

**ROMÁNSKÉ  
JAZYKY A ČEŠTINA  
VE SVĚTLE  
PARALELNÍCH  
KORPUSŮ**

**PETR ČERMÁK  
OLGA NÁDVORNÍKOVÁ  
ET ALII**

**KAROLINUM**

## Románské jazyky a čeština ve světle paralelních korpusů

Petr Čermák, Olga Nádvorníková et alii

---

Recenzovali:

prof. PhDr. František Čermák, DrSc.

prof. PhDr. Jiří Černý, CSc.

Vydala Univerzita Karlova v Praze, Nakladatelství Karolinum

Redakce Zuzana Leštinová

Grafická úprava Jan Šerých

Sazba DTP Nakladatelství Karolinum

První vydání

© Univerzita Karlova v Praze, 2015

© Petr Čermák, Olga Nádvorníková et alii, 2015

ISBN 978-80-246-2872-1

ISBN 978-80-246-2914-8 (online : pdf)



Univerzita Karlova v Praze  
Nakladatelství Karolinum 2016

[www.karolinum.cz](http://www.karolinum.cz)  
[ebooks@karolinum.cz](mailto:ebooks@karolinum.cz)



## OBSAH

- 1. SROVNÁVÁNÍ JAZYKŮ A PARALELNÍ KORPUSY ---- 11**
  - 1.1 Románské jazyky a čeština, hledání strukturních rozdílů ---- 12
  - 1.2 Metodologie vytěžování paralelních korpusů. Paralelní korpus *InterCorp* ---- 16
  - 1.3 Charakteristika jednotlivých subkorpusů ---- 25
    - 1.3.1 Španělský subkorpus ---- 25
    - 1.3.2 Italský subkorpus ---- 26
    - 1.3.3 Portugalský subkorpus ---- 26
    - 1.3.4 Francouzský subkorpus ---- 26
  
- 2. ROMÁNSKÁ KONSTRUKCE HACER/FARE/FAIRE/FAZER + INFINITIV A JEJÍ ČESKÉ PROTĚJŠKY ---- 27**
  - 2.1 Kauzativnost v románských jazycích a v češtině ---- 28
    - 2.1.1 Kauzativnost a její definice ---- 28
    - 2.1.2 Kauzativnost v románských jazycích ---- 31
    - 2.1.3 Kauzativnost v češtině ---- 31
    - 2.1.4 Typologie českých protějšků konstrukce *hacer* + infinitiv, *fare* + infinitiv, *faire* + infinitiv, *fazer* + infinitiv ---- 33
    - 2.1.5 Metodologie a postup ---- 37
  - 2.2 Španělština: konstrukce *hacer* + infinitiv a její protějšky v češtině ---- 38
    - 2.2.1 Prostředky vyjadřování kauzativnosti ve španělštině ---- 38
    - 2.2.2 K vlastnostem konstrukce *hacer* + infinitiv ---- 39
    - 2.2.3 Analýza výsledků vyhledávání v španělsko-českém korpusu ---- 41
      - 2.2.3.1 Komentář k jednotlivým typům ---- 46
      - 2.2.3.2 Možnost vkládat další slovo mezi pomocné a významové sloveso ---- 49
  - 2.3 Itálština: konstrukce *fare* + infinitiv a její protějšky v češtině ---- 50
    - 2.3.1 Prostředky vyjadřování kauzativnosti v italštině ---- 50
    - 2.3.2 K vlastnostem konstrukce *fare* + infinitiv ---- 51
    - 2.3.3 Analýza výsledků vyhledávání v italsko-českém korpusu ---- 52
      - 2.3.3.1 Komentář k jednotlivým typům ---- 56
      - 2.3.3.2 Možnost vkládat další slovo mezi pomocné a významové sloveso ---- 59
  - 2.4 Portugalština: konstrukce *fazer* + infinitiv a její protějšky v češtině ---- 59
    - 2.4.1 Prostředky vyjadřování kauzativnosti v portugalštině ---- 59
    - 2.4.2 K vlastnostem konstrukce *fazer* + infinitiv ---- 60
    - 2.4.3 Analýza výsledků vyhledávání v portugalsko-českém korpusu ---- 61
      - 2.4.3.1 Komentář k jednotlivým typům ---- 66

- 2.4.3.2 Možnost vkládat další slovo mezi pomocné a významové sloveso ---- **69**
- 2.5 Francouzština: konstrukce *faire* + infinitiv a její protějšky v češtině ---- **69**
  - 2.5.1 Prostředky vyjadřování kauzativnosti ve francouzštině ---- **69**
  - 2.5.2 K vlastnostem konstrukce *faire* + infinitiv ---- **70**
  - 2.5.3 Analýza výsledků vyhledávání ve francouzsko-českém korpusu ---- **71**
    - 2.5.3.1 komentář k jednotlivým typům ---- **76**
    - 2.5.3.2 Možnost vkládat další slovo mezi pomocné a významové sloveso ---- **79**
- 2.6 Románské kauzativní konstrukce ve světle korpusu *InterCorp* ---- **80**
  - 2.6.1 České protějšky zkoumané konstrukce ---- **80**
  - 2.6.2 Vlastnosti románské konstrukce ---- **84**

### 3. MORFOLOGICKY KOMPLEXNÍ SLOVA V ROMÁNSKÝCH JAZYCÍCH

- A JEJICH PROTĚJŠKY V ČEŠTINĚ ---- 87**
- 3.0 Afixy *ri-/re-* a *-bile/-ble/-vel* a jejich protějšky v češtině ---- **88**
- 3.1 Úvodem ---- **88**
- 3.2 Slovtvorba – definice komplexních slov ---- **88**
- 3.3 Konkrétní příklady ---- **89**
  - 3.3.1 Iterativní prefix *ri-/re-* ---- **89**
  - 3.3.2 Modální sufix *-bile/-ble/-vel* ---- **90**
- 3.4 Typologie českých protějšků ---- **90**
  - 3.4.1 Prefix *ri-/re-* ---- **90**
  - 3.4.2 Sufix *-bile/-ble/-vel* ---- **91**
- 3.5 Italská afixy *ri-* a *-bile* ---- **92**
  - 3.5.1 Prefix *ri-* ---- **92**
    - 3.5.1.1 Obecné charakteristiky ---- **92**
    - 3.5.1.2 Základní výsledky vyhledávání ---- **92**
    - 3.5.1.3 Interpretace některých sloves ---- **95**
  - 3.5.2 Sufix *-bile* ---- **97**
    - 3.5.2.1 Obecné charakteristiky ---- **97**
    - 3.5.2.2 Základní výsledky vyhledávání ---- **97**
    - 3.5.2.3 Interpretace některých adjektiv ---- **101**
- 3.6 Španělština: afixy *re-* a *-ble* ---- **102**
  - 3.6.1 Prefix *re-* ---- **103**
    - 3.6.1.1 Obecné charakteristiky ---- **103**
    - 3.6.1.2 Základní výsledky vyhledávání ---- **103**
    - 3.6.1.3 Interpretace některých sloves ---- **106**
  - 3.6.2 Sufix *-ble* ---- **108**
    - 3.6.2.1 Obecné charakteristiky ---- **108**
    - 3.6.2.2 Základní výsledky vyhledávání ---- **108**
    - 3.6.2.3 Interpretace některých adjektiv ---- **117**
- 3.7 Portugalská afixy *re-* a *-vel* ---- **119**
  - 3.7.1 Prefix *re-* ---- **119**
    - 3.7.1.1 Obecné charakteristiky ---- **119**
    - 3.7.1.2 Základní výsledky vyhledávání ---- **120**
    - 3.7.1.3 Interpretace některých sloves ---- **122**
  - 3.7.2 Sufix *-vel* ---- **123**
    - 3.7.2.1 Obecná charakteristika ---- **123**
    - 3.7.2.2 Základní výsledky vyhledávání ---- **123**
    - 3.7.2.3 Interpretace některých adjektiv ---- **128**
- 3.8 Francouzština: afixy *re-* a *-ble* ---- **130**
  - 3.8.1 Prefix *re-* ---- **130**
    - 3.8.1.1 Obecné charakteristiky ---- **130**

- 3.8.1.2 Základní výsledky vyhledávání ---- 131
- 3.8.1.3 Interpretace některých sloves ---- 135
- 3.8.2 Sufix *-ble* ---- 138
  - 3.8.2.1 Obecné charakteristiky ---- 138
  - 3.8.2.2 Základní výsledky vyhledávání ---- 139
  - 3.8.2.3 Interpretace některých adjektiv ---- 144
- 3.9 Závěrem ke slootovorbě ---- 146
  
- 4. **SLOVESNÉ PERIFRÁZE S VÝZNAMEM INICIÁLNÍ INGRESIVNOSTI V ROMÁNSKÝCH JAZYCÍCH A TYPOLOGIE JEJICH ČESKÝCH PROTĚJŠKŮ ---- 149**
  - 4.1 Úvodem k ingresivním perifrázím ---- 150
    - 4.1.1 Perifrastické konstrukce - definice ---- 151
    - 4.1.2 Rozdělení perifrastických konstrukcí ---- 152
    - 4.1.3 Cíle a metody práce ---- 153
  - 4.2 Perifráze s významem iniciální ingresivnosti: portugálština ---- 154
    - 4.2.1 Perifrastické konstrukce s významem iniciální ingresivnosti, které jsou předmětem našeho zkoumání ---- 154
    - 4.2.2 Kritéria pro vyhledávání ---- 155
    - 4.2.3 Přehled analyzovaných perifrází s kvantitativním vyjádřením ---- 156
    - 4.2.4 Závěry vyplývající z analýzy portugalského materiálu ---- 160
  - 4.3 Perifráze s významem iniciální ingresivnosti: španělština ---- 160
    - 4.3.1 Perifrastické konstrukce s významem iniciální ingresivnosti, které jsou předmětem našeho zkoumání ---- 160
    - 4.3.2 Kritéria pro vyhledávání ---- 161
    - 4.3.3 Základní klasifikace perifrází ---- 161
    - 4.3.4 Typologie českých protějšků ---- 162
    - 4.3.5 Přehled analyzovaných perifrází s kvantitativním vyjádřením ---- 164
    - 4.3.6 Závěry ---- 172
  - 4.4 Perifráze s významem iniciální ingresivnosti: itaština ---- 173
    - 4.4.1 Perifrastické konstrukce s významem iniciální ingresivnosti, které jsou předmětem našeho zkoumání ---- 173
    - 4.4.2 Obecné charakteristiky ---- 173
    - 4.4.3 Základní výsledky vyhledávání ---- 174
    - 4.4.4 Klasifikace českých protějšků ---- 176
  - 4.5 Perifráze s významem iniciální ingresivnosti: francouzština ---- 179
    - 4.5.1 Perifrastické konstrukce s významem iniciální ingresivnosti, které jsou předmětem našeho zkoumání ---- 179
    - 4.5.2 Základní klasifikace perifrází ---- 180
    - 4.5.3 Typologie českých protějšků ---- 180
    - 4.5.4 Přehled analyzovaných perifrází s kvantitativním vyjádřením ---- 181
    - 4.5.5 Závěry vyplývající z analýzy francouzského materiálu ---- 184
  - 4.6 Románské perifráze s významem iniciální ingresivnosti ve světle korpusu *InterCorp* ---- 184
  
- 5. **GERUNDIUM V ROMÁNSKÝCH JAZYCÍCH A TYPOLOGIE JEHO ČESKÝCH PROTĚJŠKŮ ---- 187**
  - 5.1 Úvodem ke gerundiu ---- 188
    - 5.1.1 Morfologie gerundia ---- 188
    - 5.1.2 Gerundium jako *converb* ---- 190
      - 5.1.2.1 Syntaktické fungování gerundia ---- 192
      - 5.1.2.2 Sémantické fungování (adverbiálního) gerundia ---- 194
    - 5.1.3 Typy respondentů gerundia v češtině ---- 196
    - 5.1.4 Závěr - cíle analýzy ---- 197

- 5.2 Gerundium ve francouzštině ---- **198**
  - 5.2.0 Morfologie gerundia ve francouzštině ---- **198**
  - 5.2.1 Vymezení zkoumaného vzorku gerundií ---- **198**
  - 5.2.2 Frekvenční analýza gerundia ve francouzštině ---- **199**
  - 5.2.3 Syntaktické a sémantické fungování gerundia ---- **202**
  - 5.2.4 České respondenty francouzského gerundia ---- **208**
    - 5.2.4.1 Plně predikační respondenty ---- **210**
      - 5.2.4.1.1 Určité sloveso v hlavní větě ---- **210**
      - 5.2.4.1.2 Určité sloveso ve vedlejší větě ---- **211**
    - 5.2.4.2 Depredikační respondenty gerundia ---- **213**
    - 5.2.4.3 Polopredikační respondent gerundia - přechodník ---- **214**
  - 5.2.5 Závěr ---- **218**
- 5.3 Gerundium ve španělštině ---- **219**
  - 5.3.0 Morfologie gerundia ve španělštině ---- **219**
  - 5.3.1 Vymezení zkoumaného vzorku gerundií ---- **220**
  - 5.3.2 Frekvenční analýza gerundia ve španělštině ---- **221**
  - 5.3.3 Syntaktické a sémantické fungování gerundia ---- **224**
  - 5.3.4 České respondenty španělského gerundia ---- **231**
    - 5.3.4.1 Plně predikační respondenty ---- **233**
      - 5.3.4.1.1 Určité sloveso v hlavní větě ---- **233**
      - 5.3.4.1.2 Určité sloveso ve vedlejší větě ---- **234**
    - 5.3.4.2 Depredikační respondenty gerundia ---- **235**
    - 5.3.4.3 Polopredikační respondenty gerundia ---- **236**
  - 5.3.5 Český přechodník a španělské gerundium ---- **237**
  - 5.3.6 Závěr ---- **240**
- 5.4 Gerundium v italštině ---- **241**
  - 5.4.0 Morfologie gerundia v italštině ---- **241**
  - 5.4.1 Vymezení zkoumaného vzorku gerundií ---- **242**
  - 5.4.2 Frekvenční analýza gerundia v italštině ---- **242**
  - 5.4.3 Syntaktické a sémantické fungování gerundia ---- **243**
  - 5.4.4 České respondenty italského gerundia ---- **251**
    - 5.4.4.1 Plně predikační respondenty ---- **252**
      - 5.4.4.1.1 Určité sloveso v hlavní větě ---- **252**
      - 5.4.4.1.2 Určité sloveso ve vedlejší větě ---- **252**
    - 5.4.4.2 Depredikační respondenty gerundia ---- **253**
    - 5.4.4.3 Polopredikační respondenty gerundia - přechodník a aktivní verbální adjektivum ---- **255**
  - 5.4.5 Závěr ---- **258**
- 5.5 Gerundium v portugalštině ---- **259**
  - 5.5.1 Morfologie gerundia v portugalštině ---- **259**
  - 5.5.2 Vymezení zkoumaného vzorku gerundií ---- **259**
  - 5.5.3 Frekvenční analýza gerundia v portugalštině ---- **260**
    - 5.5.3.1 Výskyt gerundia v i. p. m. ---- **260**
    - 5.5.3.2 Vliv složení korpusu a autorského idiolektu ---- **260**
    - 5.5.3.3 Lexikální obsazení gerundia ---- **262**
  - 5.5.4 Syntaktické fungování gerundia ---- **263**
  - 5.5.5 Sémantické typy příslovecného gerundia ---- **267**
  - 5.5.6 Shrnutí fungování českých respondentů gerundia ---- **273**
  - 5.5.7 Český přechodník a portugalské gerundium ---- **275**
  - 5.5.8 Závěr ---- **277**
- 5.6 Románské gerundium a čeština ve světle korpusu *InterCorp* ---- **278**
  - 5.6.1 Frekvence výskytu gerundia ve zkoumaných jazycích ---- **279**



- 5.6.2 Syntaktická funkce gerundia ve zkoumaných jazycích ---- **280**
- 5.6.3 Příslowečné významy gerundia ve zkoumaných jazycích ---- **282**
- 5.6.4 Protějšky gerundia v češtině ---- **284**

**6. OBRAZ ROZDÍLŮ MEZI ROMÁNSKÝMI JAZYKY A ČEŠTINOU  
V PARALELNÍM KORPUSU INTERCORP ---- 287**

Cizojazyčná shrnutí ---- **293**

Bibliografie ---- **301**

Rejstřík věcný ---- **311**

Rejstřík jmenný ---- **315**



**1. SROVNÁVÁNÍ JAZYKŮ  
A PARALELNÍ KORPUSY  
(PČ)**

Vznik elektronických jazykových korpusů a korpusové lingvistiky přinesl v lingvistice významný zlom: umožnil konfrontaci zažitých tvrzení o jazyce se statisticky zpracovanými analýzami obrovských objemů dat. Vedl též k přehodnocení role lingvistovy intuice, až do té doby základního nástroje při analýze dat: dnes je nemyslitelné opírat jakýkoli závěr o jazykovou intuici nedostatečného počtu mluvčích, nebo snad dokonce mluvčího jediného.

Pro kontrastivní lingvistiku byl problém jazykové intuice ještě palčivější – na jazykovou intuici se bylo potřeba spoléhat ve dvou, nebo dokonce ještě více srovnávaných jazycích. Lingvista byl nejčastěji rodilým mluvčím jednoho ze zkoumaných jazyků, v ostatních jazycích se musel spoléhat na svou osvojenou znalost nebo na informanty – komplikací navíc bylo to, že nešlo jen o kompetenci v příslušných jazycích, ale i o schopnost správně nahlédnout vzájemné vztahy (ekvivalence, shody a rozdíly mezi nimi). Vznik paralelních a jiných korpusů umožňujících srovnávání dat paralelně ve více jazycích přinesl i na toto pole zcela nové perspektivy. Tyto korpusy existují kratší dobu než korpusy jednojazyčné, proto je s jejich využíváním méně praktických zkušeností.

V Ústavu Českého národního korpusu Filozofické fakulty Univerzity Karlovy je již od roku 2005 budován paralelní korpus *InterCorp* (<http://ucnk.ff.cuni.cz/intercorp/>), představující i v mezinárodním kontextu špičkový korpusový nástroj svého druhu. Jeho metodologická využitelnost byla již testována na řadě dílčích studií. Tato monografie metodologii testuje v rozsáhlém měřítku, celkem na čtyřech jazycích a na čtyřech strukturně rozličných jazykových jevech. Má tak dva cíle: 1) na vybraných jevech popsat určité strukturní rozdíly mezi románskými jazyky a češtinou; 2) v praktické i teoretické rovině ověřit metodologii paralelních korpusů obecně a korpusu *InterCorp* specificky.

## **1.1 ROMÁNSKÉ JAZYKY A ČEŠTINA, HLEDÁNÍ STRUKTURNÍCH ROZDÍLŮ**

Výše jsme konstatovali, že do příchodu korpusové lingvistiky se srovnávací mezijazykové studie – tedy i studie srovnávající románské jazyky a češtinu – nutně opíraly

o víceméně intuitivní, statisticky nepodloženou analýzu dat. Tyto tradiční studie často vycházely z typologické charakteristiky jednotlivých jazyků: často se stavěla flexivní čeština proti izolační francouzštině apod. Nechme teď stranou, do jaké míry je dnes taková typologická charakteristika jazyků obhajitelná. Jisté je, že podobná tvrzení vynášela v podstatě kvantitativní, statistické soudy, aniž se opírala o jakákoli kvantifikovatelná data: tak se u kauzativních konstrukcí *hacer/fare/fazer/faire...* + infinitiv konstatovalo, že čeština jako protějšek těchto konstrukcí hojně využívá prefixu *roz-*, nebo se uvádělo, že proti bohatému španělskému systému analytických konstrukcí vyjadřujících ingresivnost stojí opět české prefixy. Nelze popřít, že takovéto české protějšky jsou možné, ale lze předpokládat, že nejsou nutně dominantní, a už vůbec ne jediné. Základní hypotézou této monografie je, že tradičně podávaný obraz shod a rozdílů mezi románskými jazyky a češtinou je nutně zjednodušený: lze očekávat, že detailní srovnávací popis, založený na materiálu z korpusů, tento obraz přinejmenším zproblematizuje. Lze předpokládat, že v každém ze zmiňovaných příkladů nabídne čeština velmi heterogenní a na první pohled nepřehledný souhrn různorodých řešení.

Abychom tuto hypotézu mohli ověřit, vybrali jsme čtyři konkrétní jazykové jevy z románských jazyků a provedli jsme výzkum jejich protějšků v paralelním korpusu *InterCorp*. Konkrétně jde o tyto jazykové jevy:

- 1) kauzativní konstrukci *hacer/fare* + infinitiv;
- 2) adjektiva se sufixem *-ble/-bile...* a slovesa s prefixem *re-/ri-*;
- 3) ingresivní perifráze;
- 4) gerundium.

Tyto jevy jsme nevybrali náhodou.

Je pochopitelné, že jejich jednotícím prvkem je skutečnost, že mají v románských jazycích stejnou/podobnou formu, byť pochopitelně využitelnost daného jazykového prostředku může být v každém jazyce trochu jiná (srov. např. velmi pestrý a bohatý repertoár španělských ingresivních vazeb s nepoměrně chudším inventářem francouzským nebo italským). Úmyslně jsme vybírali jevy z různých jazykových rovin/plánů: vedle jevu v zásadě slovotvorného (prefixy a sufixy) tak budeme zkoumat prostředky, které v sobě spojují rysy morfologické, syntaktické a lexikální. Pro zkoumání očekávané heterogenosti českých protějšků je důležité konstatovat, že i jen povšechné srovnání situace v románských jazycích a v češtině dělí zkoumané jevy do dvou skupin.

- 1) U některých jevů máme v češtině jakýsi „izomorfní“ prostředek (např. proti sufixu *-ble, -bile...* stojí český sufix *-telný*; proti románské kauzativní konstrukci české *dát si ostríhat vlasy* apod.). U těchto případů budeme mj. zjišťovat, jak často a za jakých okolností bude protějškem románského jevu právě izomorfní prostředek český.
- 2) U jiných jevů máme naopak v románských jazycích něco, co v češtině není, nebo to přinejmenším není přítomno ve stejné míře, stejně systémově (to je případ zpracovaného systému ingresivních perifrází např. ve španělštině a českého *začít pracovat* apod.). U těchto případů je rozmanitost českých protějšků očekávatelnější a naším cílem je tuto rozmanitost popsat; navíc zde platí výše řečené: je otázka, jak často je protějškem tradičně uváděný „typologický“ český protějšek.

Náš materiál ukáže, že různorodá a pestrá řešení, která se mnohdy vzpírají jednoznačně a dobře vymezené typologii, se objevují u obou typů jevů.

České protějšky vybraných jevů budeme zkoumat ze strukturního hlediska: zajímá nás jejich typologie a kvantifikace vymezených typů. Tato typologie, vytvořená částečně a priori a ověřená a zpřesněná analýzou dat, je pochopitelně zobecněním: zajímá nás, jaké obecné strukturní možnosti využívají u daných jevů zkoumané jazyky a jak si tyto strukturní možnosti navzájem odpovídají. Není tedy naším cílem detailní analýza jednotlivých případů – naše práce proto není translatologická. To se odráží i v tom, že naším cílem není ani posuzovat „správnost“ překladu. „Správnost“ překladu necháváme stranou v souladu s předpokladem korpusové lingvistiky, že na dostatečně velkém objemu dat nabude každý rozměr zkoumaného jevu náležité proporce: lze předpokládat, že „špatné“ překladatelské řešení bude zcela minoritní. Tuto otázku tedy zcela ponecháváme translatologickým studiím, byť předpokládáme, že naše analýzy strukturních souvislostí mohou být pro translatologické práce přínosem. Paralelní korpusy mohou ostatně translatologům poskytnout materiál i pro analýzu jiných závažných témat, např. tématu „návodnosti“ překladu – nás nebude otázka, do jaké míry je překladový text ovlivněn textem původním, zajímat při analýze jednotlivých příkladů (byť místy možný vliv původního textu konstatujeme), nýbrž při charakteristice korpusu jako celku: v souladu se současnými možnostmi korpusové lingvistiky provedeme ověřovací srovnání jazyka překladu s jazykem originálu (k tomu viz 1.2).

Výše zmiňovaná různorodost vybraných jevů vede k tomu, že postup analýzy musí být u každého z nich trochu jiný. Společné všem kapitolám jsou metodologie (tj. korpusová metodologie) a zmiňovaný strukturní přístup. Jak vyplývá z výše řečeného, tímto strukturním přístupem rozumíme:

- analýzu kvantitativního aspektu jednotlivých jevů, který v tradičních gramatických románských jazycích pochopitelně chybí – korpusy nám poskytují daleko větší množství dat, jejichž frekvence je samozřejmě důležitá;
- samu typologii českých protějšků; ta představuje – pokud je nám známo – výraznou novinku.

Z toho vyplývá, že naše analýza chce:

- v prvním plánu na vybraných tématech strukturně porovnat románské jazyky s češtinou;
- v druhém plánu – nijak systematicky, tj. spíše jako jakýsi vedlejší produkt – popsat daný jev v konkrétním románském jazyce, příp. naznačit rozdíly mezi zkoumanými románskými jazyky (vedlejší cíl je to také proto, že jak vysvětlíme, jednotlivé románské subkorpusy jsou nestejně velké, což srovnávání komplikuje).

Aby bylo možné zmíněných cílů dosáhnout, jednotlivé kapitoly budou mít analogickou strukturu, umožňující porovnávat situaci v jednotlivých subkorpusech:

#### 1) *Obecná definice zkoumaného jevu*

Tato podkapitola vždy stručně definuje zkoumaný jev, vymezí jeho základní charakteristiky, příp. popíše souvislosti, které mohou ovlivnit následnou korpusovou analýzu.

2) *Stručná analýza možných protějšků v češtině*

Tato podkapitola nechce být bohemistickou analýzou daného jevu – na základě relevantních bohemistických zdrojů pouze načrtává, jaké strukturní možnosti má čeština pro vyjádření zkoumaného jevu.

3) *Stanovení cíle vyhledávání a typologie českých protějšků*

Na základě poznatků z prvních dvou podkapitol je v této podkapitole přesně definováno, co je a co není cílem následující korpusové analýzy. Základ tvoří – jak už bylo řečeno – srovnání stavu v románských jazycích se situací v češtině, protože je zásadní součástí této kapitoly detailní typologie českých protějšků, jež bude základem veškerých následujících kvantifikací.

4) *Kapitoly věnované jednotlivým jazykům*

Tyto jazykové kapitoly mají analogickou strukturu, jež čtenáři umožní porovnávat data z jednotlivých románských jazyků. Analogičnost je omezena jen tím, že ne všechny subkorporusy jsou stejně velké – to vede k tomu, že např. u kauzativních konstrukcí pracujeme u španělského, italského a francouzského subkorpusu s minimální hranicí výskytu, u portugalského žádnou minimální hranici nemáme, protože bychom neměli dostatek dat. Tuto skutečnost, jež samozřejmě ovlivňuje následné srovnávání dat z různých subkorpusů, budeme, stejně jako další charakteristiky práce s korpusy, komentovat níže. Jednotlivé kapitoly mají tuto strukturu:

a) *Analýza situace v daném jazyce*

Tato podkapitola předkládá stav poznání v dané věci, vycházející z relevantní sekundární literatury. Cíl práce a prostorové omezení vede k tomu, že přehled stavu vědění je spíše stručný a u souvislostí, které nemají přímou souvislost s následnou analýzou, doplňuje stručnou charakteristiku problému odkazem na sekundární literaturu, kde zájemce najde podrobnosti.

b) *Popis vyhledávacích dotazů*

Tato podkapitola uvádí konkrétní popis postupu vyhledávání (nikoli však popis subkorpusu: ten je pro všechny kapitoly stejný a je uveden níže v podkapitole 1.2), a to včetně konkrétních vyhledávacích dotazů (je tak možné ověřit platnost dat; k tomu je však potřeba pochopitelně ovládat vyhledávací jazyk korpusu *InterCorp*).

c) *Analýza dat*

Hlavní podkapitola, obsahující popis jazykových dat získaných z korpusu *InterCorp*. V největší možné míře (tj. tak, jak to umožnila srovnatelnost subkorpusů) je zachována analogičnost výkladů s podkapitolami věnovanými ostatním jazykům, aby mohl čtenář srovnávat.

5) *Shrnutí závěrů k danému jevu*

Tato kapitola obsahuje zobecnění závěrů jazykových podkapitol věnovaných danému jevu. Předpokládá, že se čtenář seznámil s jednotlivými jazykovými podkapitolami, neopakuje totiž jejich závěry, nýbrž předkládá jejich syntézu a vyvozuje z nich obecnější konsekvence.

V našich analýzách nebyl žádný z románských jazyků důležitější než ostatní, žádný nebyl východiskem pro zkoumání jazyků ostatních. Tato skutečnost je zdůrazněna i tím, že u každého ze čtyř hlavních témat je úvodní, výchozí kapitola věnována

jinému jazyku (u kauzativ je prvním zkoumaným jazykem španělština, u slovtvorby italština, u perifrází portugalština a u gerundia francouzština). Pořadí ostatních jazyků je vždy stejné (španělština, italština, portugalština, francouzština). Je tedy možné snadno nalézt partie věnované konkrétnímu jazyku: španělštinou se zabývají kapitoly 2.2, 3.6, 4.3 a 5.3, italštinou 2.3, 3.4, 4.4 a 5.4, portugalštinou 2.4, 3.7, 4.2 a 5.5 a francouzštinou 2.5, 3.8, 4.5 a 5.2.

Tato monografie je kolektivním dílem: text je koncipován jako jednotný, vnitřně provázaný celek s jednotnými cíli, postupy a metodologií. Koordinátorem celé monografie byl Petr Čermák, koordinátory jednotlivých podkapitol Petr Čermák, Pavel Štichauer, Jaroslava Jindrová a Olga Nádvorníková, koordinátorkou jednotných metodologických postupů Olga Nádvorníková. Hlavní autoři jednotlivých podkapitol jsou uváděni šifrou: Leontýna Bratánková (LB), Petr Čermák (PČ), Štěpánka Černíková (ŠČ), Jan Hricsina (JH), Jiří Jančík (JJA), Jaroslava Jindrová (JJ), Dana Kratochvílová (DK), Zuzana Krinková (ZK), Petra Laufková (PL), Olga Nádvorníková (ON), Daniel Petřík (DP), Eliška Tříšková (ET), Pavel Štichauer (PŠ).

## 1.2 METODOLOGIE VYTĚŽOVÁNÍ PARALELNÍCH KORPUSŮ. PARALELNÍ KORPUS INTERCORP (ON)

Jak jsme uvedli výše, paralelní korpus *InterCorp* představuje svým rozsahem i kvalitou zpracování výjimečný nástroj, a to i ve světovém měřítku. V současné době (verze 8) totiž zahrnuje 40 jazyků a celkem více než 1,5 miliardy slov. Kromě zde zkoumaných románských jazyků, tj. francouzštiny, italštiny, portugalštiny a španělštiny, zahrnuje *InterCorp* také katalánštinu a rumunštinu. V korpusu jsou však zastoupeny samozřejmě i jazyky germánské, jazyky slovanské, ale i další jazykové skupiny.<sup>1</sup>

174 milionů slov z celku korpusu *InterCorp* připadá na češtinu, která je jazykem pivotním (každý text v korpusu musí mít protějšek v češtině). Čeština je centrem projektu nejen z důvodů technických (všechny jazyky jsou zarovnány podle českých textů, což umožňuje jednotné vyhledávání v několika jazycích najednou), ale především z důvodů koncepčních – *InterCorp* je součástí projektu Českého národního korpusu (<http://ucnk.ff.cuni.cz/intercorp>) a jedním z jeho hlavních cílových využití je kontrastivní výzkum vzhledem k češtině. Jak jsme zmínili výše, projekt *InterCorp* byl zahájen již v roce 2005; první verze korpusu byla uživatelům zpřístupněna v roce 2008. Výzkum

1 Plný výčet jazyků zastoupených ve verzi 6 korpusu *InterCorp*: němčina, angličtina, arabština, běloruština, bulharština, katalánština, chorvatština, dánština, španělština, estonština, finština, francouzština, řečtina, hindština, maďarština, italština, litevština a lotyština, makedonština, maltština (zatím pouze v textech EU podobně jako estonština), nizozemština, norština, polština, portugalština, rumunština, ruština, slovenština, slovinština, srbsština (v latině i cyrilici), švédština, ukrajinština a čeština jako pivotní jazyk. Verze 8 obsahuje navíc subkorpusy v hebrejštině, islandštině, japonštině, malajštině, albánštině, turečtině a vietnamštině (všechny pouze v kolekci Subtitles). Uvažuje se o připojení dalších jazyků, např. čínštiny nebo romštiny.



představovaný v této studii byl proveden v roce 2013 na verzi 6. Do nejnovější verze korpusu byla zařazena např. rozsáhlá kolekce filmových titulků (Subtitles), ale výrazně byly rozšířeny také některé subkorpusy jádra (např. subkorpus francouzský).

Porovnávání jazykových jevů v různých jazycích je v *InterCorpu* umožněno tím, že jsou k sobě v korpusu přiřazeny originály a překlady textů; jde tady o korpus *paralelní*, nikoli *srovnatelný* (*comparable*).<sup>2</sup> Každý originální text má v korpusu minimálně jeden překladový protějšek, často je jich však více; nejvíce jazykových verzí (26 včetně češtiny) má kniha *Harry Potter a kámen mudrců* J. K. Rowling, ale hned na druhém místě např. najdeme *Malého prince* A. de Saint-Exupéry, který je dostupný v 25 jazykových verzích.<sup>3</sup> Texty jsou v korpusu zarovnány na úrovni vět:<sup>4</sup>

- [FR] « *Moi, se dit le petit prince, si j'avais cinquante-trois minutes à dépenser, je marcherais tout doucement vers une fontaine ...* » (Antoine de Saint-Exupéry, *Le Petit prince*)
- [ALL] „*Wenn ich dreiundfünfzig Minuten übrig hätte*“, *sagte der kleine Prinz*, „*würde ich ganz gemächlich zu einem Brunnen laufen...*“ (přel. Grete Leitgeb – Josef Leitgeb)
- [ANG] “*As for me,*” *said the little prince to himself*, “*if I had fifty-three minutes to spend as I liked, I should walk at my leisure toward a spring of fresh water.*” (přel. Katherine Woods)
- [CS] *Kdybych já měl padesát tři minuty nazbyt, řekl si malý princ, šel bych docela pomaloučku ke studánce...* (přel. Zdeňka Stavínohová)
- [HIN] “अगर मेरे पास तिरपन मिनट बताने को होते , छोटे राजकुमार ने सोचा , तो मैं धीरे – धीरे एक जलाशय की ओर चल पड़ता ... 1” (přel. कशिशोर बलवीर, जगवंश)
- [ES] «*Si yo dispusiera de cincuenta y tres minutos —pensó el principito— caminaría suavemente hacia una fuente...*» (přel. Bonifacio del Carril)
- [IT] “*Io*”, *disse il piccolo principe*, “*se avessi cinquantatré minuti da spendere, camminerei adagio adagio verso una fontana...*” (přel. Nini Bompiani Bregoli)
- [RO] „*Eu, își spuse micul prinț, dacă aș avea de irosit cincizeci și trei de minute, aș porni în liniște spre o fântână.*“ (přel. Benedict Corlaciuc)

Pro všechny zde zkoumané románské jazyky, tj. francouzštinu, španělštinu, portugalsštinu a italštinu, je v *InterCorpu* k dispozici vyhledávání podle lemmat i podle

2 Korpusy *srovnatelné* (*comparable*) se na rozdíl od korpusů *paralelních* skládají výlučně z textů originálních. „Srovnatelnost“ jednotlivých jazykových subkorpusů pak musí být zajišťována tím, že oba subkorpusy zachovávají předem dané specifikace týkající se rozsahu, žánrového složení, komunikační funkce textů atd. Tento typ korpusů se často využívá např. pro vytváření terminologických slovníků z určité specifické oblasti (např. genetického inženýrství apod.). Je možné vytvářet i vyvážené (*balanced*) *srovnatelné* korpusy, které nemusejí být omezeny na jednu textovou oblast, to však představuje nesrovnatelně náročnější úkol. Termínu „*paralelní*“ se někdy užívá jako zastřešujícího termínu pro *srovnatelné* (*comparable*) i *překladové* korpusy (viz pozn. v Altenberg – Granger: 2002, s. 8), obvykle je však tento termín spojován pouze s posledně jmenovaným typem (viz např. McEnergy – Xiao – Tono: 2006, s. 47). K obecné charakteristice typů korpusů viz Chlumská (2014).

3 Viz celkový přehled na <https://trnka.ff.cuni.cz/ucnk/intercorp/files/IntercorpTop200>.

4 Pro další verze korpusu se plánuje i zarovnání na úrovni slov, které by velmi zkvalitnilo možnosti vyhledávání.

morfologických značek.<sup>5</sup> Předností *InterCorpu* jsou rovněž pokročilé funkce korpusového vyhledávače KonText, které umožňují získané konkordance třídit a statisticky zpracovávat např. pomocí frekvenčních seznamů. Nicméně vzhledem k tomu, že žádné morfologické značkování není naprosto spolehlivé, využili jsme v syntaxi dotazů pro jednotlivé zkoumané jazykové jevy kromě lemmatizace zejména různých kombinací regulárních výrazů.<sup>6</sup>

Jádro korpusu *InterCorp* tvoří texty, u nichž byla provedena pečlivá jazyková korektura a semi-manuální kontrola automatického zarovnání pomocí editoru paralelních textů *InterText*.<sup>7</sup> Tento základní soubor textů obsahuje především beletrii, výjimkou však nejsou ani jiné typy textů, např. texty odborné (viz níže). Vzhledem k tomu, že je *InterCorp* primárně určen pro synchronní výzkum jazyka, pocházejí tyto texty (a samozřejmě i jejich překlady) z 2. poloviny 20. století a ze století 21. Zařazení textů starších je výjimečné: jde např. o klíčové texty dané národní literatury (např. texty Karla Čapka pro subkorpus český), které existují v mnoha překladech, nebo o překlady z jazyků s menším množstvím textů přeložených do češtiny. U všech těchto textů jsou k dispozici podrobné metajazykové informace, podle nichž je buď možné specifikovat pracovní subkorpus (např. podle směru překladu), nebo které je možné zobrazit v konkordanci u jednotlivých výskytů (např. bibliografické údaje o originálu a překladu).

Tento základní soubor paralelních textů je v korpusu *InterCorp* doplněn o tzv. kolekce textů, které neprošly jazykovou korekturou v rámci projektu *InterCorp* ani ruční kontrolou automatického zarovnání na věty. Výhodou těchto subkorpusů je jednak jejich velký rozsah, kterým často několikanásobně převyšují výše zmíněné jazykové subkorpusy z jádra, jednak žánrová rozmanitost, kterou do *InterCorpu* přinášejí: vedle právních textů Evropské unie (*Acquis communautaire*) zde najdeme i texty publicistické z mnohojazyčných publicistických serverů Project Syndicate ([www.project-syndicate.org](http://www.project-syndicate.org)) a Presseurop ([www.presseurop.eu](http://www.presseurop.eu)) v deseti jazycích nebo zápisy z debat na půdě Evropského parlamentu (*EuroParl*). Nevýhodou těchto korpusových kolekcí, která rovněž přispěla k našemu rozhodnutí vyloučit je z našeho výzkumu, byly chybějící metajazykové informace, zejména informace o překladateli a směru překladu. Jak se pokusíme ukázat níže, pro kontrastivní analýzu jsou totiž tyto informace klíčové.

5 Všechny zde zkoumané románské jazyky byly morfologicky označovány pomocí programu *TreeTagger* ([www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger](http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger)), tagsety (seznamy značek) pro jednotlivé jazyky se však liší. Čeština je značkována pomocí programu *morče* (<http://ufal.mff.cuni.cz/morce>), který je schopen zohlednit její bohatou morfologii. Pro budoucí verze korpusu se plánuje harmonizace tagsetů (k této problematice viz např. Rosen: 2010; 2012).

6 Chybovost morfologického značkování není obvykle nijak významná; výjimkou je značkování minulého přechodníku v češtině (viz kapitolu o gerundi), kde se chybovost někdy pohybovala až kolem 50 % (jako minulý přechodník byly chybně označovány např. nespisovné tvary l-ového přičestí – *řek, proved* atd.).

7 Automatické zarovnání textů zajišťuje program *hunalign* (srov. Varga et al.: 2005, <http://mokk.bme.hu/recources/hunglishcorpus/>). Podrobně ke koncepci *InterCorpu* a jeho technickému zpracování viz např. Vavříň – Rosen (2008), Rosen – Vavříň (2012), obecně ke koncepci pak Čermák (2010) a zejména Čermák – Rosen (2012).

Následující tabulka ukazuje v přehledu velikost jednotlivých zkoumaných jazykových subkorpusů:

Tabulka 1. Srovnání zkoumaných subkorpusů románských jazyků v korpusu InterCorp (verze 6)

Subkorpus	Počet pozic celkem (kolekce i jádro)	Zkoumaný subkorpus (překlady z daného románského jazyka do češtiny v jádru)			
		Počet pozic	Počet textů	Počet (různých) autorů	Počet (různých) překladatelů
ŠP-CS	<b>73 002 746</b>	9 326 150	105 (36 evr. / 69 am.)	43 (21 evr. / 22 am.)	44
IT-CS	<b>54 564 618</b>	1 631 204	17	11	9
PORT-CS	<b>58 531 766</b>	1 485 541	15	11 (8 port. / 3 braz.)	9
FR-CS	<b>62 288 229</b>	<b>1 533 451</b>	31	23	24

Jak je z tabulky vidět, zkoumané subkorpusy románských jazyků jsou z hlediska rozmanitosti autorských a překladatelských idiolektů dostatečně bohaté. U portugalštiny a u španělštiny jsou zastoupeni jak autoři hispanoameričtí a brazilští, tak evropští.

Podíváme-li se však na údaje o počtu pozic v jednotlivých zkoumaných subkorpusích, musíme konstatovat, že žádný z nich rozhodně nelze považovat za reprezentativní pro celek daného jazyka.<sup>8</sup> Kromě rozsahu je potřeba vzít v úvahu i specifické žánrové složení těchto subkorpusů: jak jsme se zmínili výše, texty obsažené v jádru korpusu *InterCorp* jsou primárně beletristické (romány) a texty jiných typů jsou zastoupeny minimálně. Např. u odborných textů je významněji zastoupena pouze historie: v itaštině Procacciho *Dějiny Itálie*, ve francouzštině *Vznešené paní z XII. století* G. Dubyho nebo populárně-naučné texty pro mládež (*Marco Polo*, *Alexandr Veliký*). Ve španělštině patří mezi odborné texty *Vzpoua davů* Josého Ortegy y Gasset.

Vzhledem k nereprezentativnosti zkoumaných subkorpusů jsme frekvenční analýzy jednotlivých jazykových jevů založené na těchto subkorpusích omezili na minimum, protože by neměly dostatečnou výpovědní hodnotu. Pokud jsou frekvenční údaje v jednotlivých kapitolách uváděny, slouží primárně pro bližší určení zkoumaného souboru výskytů.<sup>9</sup> Malý rozsah subkorpusů a potenciální vlivy idiolektů (autorských i překladatelských) se pokoušíme při prezentaci analýz vyvážit rovněž tím, že u příkladů v textu uvádíme údaje o autorství originálního textu i překladu.<sup>10</sup>

8 Ostatně paralelní korpusy již ze své definice nemohou nikdy být reprezentativní pro celek daného jazyka, protože jsou omezeny na texty a textové typy, pro něž existuje překlad.

9 V některých případech, např. u gerundia, se pokoušíme frekvenční údaje získané z *InterCorpu* porovnávat s údaji z rozsáhlých jednojazyčných korpusů (viz např. **tabulku 71**).

10 Olohanová ve své knize *Introducing Corpora in Translation Studies* upozorňuje, že některé kontrastivní studie úplně opomíjejí roli překladatele, což se odráží mj. také v tom, že uvádějí pouze jména autorů původních textů, nikoli překladatelů (Olohan: 2004, s. 28).

Omezený rozsah subkorpusů vede rovněž k tomu, že jsou někdy podrobeny analýze veškeré výskyty daného jevu ve zkoumaném subkorpusu; většinou byla stanovena minimální hranice frekvence daného jevu a byly vyloučeny všechny výskyty, kde frekvence daného lemmatu tohoto limitu nedosáhla (více k tomu v příslušných kapitolách). Výjimkou je gerundium, kde se počty výskytů v italštině, portugalštině a španělštině pohybovaly od téměř 9 000 do 68 000, a kde bylo proto nutné pracovat při manuální analýze pouze se vzorky. Ve francouzštině je však gerundium méně frekventované než v ostatních zkoumaných jazycích, a proto bylo opět možné manuálně analyzovat veškeré jeho výskyty v daném subkorpusu (viz **tabulku 84** v závěru kapitoly o gerundiu).

Kromě nereprezentativnosti všech zkoumaných subkorpusů je potřeba vzít v úvahu i vzájemné rozdíly mezi nimi: např. nejrozsáhlejší španělský subkorpus obsahuje šestkrát více pozic než korpusy ostatních zkoumaných románských jazyků.

Navzdory výše uvedeným omezením však mohou podle našeho názoru zvolené subkorpusy románských jazyků korpusu *InterCorp* pomoci odhalit nové skutečnosti v kontrastivní analýze daných jevů ve srovnání s češtinou. Jak uvádí např. Altenberg – Granger (2002, s. 7), právě možnost využití rozsáhlých jazykových korpusů přispěla v posledních letech k renesanci kontrastivních analýz a kontrastivní lingvistiky. Paralelní korpusy totiž umožňují analyzovat na autentických datech tendence v úzu (*usage patterns*) a nabízejí širší možnosti výzkumu než pouhá introspekce, zejména možnost odhalit jiné, alternativní způsoby vyjádření určitého významu nebo funkce v cílovém jazyce. Jak uvidíme níže, metodologie využití paralelních korpusů v kontrastivní lingvistice se stále ještě konstituuje, je však možné uvést tři základní metodologické principy, které je potřeba při tomto typu analýz dodržovat:

1. Klíčovou otázkou kontrastivní analýzy je vymezení jazykových jednotek, které budou jejím předmětem. Jak ukazuje např. Barlow (2008), již analýza na první pohled jednoduchých lexémů (*aller* ve francouzštině a *go* v angličtině) může představovat problém, protože jejich význam je určován jejich kolokacemi.<sup>11</sup> V naší práci nejsou předmětem kontrastivní analýzy lexémy, ale gramatické nebo slovtvorné struktury, což vyžaduje komplexnější přístup než v případě analýzy na úrovni lexémů. Z tohoto důvodu předchází empirickým analýzám v každé kapitole této studie podrobný teoretický výklad, který představuje definici/fungování daného jevu jednak obecně, jednak ve všech zkoumaných románských jazycích a v češtině. Vysvětleny jsou rovněž terminologické rozdíly, dané rozdílnými lingvistickými tradicemi.

Cílem tohoto přístupu je zamezit terminologickému zmatku, který by mohl vést k záměně formální a funkční ekvivalence, protože by nebral v úvahu *valeur* analyzovaných jednotek, tedy jejich pozici v konkrétním jazykovém systému. Toto nebezpečí je zvláště výrazné u gerundia, kde téměř jednotný termín (*gerundio*, *gérondif*) a společný základ v latinském ablativu gerundia může svádět např. k podcenění rozdílů ve vztahu tohoto tvaru k jiným nefinitním tvarům ve zkoumaných jazycích. V případě gerundia

11 Máme-li být schopni data vůbec nějak popsat, musíme od jistých souvislostí odhlédnout. Extrémní případ problematické „ekvivalence“ analyzuje např. Teubert (2010): význam čínského termínu „xiao“ je totiž určován celou interpretační tradicí konfucíánské filozofie.