
- Kloss, H. (1967). Abstand languages and ausbau languages. *Anthropological Linguistics*, 9(7), 29–41.
- Král, A., & Sabol, J. (1989). *Title in Slovak: Fonetika a fonológia [Title in English: Phonetics and phonology]*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10, 707.
- Lipowski, J. (2005). *Title in Czech: Konvergence a divergence češtiny a slovenštiny v československém státě [Title in English: Convergence and divergence in the czechoslovak state]*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Lipowski, J. (2011). *Title in Slovak: Operatívna fonetika slovenčiny, češtiny a polštiny [Title in English: Czech, slovak and polish operative phonetics]*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Lotko, E. (1998). *Title in Czech: Srovnávací a bohemistické studie [Title in English: Confrontation and bohemistic studies]*. In E. Lotko (Ed.), (pp. 63–75). Olomouc: Vydavatelství Univerzity Palackého.
- Lotko, E. (1999). *Title in Czech: Srovnávací a bohemistické studie [Title in English: The confrontation and bohemistic studies]*. In E. Lotko (Ed.), (pp. 9–19). Olomouc: Vydavatelství

- Univerzity Palackého.
- Moberg, J., Gooskens, C., Nerbonne, J., & Vaillette, N. (2007). Conditional entropy measures intelligibility among related languages. *LOT Occasional Series*, 7, 51–66. Retrieved from <https://dspace.library.uu.nl/handle/1874/296747>
- Musilová, K. (2000). Title in Slovak: Mesto a jeho jazyk [Title in English: The town and its language]. In (pp. 280–289). Bratislava: Vydavateľstvo slovenskej akadémie vied.
- Nekvapil, J., Sloboda, M., & Wagner, P. (2009). *Title in Czech: Mnohojazyčnost v České republice* [Title in English: Multilingualism in the czech republic]. Prague: Lidové noviny.
- Nábělková, M. (2008). *Title in Slovak: Slovenčina a čeština v kontakte* [Title in English: Slovak and czech in contact]. Bratislava: SAV.
- Nábělková, M., & Vavřín, M. (2016). *Title in Czech: Korpus InterCorp — slovenština, verze 9 z 09.09.2016* [Title in English: The corpus InterCorp – Slovak, version 9 from 09.09.2016]. [online]. Praha: Ústav českého národního korpusu FF UK. Retrieved from <http://www.korpus.cz>
- Palková, Z. (1994). *Title in Czech: Fonetika a fonologie češtiny* [Title in English: Czech phonetics and phonology]. Praha: Karolinum.
- Rosen, A., & Vavřín, M. (2016). *Title in Czech: Korpus InterCorp — čeština, verze 9 z 09.09.2016* [Title in English: The corpus InterCorp – Czech, version 9 from 09.09.2016]. [online]. Praha: Ústav českého národního korpusu FF UK. Retrieved from <http://www.korpus.cz>
- Schüppert, A., & Gooskens, C. (2011). Investigating the link between language attitudes and intelligibility using reaction time. *Dialectologia, Special issue(2)*, 119–140. Retrieved from <http://www.publicacions.ub.edu/revistes/dialectologiaSP2011>
- Stopyra, J. (2010). Title in Polish: Semikomunikacja w skandinawskim obszarze jezykowym [Title in English: Semicommunication in the Scandinavian language area]. *Biuletyn polskiego towarzystwa jezykoznawczego*, 66, 133–141. Retrieved from http://mbc.malopolska.pl/Content/89724/biuletyn_ptj_66.pdf
- Zeevaert, L. (2007). Receptive multilingualism. In J. D. ten Thije & L. Zeevaert (Eds.), (Vol. 6, pp. 103–137). Amsterdam: John Benjamins.
- Čermák, F., & Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, 13(3), 411–427. Retrieved from http://utkl.ff.cuni.cz/~rosen/public/2012_intercorp_ijcl.pdf