

Koreference a elipsa v Pražském závislostním korpusu

Kateřina Rysová, Magdaléna Rysová

*Ústav formální a aplikované lingvistiky, Matematicko-fyzikální fakulta,
Univerzita Karlova v Praze*

rysova@ufal.mff.cuni.cz, magdalena.rysova@ufal.mff.cuni.cz

Coreference and Ellipsis in the Prague Dependency Treebank

ABSTRACT: The paper focuses on which role is given to elided sentence participants in coreference chains, i.e. whether (and to which degree) the participants that are present only implicitly in the surface layer are involved in relations of textual and grammatical coreference. Generally, the paper introduces the methods how it is possible to examine the interplays of different language phenomena in corpus data of the Prague Dependency Treebank containing multilayer annotation.

KEY WORDS: ellipsis, coreference, corpus linguistics, Prague Dependency Treebank, explicitness, implicitness

KLÍČOVÁ SLOVA: elipsa, koreference, korpusová lingvistika, Pražský závislostní korpus, explicitnost, implicitnost

1 Úvod

Každá běžná komunikace (psaná i mluvená) obsahuje určitou míru implicitnosti. Explicitní vyjádření celého informačního obsahu je v komunikaci nadbytečné a může ztěžovat celkové porozumění danému textu – čtenář či posluchač nemusí správně interpretovat komunikační záměr autora, protože nedokáže rozlišit, která informace je důležitá a která pouze dotváří kontext. Stejně tak může porozumění textu ztěžovat nadbytečná implicitnost – pokud jsou důležité informace v textu pouze implicitně, adresát nemusí vyvodit potřebný význam a může dojít k nesprávné interpretaci celého textu.

Oba extrémny tak mohou vést k neúspěšné komunikaci a každý autor by měl být schopen najít při vytváření textu správnou rovnováhu, tj. posoudit relevanci předkládaných informací (viz Griceovy maximy – Grice 1975 – či teorii relevance a komunikační normy – Sperber a Wilson 1986).

Implicitnost (resp. elipsa) se spolu s dalšími jazykovými jevy, jako je koreference, anafora či sémanticko-pragmatické vztahy, podílí na celkové koherenci textu. Všechny tyto jevy dohromady vytvářejí síť vzájemně propojených vztahů a správné porozumění těmto vztahům vede ke správné interpretaci celého textu. Z tohoto hlediska je důležité nezkoumat zmíněné jevy izolovaně, ale ve vzájemné interakci. V našem článku se zaměřujeme na vztah koreference a elidovaných větných participantů, tj. participantů přítomných v povrchové rovině věty pouze implicitně.

Pražský závislostní korpus 3.0 (PDT; Bejček a kol. 2013) obsahuje anotaci obou těchto jevů zároveň a pro výzkum jazykových interakcí tak poskytuje vhodný jazykový materiál. PDT se skládá z redigovaných autentických novinových textů, o nichž můžeme předpokládat, že jsou v komunikaci plně funkční, tj. je v nich vyvážená míra explicitnosti a implicitnosti.

2 Jazyková data

Výzkum jsme prováděli na datech Pražského závislostního korpusu 3.0, který vychází z teorie funkčního generativního popisu formulovaného Sgallem (1964) a jeho kolegy. Korpus obsahuje téměř 50 000 vět z českých novinových článků. Jednotlivé věty jsou zachycovány v závislostních stromech obsahujících různé typy lingvistických informací. Data PDT jsou anotována převážně ručně, a to na třech jazykových rovinách zároveň: morfologické, analytické (tj. povrchové syntaktické) a tektogramatické (tj. hloubkové syntakticko-sémantické).

Všechna lemmata (uzly) jsou na tektogramatické rovině opatřena značkou syntakticko-sémantické funkce nazývané funktor, jako je např. aktor, patiens, adresát, způsob, čas atd.

Vzhledem k tomu, že tektogramatická rovina zachycuje hloubkovou strukturu vět, jsou zde anotovány také elipsy – srov. příklad (1). V PDT jsou případy elipsy na tektogramatické rovině důsledně rekonstruovány, tj. do závislostních stromů jsou přidávány i uzly, které jsou v povrchové rovině věty přítomny pouze implicitně, viz obrázek (1). Takto rekonstruované (generované) uzly jsou rovněž opatřeny značkou syntakticko-sémantické funkce (funktorem) a dalšími lingvistickými informacemi, viz Mikulová a kol. (2005).

Tektogramatická rovina zároveň obsahuje anotace těchto jazykových jevů: aktuálního členění větného, asociační anafory, koreference, diskurzních vztahů (tj. sémanticko-pragmatických vztahů v rámci textu) či textových žánrů.

Korpus PDT tak poskytuje vhodný jazykový materiál pro výzkum vzájemných vztahů mezi různými jazykovými jevy (i různými jazykovými rovinami). V tomto článku se zaměříme na vztah mezi elidovanými větnými participanty (tj. implicitností v povrchové rovině) a koreferencí.

3 Anotace elipsy v PDT

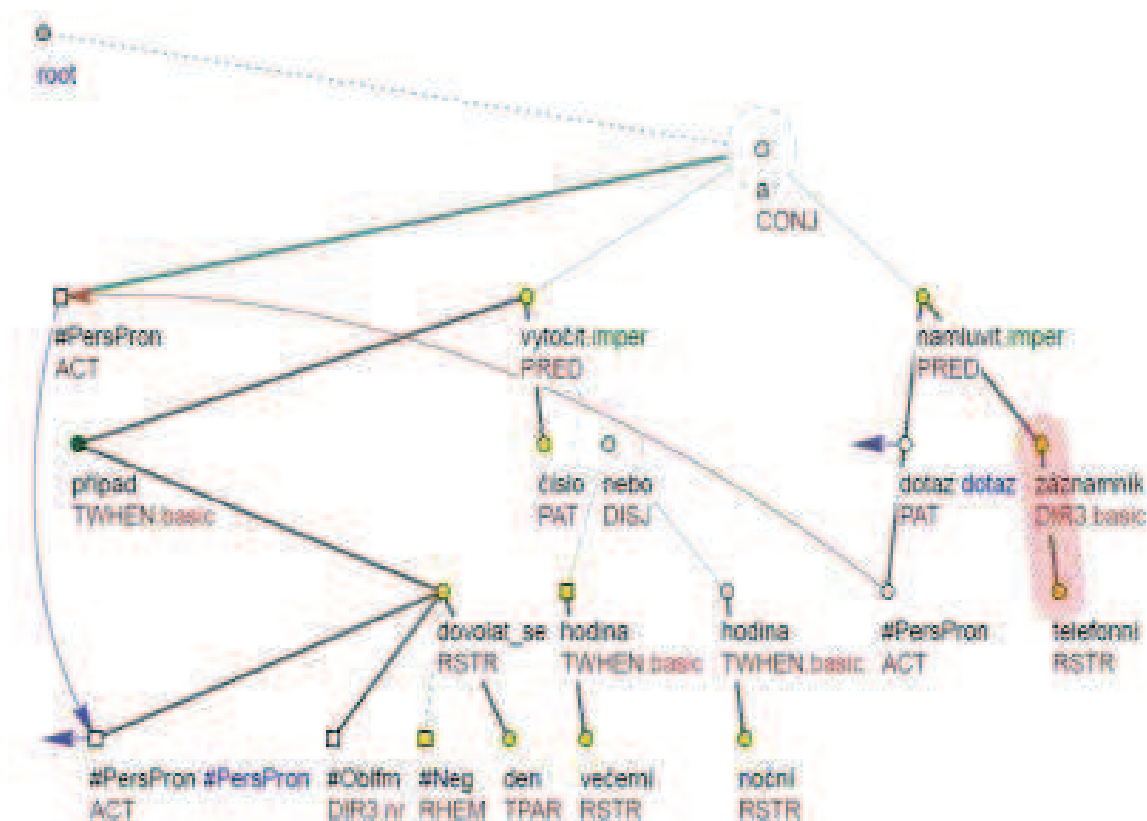
V Pražském závislostním korpusu je elipsa zachycena na hloubkové syntakticko-sémantické (tzv. tektogramatické) rovině, a to pomocí atributu *is_generated*. Jedná se o případy, kdy v povrchové rovině není určitý výraz explicitně vyjádřen, je však přítomný ve významu dané struktury. Takové výrazy tedy nebudeme v realizovaných větách, resp. výpovědích, ale jsou zachyceny v hloubkové struktuře těchto vět. Jedná se přitom o elipsy aktuální (textové) i gramatické.

Aktuálně vypuštěný výraz je vždy zřejmý (a doplnitelný) z kontextu. Příkladem aktuální vypustky je výraz *hodina* v příkladu (1).

- (1) Povrchová realizace: V případě, že se nedovoláte přes den, vytočte číslo ve večerních nebo nočních hodinách a svůj dotaz namluvte na telefonní záznamník. PDT 3.0

Hloubková struktura: ... večerních **hodinách** nebo nočních **hodinách**...

Obrázek 1 – Ukázka anotace elipsy a koreference v PDT – hloubková (tektogramatická) rovina.



První výskyt výrazu *hodina* je z povrchové realizace vypuštěn, protože je zřejmý z kontextu, je jednoznačně doplnitelný a jeho bezprostřední opakování by ztěžovalo porozumění textu. V hloubkové (významové) struktuře věty je však

přítomen. Závislostní strom pro větu z příkladu (1) zobrazuje obrázek (1). Z obrázku je patrné, že ve stromové struktuře jsou prvky přítomné v povrchové i hloubkové rovině označeny kolečkem, prvky na povrchu elidované, ale ve významu obsažené jsou označeny čtverečkem.¹

Další příklad ukazuje elipsu gramatickou – zde doplnění obligatorních valenčních rozvití sloves.

(2) Potřebujete rychle poradit? PDT 3.0

Sloveso *potřebovat* má ve svém valenčním rámci obligatorně dva aktanty – aktor a patiens. Explicitně je ale ve větě vyjádřen pouze patiens (*poradit*). Aktor je přítomen jen implicitně. Podobně sloveso *poradit* má ve valenčním rámci tři aktanty – aktor, adresát, patiens (*někdo poradí někomu s něčím*). Tyto tři aktanty opět nejsou v povrchové rovině vyjádřeny, ale najdeme je v rovině hloubkové. Příklad (2) tedy reprezentuje větu, v níž je většina výrazů přítomna pouze implicitně: v povrchové rovině jsou explicitně vyjádřené tři výrazy (*uzly*), hloubková rovina ale celkem obsahuje výrazů sedm (k elipse v PDT viz dále Mikulová a kol. 2005; Hajičová a kol. 2015; Hajič a kol. 2015).

4 Anotace koreference v PDT

Koreferenční vztahy jsou v PDT anotovány podle Nedoluzhko (2011, 2013). Anotace koreference byla provedena manuálně na celých datech tektogramatické roviny a zahrnuje vztahy mezivětné i vnitrovětné.

Koreferenční vztahy se týkají odkazování v rámci textu, tj. vytvářejí řetězce výrazů často přesahující rámec věty. V PDT se přitom rozlišují dva základní typy koreference – gramatická a textová. Kromě těchto vztahů jsou v PDT značeny také vztahy asociační anafory či exoforického odkazování.

V naší práci se zaměříme na vztahy gramatické a textové koreference.

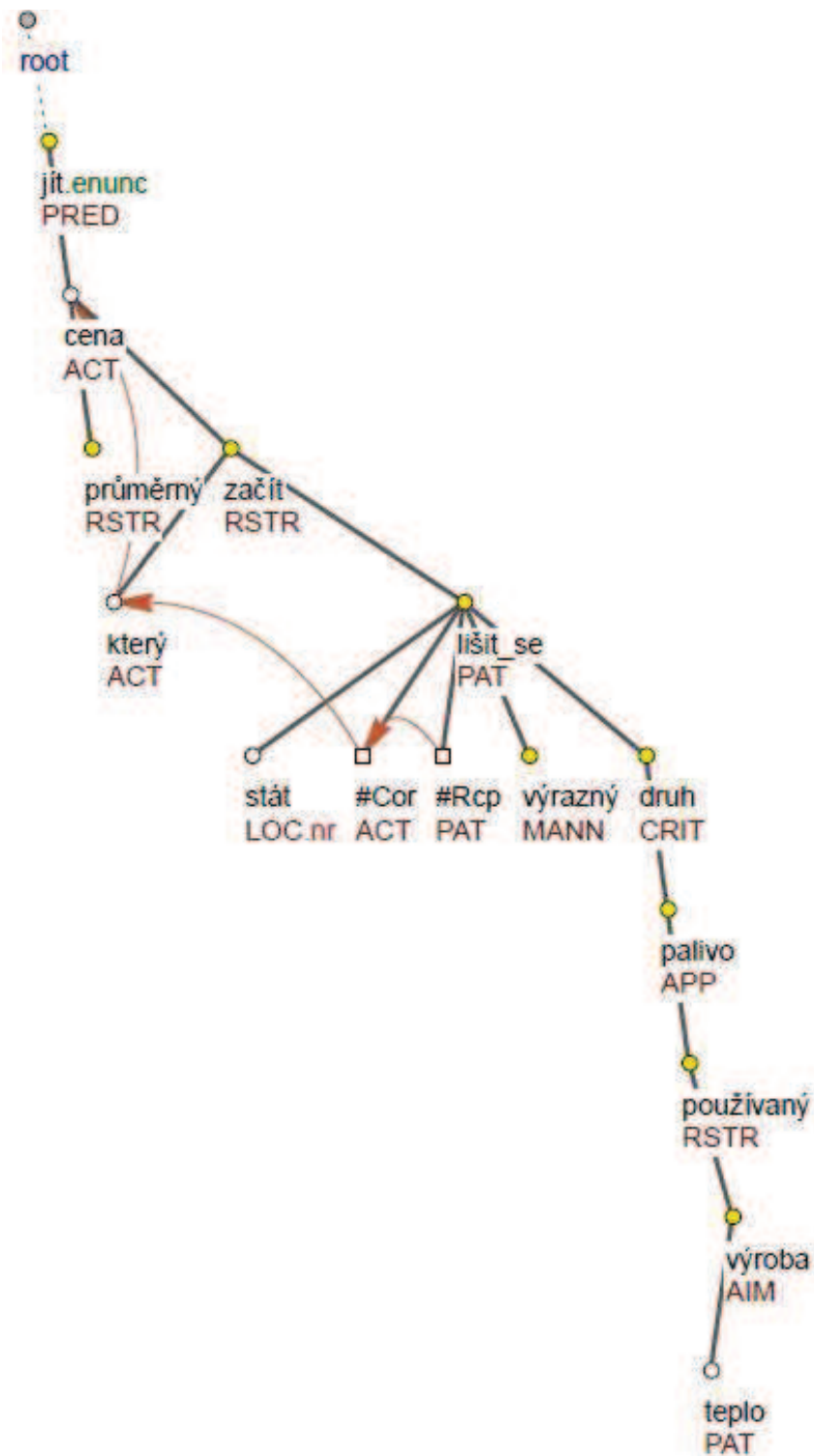
4.1 Anotace gramatické koreference v PDT

Hlavním principem koreference (gramatické i textové) je odkazování jednotlivých výrazů ke stejnému objektu – srov. příklad (3):

(3) Jde o **ceny** průměrné, **kteřé** se začnou v rámci státu výrazně lišit podle druhu paliva, používaného k výrobě tepla. PDT 3.0

¹ Šipka mezi dvěma aktory zde označuje gramatickou koreferenci, ostatní šipky na obrázku značí textovou koreferenci (ke gramatické a textové koreferenci v PDT viz oddíl 4).

Obrázek 2 – Ukázka gramatické koreference v PDT.



Antecedent *cena* odkazuje ke stejnému objektu jako výraz *které*. V tomto případě se jedná o koreferenci gramatickou mezi dvěma participanty vyjádřenými

v povrchové struktuře věty. Gramatická koreference je v PDT chápána jako takový typ vztahu, ve kterém je možné určit antecedent na základě gramatických pravidel příslušného jazyka (viz Nedoluzhko 2011).

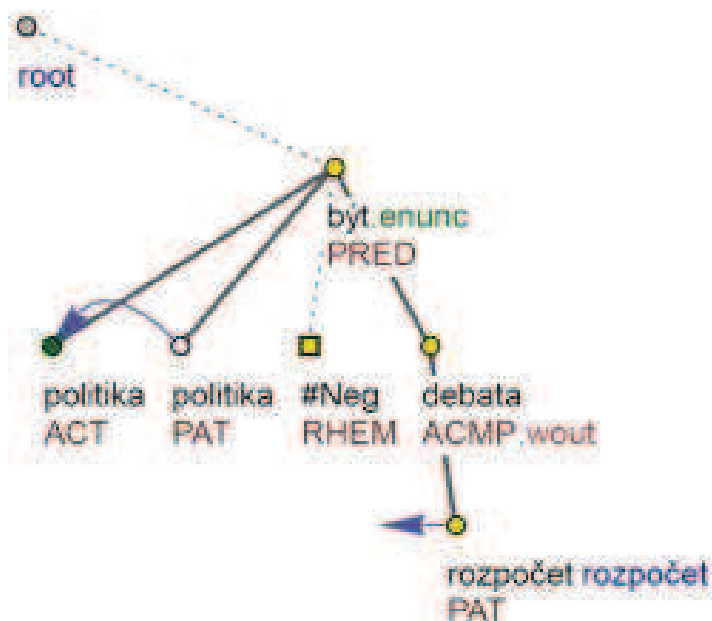
Vztahu gramatické koreference se často účastní vztažné výrazy (příklad (3)), reflexní zájmena (příklad (6)) či slovesa kontroly (tj. slovesa typu *chtít*, *dovolit*, *dokázat*, *chystat se*, *pokusit se* – viz příklady (5) a (7)) – srov. Nedoluzhko (2011).

4.2 Anotace textové koreference v PDT

Textová koreference je chápána jako vztah mezi různými jazykovými prostředky, které označují stejný objekt mimojazykové reality. Podstatou textové koreference je přitom identita obou referentů, přičemž antecedent nelze snadno odvodit pomocí gramatických pravidel daného jazyka (Nedoluzhko 2011) – srov. příklad (4) a obrázek (3) s textovým koreferenčním vztahem mezi vyjádřenými participanty *politika* a *politikou*.

(4) **Politika** by nebyla **politikou** bez debaty o rozpočtu. PDT 3.0

Obrázek 3 – Ukázka textové koreference v PDT.



Vztah koreference je zároveň symetrický (tj. pokud je první výraz koreferenční s druhým, pak druhý je koreferenční s prvním) a tranzitivní (tj. pokud je první výraz koreferenční s druhým a druhý je koreferenční s třetím, pak první je koreferenční s třetím), více viz Nedoluzhko (2011, 2013).

5 Metody a postup práce

Předmětem našeho výzkumu je zjistit vztah mezi implicitností a koreferencí. Koreferenčními vztahy jsou zcela jistě do značné míry propojeny části textu explicitně vyjádřené v povrchové realizaci (tzv. negenerované uzly). Je ale zajímavé ptát se, jak se do koreferenčních řetězců zapojují členy nevyjádřené, na povrchu elidované (tzv. generované uzly).

V rámci našeho výzkumu jsme na datech PDT 3.0 zjišťovali, mezi kterými typy uzlů (tj. generovanými či negenerovanými) se typicky objevuje vztah gramatické a textové koreference a s jakou frekvencí.

Vyhledali jsme proto všechny výskyty gramatické a textové koreference v PDT 3.0 (mezivětné i vnitrovětné), ve kterých vedl daný vztah mezi dvěma negenerovanými participanty, dvěma generovanými participanty a mezi dvěma participanty, z nichž alespoň jeden je generovaný, tj. není přítomný v povrchové, ale pouze v hloubkové rovině věty.

Výzkum byl proveden pomocí nástroje PML-TQ (Prague Markup Language Tree Query, viz Pajas – Štěpánek 2010). K přehlednému zobrazení textových souborů PDT 3.0 byl použit editor stromů TrEd (Tree Editor, viz Pajas – Štěpánek 2008).

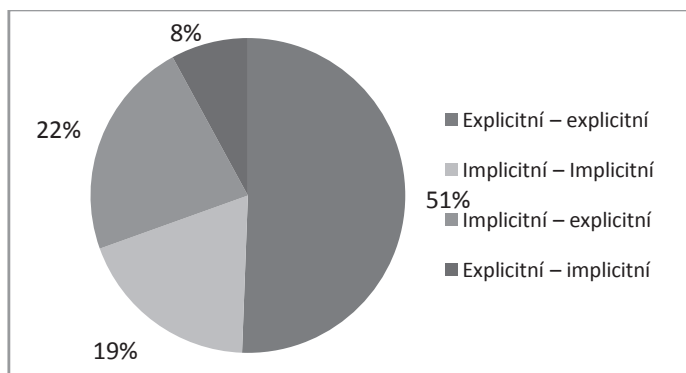
6 Výsledky a diskuse

Pro každý typ koreference jsme zjistili, kolik vztahů najdeme mezi jednotlivými výrazy, a to jak vyjádřenými explicitně, tak implicitně. Výsledky pro gramatickou koreferenci zobrazuje tabulka (1) a graf (1).

Tabulka 1 – Gramatická koreference a implicitnost v PDT.

Typ výrazu	Výskyty v PDT
Explicitní – explicitní	10519
Implicitní – implicitní	3935
Implicitní – explicitní	4688
Explicitní – implicitní	1650
Celkem vztahů	20792

Graf 1 – Gramatická koreference a implicitnost v PDT.



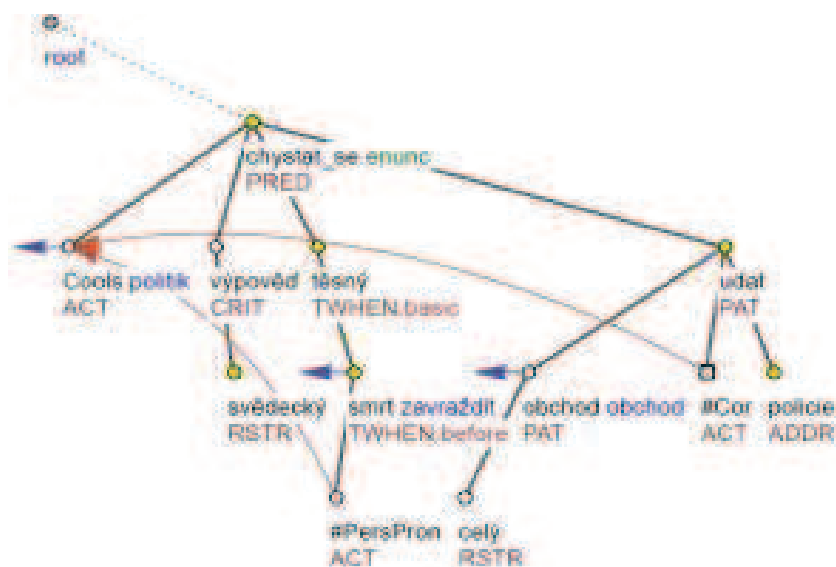
Tabulka (1) ukazuje, že v PDT najdeme celkem 20 792 vztahů gramatické koreference. Téměř přesně polovina z těchto vztahů (10 519 výskytů) propojuje výrazy, které jsou explicitně vyjádřeny v povrchové rovině. Ve druhé polovině (10 273 výskytů) se na gramatických koreferenčních vztazích podílejí určitou měrou výrazy, které jsou v povrchové realizaci textu vypuštěny. V nemalé míře nacházíme v textech koreferenční řetězce propojující dokonce dva implicitní výrazy – tyto případy tvoří přibližně pětinu všech vztahů gramatické koreference v PDT.

Příklad (5) zobrazuje případ, kdy vztah gramatické koreference vychází od výrazu implicitního a míří k výrazu explicitnímu.

- (5) Cools se podle svědeckých výpovědí chystal těsně před svou smrtí celý obchod udat policii.
PDT 3.0

Implicitním (elidovaným) výrazem je zde aktor slovesa *udat*, který koreferuje s explicitním aktorem *Cools* tzv. slovesa kontroly *chystal se* (viz obrázek (4)).

Obrázek 4 – Implicitní a explicitní participanty ve vztahu gramatické koreference v PDT.



Příklad (6) ukazuje případ opačný. Vztah gramatické koreference tentokrát vychází od výrazu explicitně vyjádřeného a směřuje k výrazu nevyjádřenému.

(6) Nikomu se nesvěří, ani své ženě Sáře. PDT 3.0

Explicitně vyjádřeným výrazem je v tomto případě zájmeno *svůj* – koreferuje s nevyjádřeným aktorem slovesa *svěřit se*.

Příklad (7) pak ukazuje dva implicitní (elidované) výrazy, mezi nimiž je vztah gramatické koreference.

(7) Mimo jiné se podle obžaloby pokusila vylákat na řádu premonstrátů v Praze na Strahově budovy v hodnotě téměř 43 miliónů korun. PDT 3.0

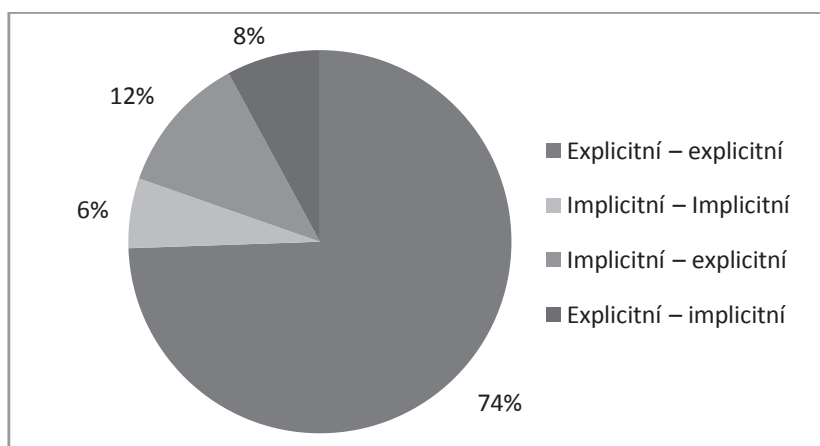
Vztahem gramatické koreference jsou v tomto případě propojeny dva nevyjádřené aktory – aktor tzv. slovesa kontroly *pokusit se* a aktor slovesa *vylákat*.

Výsledky výzkumu pro textovou koreferenci najdeme v tabulce (2) a grafu (2).

Tabulka 2 – Textová koreference a implicitnost v PDT.

Typ výrazu	Výskyty v PDT
Explicitní – explicitní	63079
Implicitní – implicitní	4997
Implicitní – explicitní	9942
Explicitní – implicitní	6702
Celkem vztahů	84720

Graf 2. Textová koreference a implicitnost v PDT.



Z porovnání tabulek (1) a (2) je patrné, že textová koreference se v PDT vyskytuje přibližně čtyřikrát častěji než koreference gramatická. Textová koreference přitom (stejně jako gramatická) vzájemně propojuje hlavně výrazy vyjádřené explicitně, tj. neelidované (63 079 výskytů). Vstupuje ale také do

vztahů s výrazy implicitními (21 641 výskytů) – tyto případy tvoří přibližně čtvrtinu všech výskytů textové koreference v PDT. V menší míře najdeme v PDT také případy, kdy jsou ve vztahu textové koreference dva výrazy implicitní, elidované (4 997 výskytů) – jedná se přibližně o 6 % výskytů textové koreference v PDT.

Příklad (8) zobrazuje případ, kdy vztah textové koreference vychází od výrazu implicitního a míří k výrazu explicitnímu.

- (8) Podporují ho lidé mladší a vzdělanější, málo příznivců má mezi důchodci a lidmi se základním vzděláním. PDT 3.0

Vztah textové koreference zde vychází od nevyjádřeného aktoru slovesa *mít*.² Tento aktor koreferuje s předcházejícím vyjádřeným patientem *on* slovesa *podporovat*.

V příkladu (9) vztah textové koreference tentokrát vychází od výrazu explicitně vyjádřeného a směřuje k výrazu nevyjádřenému.

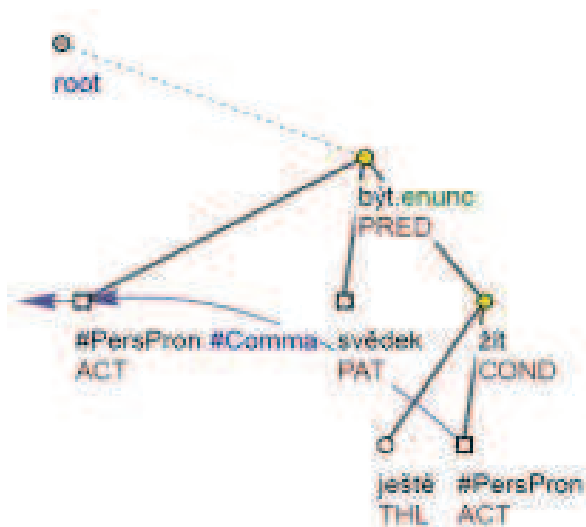
- (9) „Trváme na skutečnosti, že právo na práci je základním lidským právem garantovaným státem,“ konstatuje ČSSD v návrhu volebního programu. PDT 3.0

Explicitně vyjádřeným výrazem je zde aktor *ČSSD* slovesa *konstatovat*. Koreferuje přitom s nevyjádřeným aktorem slovesa *trvat*.

Příklad (11) ukazuje dva implicitní (elidované) výrazy propojené vztahem textové koreference.

- (11) Mohl by být – pokud by ještě žil. PDT 3.0

Obrázek 5 – Dva elidované participanty propojené vztahem textové koreference v PDT.



² K vypouštění podmětu z povrchové struktury věty viz Hajičová a kol. (2015).

V tomto případě spolu koreferují dva nevyjádřené aktory – sloves *žít* a *být* (viz obrázek (5)).

Z provedeného průzkumu je patrné, že do koreferenčních vztahů vstupují elidované výrazy významnou měrou. Přibližně polovinu vztahů gramatické koreference tvoří vztahy s generovanými uzly, u textové koreference jde přibližně o čtvrtinu vztahů.

6.1 Do jaké míry potřebuje čtenář v jazyce implicitnost?

Teorie relevance nás upozorňuje na to, že implicitnost a explicitnost musí být v jazyce vyváženy, jinak je ztíženo, nebo i ohroženo porozumění textu. Nikde se ale neuvádí, jakou míru implicitnosti adresát potřebuje pro to, aby textu bez obtíží rozuměl. Důvodem je zřejmě fakt, že tuto míru potřebné („běžné“) implicitnosti je velmi obtížné objektivně změřit. Implicitnost se v jazyce objevuje na různých rovinách a je nesnadné ji vůbec definovat, resp. určit vše, co se na významu daného textu implicitně podílí.

Jak jsme ukázali výše, PDT využívá hloubkovou rovinu, na níž rekonstruuje gramatické a aktuální (textové) elipsy. U sloves na této rovině proto například doplňuje obligatorní členy jejich valenčních rámců. Je jasné, že například u slovesa *doporučovat* čtenář ví, že obvykle *někdo doporučuje někomu něco*, ať jsou tyto výrazy v textu vyjádřené nebo nevyjádřené. Ve významu textu jsou tedy přítomné vždy. Zároveň je ovšem jasné, že doporučování probíhá někdy, někde, za nějakých podmínek, za nějakým účelem apod. Tato další doplnění už PDT na hloubkové rovině nerekonstruuje. Je tedy třeba zdůraznit, že pokud budeme chtít určovat míru implicitnosti jakéhokoli textu, vždy bude záležet na tom, jak implicitnost chápeme a kde stanovíme její hranice.

PDT nabízí jednu z možností, jak k implicitnosti přistupovat. Díky tomu, že obsahuje anotaci elipsy, nabízí jedinečnou možnost poměr implicitnosti a explicitnosti v koherentním textu vyjádřit konkrétně. V PDT najdeme 527 516 uzlů (výrazů), které jsou v povrchové rovině explicitně vyjádřeny (v atributu *is_generated* nemají anotovanou hodnotu 1) a vedle nich 72 355 uzlů, které na povrchové rovině vyjádřeny nejsou, jsou přítomné pouze v hloubkové (významové) struktuře (v atributu *is_generated* mají anotovanou hodnotu 1). Z toho vyplývá, že koherentní text (měřeno na novinových článcích)³ v průměru obsahuje přibližně 86 % informací vyjádřených explicitně a zároveň přibližně 14 % informací, které jsou přítomné pouze implicitně, tj. lze je vyčíst jen „mezi řádky“. Vzhledem k tomu, že k měření bylo použito velké množství koherentních textů (tedy textů, kterým běžný čtenář obvykle bez obtíží rozumí, je zvyklý je denně

³ Téměř 80 % textů v PDT anotovaných na hloubkové (tektogramatické) rovině tvoří texty z Lidových novin, necelých 15 % texty z Mladé fronty a zbytek textů je z Česko-moravského profitu.

číst a čerpat z nich informace), lze se domnívat, že výrazné překročení nalezených hranic poměru implicitnosti a explicitnosti jedním nebo druhým směrem (u daného typu textu), by mohlo vést k narušení čtenářského komfortu či dokonce k neporozumění celému textu.

7 Závěr

V článku jsme představili možnost využití vícevrstvé anotace korpusu PDT pro výzkum vzájemných interakcí různých jazykových jevů. Zaměřili jsme se přitom na vztah mezi koreferencí a elipsou (tj. elidovanými větnými participanty přítomnými v povrchové struktuře věty pouze implicitně).

Ukázalo se, že se většina koreferenčních vztahů vyskytuje mezi větnými participanty, které jsou v povrchové struktuře vyjádřeny. Nicméně podíl těchto vztahů mezi participanty, z nichž je alespoň jeden generovaný, není v datech PDT zanedbatelný. Přibližně polovina výskytů gramatické koreference a téměř čtvrtina výskytů textové koreference obsahuje implicitní, tj. generované participanty. Tato skutečnost zároveň potvrzuje důležitost anotace hloubkové roviny vět v závislostních korpusech. Kdyby byly anotace v PDT prováděny pouze na rovině povrchové, téměř polovina existujících vztahů gramatické koreference a čtvrtina vztahů textové koreference by nebyla zachycena.

V článku jsme se dále pokusili procentuálně vyjádřit, jaká míra implicitnosti je obvyklá v běžných novinových článcích – přibližně 86 % informací je v PDT vyjádřeno explicitně, zatímco přibližně 14 % informací implicitně (ve smyslu elidovaných větných participantů).

Naše výsledky potvrzují, že se implicitní větné participanty významně zapojují do vztahů textové koherence a celkově se tak značnou měrou podílejí na výstavbě textu.

LITERATURA

BEJČEK, E. a kol. (2013): *Prague Dependency Treebank 3.0*. Data/software. Univerzita Karlova v Praze, MFF, ÚFAL, <http://ufal.mff.cuni.cz/pdt3.0/>.

GRICE, P. (1975): *Logic and Conversation. Syntax and Semantics*, Vol. 3, *Speech Acts*. New York: Academic Press, 41–58.

HAIJČ, J. – HAIJČOVÁ, E. – MIKULOVÁ, M. – MÍROVSKÝ, J. – PANEVOVÁ, J. – ZEMAN, D. (2015): Deletions and node reconstructions in a dependency-based multilevel annotation scheme. In *Lecture Notes in Computer Science*, Vol. 9041, *16th International Conference on Computational Linguistics and Intelligent Text Processing*, Berlin / Heidelberg: Springer, 17–31.

HAIČOVÁ, E. – MIKULOVÁ, M. – J. – PANEVOVÁ, J. (2015): Reconstructions of Deletions in a Dependency-based Description of Czech: Selected Issues. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*. Uppsala, Sweden, 131–140.

MIKULOVÁ, M. a kol. (2005): *Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka*. Technická zpráva č. 2005/TR-2005-28, Univerzita Karlova v Praze, MFF, ÚFAL, 1 185 s.

NEDOLUZHKO, A. (2011): *Extended nominal coreference and bridging anaphora (An approach to annotation of Czech data in the Prague Dependency Treebank)*. Univerzita Karlova v Praze, MFF, ÚFAL.

NEDOLUZHKO, A. (2013): Generic noun phrases and annotation of coreference and bridging relations in the Prague Dependency Treebank. In *51st Annual Meeting of the Association for Computational Linguistics Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*. Sofija, Bulgaria, 103–111.

PAJAS, P. – ŠTĚPÁNEK, J. (2008): Recent advances in a feature-rich framework for treebank annotation. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*. Manchester, 673–680.

ŠTĚPÁNEK, J. – PAJAS, P. (2010): Querying Diverse Treebanks in a Uniform Way. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*. European Language Resources Association, Valletta, Malta, 1828–1835.

SGALL, P. (1964): Generativní systémy v lingvistice. *Slovo a slovesnost*, 25(4), 274–282.

SPERBER, D. – WILSON, D. (1986): *Relevance*. Harvard University Press, Cambridge, Massachusetts.

GRANTOVÁ PODPORA

Článek vznikl v rámci projektu *Automatické vyhodnocování koherence textu v češtině* č. DG16P02B016 Ministerstva kultury České republiky (program NAKI II).

Práce využívá jazykové zdroje a nástroje distribuované projektem LINDAT/CLARIN Ministerstva školství, mládeže a tělovýchovy České republiky (projekt LM2015071).