

Univerzita Hradec Králové
Přírodovědecká fakulta
Katedra českého jazyka a literatury

Semikomunikace v konceptu komunikace

Bakalářská práce

Autor:	Lukáš Kyjánek
Studijní program:	B1801 / Informatika
Studijní obor:	Český jazyk a literatura se zaměřením na vzdělávání Informatika se zaměřením na vzdělávání
Vedoucí práce:	PhDr. Jana Bílková, Ph.D.
Odborný konzultant:	Mgr. Jiří Haviger, Ph.D. UHK, Fakulta informatiky a managementu Katedra informatiky a kvantitativních metod



Zadání bakalářské práce

Autor:	Lukáš Kyjánek
Studium:	S14IN001BP
Studijní program:	B1801 Informatika
Studijní obor:	Informatika se zaměřením na vzdělávání
Název bakalářské práce:	Semikomunikace v konceptu komunikace
Název bakalářské práce AJ:	Semicomunication in the Communication Concept

Cíl, metody, literatura, předpoklady:

Bakalářská práce je zaměřena na problematiku semikomunikace v konceptu klasického přístupu ke komunikaci, jak jej definují kodaňská a pražská škola, matematika i informační technologie. Cílem práce je popsat, kromě samotné komunikace, semikomunikace a pasivního bilingvismu, také matematické přístupy skandinávské a západoslovanské jazykovědy k porozumění si jednotlivých mluvčích při semikomunikaci.

ČERNÝ, Jiří. Úvod do studia jazyka. Olomouc: Rubico, 1998. ISBN: 80-85839-24-5. JENS, Mooberg, GOOSKENS, Charlotte, NERBONNE, John, VAILLETTE, Nathan. Conditional Entropy Measures Intelligibility among Related Languages. In: Computational Linguistics in the Netherlands 2006: Selected papers from the 17th CLIN Meeting. ULTRECHT: LOT, 2007, 51-66. SLOBODA, Marián. Slovensko-česká (semi)komunikace a vzájemná (ne)srozumitelnost. Čeština doma a ve světě. Praha: Filozofická fakulta UK, 2004, XII(3-4), 15-36. ISSN: 1210-9339. THIJE, Jan D. ten, ZEEVAERT, Ludger. Receptive Multilingualism. Linguistic analyses, language policies and didactic concepts. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2007. ISBN: 978-90-272-1926-8.

Garantující pracoviště:	Katedra českého jazyka a literatury, Pedagogická fakulta
Vedoucí práce:	PhDr. Jana Bílková, Ph.D.
Oponent:	PhDr. Jiří Zeman, Ph.D.
Datum zadání závěrečné práce:	1.12.2015

Prohlášení

Prohlašuji, že tato práce je mým původním autorským dílem, které jsem vypracoval pod dohledem vedoucí práce a odborného konzultanta samostatně, a že řádně cituji všechny prameny a literaturu, z nichž jsem čerpal.

Prohlašuji, že je bakalářská práce uložena v souladu s rektorským výnosem č. 1/2013 (Řád pro nakládání se školními a některými jinými autorskými díly na UHK).

V Hradci Králové dne 26.04.2017

.....
Lukáš Kyjánek

Poděkování

Rád bych touto cestou poděkoval následujícím osobám za jejich neocenitelnou pomoc při tvorbě této bakalářské práce. Bez níže jmenovaných by nebylo možné dokončit práci v tomto rozsahu, a proto děkuji:

PhDr. Janě Bílkové, PhD., za vedení práce, za její ochotu, trpělivost a velmi přínosné rady,

Mgr. Jiřímu Havigerovi, Ph.D., za nadstandardní pomoc při realizaci a pochopení metodologie výzkumu,

Prof. Dr. Hab. Jaroslavu Lipowskému za zadání práce, počáteční vedení a podněcení zájmu o problematiku semikomunikace.

Anotace

KYJÁNEK, Lukáš. *Semikomunikace v konceptu komunikace*. Hradec Králové, 2017. Bakalářská práce. Univerzita Hradec Králové. Přírodovědecká fakulta. Katedra českého jazyka a literatury. Vedoucí práce Jana Bílková. Odborný konzultant Jiří Haviger.

Bakalářská práce sleduje jazykový fenomén semikomunikace a její vymezení. Vedle tradičního konceptu komunikace přibližuje historii definování i funkční obsah relevantních komunikačních jevů – diglosie, bilingvismu, semikomunikace – a nachází mezi nimi elementární (ne)shody za účelem pevného zasazení semikomunikace do konceptu komunikace. Používány jsou především zahraniční zdroje, neboť semikomunikace byla primárně definována na vztazích skandinávských jazyků, opomíjeny nejsou však ani zdroje západoslovanské.

Obsahem práce je také vlastní matematicko-lingvistický kvantitativní výzkum, ve kterém je na západoslovanské jazyky – češtinu, slovenštinu a polštinu – vůbec poprvé aplikována metodologie pro zkoumání vzájemného porozumění blízce příbuzných jazyků, která byla využita při výzkumu vztahů mezi skandinávskými jazyky a která spočívá v měření podmíněné entropie mezi těmito jazyky.

Klíčová slova

*semikomunikace, bilingvismus, vzájemná srozumitelnost,
podmíněná entropie, Levenštejnova metrika*

Annotation

KYJANEK, Lukas. *Semiconmunication in the Communication Concept*. Hradec Kralove, 2017. Bachelor Degree Thesis. University of Hradec Kralove. Faculty of Science. Department of Czech Language and Literature. Thesis supervisor Jana Bilkova. Thesis consultant Jiri Haviger.

The bachelor thesis deals with semiconmunication as a language phenomenon. The thesis elucidates the traditional concept of communication, the history of defining and content of the following terms – diglossia, bilingualism, semiconmunication. It also finds basic differences in order to fit semiconmunication into the overall concept of communication. Since semiconmunication was originally defined through Scandinavian languages, mainly foreign sources are used. However, not even Western Slavic sources are overlooked.

The second part of the thesis consists of author's mathematical-linguistic quantitative research based on measurement of conditional entropy. The methodology used for examination of understanding of closely related languages (originally carried out on Scandinavian languages) has been applied for the first time on West Slavic languages – Czech, Slovak and Polish language.

Keywords

*semiconmunication, bilingualism, mutual intelligibility,
conditional entropy, Levenshtein distance*

Obsah

Anotace.....	v
Annotation	vi
Obsah	vii
1 Úvod.....	ix
2 Komunikace.....	11
2.1 Typy komunikace.....	11
2.2 Formy komunikace	12
3 Proces komunikace.....	13
3.1 Informace	14
3.2 Vysílatel	14
3.3 Kódování.....	15
3.4 Kód.....	16
3.5 Signál.....	17
3.6 Prostředí.....	17
3.7 Dekódování.....	17
3.8 Příjemce	18
3.9 Zpětná vazba.....	19
4 Hranice jazyků.....	20
4.1 Genetická příbuznost jazyků.....	21
4.2 Typologická příbuznost jazyků	21
4.3 Areálová příbuznost jazyků.....	21
4.4 Abstandsprache a ausbausprache	22
5 Diglosie	23
5.1 Diglosie v západoslovanských jazycích	24
6 Multilingvismus a bilingvismus.....	25
6.1 Osvojování a výuka jazyka	25
6.2 Jazykový transfer	26
6.3 Přepínání kódů.....	26

7	Semikomunikace	27
7.1	Semikomunikace a lingua franca.....	28
7.2	Příčiny a podmínky	28
7.3	Rozvíjení a podpora	29
7.4	Skandinávská semikomunikace.....	30
7.5	Západoslovanská semikomunikace.....	31
8	Semikomunikace v konceptu komunikace	34
9	Měření podmíněné entropie mezi západoslovanskými jazyky	36
9.1	Metody.....	37
9.2	Materiály.....	39
9.3	Limity	40
9.4	Výsledky	41
9.5	Závěry a diskuze.....	45
10	Závěr	49
	Seznam obrázků	51
	Literatura	52
	 Příloha A	 55
	Příloha B	56
	Příloha C	57

1 Úvod

V současném globalizovaném světě se nevyhneme komunikaci s mluvčími, kteří vůči nám ovládají odlišný mateřský jazyk, a to tím spíše, že je Česká republika součástí Evropské unie (EU). Ta v současné době eviduje 24 oficiálně zapsaných úředních jazyků. Čím snáze se mezi sebou občané EU dorozumí, tím větší budou nejen jejich studijní a pracovní možnosti, ale také míra jejich tolerance vůči ostatním národům. Jazyky jsou tak pro EU podstatným tématem a její instituce jim proto věnují velkou pozornost

Jazyková politika EU si v první řadě zakládá na vícejazyčnosti. Podporuje se výuka cizích jazyků i mobilita evropských občanů za hranice jejich vlastních států. Cílem této jazykové politiky je, aby každý evropský občan kromě svého mateřského jazyka ovládal minimálně další dva jazyky. K naplnění tohoto cíle provádí orgány EU postupné kroky již od roku 1996 (Rinder-Schjerve a Vetter 2007).

Na základě těchto postupných kroků a dalších prognóz jazykové situace na území EU přišel australský lingvista Michale Clyne (2003) se čtyřmi potenciálními scénáři vzájemné komunikace evropských občanů:

- (a) všichni se naučí jeden jazyk jako tzv. *lingua franca* (např. angličtinu),
- (b) každý se k mateřskému jazyku naučí několik dalších příbuzných jazyků,
- (c) každý si osvojí alespoň ty jazyky, které jsou používány v sousedních státech,
- (d) všichni spolu budou komunikovat pomocí svých mateřských jazyků.

Právě poslední možnost, ačkoli se na první pohled zdá těžko proveditelná, stojí ve středu zájmu této bakalářské práce. První zmínky o problematice semikomunikace, jak byla možnost (d) poprvé pojmenována, pochází již z 60. let minulého století, a přesto ani dnes není tento termín v lingvistickém světě pevně usazen.

Navzdory terminologickým nesnázím odborné výzkumy této oblasti komunikace probíhají, zvláště pak v posledních dvou desetiletích. Zajímavým výsledkem těchto výzkumů je mimo jiné závěr, že porozumění při komunikaci, kdy spolu hovoří dvě osoby dvěma různými mateřskými jazyky, může být asymetrické. Jedna z osob, která se komunikace účastní, tedy může rozumět druhé osobě více než tato druhá osoba osobě první.

V teoretické části práce se zaměříme na otázky terminologické. Semikomunikace je popsána především na materiálech blízce příbuzných skandinávských jazyků (dánštiny, norštiny a švédštiny), a proto vedle méně početné skupiny tuzemských prací využijeme především práce zahraniční. Na základě v nich prezentovaných teorií představíme terminologii a postavení semikomunikace jakožto komunikačního fenoménu, který se nemusí nutně týkat pouze skandinávských jazyků, ale také jiných jazyků, včetně západoslovanských (češtiny, slovenštiny a polštiny).

Výklady v teoretické části práce vychází z přístupů různých lingvistických směrů a oborů (strukturalismem počínaje, současnými přístupy moderní lingvistiky konče), aby bylo výsledné zakotvení semikomunikace v lingvistické terminologii co nejpevnější. Ve druhé části práce – vlastním výzkumu – využijeme především matematicko-lingvistických metod, přestože se v praxi užívají při zkoumání semikomunikace obvykle metody sociolingvistické. Sociolingvistických výzkumů bylo však již provedeno mnoho, a proto je náš výzkum matematicko-lingvistický.

Předmětem výzkumu prezentovaného v praktické části bakalářské práce (kapitola 9) jsou předpoklady k vzájemnému porozumění blízce příbuzných západoslovanských jazyků vůči sobě a míra případné asymetrie vzájemného porozumění komunikantů. Metodologický základ čerpáme v práci *CONDITIONAL ENTROPY MEASURES INTELLIGIBILITY AMONG RELATED LANGUAGES* (Moberg et al. 2007), kterou publikovali vedle dánské lingvistky Charlotte Gooskens – jež v současnosti působí na nizozemské univerzitě v Groningenu a již léta se zabývá skandinávskou jazykovou situací – také počítačový specialista Jens Moberg, John Nerbonne a Nathan Vallette. Zmíněnou práci publikovali na materiálu jmenovaných skandinávských jazyků už roku 2007, a přesto nebyla jejich metodologie dosud aplikována na západoslovanské jazyky.

Výzkum by měl mimo jiné objasnit, zda platí předpoklady, které naznačují mnohé tuzemské sociolingvistické výzkumy, totiž že Slováci rozumí češtině lépe, než rozumí Češi slovenštině (asymetrie vzájemného porozumění). Vedle tohoto ovšem bude oproti ostatním dosud realizovaným výzkumům zapracován i třetí západoslovanský jazyk, polština, aby mohla být zodpovězena nejen otázka ohledně asymetrie vzájemného porozumění ve vztahu češtiny-polštiny a slovenštiny-polštiny, nýbrž i otázka vzájemného porozumění jako takového, např. zda polštině porozumí více Češi, nebo Slováci atd.

Práce přináší moderní pohled na relativně utišenou oblast jazykovědného zájmu. Její výsledky by mohly povzbudit další lingvistické zkoumání vzájemného porozumění blízce příbuzných jazyků. Zvláště pak výsledky vlastního výzkumu by mohly být velmi cenné. Jejich využití spatřujeme v oblasti možné budoucí „výuky“ blízce příbuzných jazyků. Tímto ovšem již přecházíme od možnosti (d) zmiňovaného Clynova scénáře k možnosti (b).

2 Komunikace

Pojem komunikace je natolik komplexním a širokým, že nepřestává být ve středu zájmu mnoha disciplín lidského zkoumání. Přitom významy tohoto pojmu při jeho užití nevyjadřují nic většího než určitý typ spojení (Kraus 2005). Při hlubším zkoumání komunikace tak lingvistika mnohdy spolupracuje s ostatními společenskovědními i přírodovědnými disciplínami v tzv. *interdisciplinárních oborech*. Právě proto, že se každý z interdisciplinárních oborů, který se komunikací zabývá, snaží uchopit tento pojem vlastním způsobem, je velmi vhodné při obecném studiu komunikace zohledňovat při našem uvažování také jejich poznatky. Obecnou definici komunikace totiž respektuje většina z nich.

Komunikace ve smyslu dorozumívání „*je přenos informací mezi minimálně dvěma účastníky (...) prostřednictvím určitého signálního systému znaků (...) uskutečňující se vždy v určitém rámci.*“ (Čermák 2011, s. 13)

Přestože se tato práce věnuje pouze mezilidské komunikaci, přenos informací neprobíhá jen mezi lidmi. Jistou komunikaci realizují mezi sebou i členové živočišné říše, ba dokonce s nástupem moderních sofistikovaných informačních technologií i stroje. Zatímco lidem slouží k dorozumívání především přirozený jazyk, pro zvířata to mohou být pachy či jiné smyslové vjemy a pro stroje pak pouze dvoustavová logická hodnota představovaná (ne)výskytem elektrického napětí.

2.1 Typy komunikace

Způsoby, jakými může mezilidská komunikace probíhat, jsou nejčastěji rozlišovány dle dvou kritérií: počtu účastníků a zvoleného signálního systému znaků (*kódu*).

Dle počtu účastníků lze komunikaci dělit na *intrapersonální*, *interpersonální* a *masovou*. Intrapersonální komunikace představuje vnitřní komunikaci jednoho člověka sama se sebou a zabývá se jí především psychologie. Komunikace interpersonální již představuje (často obousměrnou) komunikaci mezi alespoň dvěma účastníky. Jednosměrnou komunikaci k většímu množství účastníků (např. média), nazýváme masovou komunikací (Vachek 1970, Vybíral 2009).

Dle zvoleného kódu lze komunikaci dále dělit na *vizuální a akustickou*, *neverbální a verbální*. Kódem pro vizuální a akustickou komunikaci mohou být dopravní značení či sirény záchranné služby. V některých případech bývají uváděny zvlášť jako dva typy: *typ symbolický* a *typ signální*. Kódem pro neverbální komunikaci jsou doteky, pohyby těla, mimika nebo například oční kontakt. Pro verbální komunikaci je pak kódem přirozený jazyk. Verbální komunikace je tak nejrozvinutější dorozumívací typ, neboť inventář přirozeného jazyka je nesrovnatelně bohatší než inventáře předešlých dvou typů (Černý 1998).

2.2 Formy komunikace

Oblastí našeho zájmu je především komunikace interpersonální verbální, a proto je vhodné na tomto místě zmínit též její možné formy. Ty vždy vyplývají z realizací kódu. V případě přirozeného jazyka mluvíme o *formě mluvené* a *formě psané*, přičemž obě formy jsou svými vlastnostmi podstatně odlišné (Čermák 2011).

Při realizaci jazyka mluvenou formou využíváme nejen jazyk jakožto kód samotný, ale též tzv. *parajazyk*, který představují jak prozodické prvky (např. intonace), tak do určité míry prvky neverbální komunikace (Čermák 2011). Při realizaci jazyka psanou formou místo parajazyku využíváme do jisté míry omezenější možnosti, jež nám nabízí *grafika* textu. Obojí, parajazyk i grafiku, nazýváme *komunikačními prostředky*. Pro komunikaci jsou doprovodné, tudíž mají na komunikaci nepřímý podíl.

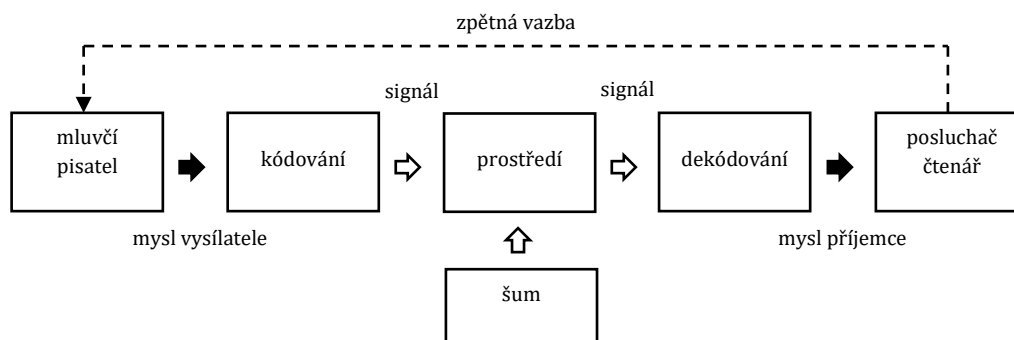
Je třeba si také uvědomit, že primární formou jazyka je forma mluvená (rozvíjená od počátku lidské existence), protože forma psaná v porovnání s mluvenou vznikla teprve nedávno (5 tisíc let zpět) (Černý 1998). Zatímco se mluvený jazyk neustále mění, psaná forma má sklony k ustrnutí (Saussure 1989), a proto není vhodné vést výzkumy více jazyků na podkladech písma. Psaný zápis jazyků se může podstatně lišit, i když mluvená forma těchto jazyků nemusí.

3 Proces komunikace

Obecné zkoumání komunikace, jak jej známe dnes, má své počátky v první polovině minulého století, kdy byly díky vzniku kybernetiky formulovány dvě teorie, *teorie komunikace* a *teorie informace*. Jejich matematické základy předložil Claude Elwood Shannon ve spolupráci s Warrenem Weaverem v práci *A MATHEMATICS THEORY OF COMMUNICATION*. Protože je předmětem obou zmíněných teorií přenos informace při komunikaci, staly se jejich poznatky zajímavými i pro ostatní vědní obory, mezi nimiž nechybí ani lingvistika.

Teorie komunikace přinesla obecný model procesu komunikace. Svým zaměřením hlavně na strukturní stránku modelu pro přenos informace, nikoli na povahu informace samotné, se její základnou staly především obory zkoumající podmínky a vlastnosti přenosu (matematika, biologie, psychologie)¹. Teorie informace oproti tomu přináší pohled na formální² reprezentaci informace, čímž se dostává do středu zájmu nejen matematiky, která je schopna například množství nebo hodnotu informace při přenosu měřit, ale také lingvistiky (Černý 1996).

Obecný model komunikace obsahuje následující objekty: *informace*, *vysílatel* (též *produktor*), *kódování* (též *produkce*), *kód*, *signál*, *prostředí* (též *kanál*), *šum*, *dekódování* (též *percepce*), *příjemce* (též *recipient*), *zpětná vazba* (Černý 1998). Konkrétní vztahy těchto objektů a lingvistickou terminologii modelu pro mluvenou i psanou formu interpersonální verbální komunikace zobrazuje obr. 1.



Obrázek 1: Model procesu komunikace mluvenou a psanou formou

Aby mohl vysílatel sdělit příjemci informaci, musí se nejdříve v mysli vysílatele informace zakódovat do určitého kódu. Tento kód se v podobě signálů přenáší skrze komunikační prostředí, v němž se k signálům přidává šum. Následně musí příjemcova

¹ Z lingvistických disciplín, které se alespoň částečně teorií komunikace zabývají, se hodí uvést například obecnou lingvistiku, neurolingvistiku či psycholingvistiku. Aplikovaně pak poznatky může využívat i rozvinout například lingvistika strojová.

² Pohled na čistě sémantickou reprezentaci informace se dosud (po půl století od vzniku obou teorií) nepodařilo důkladně prozkoumat žádnému vědnímu oboru.

mysl spojit signály opět v kód a ten dekodovat (Černý 1998). Pokud příjemce na získanou informaci dále reaguje, dochází ke zpětné vazbě, což obnáší oboustrannou výměnu rolí vysílatele s příjemcem (Čermák 2011).

Za úspěšnou lze považovat komunikaci tehdy, je-li po uskutečněním komunikačním procesu porozuměno funkci komunikace, jíž požadovala konkrétní situace (Čermák 2011). To závisí na příjemcově porozumění kódu. Kód se přenáší prostředím pomocí signálů. Čím více vyslaných signálů si příjemce složí v kód a dekoduje, tím větší je pravděpodobnost jeho porozumění původně zamýšlené funkci komunikace.

V následujících kapitolách za pomoci poznatků různých lingvistických oborů přiblížíme všechny objekty procesu komunikace podrobněji každý zvlášť.

3.1 Informace

Z pohledu kognitivní lingvistiky reprezentuje informace *mentální obraz* okolního prostředí uložený v paměti (Schwarzová 2009). K formulování informace a jejímu přenosu dochází v případě, že je zdroj této informace dostatečně motivován k uspokojení některé ze svých komunikačních potřeb (Nebeská 1992).

Při formulování informace je logicky zapotřebí každému mentálnímu obrazu přiřadit pro přenos symboly (*signály*), pomocí nichž by se v ideálním případě přenesl přesný mentální obraz z mysli vysílatele do mysli příjemce, což ovšem ve skutečnosti není možné z podstaty potřeby symbolu. Mentální obrazy (sémantika) nejsou přístupné empirickému zkoumání (Schwarzová 2009), ovšem přenos informace alespoň na rovině symbolů je předmětem teorie informace (Sedlačíková 2010).

Vhodné je též na tomto místě vyzvednout poznatek studie od H. P. Grice, který tvrdí, že účastníci komunikace (též *komunikanti*) údajně formulují informace tak, aby sdělovali vše, co je třeba, ale ne více, než je třeba (Nebeská 1992).

3.2 Vysílatel

Každý jeden komunikant by měl být schopný informace vysílat i přijímat (díky existenci zpětné vazby v modelu procesu komunikace), a proto již zde můžeme upozornit na skutečnost, že predispozice vysílatele i příjemce jsou totožné.

Přední biologickou predispozicí možnosti komunikace je kooperace obou mozkových hemisfér (zpracování jazyka) a paměť (úložiště mentálních obrazů) (Schwarzová 2009).

Při své velikosti a fungování se ustálilo základní dělení paměti na *paměť smyslovou, krátkodobou a dlouhodobou*. Z hlediska práce s jazykem se ale uplatňuje především paměť krátkodobá a dlouhodobá, protože smyslová paměť uchovává vjemy jen po velmi krátkou dobu, nanejvýš 0,5 sekundy (Čermák 2011).

Nutné na tomto místě je stručně upozornit na širokou oblast osvojování přirozeného jazyka. To se děje již od útlého věku. Právě tento proces je základem toho, jaký jazyk (jeho systémy, modely, pravidla atd.) lidé nabyli (Čermák 2011, Schwarzová 2009).

Lingvisté zabývající se osvojováním jazyka přišli mimo jiné se zajímavým poznatkem, že osvojení si potenciálního množství pravidel gramatiky trvá podstatně kratší dobu než osvojení si sémantiky (Schwarzová 2009). Člověk si totiž v průběhu života neustále upřesňuje jednotlivé sémantické rysy slov, což se v praxi projevuje tak, že po čas stárnutí dochází k rozdílu v používání jednotlivých výrazů.

Zaměříme-li se na osvojování formy přirozeného jazyka, zjistíme, že za výchozí a přirozený proces se považuje osvojení mluvené formy. Forma psaná se považuje až za odvozenou, neboť vyžaduje další intelektuální činnost (Nebeská 1992).

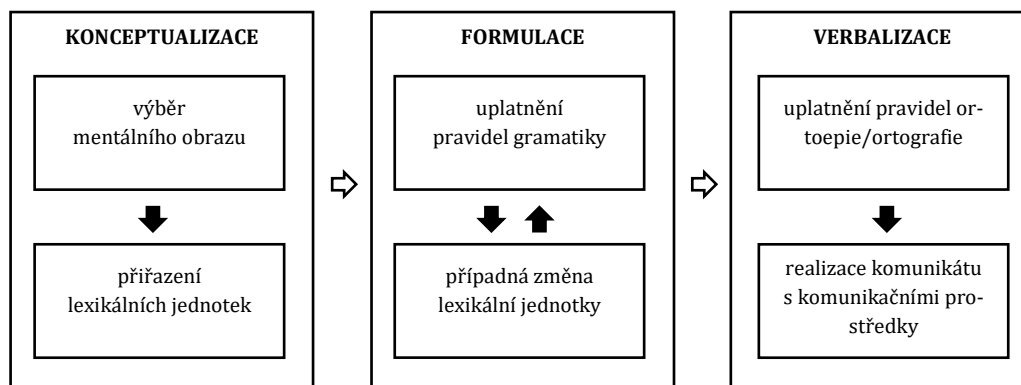
Takto zvládnutý jazyk slouží pro vysílatele ke kódování a pro příjemce k dekódování promluvy (též komunikátu). „Úspěšné kódování i dekódování je umožněno mj. znalostí společného jazykového systému.“ (Čermák 2011, s. 20)

3.3 Kódování

Kódování i dekódování lze chápat v rámci celého procesu komunikace jako samostatné podprocesy. Úkolem kódování je z paměti vysílatele převést mentální obrazy reprezentující jím vybrané informace, jež mají být formulovány, do příjemcovi srozumitelného (přijatelného) kódu (Čermák 2011). Kódování pak samozřejmě probíhá dle pravidel příslušného kódu.

Možnost přímého experimentálního studia kódování je jen ve velmi omezené míře, ale na podkladech dosavadních studií psycholingvistiky, neurolingvistiky a kognitivní lingvistiky se předpokládá, že má kódování tři základní fáze: *konceptualizaci*, *formulaci* a *verbalizaci* (Schwarzová 2009).

Při fázi konceptualizace se utváří informace, kterou vysílatel hodlá sdělit, z podoby mentálního obrazu do podoby jejího vhodného lexikálního zastoupení. Ve fázi formulace se na vybrané lexikální jednotky uplatňují nejdříve strukturní pravidla syntaxe, poté morfologie a ostatních souborů jazykových modelů, norem a konvencí. V této fázi může také dojít k úplné změně dříve zvolené lexikální jednotky. Fáze verbalizace představuje uplatnění pravidel fonetiky a fonologie, popřípadě grafiky (prakticky v podobě ortoepie, nebo ortografie) a následné pronesení vlastního komunikátu (Čermák 2011, Schwarzová 2009). Všechny fáze jsou vyobrazeny na obrázku 2.



Obrázek 2: Kódování jazyka vysílatelem

Výběr lexikálních jednotek determinují *situační podmínky* a *podmínky individuálně-specifické* (někdy též *mentální*). Situačními podmínkami se rozumí vnější faktory, které na vysílatele působí – může jít o společensko-historickou situaci, okamžité fyzické a psychické stavy komunikantů atd. Podmínky individuálně-specifické pak představují vnitřní faktory, které mají na vysílatele vliv – jedná se o jedincovy individuální vlastnosti, zkušenosti a schopnosti vnímat předměty i příjemce (Nebeská 1992).

Vlivem těchto podmínek vybírá vysílatel metodou selekce vhodnou lexikální jednotku, která by nejvíce odpovídala jeho mentálnímu obrazu. To provádí na základě výběru nejspecifičtějšího rysu zamýšleného předmětu. Je-li následně k dispozici více možných alternativ:

(1) auto, vozidlo, kára, dopravní prostředek, osobní automobil, ...,

volí vysílatel tu, jež se na základě jeho zkušeností jeví v dané situaci vůči příjemci adekvátní (Schwarzová 2009). Takto zvolenými jednotkami dokáže vysílatel vyjádřit svůj postoj k obsahu komunikátu i k příjemci (Nebeská 1992).

Na vzájemném vztahu vysílatele s příjemcem závisí ale také fáze verbalizace. Mathesius na základě dánských studií uvádí, že kupříkladu „*v rodině a vůbec v úzkých společnostech se artikuluje daleko nepřesněji a s menší silou výdechovou než v hovoru osob méně mezi sebou známých.*“ (Mathesius 1970, s. 30)

3.4 Kód

Nejrozvinutějším kódem verbální komunikace je přirozený jazyk, přičemž adjektivum „přirozený“ je velmi podstatné, neboť zvlášť rozlišujeme jazyky *přirozené*, *umělé* a *parajazyk*. Mezilidská komunikace využívá přirozený jazyk i parajazyk (Čermák 2011).

Parajazyk již byl představen v předešlých kapitolách. Přirozený jazyk lze chápat jako „*v mozku uložený systém jednotek, pravidel, modelů a konvenčních kolektivních norem k tvorbě promluv.*“ (Čermák 2011, s. 13) Z této definice je možné vyvodit, že jazyk se skládá ze tří propojených, vzájemně se ovlivňujících složek. „*Systém jednotek*“ zastupuje výrazivo jazyka (lexémy a jejich sémy), „*pravidla a modely*“ pak gramatiku jazyka (syntax, morfologii atd.) a „*konvenční kolektivní normy*“ zastupují normy ortografie, ortoepie a stylistiky (Mareš 2014).

Pohled na propojení zmíněných tří složek nabízí poznatky ženevského strukturalismu. Pro pohled na jednotky jazyka definoval strukturalismus *teorii jazykového znaku*, pro spojení znaků a gramatiky jazyka termín *langue* a pro propojení gramatiky s komunikačními prostředky jazyka termín *parole* (Saussure 1989).

3.5 Signál

Na základě kódu je informace kódována i dekódována, nicméně komunikačním prostředím je informace přenášena pomocí signálů. Za tyto signály lze považovat *znaky* jazyka. Kód slouží pouze k jejich dosazování – znaků za významy (kódování) a významů za znaky (dekódování). F. de Saussure (1989, s. 96) uvádí, že „*je lhostejné, jaká je povaha dohodnutého znaku.*“³

Znak zastupuje něco jiného (význam), obě skutečnosti se proto rozlišují. Podle Saussurovy dyadické teorie znaku zastupuje termín *signifié* označovaný význam, termín *signifiant* zastupuje označující znak (Čermák 2011, Černý 1996).

Přirozený jazyk odlišují od ostatních komunikačních systémů tzv. *figury*. L. Hjelmslev (1972) vychází z předpokladu, že promluvu lze dělit na vyšší a nižší jednotky (slova, morfémy, slabiky, fonémy, písmena), které nesou vlastní alespoň kontextový význam (znaky) a které takový význam nenesou (figury). To znamená, že výsledné signifiant se skládá z figur. Mezi figury řadíme slabiky, fonémy a písmena.

3.6 Prostředí

Neopomenutelnou složkou procesu komunikace je cesta, jíž se zakódovaná informace od vysílatele přenáší k příjemci. Právě tak lze definovat komunikační prostředí.

Důležitým faktorem prostředí, který ovlivňuje průběh komunikačního procesu, je šum (Nebeská 1992). Jedná se o vnější technologický rušivý signál narušující, či až znemožňující proces komunikace. Při působení šumu může dojít ke zkreslení informace. Praktickým příkladem může být hlučné prostředí, které podstatně oslabuje slyšitelnost mluvené komunikace.

Jako jistou formu šumu lze vnímat i tzv. *komunikační bariéru*. Ta již není představována technologickými překážkami, nýbrž vnitřními komunikačními překážkami, a to v podobě nedostatečné „připravenosti“ komunikantů (Kořenský 1992). Prakticky může jít o artikulační handicap vysílatele, příjemcovu neznalost kódu nebo například ne zkušenost v interpretování komunikačních prostředků či kontextu situace.

3.7 Dekódování

Dekódování je podproces komunikace ve své podstatě opačný kódování. Není pouhým pasivním přijímáním signálů, jedná se o proces od příjmu akustického nebo optického signálu po vytvoření a interpretaci mentálních obrazů.

Vytvoření takových mentálních obrazů nazýváme *inference* (též *vyvozování*, *usuzování*, *odvozování*). Ony mentální obrazy jsou mnohem bohatší nežli původní signální vstup (Schwarzová 2009). Jde o co největší snahu příjemce rekonstruovat komunikační záměr vysílatele (Nebeská 1992).

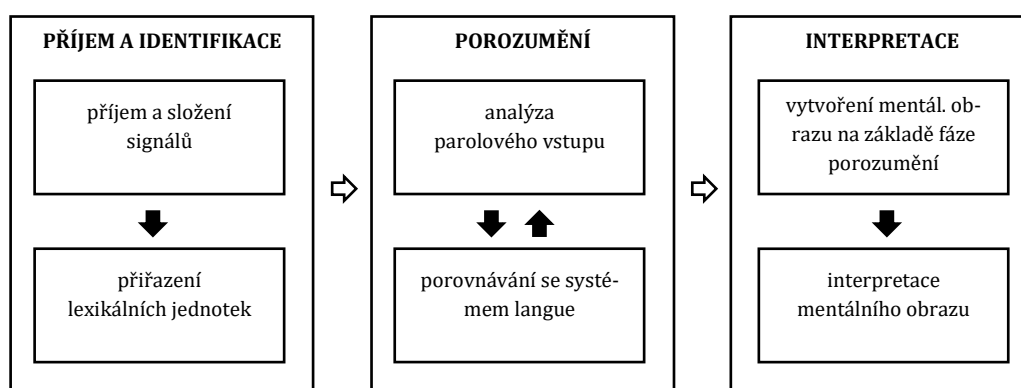
³ Současné kvantitativní studie ovšem naznačují, že *arbitrárnost*, jak bývá nemotivovanost formy vůči významu znaku nazývána, by nemusela dále být jednou ze základních vlastností znaků, neboť studie od Blasi D. E., Wichmann S., Hammarström H., Stadler P. F., Christiansen M. H.: *Sound-meaning association biases evidenced across thousands of languages* naznačuje, že napříč tisíci jazyky je možné nalézt ve slovech se stejným významem také stejné znaky.

Samotný proces dekódování lze zjednodušeně pochopit na pozadí tří fází: *příjmu signálů a identifikaci slov*, jejich *porozumění* a *interpretaci* mentálního obrazu, přičemž uplatňování jednotlivých složek těchto fází probíhá cyklicky (Čermák 2011).

V první fázi, příjmu signálů a identifikaci slov, příjemce přijímá akustické/optické vjemy (signály), z nichž následně identifikuje a složí potřebné lexikální jednotky. Současně s tím rozpozná také jejich významy (př. 2). Je prokázáno, že při nepřesném příjmu a identifikaci fonému/grafému, a tudíž váhání mezi více alternativami, vždy příjemce tento zvuk/obraz k některému fonému/písmenu zařadí (Čermák 2011).

(2) zámek vs. zámek (budova vs. zařízení k uzamykání)

Druhá a třetí fáze reprezentují nejdříve porozumění formě komunikátu, poté vytvoření a interpretaci mentálního obrazu. Dochází k celkové analýze srovnáváním parolového vstupu s příjemcovi známým systémem langue (s přihlédnutím k neverbálním složkám komunikace a situačnímu kontextu) a na podkladě této celkové analýzy si příjemce interpretuje dané sdělení i situaci (Čermák 2011). Tento obecný popis je vyobrazen na obrázku 3.



Obrázek 3: Dekódování jazyka příjemcem

Podíl na průběhu dekódování nese také příjemcem předem vytvořený předpoklad o smyslu, podobě a struktuře komunikátu, tzv. *expektace* (Nebeská 1992).

3.8 Příjemce

Pasivní účastník komunikace, příjemce, jak již bylo popsáno v kapitole o vysílateli, je determinován stejnými predispozicemi jako vysílatel.

Pro úspěšné uskutečnění komunikace se rozdíl mezi vysílatelem a příjemcem nachází v oblasti motivace. Vysílatel by měl být motivován k tomu, aby vytvořil pro příjemce co nejsrozumitelnější komunikát. Příjemce by naproti tomu měl být motivován k jeho dekódování, aby tak rekonstruoval záměr vysílatele. „Kvalita percepce závisí na tom, jak recipient je k vnímání konkrétního textu připraven, jakou pozornost mu věnuje, na kolik s produktem spolupracuje.“ (Nebeská 1992, s. 59)

3.9 Zpětná vazba

Aby neprobíhala komunikace pouze jednosměrně, považuje se za poslední člen celého procesu komunikace zpětná vazba. Představuje střídání pozice vysílatele a příjemce, čímž udržuje komunikaci v chodu.

Přestože se vysílatel vždy snaží předjímat postoje příjemce, aby tak dospěl naplnění vlastních komunikačních záměrů, příjemce nemusí vždy plně rozumět, správně vyhodnotit či interpretovat vysílatelem vyslané signály, a proto by příjemci měla být možnost zpětné vazby poskytnuta (Nebeská 1992).

4 Hranice jazyků

Při verbální komunikaci neexistuje pro všechny komunikanty na světě pouze jediný možný kód (jeden celosvětový přirozený jazyk), a tak může nastat situace, kdy se střetnou komunikanti, kteří ovládají vůči sobě rozdílné *mateřské jazyky*. Tyto různé kódy kódují též obsah různým způsobem (Černý 1998).

Rozdíly způsobů kódování odrážejí odlišnosti přírodního i kulturního prostředí, v němž se ony jazyky vyvíjely. *Sapir-Whorfova hypotéza* uvádí, že lidé, kteří mluví rozdílnými jazyky, jsou rozdílní i svou kognicí, což ovlivňuje jejich uvažování o světě. Tato skutečnost bývá označována jako tzv. *jazyková relativita* a příkladem je případ, kdy na základě důkladnějšího rozlišování sémantických rysů mají například některé asijské jazyky více možných výrazů pro rýži, než mají jazyky evropské. Přesto však lze mezi jazyky nalézt velké množství společných jevů, *jazykových univerzálií* (Sternberg 2002).

Pravděpodobnost, že se spolu setkají mluvčí s vůči sobě rozdílnými mateřskými jazyky, není malá, přestože celkový počet jazyků klesá. Počet přirozených jazyků sice není možné zjistit přesně, ale odhaduje se, že jich je mezi 3 až 7 tisíci. Snižování jejich počtu je v neprospěch jazyků s nízkým počtem mluvčích. Doposud je evidováno přibližně 2 500 takto ohrožených jazyků, přičemž v průběh 21. století by údajně mělo zaniknout až 85 % z nich (Čermák 2011). Evidováno je též 360 již vymřelých jazyků.

Každý jeden jazyk může mít navíc vnitřní *variety* (*standardní jazyk, interdialekty, dialekty, sociolekty* atp.), které mohou být vůči sobě i natolik rozdílné, že si jejich mluvčí nejsou schopni porozumět. Vyčíslování počtu jazyků se tak více ztěžuje, protože není snadné určit ostré hranice jednotlivých jazyků, natož pak nalézt hranice dvou dialektů nebo dialektu a jazyka (Saussure 1989). Významným rozlišovacím prvkem pro tyto případy je *vzájemná srozumitelnost* (Haugen 1966). Opírají se o ni různé sociolingvistické metody, které jsou v praxi užívány právě k rozlišování hranic mezi jednotlivými kódy.

Situaci nezpřehledňuje ani skutečnost, že při intenzivním kontaktu více jazyků mohou nastat dva scénáře. Buď se prosadí jeden jazyk jako *lingua franca*, dorozumívací jazyk, nebo vzniknou nové jazykové útvary zvané *sabir, pidžin* či *kreolština*, které kombinují různě vzdálené jazyky (Černý 1996).

Kritéria pro určení hranic mezi jazyky stále nejsou dostačující (Černý 1998), ustálilo se alespoň dělení jazyků dle dvou typů jejich příbuznosti: *genetické* (též *genealogické*) a *typologické*. Příbuznost některých jazyků lze označovat i tzv. *areálovou* příbuzností.

4.1 Genetická příbuznost jazyků

Genetická příbuznost ukazuje na společný historický původ jazyků (Hjelmslev 1971). Konvenčně předpokládá pro každou dvojici jazyků stejné východisko v podobě *prajazyku*, od něhož se dané jazyky odlišily *divergentním* vývojem (Mareš 2014, Vachek 1970). Takto příbuzné jazyky třídíme do společných *jazykových rodin*.

Díky genetické příbuznosti lze očekávat, že ve dvou jazycích nalezneme podobné výrazové prvky. Proto Hjelmslev (1971, s. 28–29) definuje genetickou příbuznost jako „*funkci mezi jazyky, která záleží v tom, že každý výrazový prvek v každém z jazyků má funkci k jednomu výrazovému prvku v každém z ostatních jazyků.*“

Kupříkladu čeština patří se slovenštinou, polštinou, lužickou srbštinou a kašubštinou mezi západoslovanské jazyky ve slovanské *jazykové větvi* indoevropské jazykové rodiny. Genetickou příbuznost jazyků a národů není vhodné ztotožňovat, přestože vliv národní politicko-historické situace na jazyk je neopomenutelný (Saussure 1989).

4.2 Typologická příbuznost jazyků

Typologická příbuznost jazyků si oproti genetické zakládá nikoli na společném původu jazyků, ale na jejich společných rysech, kategoriích. Proto Hjelmslev (1971, s. 88) definuje typologickou příbuznost jako „*funkci mezi jazyky, která záleží v tom, že kategorie v každém z jazyků mají funkci ke kategoriím v každém z ostatních jazyků.*“ V širším slova smyslu lze tuto příbuznost chápat jako nauku o jazykových univerzáliích (Sgall 2011).

Typologická klasifikace nerozděluje jazyky do jazykových rodin, nýbrž do pěti strukturalisty definovaných typů. V jedné jazykové rodině tak lze nalézt i jazyky rozdílných typů (Hjelmslev 1971). V. Skalička v rámci Pražského lingvistického kroužku definoval následující typy jazyků: *aglutinační*, *flektivní*, *analytický* (též *izolační*), *polysyntetický*, *introflektivní*. Na jejich podkladě charakterizujeme příbuznost morfologie, syntaxe i hláskosloví jednotlivých jazyků (Sgall 2011).

Důležité je uvědomit si, že jazykové typy jsou dynamické – příslušnost jazyků k jednotlivým typům se může měnit. Tím spíše, že žádný jazyk není pouze jedním čistým typem typologické klasifikace, přestože jeden typ u každého jazyka vždy výrazně převažuje (např. pro češtinu, slovenštinu i polštinu typ flektivní) (Černý 1996).

4.3 Areálová příbuznost jazyků

Pro některé jazyky bývá uvažována i možnost areálové příbuznosti. V konečném důsledku se jedná o rozšíření typologické příbuznosti o geografické umístění, kde se porovnávané jazyky používají. Jazyky, které jsou si areálově příbuzné, to znamená, že mají podstatné množství společných rysů (myšleno kategorií, nikoli podobných výrazových prvků) a vyskytují se v geografické blízkosti, tvoří *jazykové svazy* (Mareš 2014).

I čeština bývá zahrnována do jazykového svazu. Tento svaz s češtinou by měla tvořit též rakouská němčina, slovenština, maďarština, částečně slovinština a chorvatština, okrajově polština (Mareš 2014).

4.4 Abstandsprache a ausbausprache

V závěru kapitoly si ještě představme jeden sociolingvistický koncept sloužící ke kategorizaci autonomie jazyků. Základy tohoto konceptu, pomocí něž lze určit hranice mezi jazyky a dialekty, položil německý lingvista Heinz Kloss. Kloss (1967) definoval jako stěžejní termíny tohoto konceptu kategorie *abstandsprache* a *ausbausprache* (též *abstand* a *ausbau* jazyky).

Abstandsprache jsou jazyky, u nichž není v současnosti nutné zabývat se jejich autonomií detailně. Jde o funkčně nezávislé jazyky, u nichž se udály dostatečně velké změny na to, aby si nebyly vzájemně srozumitelné (Fishman 2008). Jako zástupce abstandsprache uvádí Kloss například němčinu a nizozemštinu (Kloss 1967).

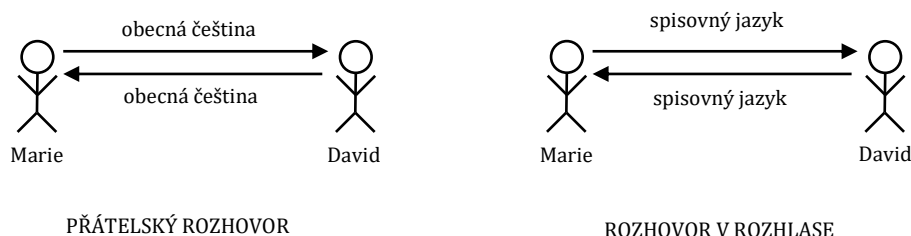
Ausbausprache jsou naopak jazyky, u nichž je detailnější zájem o jejich autonomii opodstatněný. Nezájem o tuto autonomii by totiž snadno mohl vést k podřízenosti (tohoto v současnosti již autonomního a standardizovaného) jazyka pod záminkou, že jde o dialekt jiného jazyka. Ausbau jazyky jsou totiž s některými jinými jazyky vzájemně srozumitelné, ale zároveň jsou již natolik rozvinuté, že je nelze považovat za dialekty jiného jazyka (Fishman 2008). Jako zástupce ausbausprache uvádí Kloss (1967) například češtinu a slovenštinu.

Kloss pracuje navíc s kategorií tzv. *polycentrických jazyků*. Příkladem je srbochorvatština. Sice by mohla být vnímána jako ausbausprache, neboť v současnosti vystupuje coby dva vzájemně srozumitelné jazyky: chorvatština a srbština, jenže k jejímu rozdělení došlo vlivem silného politického úsilí, nikoli přirozeným vývojem, a proto ji nelze označit přímo ausbausprache (Kloss 1967).

5 Diglosie

Komunikační jev, který bývá nazýván *diglosií*, spadá především do studií sociolingvistiky a týká se vztahů vnitřních variet jednoho jazyka. Ty se od sebe mohou lišit jak po stránce fonetické, tak lexikální či gramatické (Černý 1996). Je pro ně typická absence psané formy (Chambers a Trudgill 1998). Přijdou-li navíc jednotlivé variety do dlouhodobějšího vzájemného kontaktu, ovlivňují se (Trudgill 1986).

Diglosie je „*poměrně stabilní jazyková situace, v níž vedle dialektů existuje odlišná, vysoce kodifikovaná varianta, která je široce využívána v literatuře a psané podobě jazyka (...), avšak nikdo ji nepoužívá v běžné konverzaci.*“ (Černý 1996, s. 397)



Obrázek 4: Příklad diglosie – komunikanti volí varietu jazyka dle komunikační situace

V praxi se pro účely popisu diglosie rozdělují variety jazyka na *vysoké variety* a *nízké variety*. Vysokou varietou je ona kodifikovaná varianta s psanou podobou, zatímco nízkou varietou se rozumí například dialekt. Na základě tohoto rozdělení nalézáme komunikační situace, při nichž lze použít buď pouze vysokou varietu, nebo pouze nízkou varietu, ale jen minimum situací, kdy by na volbě variety nezáleželo, a přitom nebyl mluvčí zdrojem posměchu (Ferguson 1959).

Volba vhodné variety k nastalé komunikační situaci se v mysli mluvčího děje často nevědomě, spontánně a nazýváme ji *střídáním kódů* (Čermák 2011, Bermel 2010).

Osvojování jazyka probíhá pomocí nízkých variet. S vysokou varietou se dítě setkává až ve vzdělávacích institucích, což je též důvodem toho, že si „*lidé nikdy neosvojí vysokou varietu tak dokonale jako nízkou,*“ (Ferguson 1959, s. 105) a onu vysokou varietu pak vnímají spíše jako soubor pravidel a norem. Vysoká varieta je ale vnímána prestižněji než nízké variety, neboť za ní stojí literární dědictví. Gramatika a morfologie nízkých variet bývá jednodušší, slovní zásoba vysoké i nízké variety je zpravidla téměř totožná (mohou se lišit významy slov). Za zmínku stojí, že nízké variety mají mnohdy výrazy pro regionální předměty a skutečnosti, pro které varieta vysoká žádné pojmenování nemá (Ferguson 1959).

Diglosie bývá označována termínem *sociální bilingvismus* či pouze *bilingvismus* (Černý 1996). To je ale přijatelné pouze při vnějším pohledu na společenství jako na celek. „*Bilingvismus a diglosie jsou jednoduše dva jevy odlišné povahy,*“ (Kloss 1976, s. 136) a tak

by diglosie neměla být pod bilingvismus vůbec řazena, neboť bilingvismus je (v Sausurovském pojetí) v oblasti langue, zatímco diglosie v oblasti parole (Kloss 1976).

Takovému náhledu na rozlišování diglosie od bilingvismu nasvědčují i možnosti různých kombinování těchto jevů. Hovoříme o situacích: *diglosie s bilingvismem*, *diglosie bez bilingvismu*, *bilingvismus bez diglosii*, *ani bilingvismus a ani diglosie* (Fishman 1967). Kupříkladu diglosii s bilingvismem si můžeme přiblížit na situaci Paraguaye a Bolívie, kde sice používají vlastní jazyky, nicméně jazykem úředního styku v těchto zemích je výhradně španělština (Černý 1996).

5.1 Diglosie v západoslovanských jazycích

Situace klasifikace diglosie je v češtině, slovenštině i polštině podobná. Pro mluvčího je primárním jazykem některá z nízkých variet, ať už dialekt či interdialekt (čeština, slovenština i polština má několik různých dialektů, čeština též interdialekt v podobě obecné češtiny). Vysokou varietu (zpravidla spisovné varianty češtiny, slovenštiny a polštiny) se mluvčí učí až ve škole (Čermák 1997). Touto optikou bychom mohli považovat diglosii uvnitř všech tří jmenovaných západoslovanských jazyků za funkční, přesto se ale například mezi českými lingvisty vedou v této oblasti spory.

Část českých lingvistů totiž poukazuje na to, že se spisovná čeština vlivem kodifikací sbližuje s obecnou češtinou, čímž dochází nikoli ke střídání kódů, ale k „míšení kódů“ (Bermel 2010). Na pozadí zmíněného tuto lingvisté tvrdí, že již „o skutečnou diglosii v češtině nejde, ale některé rysy naší jazykové situace jsou diglosii blízké.“ (Hronek a Sgall 2014)

Část lingvistů v čele s Františkem Danešem a Františkem Čermákem se kloní k funkční diglosii, i když méně stabilní (Bermel 2010, Čermák 1997), a míšení kódů ospravedlňuje tím, že pisatel nemá k dispozici volbu mezi varietami, zatímco mluvčí ano, kvůli čemuž dochází ke kolísání mezi spisovnou a obecnou češtinou (Hernandezová a Suchá 2014).

6 Multilingvismus a bilingvismus

Zatímco diglosii chápeme jako vnitrojazykovou záležitost (kontakty variet se odehrávají uvnitř jednoho jazykového společenství), tato kapitola nás posouvá do oblasti komunikačního jevu, který je záležitostí kontaktu více jazyků. Ať už se shodneme, či neshodneme na tom, kde přesně leží hranice mezi jednotlivými jazyky, neoddiskutovatelnou skutečností zůstává, že tento v praxi častý kontakt komunikantů s vůči sobě rozdílnými mateřskými jazyky je klíčový pro mnohé oblasti lidské existence i lidského poznání vůbec (např. vzdělávání, kultura, technika, politika, ekonomika, ...) (Wei 2000).

Představovaný jazykový jev je označován termínem *multilingvismus* (vícejazyčnost), nebo též termínem *bilingvismus* (dvojazyčnost). Oba tyto termíny⁴ naplňují totožné obsahy kvalitativně, nikoli kvantitativně. Definován je bilingvismus/multilingvismus coby „koexistence dvou/více vybudovaných jazykových systémů.“ (Čermák 2011, s. 56)



Obrázek 5: Příklad bilingvismu – Marek ovládá a používá češtinu i němčinu

Bilingvismus může mít u každého jedince různé formy a stupně, proto níže předkládáme základní používané a pro oblast zájmu této práce relevantní označení a třídění těchto forem a stupňů (Wei 2000):

- (3) maximální bilingvismus = jedinec ovládá a používá oba jazyky na úrovni mateřských jazyků,
- (4) dominantní bilingvismus = jedinec ovládá a používá jeden jazyk častěji než druhý,
- (5) funkční bilingvismus = jedinec ovládá a používá oba jazyky, ale ne ve všech sférách,
- (6) semibilingvismus = jedinec ovládá oba jazyky nedostatečně,
- (7) asymetrický/pasivní/receptivní bilingvismus = jedinec rozumí, ale není schopen používat,
- (8) symetrický/vyvážený = jedinec ovládá a používá oba dva jazyky na stejné úrovni⁵.

6.1 Osvojování a výuka jazyka

Abychom považovali jedince za bilingvního, tzv. *bilingvistu*, musí si systémy obou jazyků vybudovat. Jak jsou u bilingvistů reprezentovány systémy dvou jazyků, na to psycholingvistika ani neurolingvistika doposud nedaly přesnou odpověď (Wei 2000).

⁴ Přestože níže již bude využíván pouze termín bilingvismus (po vzoru všech ostatních odborných publikací), představené poznatky a závěry platí stejně tak i pro multilingvismus.

⁵ Na tomto místě je vhodné upozornit na nesoulad s přístupem Čermáka (2011). Ten tvrdí, že jedinec nikdy nezná oba jazyky na stejné úrovni.

Uvažuje se hypotéza, že jsou oba jazyky reprezentovány jedním společným systémem, ale také hypotéza, že každý jazyk je reprezentován zvlášť, a to do jisté míry odlišně. „Závěr je takový, že si bilingvisté osobitě ukládají v mozku každý ze dvou jazyků, a navíc mají i všeobecný pojmový ‚sklad‘. Tyto tři odlišné ‚sklady‘ propojují silné, přímé spojovací kanály, jejichž síla závisí na úrovni ovládnutí jazyků.“ (Wei 2000, s. 31)

U bilingvistů se nezdá, že bychom se současným osvojováním obou jazyků již v dětském věku, za bilingvisty však můžeme označovat i jedince, kteří se druhý jazyk naučí až později.

Principy, pomocí nichž dochází k osvojování i výuce jednotlivých jazykových jevů popsal už R. Jakobson: *princip maximálního kontrastu, postup od jednoduchého ke složitějšímu, postup od bezpříznakového k příznakovému* (Černý 1996). Čím více se od sebe jednotlivé jazyky liší, tím složitěji jejich osvojování/učení probíhá, a zároveň čím více jazyků již známe, tím snáze se nám učí další (Weinreich 1953).

6.2 Jazykový transfer

Při současném osvojování dvou jazyků i výuce druhému jazyku v pozdějším věku (ale hlavně při jeho používání) je třeba mít na paměti vlivy již známého mateřského jazyka, což nazýváme *jazykovým transferem*. Uvažovány bývají jazykové transfery: *pozitivní, negativní a nulové* (Wode 1993).

Pozitivní jazykový transfer existuje především díky společným jevům dvou jazyků. Je velmi nápomocný při osvojování i výuce druhého jazyka. Jedná-li se o blízce příbuzný jazyk k jazyku již známému, je tento transfer navíc silným motivačním faktorem, protože již v začátcích výuky díky němu dochází k rychlým pokrokům. (Krystýnková 1995)

Negativní jazykový transfer, *interference*, je naopak „*odchylka od norem kteréhokoli z jazyků používaných v řeči bilingvistů v důsledku jejich znalosti více jak jednoho jazyka*.“ (Lehisteová 1988, s. 82) Odchyly, které vlivem interference vznikají, nazýváme *interferémy* a mohou se vyskytovat na všech rovinách jazyka. Interferémy mohou působit pro příjemce velmi rušivě a v případě nevhodné záměny kupříkladu fonémů mohou takové interferémy vést až k nesrozumitelnosti (Černý 1996, Hjelmslev 1972). Výsledný jazyk s těmito interferémy pak lze považovat za pomyslný „třetí jazyk“, tzv. *interlangue* (Krystýnková 1995, Lehisteová 1988).

Nulový jazykový transfer nepůsobí ani pozitivně, ani negativně. Do nedávna nebyl vůbec uvažován (Wode 1993).

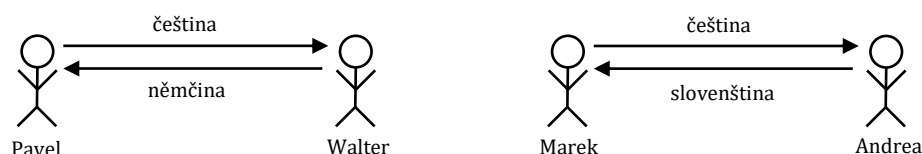
6.3 Přepínání kódů

Přepínání kódů bylo již zmíněno u diglosie. Dochází k němu ale i u bilingvistů. Přepínání kódů bilingvistům dopomáhá se vyjádřit přesně. Často k němu totiž dochází ve chvílích, kdy si bilingvista nedokáže v danou chvíli vzpomenout na právě potřebné slovo (Wei 2000).

Přepínání kódů je běžný jev a může mít mnoho podob, lze se setkat i s případy, kdy gramatiku udává jeden jazyk a jednotlivé lexémy pak jazyk druhý. V praxi je tudíž přepínání kódů často doprovázeno interferencí (Lehisteová 1988, Wei 2000).

7 Semikomunikace

Semikomunikace byla popsána na základě jazykové situace ve Skandinávii. Dánští, norští a švédští obyvatelé mnohem raději při společné komunikaci používali své mateřské jazyky, namísto lingua franca. Tento stav nazval poprvé E. Haugen v roce 1966 semikomunikací (Schüppert a Gooskens 2011).



Obrázek 6: Příklad semikomunikace – každý komunikant mluví svým mateřským jazykem; v některých případech lze předpokládat porozumění horší (Pavel a Walter), v jiných lepší (Marek a Andrea)

Odbornou lingvistickou obcí byl ale tento termín považován za zavádějící „protože naznačuje, že komunikanti budou rozumět zhruba polovině toho, co bylo řečeno, což zjevně není (obvykle) tento případ,“ (Braunmüller 2002, s. 3) a snad proto, protože jsou skandinávské jazyky pro mezinárodní komunikaci méně důležité, nebyla semikomunikace ani jako jazyková zvláštnost daného prostředí široce uznána (Braunmüller 2002).

K. Braunmüller (2002) také uvádí, že vlivem zmíněné vágnosti termínu došlo k různým metaforickým interpretacím, popřípadě neporozuměním tomuto komunikačnímu fenoménu. Příkladem uvádí pojetí semikomunikace V. Budovičovou.

Haugenem popsaná jazyková situace byla zařazena do kategorie *receptivního bilingvismu* (též *pasivní* či *asymetrický bilingvismus*)⁶, jak jej definoval Ch. F. Hockett roku 1958 (Schüppert a Gooskens 2011), a termín semikomunikace se lingvisté snažili předefinovat. Semikomunikace měla od té doby vyjadřovat divergenci mezi dvěma kódy, jakýsi druh kódového šumu (Braunmüller 2002).

Ve své podstatě tomuto předefinování odpovídá právě pojetí semikomunikace dle Budovičové (1987). Ta semikomunikaci chápe jako doprovodný jev každé komunikace. Semikomunikace má být pouhým kódovým šumem (pro příjemcovu porozumění defektní zacházení s prvky jednotlivých jazykových rovin).

Lingvisté, kteří byli obeznámeni se skandinávskou jazykovou situací, si ovšem brzy začali uvědomovat, že, aby si tito komunikanti rozuměli, nemusí se nutně ostatním dvěma jazykům zmíněné trojice učit, což je podmínkou bilingvismu, nýbrž jim stačí

⁶ Případně samozřejmě v závislosti na počtu současně takto uvažovaných jazyků také *receptivní/pasivní/asymetrický multilingvismus*.

znalost vlastního jazyka. Kontaktem s ostatními blízce příbuznými jazyky se komuni-
kanti „pouze“ zdokonalují, čímž tato situace do receptivního bilingvismu nespadá
zcela (Braunmüller 2002).

Předefinováním termínu semikomunikace, zdá se, nepostupovala široká lingvistická
obec nesprávně, ale z opačného konce. V posledním dvacetiletí vzniklo (výhradně na
materiálech dánštiny, norštiny a švédštiny) mnoho výzkumů i teorií, které pojetí se-
mikomunikace tak, jak jej popsal Haugen, připouští, podporují, studují a dále rozvíjí⁷.
Semikomunikace již není chápána toliko jako doprovodný jev komunikace, nýbrž jako
její speciální způsob dosud řazený k receptivnímu bilingvismu (Börestamuhlmann 2002).

7.1 Semikomunikace a lingua franca

Chápejme tedy semikomunikaci dle Haugenovy definice. Haugen si všiml, že „*komuni-
kace nevyžaduje u komunikantů identický kód*.“ (Haugen 1966, s. 479) Označil proto
semikomunikaci za komunikační jev, při němž se spolu komunikanti dorozumívají
každý pomocí svého mateřského jazyku, a přesto si rozumí, i když z podstaty věci vůči
sobě používané jazyky obsahují jistý šum⁸ (Haugen 1966).

Výzkumy ve Skandinávii prokázaly, že se v praxi projevuje semikomunikace také tím,
že mluvčí těchto jazyků při komunikaci mezi sebou častěji a raději sáhnou právě ke
svým mateřským jazykům než k lingua franca (Börestamuhlmann 2002). Lingua franca se
v tomto ohledu jevila pouze jako uspokojující řešení pro případy, kdy bylo pro komu-
nikanty příliš složité adaptovat se na jazyk druhého.

Používání lingua franca místo semikomunikace bylo v některých případech i vhodné.
To však pouze pro jednoduché záležitosti, např. turistiku. Pro složitější komunikační
úlohy lingua franca nebývala mezi komunikanty blízce příbuzných jazyků tak efek-
tivní jako semikomunikace (Braunmüller 2007). Snad kvůli tomu, že jedincova slovní
zásoba pro mateřský jazyk bývá zpravidla větší než jeho slovní zásoba pro jazyk cizí.

7.2 Příčiny a podmínky

Semikomunikace není funkční pro všechny komunikanty a všechny jazyky. Již Haugen
(1966) si při svém popisu semikomunikace uvědomoval, že jsou dvě základní příčiny
existence tohoto fenoménu: *vůle komunikantů a blízkost jazyků*. Sám upozorňuje na
to, že jsou komunikanti schopni překonat i velké rozdíly mezi jazyky jen díky své vůli.

Gooskens (2007, s. 446) v několika svých dlouholetých výzkumech Haugenem jmeno-
vané příčiny potvrzuje a přidává k nim ještě jednu, která semikomunikaci přibližuje
k bilingvismu, neboť tvrdí, že „*porozumění blízce příbuzných jazyků závisí především
na třech faktorech: (a) adresátově postoji k jazyku, (b) adresátově kontaktu s jazykem
a jeho dalších jazykových zkušenostech, (c) jazykové vzdálenosti jazyka mluvčího a ja-
zyka adresáta*.“

⁷ Nutno podotknout, že lingvisté užívají termínu *semikomunikace* jen zřídka a raději využívají termín
vzájemná srozumitelnost, popřípadě neprovádí „výzkumy semikomunikace“, nýbrž „výzkumy vzájemně
srozumitelných jazyků“.

⁸ Právě tento šum by byl podle Budovičové (1987) semikomunikací.

Snažíme-li se nalézt jazyky, u nichž by mohla být semikomunikace funkční, využíváme k určení blízkosti jazyků příbuznost genetickou, typologickou, případně areálovou. Je-li však nejsou kritéria k určování příbuzností stále přesně definována, často byly slyšet hlasy, že u takto blízkce srozumitelných semikomunikačních jazyků nejde toliko o autonomní jazyky, nýbrž o dialekty staršího jazyka, ať už existujícího, či nikoli. Ujasnit tuto problematiku dopomohl koncept představený Klosssem (viz kapitola 4.4) v díle *ABSTANDSPRACHEN UND AUSBAUSPRACHEN*, který semikomunikační jazyky, nezávisle na Haugenovi, kategorizuje jako *ausbausprache* s vlastními dialekty, ne za pouhé dialekty (Kloss 1967).

Semikomunikační potenciál je tak uvažován pro řadu jazyků, např.: afrikánština-fríština-nizozemština, bulharština-makedonština, bulharština-ruština, čeština-polština-slovenština, dánština-norština-švédština, faerština-islandština, hindština-urdština, chorvatština-srbština, italština-španělština, ukrajinština-ruština, ... (Haugen 1966, Stopyra 2010, Zeevaert 2007) Tento seznam je ovšem sestaven pouze na základě kritéria blízkosti jazyků, a to pouze obecně teoreticky definované, proto lze hovořit pouze o potenciálu. Některé jazyky k sobě mají navíc samozřejmě blíže (čeština-slovenština) než jiné (čeština-polština, italština-španělština).

Gooskens ve svém výzkumu (2007) také dokazuje na příkladu afrikánštiny-fríštiny-nizozemštiny, že ani teoreticky definovaná blízkost jazyků nemusí nutně znamenat vzájemnou srozumitelnost (respektive funkční semikomunikaci) a že ani vlivy zbylých faktorů, vůle a zkušenosti komunikantů, nejsou pro funkční semikomunikaci zanedbatelné (Gooskens 2007, Haugen 1966, Zeevaert a Ten Thije 2007). Proto, dokud nedojde ke konkrétním výzkumům těchto faktorů pro potenciální semikomunikační jazyky, nelze všem takovým blízkce příbuzným jazykům předpokládat pevně danou funkční semikomunikaci (Gooskens 2007).

7.3 Rozvíjení a podpora

Podstatné blízkosti konkrétně slovanských jazyků si povšimli již pražští strukturalisté, a tak do svých *TEZÍ* neopomenuli zařadit část o vhodnosti a způsobech učení slovanským jazykům na slovanských školách. Takové vyučování (včetně učebnic) by se mělo zaměřovat spíše na jevy, v nichž se jednotlivé jazykové systémy liší než na ty, které jsou jim společné (Vachek 1970).

Rozvíjení schopnosti dekodovat promluvy blízkce příbuzných jazyků, tzv. *interkomprehence*, podporuje faktor (b), jímž je podle Gooskens semikomunikace podmíněna. Tím se potenciál semikomunikace u těchto jazyků zvyšuje, ale zároveň přibližuje tento fenomén kategorii bilingvismu.

Různých zavedených metod interkomprehence je několik⁹, a tak by jejich zpracování vydalo na samostatnou práci.

⁹ Zmíňme alespoň názvy těch nejznámějších: *Galatea*, *EuRom4*, *EuRom5*, *EuroComRom*.

7.4 Skandinávská semikomunikace

Většina konceptu semikomunikace je popsána právě na materiálech skandinávského jazykového prostředí (dánština-norština-švédština). Zmíněné jazyky tak naplňují vše výše jmenované, a proto si v této kapitole uvedme již pouze některá zajímavá individuální zjištění skandinávského jazykového prostředí.

V prvé řadě bychom si měli uvědomit, že vliv na podobu jazyka má jeho vývoj, který je mimo jiné závislý na společensko-kulturní situaci. V případě vztahu severských zemí neopomeňme zmínit skutečnost, že po dobu jejich historického vývoje byly jejich osudy (a zejména vláda nad těmito krajinami) silně spjaty. Na sto let se dostalo v 16. století Norsko pod dánskou nadvládu, což výrazně poznamenalo dánštinu i norštinu – Norové se stali diglosní a bilingvní (vysokou varietou byla dánština, nízkou norština) (Braunmüller 2002). Od konce napoleonských válek do počátku 20. století pak bylo Norsko pod nadvládou Švédů.

Ostatně již od 19. století bylo vzájemné porozumění těchto jazyků chápáno jako podpora pan-skandinavismu. To vedlo ke kulturnímu a politickému sbližování těchto zemí, až vznikl Nordic Council (Zeevaert a Ten Thije 2007). Díky této instituci se sbližování severských zemí stává důležitým politicko-kulturním cílem (Braunmüller 2007), který je ovšem založen na dané jazykové situaci, již jmenovaná instituce podporuje a zařazuje do *jazykového plánování* (též *jazykové politiky*) (Börestamuhlmann 2002).

Příkladem praktického zásahu do jazykové politiky skandinávských zemí Nordic Councilem můžeme uvést například skutečnost, že každé skandinávské dítě se setkává po dobu své školní docházky nejen s angličtinou (jako lingua franca), ale též s ostatními skandinávskými jazyky. Existují i organizace, které se starají o zkoumání a publikování prací o skandinávské semikomunikaci (Börestamuhlmann 2002).

Zaměříme-li se na konkrétní výsledky proběhnuvších výzkumů semikomunikace v této geografické oblasti, zjišťujeme, že vůle rozumět blízkce příbuzným jazykům je mezi Skandinávci poměrně silná. Už děti si často myslí, že jazyk, který není jejich mateřský, je „hezčí“, a tak mají již v útlém věku větší motivaci porozumět a částečně si takový semikomunikační jazyk osvojit (Schüppert a Gooskens 2011). Mluvčí těchto tří jazyků navíc mezi svými zeměmi často cestují. Státní hranice skandinávských zemí jsou proto spíš než bariérami mezi lidmi styčnou plochou pro jejich kontakt (Börestamuhlmann 2002, Braunmüller 2002, Haugen 1966).

Jak bylo již popsáno – dánština, norština i švédština jsou blízce příbuzné jazyky geneticky (severogermánské jazyky germánské jazykové větve indoevropské jazykové rodiny), typologicky (analytické jazyky) i areálově. Označují se také jako *ausbausprache* (Braunmüller 2007).

Dánština a norština mají velkou společnou slovní zásobu, ale ostré rozdíly ve výslovnosti. Naproti tomu norština se švédštinou mají výslovnost podobnou, ale rozdíly mají ve slovní zásobě. Dánština se švédštinou pak mají vůči sobě rozdílnou jak výslovnost, tak slovní zásobu (Haugen 1966).

Zajímavou skutečností je, že vzájemné porozumění při semikomunikaci není symetrické. Opětovné sociolingvistické výzkumy potvrdily kupříkladu u dospělých Dánů

lepší porozumění mluvené švédštině než u dospělých Švédů mluvené dánštině (Schüp-pert a Gooskens 2011). Asymetrii ve vzájemném porozumění zaznamenal již Haugen (1966) a na základě vlastního sociolingvistického výzkumu vyčíslil asymetrii mezi skandinávskými jazyky takto:

- (9) 50 % Norů přiznává porozumění dánštině,
- (10) 76 % Dánů přiznává porozumění norštině,
- (11) 40 % Dánů přiznává porozumění švédštině,
- (12) 87 % Švédů přiznává porozumění dánštině,
- (13) porozumění Švédů norštině a Norů švédštině nebylo výrazně rozdílné.

Semikomunikaci i asymetrii u jmenovaných skandinávských jazyků analyzovali lingvisté několikrát i za pomoci matematického aparátu. Využívána bývala výhradně tzv. *Levenštejnova metrika*¹⁰ (Gooskens 2007).

7.5 Západoslovanská semikomunikace

Úzkou blízkost vůbec všech slovanských jazyků měli na paměti již strukturalisté Pražského lingvistického kroužku. Upozorňovali, že spolu slovanské jazyky geograficky vesměs sousedí a že rozdíly mezi těmito sousedními slovanskými jazyky jsou často menší než rozdíly mezi dialekty některých jiných jazyků (Vachek 1970). S touto úzkou blízkostí jazyků se nese i vzájemná srozumitelnost. Ostatně Haugen (1966) ve své práci, v níž představil semikomunikaci, zařadil mezi jazyky se semikomunikačním potenciálem nemálo slovanských jazyků včetně češtiny a slovenštiny, a tak se blíže zaměříme právě na tyto dva západoslovanské jazyky. K nim, na rady slovenské jazykovědkyně M. Nábělkové (2008, s. 17): „*implicitně není bezvýznamné mít při česko-slovenských jazykových úvahách polštinu jako porovnávací pozadí*“, přidejme i třetí západoslovanský jazyk – polštinu. Lužickou srbštinu ani kašubštinu, přestože jsou také západoslovanskými jazyky, pro nedostatek materiálů v této oblasti neuvažujeme, přestože by jistě šlo o zajímavé rozšíření tohoto tématu.

Jazykové kontakty češtiny a slovenštiny trvaly historicky poměrně dlouhou dobu a znamenaly jak konvergentní, tak divergentní období vývoje, čímž tyto jazyky získaly možnost průniku (*kontaminace, pozitivní/negativní transfer, fluktuace*) různých jazykových jevů (Budovičová 1974).

V současnosti jsou kontakty češtiny a slovenštiny oproti dobám minulým oslabeny. Do rozpadu Československa roku 1993 byly jazykové kontakty češtiny a slovenštiny podporovány jak ve školském, tak pracovním i kulturním prostředí jazykovým plánováním (Musilová 2000). Dnes již nejsou poznatky o slovenštině součástí českého školního vzdělávacího programu. Slovenština mizí, i když ne zcela, z veřejného života ČR a Češi se musí spokojit s pouhými nahodilými kontakty se slovenštinou – slovenské deníky se na území ČR již běžně neprodávají, televizní kanály ve slovenštině nejsou běžně přístupné, dřívější dvojjazyčné vysílání také zaniklo atd. (Mitter 2007)

Obavy o vzdalování se češtiny a slovenštiny, jež naznačují, že by se slovenština mohla pro obyvatele ČR stát cizím jazykem, se týkají především mladé generace, která již

¹⁰ Zpracována teoreticky i prakticky je tato metrika v kapitole 9, a proto zde není blíže pojednána.

nezažila společný stát Čechů a Slováků. Na druhou stranu je o slovenštinu u českých dětí patrný zájem. To slovenské děti o češtinu zájem nejeví – nezdá se jim tolik „cizí“ – a to i přes to, že dle slovenského zákona mají tvůrci programů pro děti do 12 let povinnost všechna neslovenská audiovizuální díla předabovat do slovenštiny (Svobodová 2006, Dobrotová a Musilová 2012).

Česko-polské, popřípadě slovensko-polské jazykové kontakty tak intenzivní a dlouhodobé nebyly, ani v současnosti nejsou. To ostatně dokládá také skutečnost, že Češi/Slováci nemají snadný přístup k polské kultuře (Nábělková 2008), což se ovšem netýká pohraničních oblastí, a zejména pak ne Těšínska. Tam je polština užívána už od 19. století ve velkém rozsahu nejen ve školství, ale i v ostatních oblastech (Nekvapil et al. 2009, Greň 2011).

Zatímco Češi a Slováci se mezi sebou dorozumí každý svým vlastním mateřským jazykem, s Poláky to mají Češi i Slováci v tomto ohledu složitější – musí vynaložit větší úsilí. Některé výzkumy ovšem dokonce uvádí, že ze strany Čechů je k Polákům zaznamenatelná určitá míra xenofobie (Lotko 1998).

Výzkumy semikomunikace v námi vymezeném prostředí se prováděly spíše nepřímou. Považujeme za ně konfrontační studie češtiny a slovenštiny, češtiny a polštiny, popřípadě slovenštiny a polštiny. Terminologie v této oblasti je poměrně roztříštěná, neboť semikomunikace bývá v západoslovanském prostředí označována termíny *částečný pasivní bilingvismus*¹¹ (Budovičová 1974), *diglosie*¹² (Horecký 1995), *kolektivní bilingvismus s diglosií*¹³ (Lotko 1998), nebo zjednodušeně pouze *pasivní/asymetrický/receptivní bilingvismus*, dnes nověji *percepční bilingvismus*¹⁴ (Nábělková 2008).

Z metodologického pohledu jsme v průběhu práce upozorňovali na dvě pojetí semikomunikace: sociolingvistické a matematicko-lingvistické. Zkoumání vztahů mezi západoslovanskými jazyky dosud probíhá výhradně pomocí sociolingvistických metod. Často jsou realizovány buď dotazníkově, nebo s audionahrávkami (Sloboda 2004). My zde uvedeme výsledky dvou takových snad nejcitovanějších tuzemských dotazníkových výzkumů (provedených v letech 1971 a 1997) a jednoho zahraničního, zato však nedávného (z roku 2016) a všestranného.

Zmíněné tuzemské výzkumy polštinu zpravidla opomíjí a věnují se vztahu češtiny a slovenštiny. Jejich výsledky se shodují v tom, že si vysoké procento Čechů a Slováků vzájemně rozumí, nicméně Slováci češtině rozumí lépe než Češi slovenštině (Budovičová 1974, Musilová 2000). U porozumění Čechů slovenštině byl navíc naměřen pokles (Dobrotová a Musilová 2012). Úskalím takových dotazníkových výzkumů ale je, že se dotazovaní mohou často přeceňovat.

¹¹ Tento termín navrhuje Budovičová (1974) zejména pro vztah češtiny a slovenštiny. Sama ovšem upozorňuje, že by pojem *bilingvismus* u blízce příbuzných jazyků zasloužil zpřesnění.

¹² Tento termín navrhuje J. Horecký (1995) zejména pro vztah češtiny a slovenštiny. Ospravedlňuje její existenci silné interference mezi těmito jazyky.

¹³ Tento termín navrhuje E. Lotko (1998) zejména pro vztah češtiny a polštiny na česko-polských hranicích. Zde by měl vlivem společenských, kulturních i politických faktorů plnit některý jazyk funkci vysoké variety a jiný funkci nízké variety.

¹⁴ Tento termín navrhuje Nábělková (2008) jako nejsoučasnější termín s tím, že pro semikomunikaci doporučuje užívat také termín *dvojazykovost* (nikoli *dvojazyčnost*, což by byl *bilingvismus*).

Zmíněný zahraniční výzkum (Golubovic 2016) se zaměřuje na celkové porozumění západoslovanských (čeština, slovenština, polština) a jihoslovanských (chorvatština, srbština, bulharština) jazyků se zřetelem na vztahy jazyků napříč těmito jazykovými skupinami. Výzkumu se účastnilo dohromady 10 tisíc respondentů ve formě psané i mluvené, v oblastech čtení, poslechu, doplňování slov a popisu obrázků. Prokázán byl znatelný rozdíl mezi západoslovanskými a jihoslovanskými jazyky ve všech jmenovaných oblastech.

Co se týká námi zkoumaných západoslovanských jazyků, ve všech oblastech měla k sobě čeština se slovenštinou velmi blízko, nicméně ve vztahu obou těchto jazyků k polštině měla blíže slovenština, ač ne o mnoho.

Z pohledu asymetrie vzájemného porozumění, u níž sama autorka upozorňuje na ne-signifikantní průkaznost, se u mluvené formy jedná o následující hodnoty:

- (14) Češi lépe rozumí polštině (35,34 %) než Poláci češtině (26,58 %),
- (15) Slováci lépe rozumí češtině (95,04 %) než Češi slovenštině (92,68 %),
- (16) Slováci lépe rozumí polštině (50,69 %) než Poláci slovenštině (40,69 %).

V závěru této podkapitoly připomeňme ještě základní formální (ne)shody studovaných západoslovanských jazyků. Rozdíl mluvené češtiny od slovenštiny je počítán kolem 11 %, zatímco rozdíl češtiny/slovenštiny od polštiny 26 % (Nekvapil et al. 2009).

Čeština je ze zkoumané skupiny nejfektivnější (Lotko 1999), kvůli čemuž mívá obtíže s přejímáním cizích slov (Lotko 2002). Na základě typologických rozdílů se předpokládá, že Čech porozumí polštině lépe než Polák češtině (Lotko 1998).

Slovenština má ze zkoumané skupiny nejblíže k jihoslovanským jazykům, neboť byla kodifikována na základě střední slovenštiny, ve které se vyskytuje více jevů nezápadoslovanského původu (Dudášová-Křiššáková 1999). Přesto je svou stavbou nejblíže češtině (Nekvapil et al. 2009).

Polština má pak ze zkoumané skupiny nejblíže k východoslovanským jazykům. Stupeň její fektivnosti je oproti češtině a slovenštině nejmenší, má analytické tendence (Lotko 1999). Na druhou stranu jí nečiní takové potíže jako češtině přejímat cizí slova – mluví se dokonce o evropeizaci polštiny (Lotko 2002). Blízkost češtiny/slovenštiny s polštinou byla mnohem větší, ale to ve 12. století před realizací některých hláskových změn (Lotko 1998). Ve 20. století došlo ve fonologickém systému polštiny k dalším změnám, a tak se polština češtině a slovenštině ještě více vzdálila (Dudášová-Křiššáková 1999). Také tvoření slov probíhá vlivem analytičnosti polštiny rozdílně. Polština tvoří slova morfologicko-syntakticky (např.: čs. pracoviště = pl. miejsce pracy) (Lotko 1987).

Za zmínku stojí i zrádná slova mezi všemi třemi zkoumanými jazyky. Zrádné slovo je se slovem druhého jazyka shodné formálně, nikoli však významově/stylisticky (Lotko 1987). To může příjemce velmi zavádět, protože ten se slovu snaží porozumět na základě vlastního mateřského lexika (Lipowski 2009, Stopyra 2010).

8 Semikomunikace v konceptu komunikace

V 7. kapitole byl představen stručný pohled na historii nesnadného definování termínu semikomunikace, při němž byla semikomunikace popsána Haugenem pojmána buď jako součást receptivního bilingvismu (Braunmüller 2002), nebo jako kódový šum při komunikaci (Budovičová 1987). Přestože teoretické pojetí, které nám představila Budovičová, je velmi zajímavé a bylo ve své době silné, uvažuje se dnes semikomunikace jako forma receptivního bilingvismu (Börestamuhlmann 2002). Stále ovšem existují hlasy lingvistů, kteří se domnívají, že semikomunikace je od receptivního bilingvismu odlišná (Zeevaert 2007).

L. Zeevaert (2007) proto alespoň rozlišuje receptivní bilingvismus:

- (a) u nepříbuzných, nebo vzdáleně příbuzných jazyků,
- (b) u blízce příbuzných jazyků.

Ve své podstatě jde o kompromis mezi tradičním pojetím receptivního bilingvismu (možnost (a), která se vyskytuje například v jazykovém prostředí Švýcarska) a Haugenem popsaným pojetím semikomunikace (možnost (b), která se týká například skandinávských jazyků). Doslova popisuje současný stav takto: „*termín receptivní multilingvismus může být použit k popisu komunikace mezi abstand jazyky, ale také mezi ausbau jazyky, které projevují vůči sobě malé rozdíly; termín ‚semikomunikace‘ však lze použít výhradně k popisu kontaktu mezi blízce příbuznými ausbau jazyky.*“ (Zeevaert 2007, s. 106) Na základě dalších analýz, které prokázaly, že je k semikomunikaci nutná kromě jazykové blízkosti také aktivní vůle příjemce, nedoporučuje vůbec pro semikomunikaci využívat pojmů „pasivní kompetence“, jak tomu bývá při úvahách ohledně řazení semikomunikace do receptivního/pasivního/asymetrického bilingvismu (Zeevaert a Ten Thije 2007). Současně upozorňuje také na to, že „*receptivní multilingvismus a semikomunikace jsou charakterizovány odlišnými podmínkami, a mělo by tedy být s nimi zacházeno jako s odlišnými fenomény.*“ (Zeevaert 2007, s. 106)

*

V následující části se pokusme sami vyvodit a doložit teoretické umístění semikomunikace v jazykovědném systému, a to tak, že popíšeme elementární shody a rozdíly mezi doposud představenými jevy (komunikací, diglosií, bilingvismem) a semikomunikací.

Komunikace se ve všech svých podmínkách se semikomunikací shoduje. Oba jevy lze také chápat jako způsoby dorozumívání mezi lidmi. Jediný rozdíl je v tom, že teorie komunikace po čas svého procesu pracuje s jedním kódem, zatímco semikomunikace se dvěma kódy. Tyto dva kódy jsou si ale natolik blízké, že jsou vzájemně srozumitelné

a že k nim příjemce při dekódování nepřistupuje jako k cizímu kódu, takže není nucen k přepínání kódů, nýbrž jako ke kódu mateřskému – porovnává podobu přijatého signálu se svým mateřským tak, že signál, kterému nerozumí, si nahradí za podle něj nejbližší signál, který se nachází v jeho mateřském kódu. Semikomunikaci tak lze řadit mezi způsoby vícejazyčné komunikace, která nabízí nanejvýše následující možnosti, při nichž se spolu komunikanti dorozumívají:

- (a) oba pomocí lingua franca,
- (b) první svým mateřským kódem, druhý mateřským kódem prvního,
- (c) každý svým vlastním mateřským kódem (tj. semikomunikace).

Diglosie se od semikomunikace liší nejen tím, že semikomunikace je způsobem komunikace a diglosie jazykovou situací, ale také svou vázaností na komunikační situaci. Při diglosii navíc dochází ke střídání kódů (variet jednoho jazyka), přičemž některé variety jsou vnímány jako vyšší, jiné jako nižší. Při semikomunikaci není ani jeden z kódů považován za nižší/vyšší, neboť se jedná o autonomní jazyky, které sice mohly v minulosti být součástí jednoho jazykového společenství (tedy varietami jim nadřazeného kódu), a tím pádem být i součástí diglosní jazykové situace, ale to již při jejich současném stavu, kdy se jedná o samostatné autonomní jazyky (např. češtinu a slovenštinu) nepřichází do úvahy.

Bilingvismus se zdá být semikomunikaci nejbližší, nicméně je nutné si uvědomit, že zatímco semikomunikace je způsob komunikace, tak bilingvismus je vlastnost komunikantů, nikoli způsob komunikování. Pokud bychom uvažovali nad bilingvismem jako nad způsobem komunikace (šlo by tedy o možnosti (a) a (b) viz výše), pak bilingvismus sice používá stejně jako semikomunikace různé jazyky, jenže jedinec se musí cizí jazyk naučit, vytvořit si pro něj v mozku vlastní systém, a při komunikaci, ať už je jedincův bilingvismus jakéhokoli stupně, musí mezi mateřským a cizím kódem přepínat, což se při semikomunikaci neděje.

O jistém propojení semikomunikace s bilingvismem však stále lze hovořit v případě, že k semikomunikaci přičteme interkomprehenci, kterou jedinec podstoupí, čímž pádem se vědomě vážně stává asymetricky/pasivně/receptivně bilingvním. Nicméně termín semikomunikace by nadále zůstal pojmenováním způsobu této komunikace, kdy se oba komunikanti dorozumívají svými mateřskými jazyky.

Domníváme se tedy, že tímto docházíme k témuž jako Zeevaert (2007), který tvrdí, že by měla být semikomunikace považována za fenomén odlišný od receptivního bilingvismu i od diglosie. Diglosie totiž představuje jazykovou situaci, bilingvismem označujeme vlastnost jedince a semikomunikací způsob vícejazyčné komunikace.

9 Měření podmíněné entropie mezi západoslovanskými jazyky

Gooskens (2006) se podařilo prokázat propojení mezi porozuměním a vzdáleností jazyků, čímž umožnila zaměřit se na vzájemné porozumění při semikomunikaci také z matematicko-lingvistického hlediska. Nejdříve byla za účelem měření vzájemného porozumění využívána Levenštejnova metrika, jenže ta je vhodná výhradně pro symetrické vztahy (Moberg et al. 2007), což není případ porozumění při semikomunikaci. Gooskens proto ve spolupráci s IT specialisty Mobergem, Nerbonnem a Vaillettem navrhli a na skandinávských jazycích (dánštině, norštině, švédštině) úspěšně otestovali v práci *CONDITIONAL ENTROPY MEASURES INTELLIGIBILITY AMONG RELATED LANGUAGES* novou metodu pro měření vzájemného porozumění při semikomunikaci. Její základy spočívají ve využití poznatků teorie informace. Konkrétně pracují s upravenou entropií do podoby tzv. *podmíněné entropie* (Moberg et al. 2007).

V následující části bakalářské práce se proto pokusíme replikovat metodologii jmenovaného výzkumu, ovšem na materiálech západoslovanských jazyků (češtiny, slovenštiny a polštiny). Pro používanou metodologii vytvoříme relevantní algoritmy, čímž ji celou zautomatizujeme, abychom mohli pracovat s libovolně velkými vzorky dat a minimalizovali výskyt chyb.

Základní teoretické poznatky o semikomunikaci, skandinávské i západoslovanské jazykové situaci, včetně různých dostupných výsledků sociolingvistických výzkumů, byly zevrubně popsány v 7. kapitole, a proto zde již nebudou pojednány.

Měření vzájemného porozumění v závislosti na vzdálenosti jazyků je vhodné provádět na různých jazykových rovinách. Po vzoru původního výzkumu volíme rovinu fonetickou.

Praktickým cílem našeho zájmu je odpovědět na tyto otázky týkající se komunikantů západoslovanských jazyků:

- (A) Je vzájemné porozumění komunikantů na základě jazykové blízkosti asymetrické?
- (B) Jsou z hlediska jazykové blízkosti rozdíly v celkovém vzájemném porozumění komunikantů?
- (C) Liší se z hlediska jazykové blízkosti vzájemné porozumění komunikantů u plnovýznamových a neplnovýznamových slov?
- (D) Liší se z hlediska jazykové blízkosti vzájemné porozumění komunikantů u různých typů komunikátů?

9.1 Metody

Nejčastěji používanou matematicko-lingvistickou metodou pro měření vzájemného porozumění na fonetické rovině bývá Levenštejnova metrika. Její výsledky ovšem samy o sobě nenabízí pohled na případnou asymetrii porozumění, a tak je k ní pro tyto případy využíván ještě výpočet podmíněné entropie.

Levenštejnova metrika udává rozdíl dvou řetězců. Zprostředkovává minimální počet operací (povoleno: vložení/smazání/nahrazení znaku) nutně provedených k tomu, aby byly řetězce totožné. Současně lze pomocí této metriky zarovnat znaky obou řetězců k sobě tak, aby byl počet nutných operací co nejmenší (viz obrázky 7–8).

Operacím lze nastavit váhu a každému slovnímu páru tak spočítat procentuální podobnost, míru hypotetického porozumění (na obr. 7–8 mají operace stejnou váhu). Vyčíslování vzájemného porozumění pomocí Levenštejnovy metriky je však vhodné jedině pro symetrické vztahy. Proto pomocí této metriky řetězce pouze zarovnáujeme.

	SK	k	t	ɔ	r	i:
CS	0	1	2	3	4	5
k	1	0	1	2	3	4
t	2	1	0	1	2	3
ε	3	2	1	1	2	3
r	4	3	2	2	1	2
i:	5	4	3	3	2	1

k	t	ɔ	r	i:
k	t	ε	r	i:
0	0	1	0	0

Obrázek 7: Levenštejnova metrika A,
1 nutně provedená operace
(nahrazení ɔ za ε),
porozumění 80 %

	SK	d	l	ɦ	ɔ
CS	0	1	2	3	4
d	1	0	1	2	3
l	2	1	1	2	3
ɔy	3	2	2	2	3
ɦ	4	3	3	2	3
ɔ	5	4	4	3	2

d	l	#	ɦ	ɔ
d	l	ɔy	ɦ	ɔ
0	1	1	0	0

Obrázek 8: Levenštejnova metrika B,
2 nutně provedené operace
(nahrazení l za l, vložení #¹⁵ mezi [ɦ]),
porozumění 60 %

Měření porozumění provádíme až pomocí podmíněné entropie. Ta „měří složitost zarovnání a je závislá na frekvenci a pravidelnosti zvukových souvztažností mezi dvěma jazyky.“ (Moberg et al. 2007, s. 53) Prakticky respektuje postup příjemce, jenž slyší cizí foném a zarovná jej k jinému fonému ze svého mateřského jazyka (Moberg et al. 2007).

Entropie samotná vyjadřuje míru nejistoty. Podmíněná entropie tuto míru nejistoty vyčísluje z pohledu jednotlivých komunikantů zvlášť. Vysoká entropie znamená vysokou míru nejistoty, tudíž malé porozumění, a naopak (Moberg et al. 2007).

Vzorec, jímž je podmíněná entropie vyjádřena, vypadá následovně:

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

přičemž $p(x, y)$ je pravděpodobnost výskytu znaků x a y na stejné pozici (zarovnané dvojice x, y), $p(x, y) = p(x \cap y)$. Podmíněná pravděpodobnost $p(x|y)$ vyjadřuje pravděpodobnost výskytu znaku x , pokud víme, že se na stejné pozici nachází znak y , tedy $p(x|y) = \frac{p(x \cap y)}{p(y)}$, přičemž jmenovatel zlomku zajišťuje asymetrii.

¹⁵ Operace vložení vkládá do řetězce výplňový znak (#), který nepředstavuje žádný konkrétní foném.

Podmíněná entropie $H(X|Y)$ se nemusí rovnat podmíněné entropii $H(Y|X)$. Výsledky podmíněných entropií jsou díky hodnotě základu logaritmu vyjádřeny v bitech [b].

Pro příklad výpočtu mějme databázi o velikosti dvou slovních párů, jejichž fonémy jsou již zarovnané pomocí Levenštejnovy metriky (viz obr. 9).

pozice	1	2	3	4	5		6	7	8
CS	<i>m</i>	<i>i:</i>	<i>s</i>	<i>t</i>	<i>ɔ</i>		<i>z</i>	<i>ε</i>	<i>n</i>
SK	<i>m</i>	<i>ĭε</i>	<i>s</i>	<i>t</i>	<i>ɔ</i>		<i>ʒ</i>	<i>ĭε</i>	<i>n</i>

Obrázek 9: Příklad – zarovnané fonémy Levenštejnovou metrikou

Dále jsou také vypočteny podmíněné pravděpodobnosti z pohledu obou dvou jazyků, $p(CS|SK)$ i $p(SK|CS)$, viz obr. 10. Například pozici fonému *m* ve slovenštině odpovídá pouze stejný foném v češtině, a tak jde o pravděpodobnost 100 %, $p(m|m) = 1$, a to z pohledu obou jazyků. Naproti tomu pozicím slovenského fonému *ĭε* odpovídají na české straně fonémy dva (*i:* a *ε*), a tak je pravděpodobnost z pohledu češtiny 50 % (viz př. 17). Ovšem z pohledu slovenštiny je pravděpodobnost stále rovna 100 % (viz př. 18–19), protože pozici českého fonému *i:* odpovídá pouze slovenský foném *ĭε*. Shodně tomu je u pozice českého fonému *ε*.

$$(17) \quad p(CS|SK) = p(i:|\dot{\imath}\epsilon) = p(\epsilon|\dot{\imath}\epsilon) = \frac{\frac{2}{8}}{\frac{1}{2}} = \frac{1}{2} = 0,5 \times 100 = 50 \%$$

$$(18) \quad p(SK|CS) = p(\dot{\imath}\epsilon|i:) = \frac{\frac{1}{8}}{\frac{1}{8}} = 1 \times 100 = 100 \%$$

$$(19) \quad p(SK|CS) = p(\dot{\imath}\epsilon|\epsilon) = \frac{\frac{1}{8}}{\frac{1}{8}} = 1 \times 100 = 100 \%$$

Hodnoty podmíněných pravděpodobností jsou následně zlogaritmovány. Logaritmus dvojic se 100% podm. pravděp. je roven nule, např. $-\log_2(p(m|m)) = -\log_2(1) = 0$. Oproti tomu výsledek logaritmu dvojic s jinou, v našem případě 50%, podm. pravděp. je nenulový, např. $-\log_2(p(i:|\dot{\imath}\epsilon)) = -\log_2(p(\epsilon|\dot{\imath}\epsilon)) = -\log_2(0,5) = 1$.

pozice	1	2	3	4	5		6	7	8
CS	<i>m</i>	<i>i:</i>	<i>s</i>	<i>t</i>	<i>ɔ</i>		<i>z</i>	<i>ε</i>	<i>n</i>
SK	<i>m</i>	<i>ĭε</i>	<i>s</i>	<i>t</i>	<i>ɔ</i>		<i>ʒ</i>	<i>ĭε</i>	<i>n</i>
$p(SK CS)$	1	1	1	1	1		1	1	1
$p(CS SK)$	1	0,5	1	1	1		1	0,5	1

Obrázek 10: Příklad – vypočtené podmíněné pravděpodobnosti

V této části nám k dosažení do vzorce podmíněné entropie zbývá pouze vypočítat pravděpodobnosti výskytů jednotlivých dvojic fonémů, $p(x, y)$. Naše databáze má těchto dvojic pouze osm a všechny jsou jedinečné, proto budou jejich pravděpodobnosti rovny stejné hodnotě (viz př. 20).

$$(20) \quad p(m, m) = p(i:, \dot{\imath}\epsilon) = \dots = p(\epsilon, \dot{\imath}\epsilon) = p(n, n) = \frac{1}{8} = 0,125$$

Výsledné podmíněné entropie vzorku musíme opět počítat z pohledu obou dvou jazyků, $H(CS|SK)$ i $H(SK|CS)$. Po dosazení do vzorce získáme:

$$\begin{aligned} H(CS|SK) &= (0,125 \times 0) + (0,125 \times 1) + (0,125 \times 0) + (0,125 \times 0) + (0,125 \times 0) + (0,125 \times 0) \\ &\quad + (0,125 \times 1) + (0,125 \times 0) = 0,25 \text{ b} \\ H(SK|CS) &= (0,125 \times 0) + (0,125 \times 0) + (0,125 \times 0) + (0,125 \times 0) + (0,125 \times 0) + (0,125 \times 0) \\ &\quad + (0,125 \times 0) + (0,125 \times 0) = 0 \text{ b} \end{aligned}$$

Asymetrický vztah jazyků, $H(CS|SK) > H(SK|CS)$, zkoumaných na takto malém vzorku je patrný. Uživatel češtiny si musí za jeden slovenský foném (i/ε) dosadit jeden ze dvou svých fonémů (i/ε), což se projevuje zvýšenou entropií ve vztahu $H(CS|SK)$.

9.2 Materiály

Základním požadavkem jsou databáze slov s jejich překladovými ekvivalenty – v podstatě slovní zásoba jednotlivých jazyků (Moberg et al. 2007). Zdrojem dat pro naše databáze slov češtiny, slovenštiny a polštiny je paralelní jazykový korpus InterCorp verze 9 z roku 2016 Českého národního korpusu (ČNK)¹⁶.

Pro češtinu, slovenštinu a polštinu nabízí InterCorp subkorpora: Acquis (právní texty EU), Core (beletristické texty), Europarl (zápisy z jednání Evropského parlamentu z let 2007–2011) a Subtitles (filmové titulky serveru OpenSubtitles.org) (Škrabal 2017).

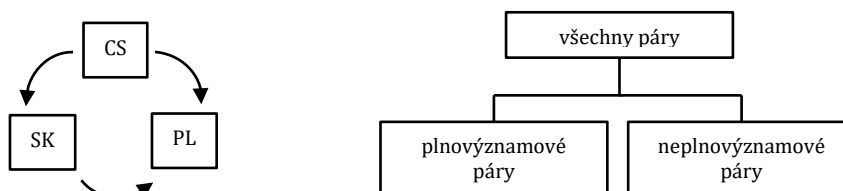
Ze všech jmenovaných subkorporů čerpáme 2 000 nejfrekventovanějších slov, k nimž hledáme vhodné překladové ekvivalenty. Počet dvou tisíc slov se ukázal jako dvou a půl násobek počtu, který poskytuje stabilní výsledek (Moberg et al. 2007).

Slova získáváme pomocí funkce „*Vlastní nastavení frekvenční distribuce*“, kterou nabízí rozhraní KonText¹⁷ ČNK. Získáváme tak nejen slova, ale i jejich morfologické značky, pomocí nichž třídíme slova na plnovýznamová a neplnovýznamová.

Překladové ekvivalenty přiřazujeme ke slovům pomocí aplikace Treq¹⁸ ČNK, která je taktéž provázána s paralelním korpusem InterCorp v9. Treq nabízí ke každému slovu zpravidla hned několik překladových ekvivalentů (viz př. 21–25). Automaticky přiřazujeme vždy ten nejfrekventovanější.

- (21) Dotaz na slovenský nelematizovaný ekvivalent z Core k českému slovu „očima“:
- (22) frekvence: 1692 procenta: 88,7 ekvivalent: „očami“
- (23) frekvence: 93 procenta: 4,9 ekvivalent: „oči“
- (24) frekvence: 47 procenta: 2,5 ekvivalent: „zrakom“
- (25) ...

Vhodné je zpracovávat ve všech třech jazykových vztazích totožnou slovní zásobu. Databáze proto vychází z českých slov, k nimž jsou nalezeny slovenské a polské ekvivalenty. Ke slovenským ekvivalentům jsou poté dohledány ekvivalenty polské (viz obr. 11 vlevo).



Obrázek 11: Doplnění překladových ekvivalentů (vlevo) a struktura databáze (vpravo)

¹⁶ (Rosen a Vavříň 2016, Nábělková a Vavříň 2016, Bańczyk et al. 2016)

¹⁷ Konkrétně ve verzi 0.9.3. Dostupné na adrese: <https://kontext.korpus.cz>.

¹⁸ Konkrétně ve verzi 1.1. Dostupné na adrese: <http://treq.korpus.cz/>.

Takto získané slovní páry zkoumáme ve třech kategoriích – všechny páry, pouze plnovýznamové páry, pouze neplnovýznamové páry (viz obr. 11 vpravo).

Protože provádíme výzkum na fonetické rovině jazyků, jsou slovní páry všech kategorií foneticky transkribovány¹⁹.

Nutno podotknout, že ne vždy se aplikaci Treq podaří nalézt překladový ekvivalent a že sestavený algoritmus fonetické transkripce češtiny, slovenštiny a polštiny není navržen pro transkripci číslic a cizích znaků (např. ř, ě, ...), a proto jsou původní počty dvou tisíc slovních párů pro každý subkorpus sníženy. Finální počty zpracovávaných slovních párů pro jednotlivé kategorie zaznamenává tabulka na obr. 12.

Subkorpus a kategorie		Dvojice jazyků a počet párů		
		CS-SK	CS-PL	SK-PL
Acquis	vše	1772	1775	1770
	plnovýznamové	1678	1681	1677
	neplnovýznamové	94	94	93
Core	vše	1986	1988	1984
	plnovýznamové	1855	1857	1854
	neplnovýznamové	131	131	130
Europarl	vše	1960	1966	1957
	plnovýznamové	1858	1864	1856
	neplnovýznamové	102	102	101
Subtitles	vše	1964	1968	1960
	plnovýznamové	1842	1846	1839
	neplnovýznamové	122	122	121
celkem zkoumaných párů		7682	7697	7671

Obrázek 12: Konečné počty zkoumaných slovních párů

9.3 Limity

Zaměřme se před popisem výsledků na jistá omezení, která se vyskytla při zpracovávání jednotlivých slovních párů.

Určitým omezením je přiřazování překladových ekvivalentů pomocí aplikace Treq ČNK. V některých případech (viz př. 26–30) se ukázalo, že jsou častěji užívány ekvivalenty, které si nejsou formálně podobné, přestože by takové ekvivalenty jazyk poskytnout mohl. Řešením je úprava algoritmu tak, aby porovnával formální podobnost těchto potenciálních ekvivalentů a teprve až na základě toho vybíral „vhodné“ ekvivalenty. Jenže častější užití formálně nepodobných výrazů může být požadováno stylisticky, a proto ke slovu automaticky přiřazujeme vždy ten nejfrekventovanější ekvivalent.

- (26) Dotaz na slovenský nelematizovaný ekvivalent z Core k českému slovu „prostě“:
- (27) frekvence: 378 procenta: 57,2 ekvivalent: „jednoducho“
- (28) frekvence: 126 procenta: 19,1 ekvivalent: „proste“
- (29) frekvence: 44 procenta: 6,7 ekvivalent: „prosto“
- (30) ...

¹⁹ Formální stránka pravidel ortoepie, obsahy i přístupy k fonetikám a fonologiím češtiny, slovenštiny i polštiny byly zpracovány na základě publikací: Balowski 1993, Kajanová-Schulzová 1970, Král' a Sabol 1989, Lipowski 2011, Palková 1994, Pauliny 1979, Sokolová et al. 2005, Vachek 1970, Zeman 2008.

Nastaly ovšem i situace, kdy aplikace Treq nabídla úplně nevhodné překladové ekvivalenty. Takových případů jsme ale zaznamenali jen jednotky, zpravidla u složitějších morfologicko-syntaktických jevů, např. u polské a slovenské lexikální negace atd.

Svoje limity mají též vytvořené algoritmy fonetických transkripcí češtiny, slovenštiny a polštiny. Transkripce bylo nutné zpracovat pomocí znaků mezinárodně užívané fonetické abecedy (IPA), která se ale pro slovanské jazyky téměř vůbec nevyužívá, a tak bylo zapotřebí řešit běžné nesnáze jednotlivých přístupů ke zvukovým stránkám všech tří zkoumaných jazyků.

Konkrétně máme na mysli nutnost volby adekvátního zaznamenávání fonémů, například zda neslabikotvorné diftongické /u/ a /i/ transkribovat jako *ʊ* a *i*, či *w* a *j* (zvolena první možnost). Dále také fonologické spory, například zda diftongy vnímat jako monofonémy, či bifonémy (zvolena první možnost), nebo zda zpracovávat všechny alofony (nakonec zpracovány pouze ty alofony, které jsou nutné k zachování binárního protikladu znělost-neznělost, např. *ɣ*, *ʃ*, a které by mohly mít jisté percepční důsledky, např. *w*, *ɔŋ*, ...) atd.

Potíže transkripčním algoritmům způsobuje též výslovnost přejatých slov a zkratk, což nebylo možné postihnout. Stoprocentní není algoritmus ani v případech výrazů, u nichž je potřeba přihlídnout k morfologické stavbě (např. [*pɔɥpɛ*] vs. [*pɔʔupɛvɪt*]). Pro některé předpony byla tato problematika ale ošetřena.

Přes všechna výše jmenovaná omezení se při čtyřech ručních kontrolách 500 takto automaticky zpracovaných slovních párů objevily chyby v méně než 5 % případů, což bývá obecně považováno za akceptovatelnou míru chyby.

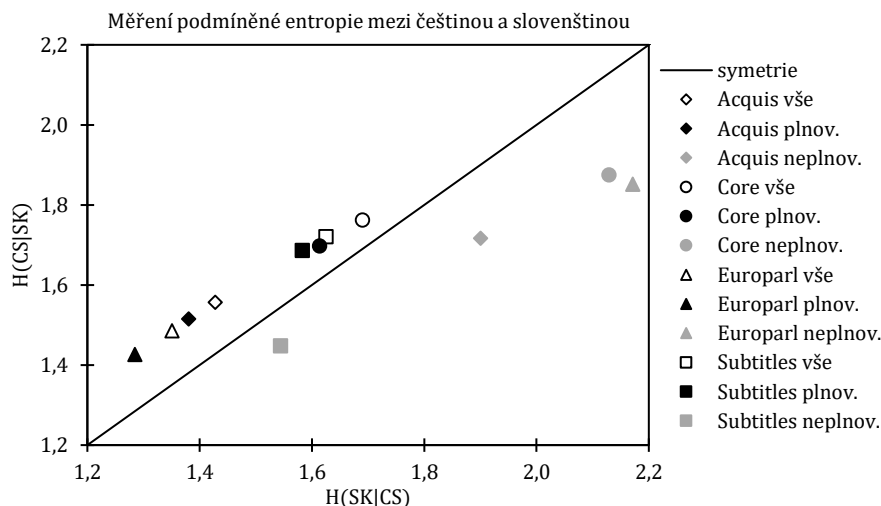
9.4 Výsledky

Uskutečněným měřením podmíněné entropie mezi západoslovanskými jazyky jsme provedli výzkum jednoho ze tří faktorů semikomunikace. Námi studovaným faktorem byla jazyková vzdálenost na fonetické rovině.

Vliv žádného z faktorů ovšem není zanedbatelný (Gooskens 2007). Gooskens (2006) ale potvrdila, že existuje vztah mezi jazykovou vzdáleností a vzájemným porozuměním, a proto tak naměřené výsledky interpretujeme, přestože si uvědomujeme skutečnost, že se na výsledném porozumění podílí též ostatní dva faktory i mimojazykové prvky.

Na následujících třech stránkách jsou v grafech (obr. 13–15) vyneseny naměřené hodnoty podmíněných entropií. Osy grafů představují vždy vztah dvou jazyků ve tvaru výpočtu $H(\text{vysílatel}|\text{příjemce})$, diagonální přímka, která rozděluje oblast grafu na dvě poloviny, prochází hodnotami přesně symetrického vzájemného porozumění obou jazyků.

Před samotnou interpretací výsledků připomeňme, že vysoká podmíněná entropie je spojena s vyšší mírou nejistoty, a tudíž horším porozuměním. Nízká hodnota podmíněné entropie znamená nižší míru nejistoty, a tak lepší porozumění.



Obrázek 13: Graf podmíněných entropií mezi češtinou a slovenštinou²⁰

(A) Na základě dostupných sociolingvistických výzkumů (Budovičová 1974, Musilová 2000, Golubovic 2016) jsme předpokládali, že lépe porozumí Slováci češtině než naopak. Tento předpoklad se naším měřením potvrdil. Zpravidla platilo $H(CS|SK) > H(SK|CS)$.

Na ose x jsou hodnoty entropií pro případy slovenštiny zarovnané k češtině. Na ose y jsou hodnoty entropií pro případy češtiny zarovnané ke slovenštině. Jelikož diagonální přímka grafu představuje přesně symetrické vzájemné porozumění, tak body, které jsou vykresleny mimo ni, vyjadřují asymetrii. Body nad diagonálou zaznamenávají lepší porozumění Slováků češtině a body pod diagonálou lepší porozumění Čechů slovenštině.

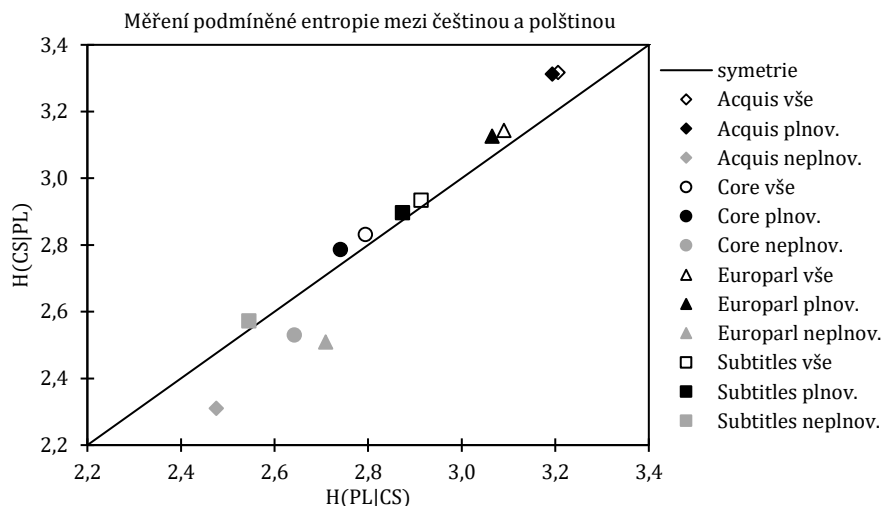
(B) Jednotlivé hodnoty podmíněných entropií mezi češtinou a slovenštinou byly v porovnání s hodnotami měření mezi skandinávskými (Moberg et al. 2007) i ostatními západoslovanskými jazyky mimořádně nízké (na ose x mezi 1,2850 a 2,1719 bitů, na ose y mezi 1,4259 a 1,8753 bitů).

(C) Zajímavé byly výsledky měření podmíněných entropií u kategorií s plnovýznamovými a neplnovýznamovými slovy. Zatímco při měření podmíněných entropií mezi skandinávskými i ostatními západoslovanskými jazyky měly nejvyšší hodnoty podmíněných entropií kategorie se všemi slovy, u vztahu češtiny a slovenštiny to byly kategorie s neplnovýznamovými slovy (vyjma kategorie subkorpusu Subtitles).

U všech kategorií s neplnovýznamovými slovy navíc došlo při měření ke snaze obrátit trend asymetrie. Hodnoty podmíněných entropií všech kategorií s neplnovýznamovými slovy se na rozdíl od ostatních totiž nacházely na opačné straně od diagonály.

(D) Z pohledu jednotlivých typů komunikátů měly nejvyšší hodnoty podmíněných entropií komunikáty beletristické (Core) a nejnižší hodnoty podmíněných entropií komunikáty administrativní a právnické (Europarl a Acquis).

²⁰ Konkrétní naměřené hodnoty jsou k dispozici v tabulce 1, viz Příloha A.



Obrázek 14: Graf podmíněných entropií mezi češtinou a polštinou²¹

(A) Na základě dostupných výzkumů (Golubovic 2016, Lotko 1998) jsme předpokládali, že lépe porozumí Češi polštině než naopak. Tento předpoklad se naším měřením však nepotvrdil. Zpravidla platilo $H(CS|PL) > H(PL|CS)$.

Nejednalo se tedy o symetrické porozumění češtiny a polštiny, přestože jednotlivé hodnoty podmíněných entropií se přímce symetrie velmi blížily.

(B) Jednotlivé hodnoty podmíněných entropií mezi češtinou a polštinou byly v porovnání s hodnotami měření mezi skandinávskými (Moberg et al. 2007) i ostatními západoslovanskými jazyky nejvyšší (na ose x mezi 2,4756 a 3,2059 bitů, na ose y mezi 2,3098 a 3,3167 bitů).

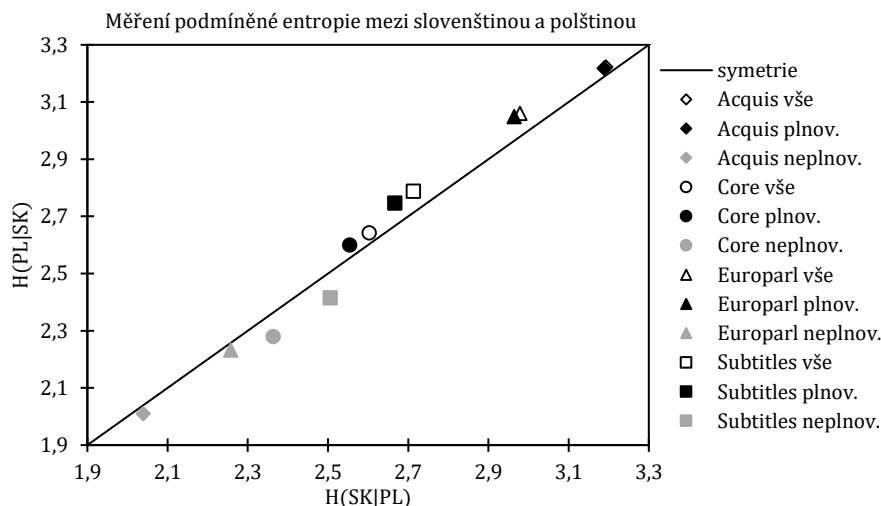
Předpoklad, že jazyková vzdálenost češtiny a polštiny je větší než tato vzdálenost češtiny se slovenštinou, se tedy potvrdil.

(C) Z pohledu jednotlivých kategorií slov se jednalo o předpokladatelné výsledky. Nejvyšší podmíněné entropie byly pro každý subkorpus naměřeny u kategorií se všemi slovy, nižší podmíněné entropie u kategorií s plnovýznamovými slovy a nejnižší podmíněné entropie u kategorií s neplnovýznamovými slovy.

U všech kategorií s neplnovýznamovými slovy (vyjma subkorpusu Subtitles) navíc při měření došlo ke snaze obrátit trend asymetrie. Hodnoty podmíněných entropií všech kategorií s neplnovýznamovými slovy (vyjma subkorpusu Subtitles) se na rozdíl od ostatních nacházely opět na opačné straně od diagonály.

(D) Zajímavé byly výsledky měření podmíněných entropií u typů komunikátů. Nejvyšší hodnoty podmíněných entropií jsme naměřili u administrativních a právnických komunikátů (Europarl a Acquis), zatímco nejnižší hodnoty podmíněných entropií u beletristických komunikátů (Core).

²¹ Konkrétní naměřené hodnoty jsou k dispozici v tabulce 2, viz Příloha A.



Obrázek 15: Graf podmíněných entropií mezi slovenštinou a polštinou²²

(A) Na základě dostupného sociolingvistického výzkumu (Golubovic 2016) jsme předpokládali, že lépe porozumí Slováci polštině než naopak. Tento předpoklad se naším měřením potvrdil. Zpravidla platilo $H(PL|SK) > H(SK|PL)$.

O symetrické vzájemné porozumění se tedy nejednalo, přestože jednotlivé hodnoty podmíněných entropií se přímce symetrie velmi blížily a kopírovaly ji podobně jako při měření mezi češtinou a polštinou.

(B) Jednotlivé hodnoty podmíněných entropií mezi slovenštinou a polštinou byly vysoké (na ose x mezi 2,0398 a 3,1926 bitů, na ose y mezi 2,0102 a 3,2221 bitů), ne ale tolik jako u měření mezi češtinou a polštinou.

Lze tedy tvrdit, že jazyková vzdálenost slovenštiny a polštiny je menší než tato vzdálenost češtiny a polštiny, ale stále je podstatně větší než jazyková vzdálenost češtiny a slovenštiny.

(C) Z pohledu jednotlivých kategorií slov se jednalo o předpokladatelné výsledky. Nejvyšší podmíněné entropie byly pro každý subkorpus naměřeny u kategorií se všemi slovy, nižší podmíněné entropie u kategorií s plnovýznamovými slovy a nejnižší podmíněné entropie u kategorií s neplnovýznamovými slovy.

Tentokrát u všech kategorií s neplnovýznamovými slovy při měření došlo ke snaze obrátit trend asymetrie. Hodnoty podmíněných entropií všech kategorií s neplnovýznamovými slovy se na rozdíl od ostatních nacházely znovu na opačné straně od diagonály.

(D) Výsledky měření podmíněných entropií podle typů komunikátů byly totožné jako při měření mezi češtinou a polštinou. Nejvyšší hodnoty podmíněných entropií jsme naměřili u administrativních a právnických komunikátů (Europarl a Acquis), zatímco nejnižší hodnoty podmíněných entropií u beletristických komunikátů (Core).

²² Konkrétní naměřené hodnoty jsou k dispozici v tabulce 3, viz Příloha A.

9.5 Závěry a diskuze

Výzkum (Moberg et al. 2007), z něhož vycházíme, potvrdil, že měření podmíněné entropie mezi jazyky je vhodným matematickým aparátem pro zjišťování míry asymetrie vzájemného porozumění dvou jazyků na základě jejich vzájemné blízkosti, tj. jazykových předpokladů. Zároveň prokázal, že je mezi skandinávskými jazyky patrný rozdíl v hodnotách podmíněných entropií u plnovýznamových a neplnovýznamových slov.

Předmětem našeho výzkumu nebylo potvrdit, či vyvrátit vhodnost měření podmíněné entropie ke zjišťování asymetrie vzájemného porozumění, nýbrž tento matematicko-lingvistický přístup aplikovat na západoslovanské jazyky, a tím odpovědět na otázky uvedené v úvodu této kapitoly.

(A) Naše měření potvrzují asymetrii ve vzájemném porozumění u všech zkoumaných vztahů (češtiny-slovenštiny, češtiny-polštiny i slovenštiny-polštiny), avšak v rozdílné míře. V grafech tuto míru zaznamenávají vzdálenosti²³ bodů od přímky symetrie. Velmi nízké asymetrie jsou u vztahů s polštinou.

Výsledky měření asymetrie vzájemného porozumění jsou následující:

- (31) Slováci lépe rozumí češtině než Češi slovenštině,
- (32) Poláci lépe rozumí češtině než Češi polštině,
- (33) Slováci lépe rozumí polštině než Poláci slovenštině,

přičemž původní výzkum uvádí, že „vztah [podmíněné entropie a vzájemného porozumění, pozn. autora] je nejčistší, když jsou zahrnuta právě všechna slova,“ (Moberg et al. 2007, s. 64). Proto výsledky 31 a 33 potvrzují předpoklady vycházející z dosud provedených výzkumů (Golubovic 2016, Budovičová 1974, Musilová 2000), zatímco výsledek 32 předpoklady (Golubovic 2016, Lotko 1998) nepotvrzuje. Nutno ovšem upozornit na skutečnost, že J. Golubovic (2016, s. 84–85) sama upozorňuje na nesignifikantnost svých výsledků asymetrie vzájemného porozumění mezi západoslovanskými jazyky.

Příčin, kvůli kterým výsledek 32 nepotvrdil předpoklady, by mohlo být několik (pojednány byly zevrubně v kapitole 9.3). Do úvahy přichází nepřesnosti v oblastech:

- (a) přiřazování překladových ekvivalentů,
- (b) fonetické transkripce češtiny a polštiny,
- (c) zarovnávání fonémů,
- (d) provedených sociolingvistických výzkumů.

(a) Přiřazování překladových ekvivalentů prováděla aplikace Treq ČNK, a tak nebylo možné ovlivnit případy, kdy byl ke slovu přiřazen sémanticky nevhodný ekvivalent. Formální podobnost by bylo možné postihnout, ale tím by mohly být narušeny stylové faktory komunikátu.

(b) Fonetická transkripce češtiny a polštiny byla zpracována syntézou několika zdrojů²⁴. IPA znaky respektují jednotlivé vlastnosti fonémů. Alofony fonémů nebyly

²³ Zároveň jsou tyto vzdálenosti spočítány a zaznamenány v tabulkách 1–3, viz Příloha A.

²⁴ (Balowski 1993, Lipowski 2011, Palková 1994, Vachek 1970, Zeman 2008)

až na výjimky zapracovány, což zejména u polštiny výrazně snížilo počet IPA znaků. Tyto nezpracované alofony (např. \tilde{b} , b^j , η , ...) ale nemají na percepci výrazný vliv.

(c) Zarovnávání fonémů pomocí Levenštejnovy metriky není ideální, protože nerespektuje vzájemnou blízkost fonémů, ale zarovnává fonémy binární logikou (principem „je“/„není“ znak fonému stejný). Přesto však původní výzkum (Moberg et al. 2007) potvrdil, že je Levenštejnova metrika pro zarovnávání fonémů dostatečná. Nutno podotknout, že podobné kvantitativní výzkumy na fonetické rovině jazyka byly dosud prováděny výhradně a pouze s fonémy zarovnanými Levenštejnovou metrikou.

Potíže v oblastech (a) i (b) se objevily v méně než 5 % případů, proto považujeme tyto příčiny za zanedbatelné. Dostatečnost oblasti (c) prokázal výzkum, z něhož vycházíme, i praxe ostatních dosud provedených kvantitativních výzkumů. Lze navíc předpokládat, že kdyby byla příčinou nepotvrzení předpokladu 32 některá z oblastí (a), (b), (c), pak by se toto jistě projevilo i při ostatních měřeních. Ilustrativní příklady zpracování (a), (b) a (c) ve vztahu češtiny-polštiny jsou k náhledu v Příloze C.

(d) Provedených sociolingvistických výzkumů mezi češtinou a polštinou bylo doposud realizováno jen minimum. Přestože by právě tyto výzkumy mohly situaci vyjasnit, je jazyková blízkost češtiny a polštiny zkoumána především teoreticky, například na podkladech typologických rysů obou jazyků. Zdá se tak, že se zde otevírá oblast více než vhodná k dalšímu nejen sociolingvistickému zkoumání.

Opomíjeny by neměly být ani ostatní faktory semikomunikace. Čím nižší asymetrie je, tím větší totiž budou mít tyto faktory vliv. Náš výzkum zkoumá jazykovou vzdálenost, tj. systémové jazykové předpoklady ke vzájemnému porozumění, ale nelze přehlížet vůli komunikantů ani zkušenosti komunikantů s jazykem.

(B) Rozdíly celkového vzájemného porozumění vhodně vykresluje rozmístění bodů v grafech se stejnými měřítky os x a y (viz Příloha B). Čím výše se body nachází, tím horší je vzájemné porozumění.

Naše měření potvrzují předpoklad (Nekvapil et al. 2009), že rozdíl jazykové vzdálenosti češtiny a slovenštiny je menší než češtiny a polštiny, popřípadě slovenštiny a polštiny.

Měření u vztahu češtiny a slovenštiny ostatně prokázalo mimořádnou blízkost těchto jazyků. Taková blízkost nebyla zaznamenána ani mezi západoslovanskými, ani mezi skandinávskými jazyky. Lze se proto s přihlédnutím k výsledkům sociolingvistických výzkumů domnívat, že mezi Čechy a Slováky je funkční semikomunikace, nebo je alespoň vysoce pravděpodobná.

Zato v kombinaci obou jazyků, jak češtiny, tak slovenštiny s polštinou byly hodnoty podmíněných entropií velmi vysoké. Tyto hodnoty byly nejvyšší z dosud měřených západoslovanských i skandinávských jazyků. Z komunikační praxe ovšem víme, že s vynaložením určitého úsilí je přesto semikomunikace Čechů/Slováků s Poláky do jisté míry možná. Považujeme to za potvrzení toho, že oba zbylé faktory semikomunikace mají na komunikační praxi taktéž silný vliv.

Nejnižší hodnoty podmíněných entropií, tj. nejlepší porozumění, jsme tedy naměřili mezi češtinou a slovenštinou, vyšší hodnoty mezi slovenštinou a polštinou, nejvyšší

hodnoty, tj. nejhorší porozumění, mezi češtinou a polštinou. Nutno podotknout, že rozdíl mezi porozuměním slovenštiny-polštiny a češtiny-polštiny není tak velký, zatímco rozdíl mezi oběma těmito vztahy a vztahem češtiny-slovenštiny je markantní.

(C) Naše měření potvrzuje závěr původního výzkumu (Moberg et al. 2007), že mezi hodnotami podmíněných entropií u kategorií s plnovýznamovými slovy a u kategorií s neplnovýznamovými slovy je výrazný rozdíl.

Mezi západoslovanskými jazyky je tento rozdíl dokonce tak velký, že hodnoty podmíněných entropií u kategorií s neplnovýznamovými slovy zpravidla stojí na opačné straně asymetrie než kategorie ostatní. Zdá se tak, že neplnovýznamová slova jsou foneticky více rozdílná než slova plnovýznamová, čímž u nich při měření dochází ke snaze obrátit trend asymetrie. Neplnovýznamová slova však svůj význam získávají až ve spojení se slovy plnovýznamovými, a tak mají na celkový význam komunikátu nižší vliv než slova plnovýznamová.

Obzvláště netypicky se chovají neplnovýznamová slova v případě měření podmíněné entropie mezi češtinou a slovenštinou. Kategorie s neplnovýznamovými slovy (kromě subkorpusu Subtitles) zde mají nejvyšší hodnoty entropií ze všech pro tento vztah měřených kategorií. To není pozorovatelné ani mezi západoslovanskými jazyky, ani mezi skandinávskými jazyky.

Přestože by na tom jistý vliv mohla mít velikost vzorku neplnovýznamových slov, vysvětlujeme si tento stav nízkými celkovými velikostmi podmíněných entropií, tzn. mimořádnou blízkostí češtiny a slovenštiny oproti všem ostatním měřeným dvojicím. Při tak mimořádně nízkých hodnotách se totiž i malé odchylky mohou projevit jako zásadní.

(D) Současně jsme objevili zajímavou skutečnost, která jistě dále stojí za bližší průzkum, a to sice, že hodnoty podmíněných entropií, tedy vzájemné porozumění, se liší v závislosti na stylu komunikátu.

Ve vztahu češtiny a slovenštiny jsme naměřili nejnížší hodnoty podmíněných entropií, tj. nejlepší vzájemné porozumění, u administrativních a právnických komunikátů (Acquis, Europarl). Naopak nejvyšší hodnoty, tj. nejhorší vzájemné porozumění, u komunikátů beletristických (Core).

Pro vztahy s polštinou, ať už čeština-polština, nebo slovenština-polština jsme naměřili hodnoty podmíněných entropií přesně opačně – nejnížší měly vždy beletristické komunikáty, nejvyšší hodnoty měly komunikáty administrativní a právníkové.

Ještě zajímavější je pak skutečnost, že hodnoty podmíněných entropií u polo-beletristických komunikátů (subkorpus Subtitles) se zpravidla stabilně pro všechny měřené vztahy umísťovaly uprostřed spektra.

Příčin výše konstatovaných stavů lze spatřovat hned několik. U případu rozdílu porozumění různým komunikátům u češtiny-slovenštiny a češtiny-polštiny/slovenštiny-polštiny vyzdvihneme například již v kapitole 7.5 zmiňovanou evropeizaci polštiny. Případ subkorpusu Subtitles si vysvětlujeme tím, že jde o polo-beletristický komunikát, který je z podstaty své funkce koncipovaný jako běžná mluvená řeč, a tak s sebou

nenese tolik stylových prvků jako ryze beletristické, administrativní či právnické komunikáty, které se objevovaly na okrajích měřeného spektra.

*

Do budoucna ovšem nabízí zde použitá metodologie pro západoslovanské jazyky mnohem více možností. Nabízí se možnost zkoumání rozdílů slov rozdělených do různých dalších kategorií (dle příbuznosti slov, dle jejich původu atp.), ale i možnost využití této metodologie na dalších jazykových rovinách. I na rovině fonetické však stále ještě alespoň části této metodologie nabízí další zajímavé záležitosti, kupříkladu pravděpodobnosti (frekvence) zarovnávání fonémů vůči sobě.

Za zmínku stojí i potenciální možnost zpřesnění výsledků měření pomocí úpravy dosud používaného vzorce pro výpočet podmíněné entropie. Kdyby vedle prosté pravděpodobnosti výskytu konkrétní zarovnané dvojice fonémů, $p(x, y)$, respektoval vzorec též jejich potenciální pravděpodobnost výskytu na základě předešlého fonému, tzn. *Markovovy řetězce*, jistě bychom se dobrali ještě přesnějších výsledků. Podobně bychom výsledky zpřesnili, kdybychom byli schopni zarovnávat fonémy k sobě tak, aby byla více respektována jejich vzájemná blízkost, což by ovšem znamenalo nutnost navrhnout jinou než Levenštejnovu metriku.

10 Závěr

Komunikace je pro lidské společenství nepostradatelnou složkou vzájemného soužití. Doby, kdy měly od sebe jednotlivé státy ostře oddělené hranice, jsou již i v našem středoevropském prostředí pryč, životní styl lidí ovládá globalizace a s ní také spojená potřeba komunikovat s mluvčími jiného jazyka.

Problematika vícejazyčné mezilidské komunikace, konkrétně mezi jazyky, které jsou si blízce příbuzné, se stala hlavním tématem této bakalářské práce. Ukázalo se totiž, že v takto blízkých jazykových společenstvech se setkáváme se specifickým způsobem komunikace, tzv. semikomunikací či receptivním bilingvismem, který spočívá v tom, že se spolu osoby dorozumívají každý za pomoci svého mateřského jazyka, a přitom si víceméně rozumí.

Abychom tuto problematiku mohli blíže zkoumat, představujeme (v kapitolách 2–3) nejdříve samotné pojetí, koncept komunikace, a to za pomoci tuzemským lingvistických tradic. Využíváme poznatků strukturalismu pražské, kodaňské i ženevské školy, neopomíjíme však ani znalosti moderní lingvistiky a jejích interdisciplinárních oborů. Současné studium komunikace navíc, víc než kdy jindy, vyžaduje též hledět na obory matematické a inženýrské, které na podkladech teorie informace a teorie komunikace nabízí různé způsoby exaktního nahlížení této problematiky.

Následujícími kapitolami (4–7) čtenáře seznamujeme s komunikačními jevy, které se dotýkají tématu vícejazyčné komunikace. Zda se při našem zkoumání opravdu jedná o dva různé jazyky, či nikoli, čtenáři pomůže pochopit kapitola 4. V ní popisujeme rozdíly mezi jazykem a jeho varietami, jejich vzájemné rozpoznávání pomocí konceptu Heinze Klose (ausbau a abstand jazyky) a různé typy příbuzností jazyků (genetickou, typologickou, areálovou).

Diglosie (kapitola 5) je jako komunikační jev pouze zmíněna a není středem našeho zájmu, neboť se v tomto případě jedná o vnitrojazykovou situaci, která se více než různých jazyků týká vnitřních variet jednoho jazyka.

Multilingvismus (kapitola 6) a všechny jeho podoby (bilingvismus, trilingvismus atd.) je již námi zkoumané problematice blíže. Tomu nasvědčuje i skutečnost, že způsob komunikace, kdy mezi sebou osoby mluví každý svým mateřským jazykem a rozumí si, bývá často nazýván receptivním multilingvismem.

Semikomunikace (kapitola 7) je středem našeho zájmu. Definování a upevňování terminologie s ní spojené s sebou ovšem neslo a dodnes nese mnohé nesnáze a pro námi nastíněný způsob komunikace (každý komunikant mluví vlastním mateřským jazykem) nebývá tento termín často užíván. V bakalářské práci je pojednána jak historie těchto peripetií definování semikomunikace, tak také jednotlivé faktory, které s ní neoddělitelně souvisí.

Závěrečná teoretická kapitola (8) všechny poznatky shrnuje a po vzoru lingvistických teoretických prací nachází pro semikomunikaci vhodné místo v rámci jazykovědných teorií a přístupů. Ukazuje se, že nelze nahlížet problematiku černobíle, neboť semikomunikace sice sama o sobě („čistá“) má s multilingvismem jen málo společného, ovšem každou skutečně semikomunikací se komunikanti zdokonalují, učí, čímž se konkrétně fenomén semikomunikace přibližuje multilingvistice na úrovni pasivního/receptivního. Mějme ovšem též na paměti, že při semikomunikaci se kódy nestřídají, nýbrž porovnávají (Stopyra 2010), což ji od multilingvistice opět oddaluje, nicméně ji do jisté míry předurčuje k tomu, aby mohla být zkoumána exaktně.

Poslední kapitola (9) využívá právě exaktního přístupu ve zkoumání semikomunikace, neboť obsahuje avizovaný vlastní, pro západoslovanské prostředí do jisté míry jedinečný, matematicko-lingvistický kvantitativní výzkum. Záměrně je realizovaný výzkum čistě kvantitativní bez přihlížení k sociolingvistickým faktorům (ty jsou všechny jmenovány a popsány v kapitole 7), neboť čistě sociolingvistických výzkumů bylo na rozdíl od kvantitativních mezi západoslovanskými jazyky, zejména mezi češtinou a slovenštinou, dosud realizováno mnoho.

Metodologicky vycházíme z výzkumu realizovaného na materiálech skandinávského jazykového prostředí, kde bývá semikomunikace skloňována nejčastěji. Naše realizace ale zpracovává západoslovanské jazyky (češtinu, slovenštinu a polštinu), a to z materiálů korpusu InterCorp verze 9 z roku 2016.

Výzkum je prováděn na fonetické jazykové rovině. Metodologie je navíc plně zautomatizována²⁵. Vytvořeny byly algoritmy pro třídění slov na plnovýznamová a neplnovýznamová, pro získávání překladových ekvivalentů z webové aplikace Treq ČNK, pro fonetické transkripce všech tří jmenovaných západoslovanských jazyků zvláště, pro zarovnání fonémů Levenštejnovou metrikou a také algoritmy pro výpočty pravděpodobností a samotné podmíněné entropie.

Jisté limity námi realizovaného výzkumu jsme zmínili v podkapitole 9.3, například neexistence unifikace fonetické transkripce slovanských jazyků, což si slavisté uvědomují sice již od dob Pražského lingvistického kroužku (Vachek 1970), nicméně doposud s tímto nebylo moc provedeno. Výsledky našeho výzkumu přesto často potvrzují sociolingvistické předpoklady a v mnohém je zpřesňují.

Lze konstatovat, že cíle bakalářské práce byly splněny. V kapitole 9.5 se nám podařilo nastínit i některé další možnosti využití měření podmíněné entropie mezi blízce příbuznými jazyky. Výsledky těchto výzkumů by mohly být prospěšné nejen lingvistice a jejím odvětvím obecně (teorie zejména obecné lingvistiky, metodologie matematické lingvistiky a sociolingvistiky, algoritmy strojové a jinak aplikované lingvistiky), ale takto zmapované jazykové prostředí by mohlo přispět také k dalšímu zkoumání jazykového vývoje a příbuznosti jazyků. V neposlední řadě lze tyto výzkumy považovat za základní (badatelské) výzkumy pro budoucí výuku blízce příbuzných jazyků.

²⁵ Algoritmy sice nejsou součástí bakalářské práce, ale jsou na vyžádání k dispozici.

Seznam obrázků

Obrázek 1: Model procesu komunikace mluvenou a psanou formou.....	13
Obrázek 2: Kódování jazyka vysílatelem	15
Obrázek 3: Dekódování jazyka příjemcem.....	18
Obrázek 4: Příklad diglosie.....	23
Obrázek 5: Příklad bilingvismu	25
Obrázek 6: Příklad semikomunikace.....	27
Obrázek 7: Levenštejnova metrika A.....	37
Obrázek 8: Levenštejnova metrika B.....	37
Obrázek 9: Příklad – zarovnané fonémy Levenštejnovou metrikou	38
Obrázek 10: Příklad – vypočtené podmíněné pravděpodobnosti.....	38
Obrázek 11: Doplnění překladových ekvivalentů a struktura databáze	39
Obrázek 12: Konečné počty zkoumaných slovních párů	40
Obrázek 13: Graf podmíněných entropií mezi češtinou a slovenštinou	42
Obrázek 14: Graf podmíněných entropií mezi češtinou a polštinou.....	43
Obrázek 15: Graf podmíněných entropií mezi slovenštinou a polštinou	44

Literatura

BALOWSKI, Mieczysław, 1993. *Fonetika a fonologie současné polštiny*. Praha: Karolinum. ISBN: 80-7066-793-1.

BAŇCZYK, Łucja, DYBALSKA, Renata Rusin, VAVŘÍN, Martin, 2016. *Korpus InterCorp – polština, verze 9 z 9. 9. 2016*. Praha: Ústav Českého národního korpusu FF UK. Dostupný z: <http://www.korpus.cz>.

BERMEL, Neil, 2010. O tzv. české diglosii v současném světě. *Slovo a slovesnost*. **70**(1), s. 5-30. ISSN: 0037-7031.

BÖRESTAMUHLMANN, Ulla, 2002. Interscandinavian Comprehension and Scandinavian Language Community: Real or Ideal? In: *Pour une modélisation de l'apprentissage simultané de plusieurs langues apparentées ou voisines* [online]. Nice: Publications de la Faculté des LASH de Nice, s. 129-136 [cit. 12.12.2016]. Dostupné z: <http://logatome.eu/pmasplav2001.pdf>.

BRAUNMÜLLER, Kurt, 2002. Semicommunication and Accommodation. *International Journal of Applied Linguistics*. **12**(1), s. 1-23. ISSN: 1473-4192.

BRAUNMÜLLER, Kurt, 2007. Receptive Multilingualism in Northern Europe in the Middle Ages. In: *Receptive Multilingualism*. Amsterdam: John Benjamins Publishing Company, s. 25-49. ISBN: 978-90-272-1926-8.

BUDOVIČOVÁ, Viera, 1974. Spisovné jazyky v kontakte. *Slovo a Slovesnost*. **35**(3), s. 171-181. ISSN: 0037-7031.

BUDOVIČOVÁ, Viera, 1987. Semikomunikácia ako lingvistický problém. *Studia Academica Slovaca*. **16**, s. 49-66.

CLYNE, Michael, 2003. Towards Inter-Cultural Communication in Europe Without Linguistic Homogenization. In: *The Cost of Multilingualism – Globalisation and Linguistic Diversity*. Vídeň: Verlag der Österreichischen Akademie der Wissenschaften, s. 39-48. ISBN: 978-3-7001-3183-0.

ČERMÁK, František, 1997. Obecná čeština je součástí české diglosie. *Jazykovědné aktuality*. **34**(3, 4), s. 34-43. ISSN: 1212-5326.

ČERMÁK, František, 2011. *Jazyk a jazykověda*. Praha: Karolinum. ISBN: 978-80-246-1946-0.

ČERMÁK, František, ROSEN, Alexandr, 2012. The case of InterCorp, a multilingual parallel corpus. In *International Journal of Corpus Linguistics*. **17**(3), s. 411-427. ISSN: 1384-6655.

ČERNÝ, Jiří, 1996. *Dějiny lingvistiky*. Olomouc: Votobia. ISBN: 80-85885-96-4.

ČERNÝ, Jiří, 1998. *Úvod do studia jazyka*. Olomouc: Rubico. ISBN: 80-85839-24-5.

DOBROTOVÁ, Ivana, MUSILOVÁ, Květoslava, 2012. Z dějin kontaktu dvou blízkých slovanských jazyků. *Bohemistika*. **2012**(1), s. 35-60. ISSN: 1642-9893.

DUDÁŠOVÁ-KRIŠŠÁKOVÁ, Júlia, 1999. Fonologický systém spisovnej slovenčiny a polštiny z typologického hľadiska. *Slavica Slovaca*. **34**(1), s. 16-24. ISSN: 0037-6787.

FERGUSON, Charles A., 1959. Diglosia. In: *Antológia bilingvizmu*. Bratislava: Academic Electronic Press, s. 101-113. ISBN: 80-88880-54-8.

FISHMAN, Joshua A., 1967. Bilingualism With and Without Diglossia; Diglossia With and Without Bilingualism. *Journal of Social Issues*. **23**(2), s. 29-38. ISSN: 1540-4560.

FISHMAN, Joshua A., 2008. Rethinking the Ausbau-Abstand Dichotomy into a Continuous and Multivariate System. *International Journal of the Sociology of Language*. **2008**(191), s. 17-26. ISSN: 1613-3668.

GOLUBOVIC, Jelena, 2016. *Mutual intelligibility in the Slavic language area*. Groningen. Disertační práce. University of Groningen. Vedoucí práce John Nerborne, Vincent van Heuven a Charlotte Gooskens.

GOOSKENS, Charlotte, 2006. Linguistic and Extra-Linguistic Predictors of Inter-Scandinavian Intelligibility. In: *Linguistics in the Netherlands 2006*. Amsterdam: John Benjamins Publishing Company, s. 101-113. ISBN: 978-9-027-23166-6.

GOOSKENS, Charlotte, 2007. The Contribution of Linguistic Factors to the Intelligibility of Closely Related Languages. *Journal of Multilingual and Multicultural Development*. **28**(6), s. 445-467. ISSN: 0143-4632.

- GOOSKENS, Charlotte, KÜRSCHNER, Sebastian, BEZOOIJEN van, Renée, 2012. Intelligibility of Swedish for Danes: loan words compared with inherited words. In: *Language for its own sake*. Amsterdam: Amsterdam Contributions to Scandinavian Studies, s. 435-455. ISBN: 978-90-809186-0-3.
- GREŇ, Zbigniew, 2011. Tożsamość na pograniczu blisko spokrewnionych etnosów – przypadek w trójkącie polsko-czesko-słowackim. *Prace komisji środkowoeuropejskiej pau*. **19**, s. 57-73. ISSN: 1233-0558.
- HAUGEN, Einar, 1966. Semicommunication: The Language Gap in Scandinavia. In: *Studies by Einar Haugen: Presented on the occasion of his 65th birthday, April 19, 1971*. Haag: Mouton 1972, s. 479-496.
- HERNANDEZOVÁ, Pavla, SUCHÁ, Karolína, 2014. Bilingvismus, diglosie a „smíšené“ jazyky. In: *Encyklopedie lingvistiky* [online]. Olomouc: Univerzita Palackého v Olomouci, 25. 10. 2014. [cit. 20. 11. 2016]. Dostupné z: http://oltk.upol.cz/encyklopedie/index.php5/Bilingvismus_diglosie_a_smíšené_jazyky.
- HJELMSLEV, Louis, 1971. *Jazyk*. Praha: Academia.
- HJELMSLEV, Louis, 1972. *O základech teorie jazyka*. Praha: Academia.
- HORECKÝ, Ján, 1995. Slovensko-česká diglosná komunikácia. In: *Sociolinguistica Slovaca*. Bratislava: Vydavateľstvo slovenskej akadémie vied, 1995, s. 183-188. ISBN: 80-224-0160-9.
- HRONEK, Jiří a SGALL, Petr, 2014. *Čeština bez příkras*. Praha: Karolinum. ISBN: 978-80-246-2459-4.
- CHAMBERS, Jack K. a TRUDGILL, Peter, 1998. *Dialectology*. Cambridge: Cambridge University Press, 1998. ISBN: 0-521-59378-6.
- KAJANOVÁ-SCHULZOVÁ, Ol'ga, 1970. *Úvod do fonetiky slovenčiny*. Bratislava: Slovenské pedagogické nakladateľstvo.
- KLOSS, Heinz, 1967. Abstand languages and ausbau languages. *Anthropological linguistics*. **9**(7), s. 29-41. ISSN: 0003-5483.
- KLOSS, Heinz, 1976. O ‚diglosii‘. In: *Antológia bilingvizmu*. Bratislava: Academic Electronic Press, s. 129-138. ISBN: 80-88880-54-8.
- KOŘENSKÝ, Jan, 1992. *Komunikace a čeština*. Praha: H&H. ISBN: 80-85467-92-5.
- KRAUS, Jiří, 2005. *Nový akademický slovník cizích slov: A-Ž*. Praha: Academia. ISBN: 80-200-1351-2.
- KRÁL, Ābel, SABOL Ján, 1989. *Fonetika a fonológia*. Bratislava: Slovenské pedagogické nakladateľstvo. ISBN: 80-08-00036-8.
- KRYSTÝNKOVÁ, Jarmila, 1995. K problematice pozitivního a negativního transferu při vyučování ruskému jazyku. *Opera Slavica*. **5**(4). ISSN: 1211-7676.
- LEHISTEOVÁ, Ilse, 1988. Pojem interference. In: *Antológia bilingvizmu*. Bratislava: Academic Electronic Press, s. 129-138. ISBN: 80-88880-54-8.
- LIPOWSKI, Jaroslav, 2009. Percepčné dôsledky slovensko-pol'sko-českých fonetických transferov. In: *Od-kazy a výzvy modernej jazykovej komunikácie*. Banská Bystrica: Univerzita Mateja Bela v Banskej Bystrici, s. 72-84. ISBN: 978-80-8083-960-4.
- LIPOWSKI, Jaroslav, 2011. *Operatívna fonetika slovenčiny, češtiny a polštiny*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego. ISBN: 978-80-7294-511-5.
- LOTKO, Edvard, 1987. O zrádných slovech v polštině a češtině. In: *Srovnávací a bohemistické studie*. Olomouc: Vydavatelství Univerzity Palackého, s. 147-155. ISBN: 978-80-244-2201-5.
- LOTKO, Edvard, 1998. Čeština a polština. In: *Srovnávací a bohemistické studie*. Olomouc: Vydavatelství Univerzity Palackého, s. 63-75. ISBN: 978-80-244-2201-5.
- LOTKO, Edvard, 1999. Ke konfrontaci příbuzných jazyků. In: *Srovnávací a bohemistické studie*. Olomouc: Vydavatelství Univerzity Palackého, s. 9-19. ISBN: 978-80-244-2201-5.
- MAREŠ, Petr, 2014. *Úvod do lingvistiky a lingvistické bohemistiky*. Praha: Karolinum. ISBN: 978-80-246-2640-6.
- MATHESIUS, Vilém, 1911. O potenciálnosti jevů jazykových. In: VACHEK, Josef, 1970. *U základů pražské jazykovědné školy*. Praha: Academia.
- MITTER, Patrik, 2007. Srovnání recepce slovenštiny v ČR a češtiny v SR. *Jazykovědné aktuality*. **44**, s. 45-54. ISSN: 1212-5326.
- MOBERG, Jens, GOOSKENS, Charlotte, NERBONNE, John, VAILLETTE, Nathan, 2007. Conditional Entropy Measures Intelligibility among Related Languages. In: *Proceedings of the 17th Meeting of Computational Linguistics in the Netherlands*. Utrecht: LOT, s. 51-67. ISBN: 978-90-78328-41-4.
- MUSILOVÁ, Květoslava, 2000. Česko-slovenský pasivní bilingvismus. In: *Mesto a jeho jazyk*. Bratislava: Vydavateľstvo slovenskej akadémie vied, s. 280-289. ISBN: 80-224-0605-8.

- NÁBĚLKOVÁ, Mira, 2008. *Slovenčina a čeština v kontakte*. Bratislava: SAV. ISBN: 978-80-224-1060-1.
- NÁBĚLKOVÁ, Mira, VAVŘÍN, Martin, 2016. *Korpus InterCorp – slovenština, verze 9 z 9. 9. 2016*. Praha: Ústav Českého národního korpusu FF UK. Dostupný z: <http://www.korpus.cz>.
- NEBESKÁ, Iva, 1992. *Úvod do psycholingvistiky*. Praha: H&H. ISBN: 80-85467-75-5.
- NEKVAPIL, Jiří, SLOBODA, Marián, WAGNER, Petr, 2009. *Mnohojazyčnost v České republice*. Praha: Lidové noviny, 2009. ISBN: 978-80-7106-581-4.
- PALKOVÁ, Zdena, 1994. *Fonetika a fonologie češtiny*. Praha: Karolinum. ISBN: 80-7066-843-1.
- PAULINY, Eugen, 1979. *Slovenská fonológia*. Bratislava: Slovenské pedagogické nakladateľstvo.
- RINDLER-SCHJERVE, Rosita, VETTER, Eva, 2007. Linguistic Diversity in Hamburg Austria as a Model for Modern European Language Policy. In: *Receptive Multilingualism*. Amsterdam: John Benjamins Publishing Company, s. 49-73. ISBN: 978-90-272-1926-8.
- ROSEN, Alexandr, VAVŘÍN, Martin, 2016. *Korpus InterCorp – čeština, verze 9 z 9. 9. 2016*. Praha: Ústav Českého národního korpusu FF UK. Dostupný z: <http://www.korpus.cz>.
- SAUSSURE, Ferdinand de, 1989. *Kurs obecné lingvistiky*. Praha: Odeon. ISBN: 80-207-0070-6.
- SEDLAČÍKOVÁ, Blanka, 2010. *Historie matematické lingvistiky*. Brno. Disertační práce. Masarykova univerzita. Přírodovědecká fakulta. Vedoucí práce Eduard Fuchs.
- SGALL, Petr, 2011. *Jazyk, mluvení, psaní*. Praha: Karolinum. ISBN: 978-80-246-1903-3.
- SCHÜPPERT, Anja, GOOSKENS, Charlotte, 2011. Investigating the Link Between Language Attitudes and Intelligibility Using Reaction Time. *Dialectologia*. 7, s. 119-140. ISSN: 2013-2247.
- SCHWARZOVÁ, Monika, 2009. *Úvod do kognitivní lingvistiky*. Praha: Dauphin. ISBN: 978-80-7272-155-9.
- SLOBODA, Marián, 2004. Slovensko-česká (semi)komunikace a vzájemná (ne)srozumitelnost. *Čeština doma a ve světě*. 12(3, 4), s. 208-220. ISSN: 1210-9339.
- SOKOLOVÁ, Miloslava, MUSILOVÁ, Květoslava, SLANČOVÁ, Daniela, 2005. *Slovenčina a čeština*. Bratislava: Univerzita Komenského v Bratislave. ISBN: 80-223-2358-1.
- STERNBERG, Robert J., 2002. *Kognitivní psychologie*. Praha: Portál. ISBN: 80-7178-376-5.
- STOPYRA, Janusz, 2010. Semikomunikacja w skandynawskim obszarze językowym. In: *Biuletyn polskiego towarzystwa językoznawczego*. 66, s. 133-141. ISSN: 0032-3802.
- SVOBODOVÁ, Jana, 2006. K pasivnímu bilingvismu dětí v česko-slovenském prostoru. In: *Slovo o slove*. Prešov: Prešovská univerzita v Prešově, s. 17-23. ISBN: 80-8068-491-X.
- ŠKRABAL, Michael, 2017. Korpus InterCorp verze 9. In: *WIKI Český národní korpus* [online]. [cit. 05.02.2013]. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:intercorp:verze9>.
- TRUDGILL, Peter, 1986. *Dialects in contact*. Oxford: Basil Blackwell. ISBN: 0-631-12691-0.
- VACHEK, Josef, 1970. *U základů pražské jazykovědné školy*. Praha: Academia.
- VYBÍRAL, Zbyněk, 2009. *Psychologie komunikace*. Praha: Portál. ISBN: 978-80-7367-387-1.
- WEI, Li, 2000. Dimenzie bilingvizmu. In: *Antológia bilingvizmu*. Bratislava: Academic Electronic Press, s. 129-138. ISBN: 80-88880-54-8.
- WEINREICH, Uriel, 1953. Jazyky v kontakte. In: *Antológia bilingvizmu*. Bratislava: Academic Electronic Press, s. 129-138. ISBN: 80-88880-54-8.
- WODE, Henning, 1993. Základné pojmy. In: *Antológia bilingvizmu*. Bratislava: Academic Electronic Press, s. 129-138. ISBN: 80-88880-54-8.
- ZEEVAERT, Ludger, 2007. Receptive Multilingualism and Inter-Scandinavian Semicommunication. In: *Receptive Multilingualism*. Amsterdam: John Benjamins Publishing Company, s. 103-137. ISBN: 978-90-272-1926-8.
- ZEEVAERT, Ludger a THIJE Ten D., Jan, 2007. Introduction. In: *Receptive Multilingualism*. Amsterdam: John Benjamins Publishing Company, s. 1-25. ISBN: 978-90-272-1926-8.
- ZEMAN, Jiří, 2008. *Základy české ortoepie*. Hradec Králové: Gaudeamus. ISBN: 978-80-7041-778-2.

Příloha A

Tabulka 1: Hodnoty podmíněných entropií a asymetrie mezi češtinou a slovenštinou

Subkorporus a kategorie		Podmíněná entropie		Asymetrie
		$H(CS SK)$	$H(SK CS)$	$\frac{H(CS SK) - H(SK CS)}{H(CS SK)}$
Acquis	vše	1,5569	1,4281	0,0827
	plnovýznamové	1,5149	1,3808	0,0885
	neplnovýznamové	1,7170	1,9005	- 0,1069
Core	vše	1,7626	1,6901	0,0411
	plnovýznamové	1,6976	1,6135	0,0495
	neplnovýznamové	1,8753	2,1292	- 0,1354
Europarl	vše	1,4853	1,3509	0,0905
	plnovýznamové	1,4259	1,2850	0,0988
	neplnovýznamové	1,8513	2,1719	- 0,1732
Subtitles	vše	1,7205	1,6251	0,0554
	plnovýznamové	1,6866	1,5829	0,0615
	neplnovýznamové	1,4478	1,5444	- 0,0667

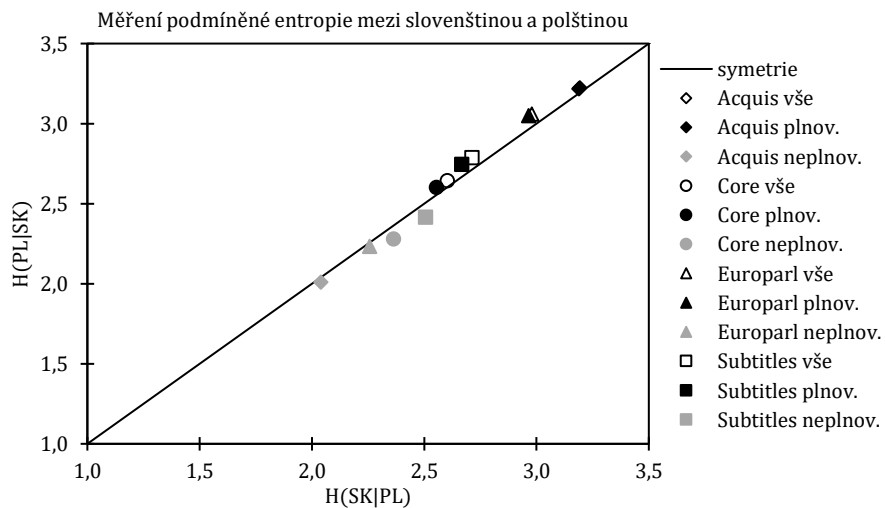
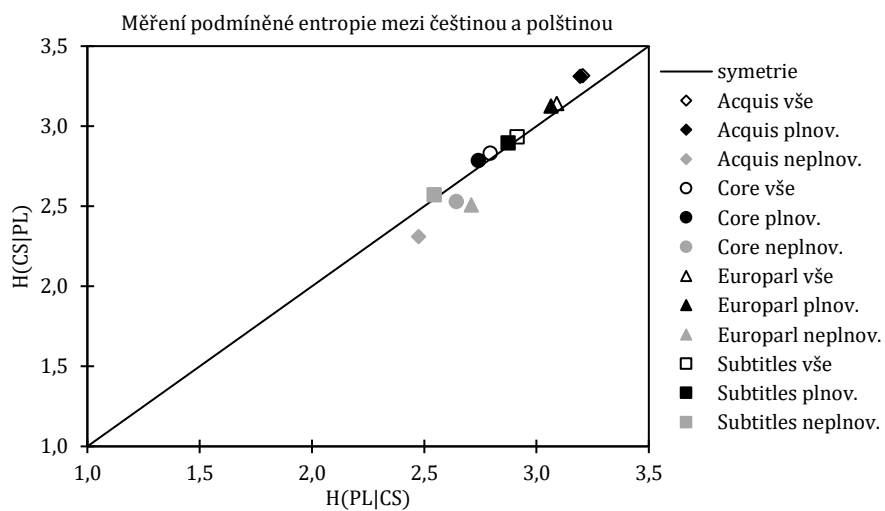
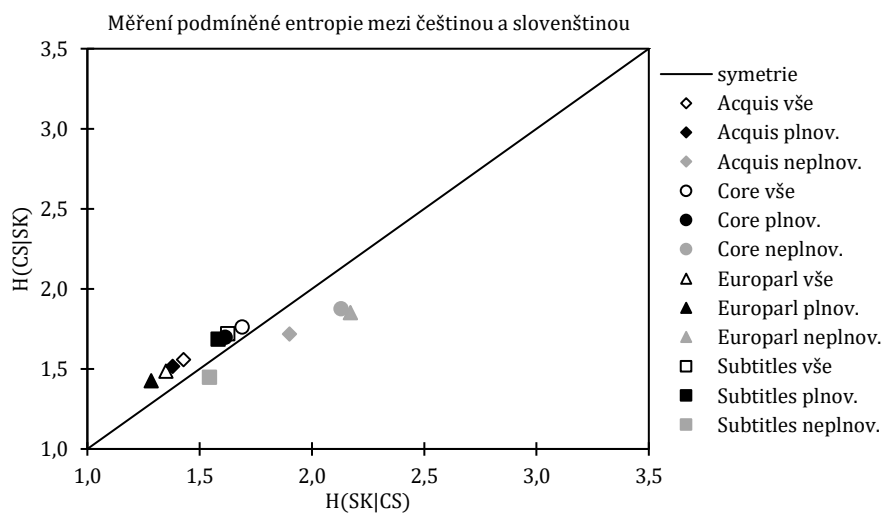
Tabulka 2: Hodnoty podmíněných entropií a asymetrie mezi češtinou a polštinou

Subkorporus a kategorie		Podmíněná entropie		Asymetrie
		$H(CS PL)$	$H(PL CS)$	$\frac{H(CS PL) - H(PL CS)}{H(CS PL)}$
Acquis	vše	3,3167	3,2059	0,0334
	plnovýznamové	3,3124	3,1940	0,0357
	neplnovýznamové	2,3098	2,4756	- 0,0718
Core	vše	2,8313	2,7942	0,0131
	plnovýznamové	2,7862	2,7407	0,0163
	neplnovýznamové	2,5305	2,6426	- 0,0443
Europarl	vše	3,1431	3,0904	0,0168
	plnovýznamové	3,1262	3,0652	0,0195
	neplnovýznamové	2,5088	2,7095	- 0,0800
Subtitles	vše	2,9344	2,9137	0,0071
	plnovýznamové	2,8970	2,8738	0,0080
	neplnovýznamové	2,5727	2,5451	- 0,0107

Tabulka 3: Hodnoty podmíněných entropií a asymetrie mezi slovenštinou a polštinou

Subkorporus a kategorie		Podmíněná entropie		Asymetrie
		$H(SK PL)$	$H(PL SK)$	$\frac{H(PL SK) - H(SK PL)}{H(PL SK)}$
Acquis	vše	3,1926	3,2221	0,0092
	plnovýznamové	3,1887	3,2175	0,0090
	neplnovýznamové	2,0398	2,0102	- 0,0147
Core	vše	2,6028	2,6425	0,0150
	plnovýznamové	2,5543	2,6003	0,0177
	neplnovýznamové	2,3639	2,2791	- 0,0372
Europarl	vše	2,9789	3,0599	0,0265
	plnovýznamové	2,9647	3,0485	0,0275
	neplnovýznamové	2,2580	2,2322	- 0,0116
Subtitles	vše	2,7134	2,7880	0,0268
	plnovýznamové	2,6667	2,7460	0,0289
	neplnovýznamové	2,5057	2,4151	- 0,0375

Příloha B



Příloha C

jazyk	transkripce	zarovnnání	jazyk	transkripce	zarovnnání
česky	sɛ	s ɛ	česky	smlɔvɪ	# s m l ɔv v ɪ
polsky	z	# z	polsky	ʔumɔvi	ʔ u m # ɔ v i
česky	jɛ	j ɛ # #	česky	naʀi:zɛɲi:	# # # n a ɾ i: z ɛ ɲ i:
polsky	jest	j ɛ s t	polsky	ɾɔspɔʒɔŋdʒɛɲa	ɾ ɔ s p ɔ ʒ ɔŋ dʒ ɛ ɲ a
česky	tɛla:nɛk	tɛ a:n n ɛ k	česky	sɔɣladu	s ɔɣ a d # u
polsky	ʔart	# # ʔ a r t	polsky	zgɔdɲɛ	# z g ɔ d ɲ ɛ
česky	ʔɛs	ʔ ɛ s	česky	tɔɦɔɔ	# # # # # # t ɔ ɦ ɔ t ɔ
polsky	ʔɛs	ʔ ɛ s	polsky	ɲiɲɛjʃɛgɔ	ɲ i ɲ ɛ j ʃ ɛ g ɔ # # # #
česky	ma:	m a:	česky	zejmɛ:na	# z ɛ j m ɛ:n # # # a
polsky	ma	m a	polsky	ʃfɛgɔɫnɔɛtɛi	ʃ f ɛ g ɔ ɫ n ɔ ɛ tɛ i
česky	dɲɛ	d ɲ ɛ	česky	ʔɛxs	ʔ ɛ x s
polsky	dɲa	d ɲ a	polsky	ʔɛfk	ʔ ɛ f k
česky	bɪ:t	b i: t	česky	tɛ	# # tɛ
polsky	bɪtɛ	b i tɛ	polsky	ʔart	ʔ a r t
česky	spɔlɛtɛɛnstvi:	# # s p ɔ l ɛ tɛ ɛ n s t v i:	česky	tɔmu	# # # # # # t ɔ m u
polsky	ʔɛurɔpɛjskɔɣ	ʔ ɛ u r ɔ p ɛ # j s # k ɔɣ	polsky	pʃɛvidʒanɛ	p ʃ ɛ v i dʒ a n ɛ
česky	jsɔɣ	j s ɔɣ	česky	sta:tɪ	# # # # # # s t a:t i
polsky	sɔɣ	# s ɔɣ	polsky	ʃwɔnkɔfskjɛ	ʃ w ɔ n k ɔ f s # k j ɛ
česky	bɪ	b i	česky	tɛlɛnskɛ:	tɛ ɛ n s # k ɛ:
polsky	bɪ	b i	polsky	paɲstʃa	# p a ɲ s t ʃ a
česky	ʔɔtst	ʔ ɔ t s t	česky	ʔa	ʔ a
polsky	ʔust	ʔ # u s t	polsky	ʔi	ʔ i
česky	naʀi:zɛɲi:	# # # n a ɾ i: z ɛ ɲ i:	česky	musi:	m u s i:
polsky	ɾɔspɔʒɔŋdʒɛɲa	ɾ ɔ s p ɔ ʒ ɔŋ dʒ ɛ ɲ a	polsky	muɕi	m u ɕ i
česky	radɪ	r a d i	česky	smɲɛrɲutɛ	s m ɲ ɛ r ɲ ɪ tɛ ɛ # # # #
polsky	radɨ	r a d i i	polsky	dɪrɛktɪvi	# # d i i r # # # ɛ k t i v i i
česky	sɛ	s ɛ	česky	ktɛrɛ:	k t ɛ r ɛ:
polsky	z	# z	polsky	kturɛ	k t u r ɛ
česky	kɔmɪsɛ	k ɔ m i s # ɛ	česky	ʔɛɣr	ʔ # ɛɣ r
polsky	kɔmisji	k ɔ m i s j i i	polsky	ʔɛur	ʔ ɛ u r
česky	ʔɔɦlɛdɛm	ʔ # # ɔ ɦ ɛ d # # # ɛ m	česky	prɔduktu:	p ɾ ɔ d u k t # u:
polsky	ʔuvzɣlɛŋdɲajɔŋtɔ	ʔ u v z ɣ l ɛŋ d ɲ a j ɔŋ tɔ	polsky	prɔduktuf	p ɾ ɔ d u k t u f

jazyk	transkripce	zarovnáání	jazyk	transkripce	zarovnáání
česky	<i>mu:ze</i>	m u: z ε	česky	<i>pɔtpɔrɪ</i>	p ɔ t p ɔ r ɪ
polsky	<i>mɔʒɛ</i>	m ɔ ʒ ε	polsky	<i>pɔmɔʦɪ</i>	p ɔ # m ɔ ʦ ɪ
česky	<i>tɛ:tɔ</i>	# # # t ε: t ɔ	česky	<i>pʃɛtpɪsʊ:</i>	p ʃ ɛ t p ɪ s ʊ:
polsky	<i>niɲɛjʃɛj</i>	ɲ i ɲ ɛ j ʃ ɛ j	polsky	<i>pʃɛpɪsʊf</i>	p ʃ ɛ # p ɪ s ʊ f
česky	<i>za:klaʝɛ</i>	# z a: k l a # ʝ ɛ	česky	<i>tatɔ</i>	t a t ɔ
polsky	<i>pɔʦʦavʝɛ</i>	p ɔ t s t a v ʝ ɛ	polsky	<i>tɛ</i>	t # # ɛ
česky	<i>kʦɛrɪ:m</i>	# k t ɛ # # r ɪ: m	česky	<i>maji:</i>	m a j i:
polsky	<i>zmjɛnaʝɔɲʦɛ</i>	z m j ɛ ɲ a j ɔɲ ʦ ɛ	polsky	<i>majɔɯ</i>	m a j ɔɯ
česky	<i>pʃi:paʝɛ</i>	p ʃ i: p a # ʝ ɛ	česky	<i>ʔɛvrɔpskɛ:ɦɔ</i>	ʔ ɛ v r ɔ p # # s k # ɛ: ɦ ɔ
polsky	<i>pʃɪpatku</i>	p ʃ i p a t k ʊ	polsky	<i>ʔɛvrɔpɛjskʝɛɡɔ</i>	ʔ ɛ ʊ r ɔ p ɛ j s k j ɛ ɡ ɔ
česky	<i>zalnɔzɛɲi:</i>	# z a ɔ z ɛ ɲ i:	česky	<i>pʃɪjɔla</i>	p ʃ i j # # a a
polsky	<i>ʔʊʦʦɛɲa</i>	ʔ ʊ ʦ ʦ ɔ z ɛ ɲ a	polsky	<i>pʃɪjɛwa</i>	p ʃ i j ɛ w a # #
česky	<i>stanɔvi:</i>	# # s t a n ɔ v # i:	česky	<i>rɔku</i>	r ɔ k ʊ
polsky	<i>ʔʊstanavja</i>	ʔ ʊ s t a n a v j a	polsky	<i>rɔku</i>	r ɔ k ʊ
česky	<i>pʃi:lɔzɛ</i>	p ʃ i:l ɔ z # # # # # ɛ	česky	<i>ʔɔblasɔɪ</i>	ʔ ɔ b l a s c ɪ
polsky	<i>zawɔɲʦɪɲiku</i>	# # # # # z a w ɔɲ ʦ ɪ k ʊ	polsky	<i>ɖɛɖɪɲɛ</i>	# # ɖɛ ɛ ɖɛ i ɲ ɛ
česky	<i>tɔtɔ</i>	# # # t ɔ t ɔ	česky	<i>lɛdna</i>	# l ɛ d n a
polsky	<i>niɲɛjʃɛ</i>	ɲ i ɲ ɛ j ʃ ɛ	polsky	<i>stɪʦɪɲa</i>	s t ɪ ʦ ɲ a
česky	<i>mɔɦɔɯ</i>	m ɔ ɦ ɔɯ	česky	<i>nɛjsɔɯ</i>	n ɛ j s ɔɯ
polsky	<i>mɔɡɔɯ</i>	m ɔ ɡ ɔɯ	polsky	<i>ɲɛ</i>	ɲ ɛ # # #
česky	<i>ra:mtsɪ</i>	r a: m ʦ ɪ	česky	<i>vi:bɔru</i>	v i: b ɔ # # # r ʊ
polsky	<i>ramax</i>	r a m a x	polsky	<i>kɔmitɛtu</i>	# # k ɔ m i t ɛ t ʊ
česky	<i>naɦrazʊɛ</i>	n a ɦ r a z ʊ j ɛ	česky	<i>budɛ</i>	b ʊ d ɛ
polsky	<i>ʔɔɖʒɪmuʝɛ</i>	ʔ ɔ d ʒ i m ʊ j ɛ	polsky	<i>bɛɲɖɛ</i>	b ɛɲ ɖɛ ɛ
česky	<i>bɪlɔ</i>	b ɪ ɔ # # # #	česky	<i>zvi:ʃat</i>	z v # i: ʃ a t
polsky	<i>zɔstawɔ</i>	# # z ɔ s t a w ɔ	polsky	<i>zvʝɛʒɔɲt</i>	z v j ɛ ʒ ɔɲ t
česky	<i>pɔʦʦupɛm</i>	p # ɔ # s t ʊ p ɛ m	česky	<i>prɔsɪntɛ</i>	p r ɔ s ɪ n ʦ ɛ
polsky	<i>prɔʦɛdurɔɯ</i>	p r ɔ ʦ ɛ d ʊ # r ɔɯ	polsky	<i>ɡrudɲa</i>	ɡ r # # ʊ d ɲ a
česky	<i>naʃi:zɛɲi:</i>	# # # n a ʃ i: z ɛ ɲ i:	česky	<i>platnɔst</i>	p l a t n ɔ s t
polsky	<i>rɔʦpɔʒɔɲɖɛɲa</i>	r ɔ s p ɔ ʒ ɔɲ ɖɛ ɲ a	polsky	<i>ʒɪʦɛ</i>	# # # # ʒ i ʦ ɛ
česky	<i>pɔɖmi:nki</i>	p ɔ d m i: n k ɪ	česky	<i>spɔlɛʦɛnstvi:</i>	# # s p ɔ ɛ ʦɛ ɛ n s t v i:
polsky	<i>varunki</i>	# v a r ʊ n k i	polsky	<i>ʔɛvrɔpɛjskɔɯ</i>	ʔ ɛ ʊ r ɔ p ɛ # j s # k ɔɯ
česky	<i>vi:ɕɛ</i>	v i: ɕ ɛ #	česky	<i>ʔɛvrɔpski:x</i>	ʔ ɛ v r ɔ p # # s k i:x
polsky	<i>viʒɛj</i>	v i ʒ ɛ j	polsky	<i>ʔɛvrɔpɛjskix</i>	ʔ ɛ ʊ r ɔ p ɛ j s k i x
česky	<i>rɔʒɦɔɖnuci:</i>	r ɔ z ɦ ɔ d # # n ʊ c i:	česky	<i>mɲɛɲi:</i>	# m ɲ ɛ ɲ i:
polsky	<i>ɖɛʦɪʒi</i>	# # # # ɖ ɛ ʦ i ʒ i i	polsky	<i>zmjani</i>	z m j a n i
česky	<i>ɡɔɪ</i>	ɡ d ɪ	česky	<i>tʃɛba</i>	# t ʃ ɛ b a
polsky	<i>ɡɔɪ</i>	ɡ d i	polsky	<i>nalɛʒɪ</i>	n a ɛ ʒ i