10. 5. Korpus jako zdroj dat pro výzkum syntaxe
Slovnědruhový přechod – mění se syntaktická funkce slovního druhu

6.1 Ještě několik slov ke značkování neohebných slovních druhů s ohledem na slovnědruhové transpozice

6.1.A Motivační úvod

Jednou z oblastí slovnědruhových přechodů je tzv. prepozicionalizace, při níž se z adverbia (mnohdy vzniklého ustrnutím pádu jména) stane nevlastní předložka. Mějme určit slovnědruhovou platnost tvaru *místo* v následující větě:

<Místo> *vašich schůzek sis měl lépe vybírat.*

Tato věta není z korpusů, ale demonstrujeme na ní řídký případ, kdy slovnědruhová homonymie na úrovni jednotlivého tvaru způsobí dvojí možnou interpretaci na úrovni celé věty. Z korpusu je následující doklad: *První muž se ho pokusil obelstít na trajektu Star, což bylo* ***<místo>*** *dohodnuté schůzky.* V tomto případě vyloučíme asi možnost, že jde o předložku. K vyloučení interpretace je třeba analyzovat celé souvětí vč. koreference.

6.1.B Nastínění problému

Homonymie může nastat tam, kde je na základě kontextu možné obojí čtení. Kontext je dvojí, jazykový a mimojazykový. Z hlediska jazykového kontextu je pro interpretaci tvaru jako předložky třeba splnit podmínku, že za tímto tvarem následuje tvar jména / tvary jmen (jmenná skupina) v tom pádě, se kterým se předložka pojí. To je podmínka nutná, nikoli postačující.

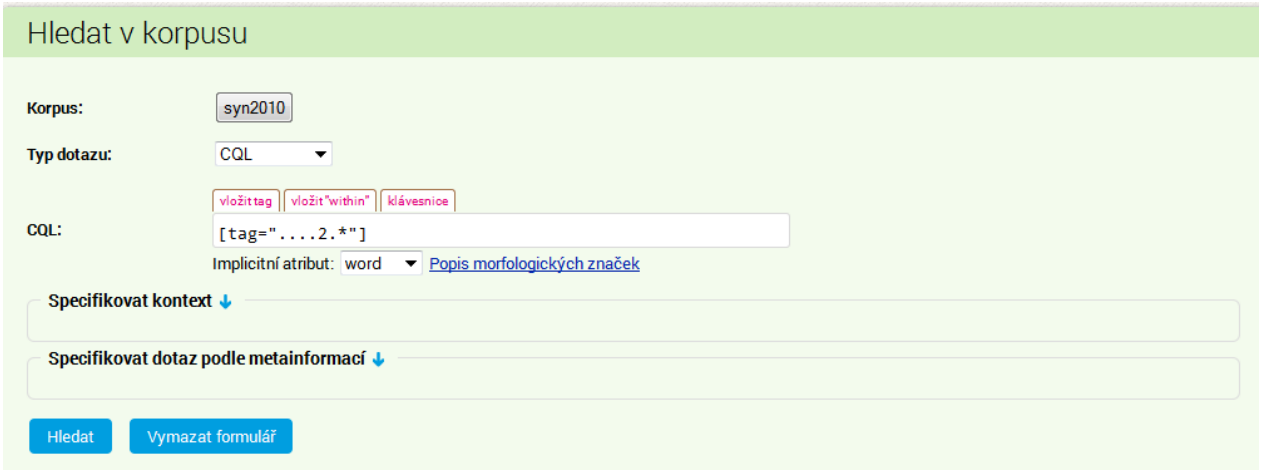
6.1.C Otázky

Podívejme se na to, jak jsou v korpusech značkovány nepůvodní předložky (seznam najděte např. v mluvnicích). Pokusme se najít případy, kdy tvarová homonymie na úrovni slovního druhu je důvodem chybné desambiguace.

6.1.D Formulace dotazu pro získání dat z korpusů

Zvolíme **Typ dotazu slovní tvar** a do dotazovacího řádku zapíšeme *kolem*. Zvolíme filtr (pozitivní), rozsah hledání <–1,–1>, **Typ dotazu cql** a do dotazovacího řádku zapíšeme *[tag="[APC].N.7.*"]*. Chceme tak získat řádky, na nichž se může vyskytnout substantivum *kolo* rozvité shodným přívlastkem (adjektivem, zájmenem nebo číslovkou v 7. pádě středního rodu).

Poté ponecháme filtr (pozitivní), zadáme rozsah hledání <1,1>, **Typ dotazu cql** a do dotazovacího řádku zapíšeme *[tag="....2.*"]*. Chceme získat pouze ty případy, kdy za tvarem *kolem* následuje jméno ve 2. pádě (předložka *kolem* se pojí se 2. pádem).



Týž postup zopakujeme pro vyhledání tvaru *místo,* přičemž omezíme výběr konkordančních řádků na ty případy, kdy bezprostředně vpravo (<1,1>) za tímto tvarem stojí tvar označkovaný jako 2. pád (*....2.**) a bezprostředně vlevo (<–1,–1>) před tímto tvarem stojí tvar označkovaný jako adjektivum, zájmeno nebo číslovka středního rodu a jednotného čísla v 1. nebo 4. pádě (*[APC].NS[14].**).

6.1.E Třídění a pozorování dat získaných z korpusů

Nyní si pozorně prohlédneme konkordance a zjistíme, že v desambiguaci se vyskytují chyby.

čekat nedokážu . " Stál tam s bílým ručníkem přehozeným kolem /kolo/NNNS7-----A----- krku a rukama svíral jeho konce . Pustila košili zpátky

případných smrtících látek . " Přenesla pohled ke knihám rozloženým kolem /kolo/NNNS7-----A----- sebe . " Zdá se , že jsi měla napilno

květinou , kde nebyla růží . S copem třikrát pleteným kolem /kolo/NNNS7-----A----- čela , hořekala veselím a radostněla zpěvem . Říkávala :

Chybnou desambiguaci pozorujeme i v případě konkordancí tvaru *místo*.

zná název Khao Lak . V únoru ves jako nejpostiženější místo /místo/RR--2---------- Thajska navštívili američtí prezidenti George Bush starší a Bill Clinton

6.1.F Formulace závěrů

Je patrné, že slovnědruhové přechody, které mají za následek vícero interpretací na úrovni lemmatizace, slovního druhu i dalších slovnědruhově závislých kategorií, značně ovlivňují obtíže a chyby na úrovni desambiguace. S výsledky automatické analýzy je tudíž třeba pracovat opatrně s vědomím toho, že je třeba údaje vždy zkontrolovat.

6.1.G Formulace dalších otázek vyplynuvších ze zkoumání daného jevu

Ve výběru konkordancí, na nichž jsme zkoumali úspěšnost desambiguace, jsme vycházeli z toho, že ve zvolených případech by mohlo dojít k problémům při aplikaci desambiguačních pravidel založených na lingvistických předpokladech. Nezkoumali jsme úspěšnost všech konkordančních řádků. Vybrali jsme pouze ty, pro které platilo, že tvar lze na základě lingvistické analýzy bezprostředního kontextu interpretovat obojím způsobem. Pravidlová desambiguace byla použita pro odstranění některých chyb desambiguace stochastické. Zajímavé by bylo porovnat chybovost v jednotlivých korpusech řady (více Jelínek 2008, Skoumalová 2011).

6.1.H Zadání cvičení, v nichž lze uplatnit analogické postupy

Sledujte, jak je provedena desambiguace u tvarů *během, bokem, úderem, stranou, ...* Postupujte podobným způsobem, jak bylo naznačeno.

17. 5. Dú: Sleduj v korpusu slovnědruhové značkování tvaru *díky*, který lze interpretovat vícero způsoby a pokus se navrhnout, jak postupovat při odhalení chyb v disambiguaci.