Univerzita Karlova v Praze

Filozofická fakulta

Ústav českého jazyka a teorie komunikace

Filologie – Český jazyk

K a t e ř i n a R y s o v á

O slovosledu z komunikačního pohledu

On Word Order from the Communicative Point of View

Teze disertační práce

Vedoucí práce: PhDr. Šárka Zikánová, Ph.D.

2013

Abstrakt

Práce se zabývá popisem českého slovosledu kontextově nezapojených participantů. Sleduje, zda v povrchovém slovosledu existuje jejich základní (frekvenčně výrazně převažující) pořadí (srov. *narodit se v Brně v roce 1950* vs. *narodit se v roce 1950 v Brně*). Zároveň se věnuje sledování faktorů, které slovosled ovlivňují (např. formě participantu, způsobu jeho lexikálního vyjádření nebo vlivu slovesné valence). V závěru krátce srovnává slovosledné tendence v češtině a v němčině. Pro ověření stanovených cílů užívá zejména data *Pražského závislostního korpusu*. Práce teoreticky vychází ze zásad funkčního generativního popisu. Výsledky výzkumu ukazují, že v českém povrchovém slovosledu lze alespoň v některých případech vysledovat určité obecnější tendence k jednomu ze dvou možných slovosledných pořadí.

Klíčová slova

slovosled, syntax, aktuální členění, informační struktura věty

Abstract

The presented thesis is focused on the Czech word order of contextually non-bound verbal modifications. It monitors whether there is a basic order in the contextually non-bound part of the sentence (significantly predominant in frequency) in the surface word order (cf. *narodit se v Brně v roce 1950* vs. *narodit se v roce 1950 v Brně*; literally *to be born in Brno in 1950* vs. *to be born in 1950 in Brno*). At the same time, we try to find out the factors influencing the word order (such as the form of modifications, their lexical expression or the effect of verbal valency). Finally, we briefly compare the word order tendencies in Czech and German. For the verification of the objectives, mainly the data from the *Prague Dependency Treebank* are used. The work is based on the theoretical principles of Functional Generative Description. Research results demonstrate that, at least in some cases, it is possible to detect certain general tendencies to use preferably one of two possible surface word order sequences in Czech.

Key words

word order, syntax, topic-focus articulation, sentence information structure

1 Cíle práce a stručný nástin metod výzkumu

Hlavním cílem předkládané disertační práce je přispět k popisu českého slovosledu. Nejprve je v ní věnována pozornost větě jako takové – její kontextově zapojené i nezapojené části. Práce ověřuje teorii Ludmily Uhlířové (1974), která předpokládá, že některé větné participanty jsou ze své podstaty předurčeny k tomu, aby vystupovaly spíše v roli tématu nebo rématu,[1] tedy k tomu, aby v nepříznakovém objektivním slovosledu stály spíše na začátku nebo spíše na konci věty.

Poté se centrum pozornosti přesouvá na kontextově nezapojenou část věty. V rozsáhlém korpusu téměř 40 000 vět – v *Pražském závislostním korpusu* – jsou vyhledávány všechny dostupné kombinace dvojic kontextově nezapojených participantů s cílem zjistit, zda tyto participanty tíhnou ve dvojici k určitému vzájemnému slovoslednému pořadí.

V dalším kroku je ověřován možný vliv valence na český slovosled (v návaznosti na předpoklad německé gramatiky W. Flämiga /1991/ o tom, že valence je hlavní slovosledný činitel v němčině).

Na závěr je na jiném jazykovém materiálu než v předchozích případech ověřováno vzájemné slovosledné postavení participantů času, místa a způsobu – vyjádřených jednotnou formou zájmenných příslovcí – v češtině a v němčině (např. *někde – nějak* vs. *irgendwo – irgendwie*).

1.1 Dílčí kroky pro ověření jednotlivých cílů

1. Sestavit stupnici kontextové zapojenosti pro nejčastěji se vyskytující participanty v datech *Pražského závislostního korpusu* – tj. zjistit, zda některé participanty výrazně častěji než jiné vystupují jako kontextově zapojené v porovnání s tím, jak často vystupují jako kontextově nezapojené. Stupnice kontextové zapojenosti je sestavena zvlášť pro participanty vyjádřené nevětně a větně. Získaná stupnice je porovnávána s obdobným výzkumem L. Uhlířové (1974). Kontextová zapojenost participantů je hodnocena v souladu s anotacemi aktuálního členění v *Pražském závislostním korpusu*.

[1] Předkládaná práce využívá teorii aktuálního členění, jak je popsána ve funkčním generativním popisu, pracuje tedy s pojmy základ a ohnisko. Tam, kde komentuje či pro srovnání využívá teorii jinou, užívá pojmy aktuálního členění tak, jak jsou nastaveny v příslušné teorii. L. Uhlířová např. používá pojmy téma a réma, proto jsou uvedeny i na tomto místě práce.

Rozlišení větných a nevětných participantů umožnilo zároveň popsat typickou formu výskytu jednotlivých aktantů a volných doplnění (větnou vs. nevětnou), čehož je využito pro sledování slovosledného chování jednotlivých typů participantů v závislosti na formě jejich vyjádření (zjištěné výsledky jsou porovnávány s výzkumem I. Nebeské /1984/ věnujícím se způsobu vyjádření vybraných typů příslovečných určení).

2. Na dvojicích kontextově nezapojených participantů ověřit v datech *Pražského závislostního korpusu*, zda jedno z jejich slovosledných uspořádání je výrazně frekvenčně převažující. Je přitom rozlišováno mezi participanty vyjádřenými větně a nevětně. Pro statistické vyhodnocení získaných dat byl použit test dobré shody (Pearsonův chí-kvadrát test).

Na třech dvojicích kontextově nezapojených participantů vyjádřených nevětně, které mají ve zvoleném korpusu nejvyšší počet výskytů, je dále podrobně sledováno, jaké faktory ovlivňují jejich povrchové pořadí. Kontextová nezapojenost participantů je hodnocena v souladu s anotacemi aktuálního členění v *Pražském závislostním korpusu*.

3. V přípravné fázi – pro pozdější zjišťování vlivu valence na český slovosled – ověřit, které typy volných slovesných doplnění v češtině mohou vystupovat jako valenčně obligatorní (a tedy u kterých typů participantů je možné ověřit, zda na jejich postavení ve větě má vliv jejich valenčnost). Jejich obligatornost je posuzována podle elektronického slovníku *Vallex* (v tištěné formě *Valenční slovník českých sloves* /2008/, slovník je vystavěn na teoretickém základě funkčního generativního popisu).

Na datech *Pražského závislostního korpusu* ověřit, zda (popř. do jaké míry) ovlivňuje povrchový slovosled kontextově nezapojených, nevětně vyjádřených participantů jejich vztah k řídícímu slovesu (slovesná valence). Valenčnost participantů je hodnocena podle elektronického valenčního slovníku *PDT-Vallex*, který je propojen s *Pražským závislostním korpusem*.

4. Na dvojicích vybraných typů participantů vyjádřených zájmenným příslovcem (*někdy*, *někde*, *nějak*) ověřit, zda existuje jejich frekvenčně výrazně převažující pořadí. Tento cíl je ověřován na datech *Českého národního korpusu*.

Doplňkově je podobný průzkum proveden i na německém jazykovém materiálu (na datech německého korpusu *Digitales Wörterbuch der deutschen Sprache*), což umožňuje alespoň dílčí srovnání s předpoklady gramatiky W. Flämiga (1991) o německém slovosledu.

2 Teoretické východisko

Pro popis slovosledu a aktuálního členění větného existují v české lingvistice dvě základní pojetí, obě do jisté míry navazují na Viléma Mathesia (souhrnně 1947). Prvním z nich je tzv. brněnská teorie aktuálního členění větného (funkční perspektivy větné) Jana Firbase (např. 1971, 1979, 1992) a jeho spolupracovníků. Z Firbasova pojetí do značné míry vychází v *Mluvnici češtiny 3* (1987) Ludmila Uhlířová, která se tématu slovosledu věnuje v četných studiích (srov. např. 1974, 1987 aj.). Tato teorie slovosledu a aktuálního členění je založena zejména na pohledu komunikačně orientované lingvistiky.

Druhým pohledem je funkční generativní popis jazyka (FGP) zformulovaný v 60. letech 20. století Petrem Sgallem (1964, 1967). Funkčně-generativní pojetí aktuálního členění se Sgallem zformulovala a dále významně rozvinula zejména Eva Hajičová (např. 1980, 1998, 2005, 2012a, 2012b aj.). Tématu slovosledu z pohledu FGP se dále věnuje Šárka Zikánová (např. 2006, 2009). Pojetí FGP se orientuje na slovosled a aktuální členění z pohledu gramatiky jako nástroje generujícího věty daného jazyka. Na tomto teoretickém základě bylo také vytvořeno anotační schéma *Pražského závislostního korpusu*, který slouží jako hlavní zdroj jazykových dat pro předkládanou práci. Tato práce tedy navazuje zejména na pojetí slovosledu a aktuálního členění funkčního generativního popisu.

Komunikačně orientovaný popis slovosledu a aktuálního členění i FGP sestavují předpokládanou stupnici výpovědní dynamičnosti (VD) pro kontextově nezapojené členy věty, viz přehledové srovnání níže. FGP nazývá stupnici VD systémovým uspořádáním, viz P. Sgall a kol. (1980). Ke stupnici VD ve druhém pojetí souhrnně viz zejména kapitolu L. Uhlířové *Základní rozložení VD* v *Mluvnici češtiny 3* /1987/ vycházející z pojetí J. Firbase (1975 aj.). Obě stupnice mají do jisté míry odrážet slovosled realizovaných výpovědí (nemusí se s ním ale vždy shodovat zcela – proto FGP rozlišuje tzv. slovosled hloubkový, zachovávající systémové uspořádání, a slovosled povrchový, který se od hloubkového může v některých případech lišit).

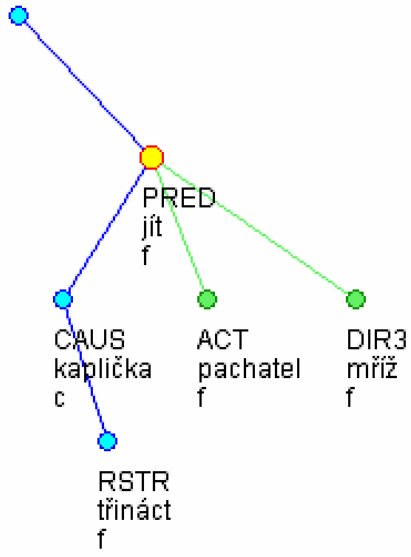
Předkládaná disertační práce se zaměřuje na slovosled povrchový, tj. zkoumá na dvojicích participantů objevujících se ve skutečně realizovaných výpovědích, jaké obvyklé slovosledné postavení ve výpovědích vůči sobě zaujímají, pokud jsou kontextově nezapojené. Jejím cílem není rozhodnout, které ze zjištěných pořadí je základní, ale klade si otázku, zda je možné v povrchovém slovosledu sledovat určité obecnější tendence.

Následující přehled porovnává stupnici systémového uspořádání FGP a stupnici výpovědní dynamičnosti v nescénických větách, jak ji prezentuje *Mluvnice češtiny 3*. Do srovnání byly zahrnuty jen ty participanty, které se nacházejí v obou stupnicích.

Systémové uspořádání	**VD v nescénických větách v *Mluvnici češtiny 3***
konatel	agens
čas (kdy)	čas (kdy)
čas (odkdy)	čas (odkdy)
čas (dokdy)	čas (dokdy)
čas (jak často)	čas (jak často)
čas (jak dlouho)	čas (jak dlouho)
místo	místo
nástroj (prostředek)	prostředek
adresát	beneficiens
patiens	patiens
podmínka	podmínka
účel	příčina
příčina	účel

3 Jazykový materiál

Zdrojem jazykového materiálu pro vlastní analýzu je zejména *Pražský závislostní korpus 2.0* (PDT; J. Hajič a kol., 2006b), který čítá 38 727 výpovědí využitelných pro provedenou slovoslednou analýzu. Tento korpus obsahuje (na tzv. tektogramatické rovině, tj. na hloubkové, sémantické rovině) články z deníků *Lidové noviny* a *Mladá fronta Dnes* a z ekonomického týdeníku *Českomoravský profit*. PDT obsahuje anotaci kontextové zapojenosti, viz obrázek 1.



Obrázek 1: Ukázka závislostního stromu s anotací kontextové zapojenosti z *Pražského závislostního korpusu*. Závislostní strom reprezentuje větu *Za třináct kapiček jdou pachatelé za mříže* (jedná se o první větu textu). Symbol *c* v obrázku označuje uzel kontrastivně kontextově zapojený, symbol *f* uzly kontextově nezapojené.

Doplňkově je využit i *Český národní korpus* (synchronní část) a pro srovnání některých slovosledných tendencí češtiny a němčiny i synchronní texty německého korpusu *Digitales Wörterbuch der deutschen Sprache*.

4 Stručný popis závěrů práce

Cílem práce bylo popsat dílčí aspekty českého (a zčásti také německého) slovosledného uspořádání výpovědí pocházejících převážně z novinových deníků.

První část práce zkoumala, jaké je obvyklé místo jednotlivých typů participantů v aktuálním členění větném. Na základě toho, v kolika procentech se členy věty vyskytovaly v datech *Pražského závislostního korpusu 2.0* jako kontextově zapojené a nezapojené,[2] byla sestavena stupnice kontextové zapojenosti (nejprve jen pro participanty vyjádřené nevětně). Z této stupnice vyplývá zjištění, že obvykle kontextově zapojená bývají (nevětně vyjádřená) volná doplnění vyjadřující měřítko, přípustku, čas (kdy), výjimku, čas (současně s čím / během jaké doby) a aktant aktor, který často zastává syntaktickou funkci podmětu. Naopak často kontextově nezapojená bývají (nevětně vyjádřená) volná doplnění vyjadřující míru, způsob, dědictví, účinek, záměr, směr (kudy) a účel.

Získaná stupnice byla porovnána se stupnicí tematičnosti Ludmily Uhlířové (1974). Výrazně jiné je ve stupnicích postavení účelu. L. Uhlířová jej klade mezi členy, které často zaujímají pozici tématu, data z *Pražského závislostního korpusu* jej ale řadí mezi členy, které jsou častěji kontextově nezapojené. Podobné je to i u určení vyjadřujících zřetel. Nižší míru tematičnosti / kontextové zapojenosti než ve stupnici L. Uhlířové vykazují data PDT i u určení podmínky. Naopak vyšší míru tematičnosti / kontextové zapojenosti oproti L. Uhlířové ukazují data PDT u původce (resp. autora) a u členů vyjadřujících výsledek a omezení (PDT má však jen kategorii výjimka).

Podobnou míru tematičnosti / kontextové zapojenosti shledaly oba průzkumy u určení vyjadřujících přípustku, čas, příčinu, (široce pojaté) místo, původ, prostředek a míru.

U určení míry tematičnosti či rematičnosti, resp. kontextové zapojenosti či nezapojenosti jednotlivých sémantických kategorií se tedy oba výzkumy zhruba v polovině případů víceméně shodují, ve druhé polovině se liší (nebo nejsou zcela porovnatelné, zejména kvůli jinému nastavení kategorií, srov. omezení vs. výjimka, původce vs. autor).

Stupnice kontextové zapojenosti byla posléze sestavena i pro členy vyjádřené větně. Ukázalo se, že podle dat *Pražského závislostního korpusu* má participant vyjádřený větou silnou tendenci k tomu, aby v aktuálním členění fungoval jako kontextově nezapojený prvek. Jako nejčastěji kontextově nezapojené vystupují věty vyjadřující zejména způsob, patiens, výsledek a účinek.

[2] Kontextová zapojenost je hodnocena v souladu s její anotací v *Pražském závislostním korpusu*.

Sestavené stupnice kontextové zapojenosti mohou do jisté míry odrážet obecnější tendence ve slovosledném uspořádání české výpovědi (s objektivním slovosledem, v němž ohniskové prvky následují za základovými).

Dále byla pozornost zaměřena na kontextově nezapojenou část výpovědi. Na dvojicích kontextově nezapojených participantů bylo na datech *Pražského závislostního korpusu 2.0* ověřováno, zda je jedno z jejich slovosledných uspořádání výrazně frekvenčně převažující (zda je častější užívat např. slovosled *Narodil se v roce 1950 v Brně* nebo *Narodil se v Brně v roce 1950*). Opět bylo rozlišováno mezi participanty vyjádřenými větně a nevětně.

Z výsledků výzkumu je patrné, že v některých případech v kontextově nezapojené části věty skutečně můžeme (na nevětně vyjádřených participantech) pozorovat (alespoň v datech *Pražského závislostního korpusu*) určitou tendenci k jistému slovosslednému postavení (např. míra – patiens; způsob – patiens; čas – místo; čas – patiens; adresát – patiens), viz tabulka 1 a graf 1. V jiných případech se však z pohledu povrchové syntaxe zdá, že určité dvojice participantů preferované slovosledné pořadí (z hlediska frekvence výskytu) nemají (např. patiens / doplněk; patiens / prostředek).

Z pohledu povrchové syntaxe najdeme pro většinu dvojic kontextově nezapojených, nevětně vyjádřených aktantů a volných doplnění obě slovosledná pořadí (více či méně zastoupená). Jednotlivé dvojice tak vykazují pouze jistou (různě silnou) tendenci k určitému slovosslednému pořadí.

Výsledky této části práce jsou porovnávány se stupnicí výpovědní dynamičnosti z *Mluvnice češtiny 3* a s tzv. systémovým uspořádáním Petra Sgalla, Evy Hajičové a Evy Buráňové (1980) s vědomím toho, že systémové uspořádání bylo stanoveno pro tzv. slovosled hloubkový, zatímco tato práce se zaměřuje na slovosled povrchový.

Z 20 získaných dvojic (které se v PDT vyskytly alespoň ve 100 případech) nelze se systémovým uspořádáním srovnat 3 (systémové uspořádání nezahrnuje kategorie doplněk, ne/prospěch a doprovod). Zbývá tedy 17 dvojic. U 10 z nich data *Pražského závislostního korpusu* potvrdila vzájemné pořadí, které předpokládá stupnice systémového uspořádání, u 7 z nich systémové uspořádání nepotvrdila.

Pro větně vyjádřené, kontextově nezapojené participanty data *Pražského závislostního korpusu 2.0* ukazují, že (alespoň u některých typů vět) existuje tendence, aby věta stála spíše za nevětně vyjádřeným participantem. Tato tendence koresponduje s již známým slovosledným jevem – delší členy obvykle následují po kratších (srov. k tomu např. *Mluvnice češtiny 3*, 1987; P. Sgall a kol., 1980; Š. Zikánová, 2006). Podobnou tendenci podle získaného materiálu najdeme zejména u kontextově nezapojených závislých vět v roli aktoru, patientu,

účelu nebo příčiny. Ne vždy ale tento jev můžeme pozorovat. Například u kontextově nezapojených závislých vět vyjadřujících výsledek děje nebo podmínku najdeme významně zastoupená obě pořadí. Větná forma tedy nemusí být jako slovosledný faktor rozhodující.

V další části práce bylo pro potřeby výzkumu vlivu valence na slovosled s pomocí *Valenčního slovníku českých sloves* (2008) ověřováno, které větné participanty mohou ve větě vystupovat jako obligatorní (tedy ze sémantického hlediska nutně přítomné v hloubkové větné struktuře). Získané výsledky byly porovnány s popisem valence větných členů W. Flämiga (1991) pro němčinu.

Ve Flämigově teorii je valence považována za jeden z hlavních slovosledných faktorů v němčině. Sledování slovosledných faktorů, které se výrazně projevují v němčině, může být užitečné i pro popis slovosledu českého, protože zejména v minulosti docházelo ke zřetelnému kontaktnímu ovlivňování češtiny němčinou (a to i v oblasti slovosledu). V češtině například není přísudkové sloveso považováno za hranici mezi základem a ohniskem, pokud stojí ve větě na druhém místě. Čeština nemá postavení slovesa pevně dané gramatikou, přesto se přísudek na druhém místě objevuje poměrně často. Toto postavení se někdy vysvětluje právě kontaktem s němčinou (srov. např. E. Hajičová a kol., 2002: 105), v níž je v hlavní větě pozice určitého slovesného tvaru na druhém místě gramatikalizována.

Ukázalo se, že ne/možnosti být obligatorním volným slovesným doplněním ve větě jsou v němčině a v češtině velice podobné, téměř shodné.

Na základě frekvence výskytu volných doplnění ve valenčních rámcích sloves v roli obligatorního členu byla pro češtinu sestavena následující „stupnice obligatornosti“[3]:

směr (kam) – směr (odkud) – místo (kde) – způsob (jak) – směr (kudy) – míra – čas (ze kdy / na kdy).

Zároveň se ukázalo, že se v jednom slovesném valenčním rámci může (alespoň podle *Valenčního slovníku českých sloves*) zároveň vyskytovat několik valenčně obligatorních aktantů (aktanty rozumíme v souladu s teorií funkčního generativního popisu aktor, patiens, adresát, výsledek děje a původ), ale ne několik valenčně obligatorních volných doplnění (jako jsou např. volné doplnění času, místa, způsobu, přípustky, příčiny ap.) – s výjimkou volných doplnění vyjadřujících „ze kdy na kdy“ a „odkud kam“, která ale lze považovat za jedno složené časové a jedno složené směrové doplnění. Tento poznatek by se mohl stát dalším kritériem pro rozlišení aktantů a volných doplnění.

[3] Na začátku jsou volná doplnění, která jsou obligatorní u nejvíce slovesných lexií, na konci ta, která se jako obligatorní objevují nejméně.

Dalším z cílů práce bylo ověřit na datech *Pražského závislostního korpusu 2.0* (PDT), jak funguje slovesná valence jako slovosledný faktor. Data PDT například ukazují, že nevětně vyjádřené, kontextově nezapojené doplnění směru (kam) se v povrchovém slovosledu chová podobně v obou případech, tj. ať je ve větě obligatorně vázáno slovesnou valencí nebo ať je členem fakultativním.[4] V obou případech má tendenci stát za fakultativními volnými slovesnými doplněními, tedy spíše ke konci věty. Podrobnější popis slovosledu ostatních typů obligatorních slovesných doplnění je uveden v závěrečných výsledcích vlastní práce. Vyhodnocení vlivu valence na slovosled bylo v některých případech limitováno relativně nízkým výskytem valenčně obligatorních doplnění v korpusu. Podle získaných dat se však zdá, že slovesná valence nemusí být vždy nejsilnějším slovosledným faktorem.

Posledním z cílů práce bylo ověřit na vybraných typech participantů (na volném doplnění místa, času a způsobu) vyjádřených zájmenným příslovcem (*někde, někdy, nějak*), zda existuje jejich frekvenčně výrazně převažující pořadí (zmíněná lemmata byla vybrána kvůli tomu, že participanty jimi vyjádřené mají jednotnou formu, délku, velmi podobnou míru lexikálního významu a pravděpodobně ve většině případů jsou kontextově nezapojené). Tento cíl byl ověřován pro češtinu i pro němčinu. Pro češtinu na datech *Českého národního korpusu*, pro němčinu na datech korpusu *Digitales Wörterbuch der deutschen Sprache*.

Z provedené sondy vyplývá, že v češtině i v němčině lze použít obě slovosledná pořadí zkoumaných dvojic slov. Zdá se však, že zejména u některých dvojic je jedno z pořadí upřednostňováno. V obou jazycích je zřejmě například běžnější používat fráze s pořadím čas – místo, tedy „někdy někde“.

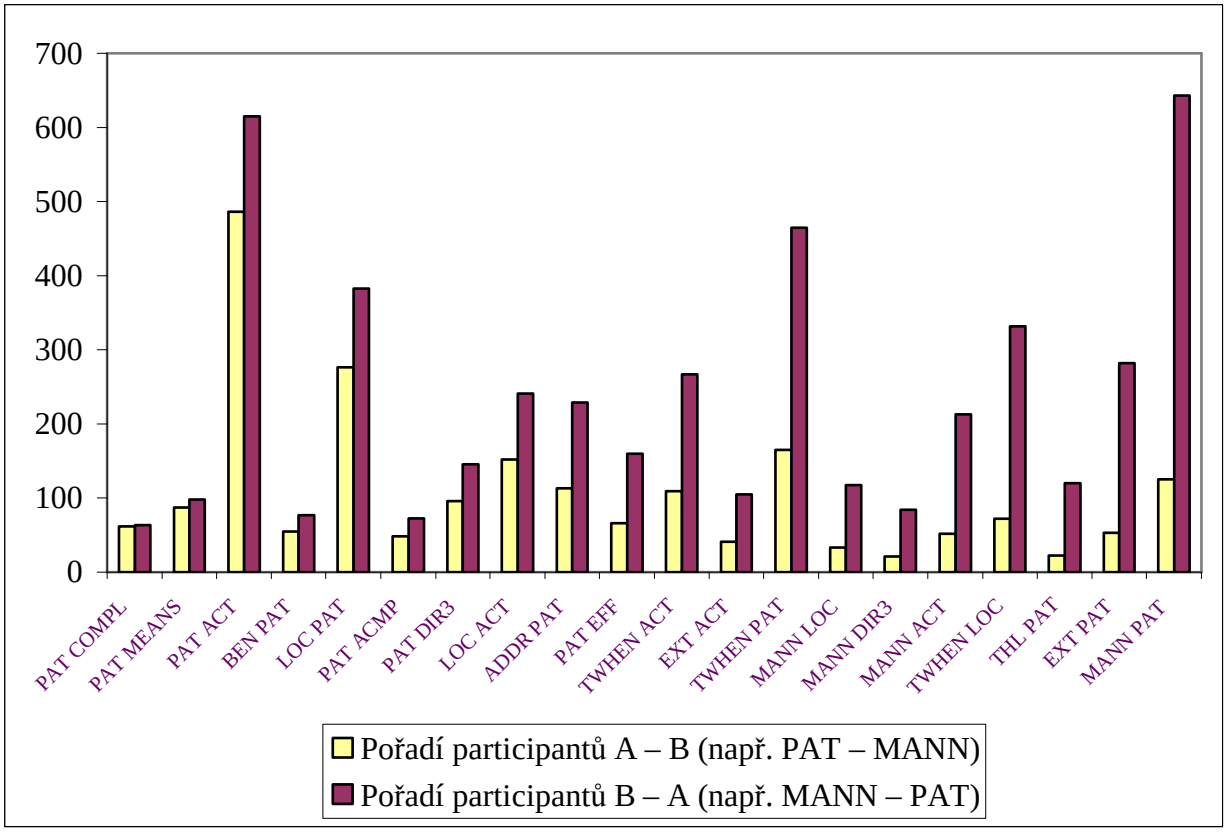
Práce celkově ukazuje, že v českém povrchovém slovosledu publicistických textů lze v některých případech nalézt obecnější slovosledné tendence, resp. že některé kontextově nezapojené participanty tíhnou ve výpovědi k určitému typu řazení, čehož by bylo možné využít například při automatických zpracováních textu.

[4] Obligatornost a fakultativnost participantů byla hodnocena podle valenčního slovníku *PDT-Vallex*.

Pořadí participantů A – B	**Počet výskytů dvojice A – B**	**Pořadí participantů B – A**	**Počet výskytů dvojice B – A**	**Celkem**	**Pravděpodobnost (B – A)**	**Reject (0.95)**	**Reject (0.999)**	**Reject (0.9999)**
COMPL doplněk – PAT patiens	62	PAT patiens – COMPL doplněk	63	125	0,50	0	0	0
MEANS prostředek – PAT patiens	87	PAT patiens – MEANS prostředek	98	185	0,53	0	0	0
ACT aktor – PAT patiens	486	PAT patiens – ACT aktor	615	1 101	0,56	1	1	0
PAT patiens – BEN ne/prospěch	55	BEN ne/prospěch – PAT patiens	77	132	0,58	0	0	0
PAT patiens – LOC místo	276	LOC místo – PAT patiens	383	659	0,58	1	1	1
ACMP doprovod – PAT patiens	48	PAT patiens – ACMP doprovod	73	121	0,60	1	0	0
DIR3 směr kam – PAT patiens	96	PAT patiens – DIR3 směr kam	145	241	0,60	1	0	0
ACT aktor – LOC místo	152	LOC místo – ACT aktor	241	393	0,61	1	1	1
PAT patiens – ADDR adresát	113	ADDR adresát – PAT patiens	229	342	0,67	1	1	1
EFF výsledek – PAT patiens	66	PAT patiens – EFF výsledek	160	226	0,71	1	1	1
ACT aktor – TWHEN čas kdy	109	TWHEN čas kdy – ACT aktor	267	376	0,71	1	1	1
ACT aktor – EXT míra	41	EXT míra – ACT aktor	105	146	0,72	1	1	1
PAT patiens – TWHEN čas kdy	165	TWHEN čas kdy – PAT patiens	465	630	0,74	1	1	1
LOC místo – MANN způsob	33	MANN způsob – LOC místo	117	150	0,78	1	1	1
DIR3 směr kam – MANN způsob	21	MANN způsob – DIR3 směr kam	84	105	0,80	1	1	1
ACT aktor – MANN způsob	52	MANN způsob – ACT aktor	213	265	0,80	1	1	1
LOC místo – TWHEN čas kdy	72	TWHEN čas kdy – místo LOC	332	404	0,82	1	1	1
PAT patiens – THL čas jak dlouho	22	THL čas jak dlouho – PAT patiens	120	142	0,83	1	1	1
PAT patiens – EXT míra	53	EXT míra – PAT patiens	282	335	0,84	1	1	1
PAT patiens – MANN způsob	125	MANN způsob – PAT patiens	634	759	0,84	1	1	1

Tabulka 1: Počet výskytů nevětně vyjádřených, kontextově nezapojených participantů ve dvojici: v pořadí A – B a v pořadí B – A (na základě dat *Pražského závislostního korpusu 2.0*), viz také graf 1

Poslední tři sloupce tabulky se záhlavím „Reject“ zobrazují výsledek testu dobré shody (Pearsonova chí-kvadrát testu), který je využíván jako statistická metoda k vyhodnocení získaných dat. Je-li v těchto sloupcích hodnota 1, znamená to, že ověřovaná hypotéza (testovaný vzorek je do skupin A – B a B – A rozložen rovnoměrně, odchylka je daná pouhou náhodou) byla (na dané úrovni spolehlivosti) zamítnuta. Pokud je v nich hodnota 0, znamená to, že tato hypotéza zamítnuta nebyla. Test byl proveden pro tři různé úrovně jistoty: 0.95, 0.999 a 0.9999 (čím vyšší hodnota úrovně, tím je výsledek testu spolehlivější). Obvyklá míra pravděpodobnosti chyby, pro kterou se test dobré shody obecně počítá, je 5 %.



Graf 1: Počet výskytů nevětně vyjádřených, kontextově nezapojených participantů ve dvojici: v pořadí A – B a v pořadí B – A (na základě dat *Pražského závislostního korpusu 2.0*); zobrazeny jen dvojice, které se v korpusu vyskytly celkem alespoň 100krát

Pouze ve třech případech (u dvojic kontextově nezapojených, nevětně vyjádřených participantů, které se v *Pražském závislostním korpusu* celkem v obou slovosledných pořadích vyskytly alespoň 100krát) test dobré shody nezamítl (na hladině spolehlivosti 95 %) hypotézu, že daný vzorek je rozložený rovnoměrně, a to u dvojic:

patiens / doplněk;
patiens / prostředek;
patiens / (ne)prospěch.

U ostatních 17 dvojic lze předpokládat, že rozložení testovaného vzorku není rovnoměrné. Jejich frekvenčně převažující pořadí je tedy podle dat *Pražského závislostního korpusu* následující:

adjektivní patiens – infinitivní aktor;
substantivní aktor – substantivní patiens;
místo – patiens;
patiens – doprovod;
patiens – směr (kam);
místo – aktor;
adresát – patiens;
patiens – výsledek;
čas (kdy) – aktor;
míra – aktor;
čas (kdy) – patiens;
způsob – místo;
způsob – směr (kam);
způsob – aktor;
čas (kdy) – místo;
čas (jak dlouho) – patiens;
míra – patiens;
způsob – patiens.

5 Závěrečný komentář

Práce celkově ukazuje, že v českém povrchovém slovosledu publicistických textů lze v některých případech nalézt obecnější slovosledné tendence, resp. že některé kontextově nezapojené participanty tíhnou ve výpovědi k určitému typu řazení. Zjištěné tendence by bylo možné využít při automatickém zpracování textu. V současné době vzniká mnoho strojových překladačů a ukazuje se, že jedním ze závažných problémů, které komplikují kvalitu jejich překladu, je jejich zacházení se slovosledem (zejména v jazycích, v nichž slovosled není pevně vázán gramatickými pravidly). Získaná slovosledná schémata by tak mohla pomoci jejich práci zdokonalit.

Literatura

DANEŠ, František. *Intonace a věta ve spisovné češtině*. 1. vyd. Praha: ČSAV, 1957.

DANEŠ, František. Sémantická struktura slovesná a struktura věty. In *Zborník filozofickej fakulty Univerzity Komenského, Philologica 1971–2*. Bratislava: 1974, s. 17–25.

DANEŠ, František. Typy tematických posloupností v textu (na materiále českého textu odborného). *Slovo a slovesnost*. 1968, roč. 29, č. 2, s. 125–141. ISSN 0037-7031.

DANEŠ, František; HLAVSA, Zdeněk et al. *Větné vzorce v češtině*. 1. vyd. Praha: Academia, 1981.

DANEŠ, František; HLAVSA, Zdeněk; GREPL, Miroslav. *Mluvnice češtiny: vysokoškolská učebnice pro studenty filozofických a pedagogických fakult. Díl 3. Skladba.* 1. vyd. Praha: Academia, 1987.

FIRBAS, Jan. A functional view of „ordo naturalis“. *Brno Studies in English.* 1979, roč. 13, s. 29–59. ISSN 0524-6881.

FIRBAS, Jan. A Note on Transition Proper in Functional Sentence Analysis. *Philologica Pragensia.* 1965, roč. 8, s. 170–176.

FIRBAS, Jan. *Functional Sentence Perspective in Written and Spoken Communication*. 1. vyd. Cambridge: Cambridge University Press, 1992. ISBN 0-521-03182-6. ISBN 978-0-521-03182-0. ISBN 0-521-37308-5. ISBN 978-0-521-37308-5.

FIRBAS, Jan. On 'existence/appearance on the scene' in functional sentence perspektive. *Prague Sludies in English*. 1975, roč. 16, s. 47–70.

FIRBAS, Jan. On the Communicative Value of the Modern English Finite Verb. *Brno Studies in English.* 1961, roč. 3, s. 97–104.

FIRBAS, Jan. On the concept of communicative dynamism in the theory of functional sentence perspective. *Sborník prací filozofické fakulty brněnské univerzity.* 1971, A 19, s. 135–144. ISSN 0231-7710.

FIRBAS, Jan. On the Prosodic Features of the Modern English Finite Verb as Means of Functional Sentence Perspektive. *Brno Studies in English.* 1968, roč. 7, s. 12–47.

FLÄMIG, Walter: *Grammatik des Deutschen: Einführung in Struktur- und Wirkungszusammenhänge*. 1. vyd. Berlin: Akademie Verlag, 1991. ISBN 3-05-000686-2.

HAJIČ, Jan et al. *Průvodce PDT 2.0* [online]. Praha: 2006a [cit. 24. 07. 2012]. Dostupné z WWW: <http://ufal.mff.cuni.cz/pdt2.0/doc/pdt-guide/cz/html/index.html>.

HAJIČOVÁ, Eva. *Negace a presupozice ve významové stavbě věty*. 1. vyd. Praha: Academia, 1975.

HAJIČOVÁ, Eva. Negation and Topic vs. Comment. *Philologica Pragensia.* 1973, roč. 16, s. 91–93.

HAJIČOVÁ, Eva. What we have learned from complex annotation of topic-focus articulation in a large Czech corpus. *Écho des études romanes*. 2012a, roč. 8, č. 1, s. 51–64. ISSN 1801-0865.

HAJIČOVÁ, Eva; HAVELKA, Jiří; VESELÁ, Kateřina. Corpus Evidence of Contextual Boundness and Focus [online]. In *Proceedings from The Corpus Linguistics Conference Series.* Birmingham: 2005, roč. 1, č. 1. Dostupné z WWW: <http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-birmingham.aspx>. ISSN 1747-939.

HAJIČOVÁ, Eva; PANEVOVÁ, Jarmila; SGALL, Petr. *Úvod do teoretické a počítačové lingvistiky*. Svazek 1. 1. vyd. Praha: Karolinum, 2002. ISBN 80-246-0470-1.

HAJIČOVÁ, Eva; PARTEE, Barbara H.; SGALL, Petr. *Topic-Focus Articulation, Tripartite Structures and Semantic Content*. 1. vyd. Dordrecht – Boston: Kluwer Academic Publishers, 1998. ISBN 0-7923-5289-0.

HAJIČOVÁ, Eva; SGALL, Petr. Degrees of Contrast and the Topic-Focus Articulation. In STEUBE, A. (ed.). *Information Structure: Theoretical and Empirical Aspects*. Berlin – New York: de Gruyter, 2004, s. 1–13.

HAJIČOVÁ, Eva; SGALL, Petr; SKOUMALOVÁ, Hana. Identifying Topic and Focus by an Automatic Procedure. In KRAUWER, S. et al. (eds.). *Sixth Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, 21–23 April 1993*. Utrecht: 1993, s. 178–182. ISBN 90-5434-014-2.

HAJIČOVÁ, Eva. Topic-Focus revisited (Through the eyes of the Prague Dependency Treebank). In: Смыслы, тексты и другие захватывающие сюжеты. Сборник статей в честь 80-летия Игоря Александровича Мельчука. Rusko, Moskva: 2012b, s. 218–232. ISBN 978-5-9551-0593-2.

LOPATKOVÁ, Markéta; ŽABOKRTSKÝ, Zdeněk; KETTNEROVÁ, Václava et al. *Valenční slovník českých sloves*. 1. vyd. Praha: Karolinum, 2008. Dostupné také z WWW: <http://ufal.mff.cuni.cz/vallex/2.5/doc/home.html>. ISBN 978-80-246-1467-0.

MATHESIUS, Vilém. *Čeština a obecný jazykozpyt: soubor statí*. 1. vyd. Praha: Melantrich, 1947.

MIKULOVÁ, Marie et al. *Anotace na tektogramatické rovině Pražského závislostního korpusu: referenční příručka*. 1. vyd. Praha: Universitas Carolina Pragensia, 2008. Dostupné také z WWW: <http://ufal.mff.cuni.cz/pdt2.0update/doc/tr-ref-cz-cz.pdf>. ISBN 978-80-254-3769-8. ISBN 978-80-7378-063-0.

MIKULOVÁ, Marie et al. *Anotace na tektogramatické rovině Pražského závislostního korpusu: anotátorská příručka*. 1. vyd. Praha: Universitas Carolina Pragensia, 2005. Dostupné také z WWW: <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/t-layer/html/index.html>. ISBN 80-254-3087-1.

NEBESKÁ, Iva. K vyjadřování předmětu a příslovečných určení v češtině. *Naše řeč.* 1984, roč. 67, s. 10–17.

PANEVOVÁ, Jarmila. *Formy a funkce ve stavbě české věty*. 1. vyd. Praha: Academia, 1980.

PANEVOVÁ, Jarmila. On Verbal Frames in Functional Generative Description. *The Prague Bulletin of Mathematical Linguistics.* 1974, roč. 22, s. 3–40.

PANEVOVÁ, Jarmila. Rozvití předmětová a příslovečná, doplňující a určující. *Naše řeč.* 1975, roč. 58, s. 61–66.

PANEVOVÁ, Jarmila. Valency Frames and the Meaning of the Sentence. In LUELSDORFF, P. A. (ed.). *The Prague School of Structural and Functional Linguistics*. Amsterdam – Philadelphia: John Benjamins Publishing Copany, 1994, s. 223–243.

RYSOVÁ, Kateřina. Ke slovosledu v konstrukcích s obligatorními participanty (se zaměřením na slovosled patientu a způsobového slovesného doplnění). In PETKEVIČ, V.; ROSEN, A. (eds.). *Korpusová lingvistika Praha 2011. 3. Gramatika a značkování korpusů.* Praha: Ústav českého národního korpusu FF UK; Nakladatelství Lidové noviny, 2011a, s. 62–69. ISBN 978-80-7422-116-3.

RYSOVÁ, Kateřina. Možnosti jednotlivých volných slovesných doplnění být obligatorním členem věty. In ČMEJRKOVÁ, S.; HOFFMANNOVÁ, J.; KLÍMOVÁ, J. (eds.). *Čeština v pohledu synchronním a diachronním: Stoleté kořeny Ústavu pro jazyk český*. Praha: Karolinum, 2012a, s. 615–620. ISBN 978-80-246-2121-0.

RYSOVÁ, Kateřina. *Někdy, někde, nějak* v češtině a v němčině (Ke slovosledu pořadí kategorií času, místa a způsobu). In PALECSKOVÁ, D.; KUMOROVÁ, Z.; GREGORÍK, P. (eds.). *RARA AVIS IX. 1. diel – jazykoveda*. Trnava: Filozofická fakulta Univerzity sv. Cyrila a Metoda v Trnave, Katedra slovenského jazyka a literatúry, 2012b, s. 168–177. ISBN 978-80-8105-403-7.

RYSOVÁ, Kateřina. The Word Order of Inner Participants in Czech, Considering the Systemic Ordering of Actor and Patient [online]. In GERDES, K.; HAJIČOVÁ, E.; WANNER, L. (eds.). *Proceedings of the International Conference on Dependency Linguistics (Depling 2011)*. Barcelona, Španělsko: Universitat Pompeu Fabra, 2011b, s. 183–192. Dostupné z WWW: <http://depling.org/proceedingsDepling2011/>. ISBN 978-84-615-1834-0.

RYSOVÁ, Kateřina. Umístění valenčního doplnění místa a patientu v české výpovědi. In NESWETHA, J.; CZERWONKA, Z.; RYSOVÁ, K. (eds.). *Grenzüberschreitungen: Polnische, tschechische und deutsche Sprache, Literatur und Kultur.* Hildesheim – Zürich – New York: Georg Koms Verlag, 2013, s. 321–330. ISBN: 978-3-487-15004-8. ISSN 1860-4692.

SGALL, Petr. *Generativní popis jazyka a česká deklinace*. 1. vyd. Praha: Academia, 1967.

SGALL, Petr. Generativní systémy v lingvistice. *Slovo a slovesnost*. 1964, roč. 25, č. 4, s. 274–282.

SGALL, Petr; HAJIČOVÁ, Eva. The Position of Information Structure in the Core of Language. In CARLSON, G. N. a PELLETIER, F. J. (eds.). *Reference and Quantification: The Partee Effect.* Stanford (California): CSLI Publications, 2005, s. 289–302.

SGALL, Petr; HAJIČOVÁ, Eva; BURÁŇOVÁ, Eva. *Aktuální členění věty v češtině*. 1. vyd. Praha: Academia, 1980.

SGALL, Petr; HAJIČOVÁ, Eva; PANEVOVÁ, Jarmila. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspect.* Dordrecht: Reidel Publishing Company, Praha: Academia, 1986.

SGALL, Petr. Towards a definition of Focus and Topic. *Prague Bulletin of Mathematical Linguistics*. 1979, roč. 31, s. 3–27.

SHEVCHUKOVA, Antonina. *Ausklammerung und Relativsatzbildung als Rhematisierungsstrategie im Deutschen.* Gießen, 2009. Disertační práce. Justus-Liebig-Universität Gießen. Dostupné z WWW: <http://geb.uni-giessen.de/geb/volltexte/2011/8183/pdf/ShevchukovaAntonina_2009_07_08.pdf>.

TESNIÈRE, Lucien. *Éléments de syntaxe structurale.* 1. vyd. Paris: Klincksieck, 1959. ISBN 2-252-01861-5.

UHLÍŘOVÁ, Ludmila. *Knížka o slovosledu*. 1. vyd. Praha: Academia, 1987.

UHLÍŘOVÁ, Ludmila. O vztahu sémantiky příslovečného určení k aktuálnímu členění. *Slovo a slovesnost*. 1974, roč. 35, s. 99–106.

UREŠOVÁ, Zdena. *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)* [online]. Příloha k disertační práci. Praha: 2011 [cit. 27. 07. 2012]. Dostupné z WWW: <http://ufal.mff.cuni.cz/PDT-Vallex/PDT-Vallex-plna-verze.pdf>.

VESELÁ, Kateřina; HAVELKA, Jiří. *Anotování aktuálního členění věty v Pražském závislostním korpusu*. Technická zpráva. Praha: ÚFAL MFF UK, 2003. Dostupné z WWW: <http://ufal.mff.cuni.cz/pdt2.0/publications/VeselaHavelkaTR2003.pdf >.

ZIKÁNOVÁ, Šárka. *Postavení slovesného přísudku ve starší češtině (1500–1620)*. 1. vyd. Praha: Karolinum, 2009. ISBN 978-80-246-1381-9.

ZIKÁNOVÁ, Šárka. Problematické syntaktické struktury: k rozborům aktuálního členění v Pražském závislostním korpusu. In POLÁCH, V. P. (ed.). *Svět za slovy a jejich tvary, svět za spojením slov*. Olomouc: Vydavatelství Univerzity Palackého v Olomouci, 2008, s. 233–240. ISBN 978-80-244-1984-8.

ZIKÁNOVÁ, Šárka. What do the data in Prague Dependency Treebank say about systemic ordering in Czech? *The Prague Bulletin of Mathematical Linguistics.* 2006, roč. 86, s. 39–46.

ZIKÁNOVÁ, Šárka; TÝNOVSKÝ, Miroslav; HAVELKA, Jiří. Identification of Topic and Focus in Czech: Evaluation of Manual Parallel Annotations. *The Prague Bulletin of Mathematical Linguistics*. 2007, roč. 87, s. 61–70.

Zdroje

CIMRMAN, Jára da; SMOLJAK, Ladislav; SVĚRÁK, Zdeněk. *Posel z Liptákova*. Praha: Paseka, 1992. ISBN 80-85192-41-1.

Český národní korpus. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <http://www.korpus.cz>.

Digitales Wörterbuch der deutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften. Berlin. Dostupný z WWW: <http://www.dwds.de/>

HAJIČ, Jan et al. *Prague Dependency Treebank 2.0*. Philadelphia: Linguistic Data Consortium, 2006b.

PDT-Vallex. Ústav formální a aplikované lingvistiky Matematicko-fyzikální fakulty Univerzity Karlovy v Praze. Praha 2012. Dostupný z WWW: <http://ufal.mff.cuni.cz/lindat/PDT-Vallex.html>.