

Klára Jágrová

kjagrova@coli.uni-saarland.de

Previous research:

I) Linguistic distances between CS and PL – comparative analysis of 100 most frequent nouns

ENG	original CS	(cognate) translations		
		PL	RU	BG
year	rok	rok	год	година
human	člověk	człowiek	человек	човек
day	den	dzień	день	ден
time	doba	doba	пора	доба
child	dítě	dziecko	дитя	дете
life	život	życie	жизнь	живот
place	místo	miejsce	место	место
hand	ruka	ręka	рука	ръка
work	práce	praca	работа	работа
world	svět	świat	свет	свят
side	strana	strona	сторона	страна

Lexical distance as the percentage of non-cognates

		reader			
		BG	RU	CS	PL
stimulus	BG		12	28	35
	RU	12		19	23
	CS	29	26		16
	PL	36	30	14	

Orthographic distance of cognates (LD in %)

		reader			
		BG	RU	CS	PL
stimulus	BG		14	24	32
	RU	13		27	34
	CS	24	27		36
	PL	34	35	36	

(Jágrová et al., forthcoming)

large discrepancy → potential?

II) Diachronically-based cross-lingual orthographic correlates (CS-PL, examples:)

LA – ŁO		ŘE – RZE	
zlato	złoto (gold)	břeh	brzeg (shore)
hlava	głowa (head)	před	przed (in front of)

(Fischer et al., 2015)

Context: Receptive Multilingualism

- Slavic languages – expected to be mutually intelligible
- depending on language combination and various factors: orthography, lexis, morphology, syntax
→ different degrees of linguistic distance

Hypotheses

- The distinct PL orthography impairs reading intercomprehension of PL for CS readers.
- Speakers of other Slavic languages are likely to understand PL better when the orthography is adapted to the readers.

Questions

- Which of the possible linguistic adaptations holds the greatest potential to improve intercomprehension?
 - theoretically (see example below)
 - in practice (future experiments)
→ correlation?

Example & demonstration of method: different adaptations of PL towards CS

ADAPTATION METHODS	In the European Parliament, all official languages are equally important.									reduction of distances in %																																																											
	[In Parliament European all languages official are equally important.]									orth	lex	synt																																																									
	ORIGINAL PL	W	P	a	r	a	m	e	n	c	i	e	E	u	r	o	p	e	j	s	k	i	m	w	s	z	y	s	t	k	i	e	j	ę	z	y	k	i	u	r	z	ę	d	o	w	e	s	ą	r	ó	w	n	i	e	i	s	t	o	t	n	ę	.							
orth	V	P	a	r	a	m	e	n	c	i	e	Ě	e	v	r	o	p	e	j	s	k	ý	m	v	š	i	s	t	k	y	j	ę	z	y	k	y	ú	ř	e	d	o	v	é	j	s	o	r	ó	v	n	ě	j	i	s	t	o	t	n	é	.	-28	0	0						
morph	W	P	a	r	a	m	e	n	c	i	e	u	E	u	r	o	p	s	k	é	m	w	s	z	y	s	t	k	y	j	ę	z	y	k	y	u	r	z	ę	d	n	í	s	ą	r	ó	w	n	ě	i	s	t	o	t	n	é	.	-21	0	0									
closed	V	P	a	r	a	m	e	n	c	i	e	u	E	u	r	o	p	e	j	s	k	i	m	v	š	e	c	h	n	y	j	ę	z	y	k	y	u	r	z	ę	d	o	w	e	j	s	o	r	ó	w	n	i	e	i	s	t	o	t	n	ę	.	-19	0	0					
lex	W	P	a	r	a	m	e	n	c	i	e	u	E	u	r	o	p	e	j	s	k	i	m	w	s	z	y	s	t	k	i	e	j	ę	z	y	k	y	u	r	z	ę	d	o	w	e	s	ą	ś	c	i	e	j	n	i	e	*	w	y	z	n	a	m	n	ę	.	-8	-22	0
all without order	V	P	a	r	a	m	e	n	c	i	e	u	e	v	r	o	p	s	k	é	m	v	š	e	c	h	n	y	j	ę	z	y	k	y	ú	ř	e	d	n	í	j	s	o	s	t	e	j	n	ě	v	ý	z	n	a	m	n	ę	.	-60	-22	0								
order	W	E	u	r	o	p	e	j	s	k	i	m	P	a	r	a	m	e	n	c	i	e	s	ą	w	s	z	y	s	t	k	i	e	u	r	z	ę	d	o	w	e	j	ę	z	y	k	i	r	ó	w	n	i	e	i	s	t	o	t	n	ę	.	0	0	-56					
all without lex	V	E	u	r	o	p	s	k	é	m	p	a	r	a	m	e	n	c	i	e	u	j	s	o	v	š	e	c	h	n	y	ú	ř	e	d	n	í	j	ę	z	y	k	y	r	ó	w	n	ě	j	i	s	t	o	t	n	é	.	-46	0	-56									
all without orth	V	E	u	r	o	p	s	k	é	m	p	a	r	a	m	e	n	c	i	e	u	j	s	o	v	š	e	c	h	n	y	u	r	z	ę	d	n	í	j	ę	z	y	k	y	ś	c	i	e	j	n	ě	*	v	ý	z	n	a	m	n	ę	.	-50	-22	-56					
all without morph	V	E	u	r	o	p	s	k	ý	m	p	a	r	a	m	e	n	c	i	e	Ě	j	s	o	v	š	e	c	h	n	y	ú	ř	e	d	o	v	é	j	ę	z	y	k	y	s	t	e	j	n	ě	v	ý	z	n	a	m	n	ę	.	-52	-22	-56							
all without closed	V	E	u	r	o	p	s	k	é	m	p	a	r	a	m	e	n	c	i	e	u	j	s	o	v	š	i	s	t	k	y	ú	ř	e	d	n	í	j	ę	z	y	k	y	s	t	e	j	n	ě	v	ý	z	n	a	m	n	ę	.	-54	-22	-56								
all	V	E	u	r	o	p	s	k	é	m	p	a	r	a	m	e	n	c	i	e	u	j	s	o	v	š	e	c	h	n	y	ú	ř	e	d	n	í	j	ę	z	y	k	y	s	t	e	j	n	ě	v	ý	z	n	a	m	n	ę	.	-60	-22	-56								
TARGET CS	V	E	u	r	o	p	s	k	é	m	p	a	r	a	m	e	n	c	i	e	u	j	s	o	v	š	e	c	h	n	y	ú	ř	e	d	n	í	j	ę	z	y	k	y	s	t	e	j	n	ě	v	ý	z	n	a	m	n	ę	.	-60	-22	-56								

Example sentence from:
<http://www.europarl.europa.eu/aboutparliament/en/20150201PVL00013/Multilingualism/> [retrieved: 28.08.2016]

- Legend:**
- **orth** = orthographic transformation by PL-CS orthographic correspondence rules
 - **morph** = exchange of morphological units (inflection + derivation)
 - **closed** = exchange of closed class words
 - **lex** = substitution of lexeme if not cognate to CS target; closed class words excluded; orthography of substituted CS words is PL
 - **order** = adaptation of word order
 - **all** = orth+morph+closed+lex+order
 - **diff** = number of different items mentioned towards Czech target
 - **LD** = Levenshtein distance (number of deletions, insertions, substitutions divided by number of alignment slots; differences in diacritics cost 0.5)
 - **blue** = different characters towards target (cost of 1)
 - **lighter blue** = differences in characters with cost of only 0.5

Aim: Determine potential for improvement in each adaptation method

1. Count differences to target

→ theoretical measure of potential for each adaptation

- greatest reductions in distance when applied individually: **orth, lex, order**
- when **morph** or **closed** is not applied: combination of all other methods still relatively powerful in reducing linguistic distance
- combination of all adaptation methods: potential to reduce linguistic distance to zero

2. Compare data with results from reading experiments with CS native speakers with no knowledge of PL (to be launched)

- sentences in all adaptation methods are experimental stimuli
- Which units of language carry which information in concrete intercomprehension situations?

3. Which units are CS readers able to decode from the different adaptations?

- find suitable prediction formula(e)
- compare to success of other linguistic distance measures as predictors