

Posudek na disertační práci Mgr. L. Polákové *Discourse relations in Czech*

(předloženo na MF UK v Praze, r. 2015, 185 stran)

Práce L. Polákové se zabývá analýzou tzv. diskurzních vztahů, které chápe jako jeden z aspektů textové koherence. Je založena na anotačním schématu Pražského závislostního korpusu: syntaktická anotace textů korpusu byla rozšířena o anotaci na úrovni diskurzu. Výstupem práce je nejen návrh anotačních zásad diskurzních vztahů, ale i jazykovědná reflexe této problematiky. Práce tak jasně ukazuje jednu z podstatných přidaných hodnot anotačního procesu - důsledná aplikace teorie na konkrétní data vede k nutné revizi vstupních teoretických principů, které tak v průběhu anotačního procesu doznají změn. Anotace tak není jen mechanickou aplikací předem daných postupů, ale stává se impulsem pro tvůrčí zásahy do jazykovědné teorie. Je ovšem třeba mít na paměti existenci určitých limitujících faktorů: vývoj či výběr teorie, na níž je anotace založena, bývá nutně omezen svým cílem, a tím je právě anotace a její technické parametry.

V úvodní části práce autorka představuje klíčové pojmy a metody, z nichž její výzkum vychází. Orientuje se přitom na ta pojetí diskurzu, která doznala aplikací v korpusových anotacích. Vzhledem k cílům a charakteru předkládané dizertační práce je tato volba pochopitelná a legitimní. Diskurz a vztahy, které se v jeho rámci konstituují, chápe autorka v intencích pražského funkčně-generativního přístupu. Diskurz tak není další rovinou systému (v rámci hierarchických vztahů forma-funkce), ale je pojímán jako samostatná entita mimo systém. Rozhodnutí anotovat diskurzní vztahy na úrovni tektogramatické je pak dáno technickými důvody. Autorka v rámci této kapitoly zmiňuje termíny jako koheze a koherence textu a stručně pojednává o pojmech dalších, které však vzhledem k zaměření své práce podrobně nerozporuje.

Klíčovým rozhodnutím, které anotátoři PDiT museli učinit, byl výběr základního přístupu k analýze diskurzních vztahů. L. Poláková ve své práci konfrontuje dva takové přístupy uplatňované ve světové lingvistice: přístup globální reprezentovaný tzv. Rhetorical Structure Theory a přístup lokální, uplatňovaný v anotacích Penn Discourse Treebank. Globální přístup odvozuje diskurzní vztahy z rétorické struktury textu jakožto celku, jehož části spoluvytvářejí globální diskurzní efekt, přístup lokální pak vychází z konektivních lexikálních signálů a na základě jejich funkcí definuje lokální vztahy mezi spojovanými entitami. Volba anotátorů PDiT směřuje ke druhé variantě, tedy lokálně-lexikální, které je výhodná i z toho důvodu, že může využívat informace obsažené na tektogramatické úrovni.

V závěrečných podkapitolách úvodní části práce pak L. Poláková nastoluje problematická témata, k nimž se její výzkum dále vyjádří. Jedná se např. o rozdíl mezi vztahy označovanými jako sémantické a vztahy označovanými jako pragmatické. Zdůrazňuje vágnost tohoto rozlišování, které je podle některých lingvistů jednou z brzd vývoje jazykovědy. Dle mého názoru je velmi sporné, zda faktory jako explicitnost versus implicitnost (a tedy nutnost inference) či „objektivní“ ontologická reference versus subjektivita opravňují postulovat kvalitativně odlišné domény. Dalším z problematických témat je otázka vztahu mezi syntaktickými strukturami a diskurzními entitami (vztah je jak známo asymetrický), či otázka počtu odlišných diskurzních vztahů, které pak anotace na konkrétní texty uplatňuje.

Třetí a čtvrtá kapitola práce je věnována vlastnímu procesu anotace, jeho technickým parametrům, ale i problematickým aspektům. Jedná se o ucelenou a originální reflexi zkušeností z anotačního procesu. Postupně zmíním podněty k diskuzi, které se týkají jednotlivých částí této velmi kvalitní statě. Níže uvedené podněty mají spíše dílčí charakter a týkají se konkrétních tvrzení:

1) str. 52 - 53: uvedené slovnědruhové dělení konektorů jistě není bezproblémové. I když to není vlastním tématem práce, zajímalo by mě, na základě jakých kritérií bylo rozlišováno mezi kategoriemi a), b), c): srov. řazení výrazů jako *však, ovšem, totiž, proto, přesto, potom, dokonce...*

2) str. 54 - 55: problém identifikace diskurzních struktur ve vztahu k jejich syntaktické formě. Typicky se jedná o problém nominalizací. Podle tvrzení na str. 58 nebyla žádná nominální predikativní syntagmata anotována jako diskurzní jednotky. Bylo by možné uvést argumenty pro toto řešení? Viz např. následující text (jako variace na příklad 25):

Nestihl to kvůli pozdnímu odjezdu vlaku. Kromě toho také nejelo metro.

Jak by byl anotován levostranný diskurzní argument výrazu *kromě toho*?

3) k úvahám na str. 64 (3.6.3., 3.6.4.): jak byly anotovány tzv. nepravé vztahné věty? Např.:

Dal ránu pěstí Petrovi, který zůstal stát nezraněn.

4) str. 67: hledíme-li na text z hlediska jeho organizace (tedy nikoli z hlediska úžeji chápaných diskurzně-rétorických vztahů), i referenční časové výrazy mohou fungovat jako organizátory. Typicky například ve vyprávění či popisu procesů, kdy referenční časová adverbiala v iniciálním nerematickém postavení signalizují jednotlivé etapy (konstitutivní části narativní sekvence) a tím de facto strukturují text obdobně jako např. výčtové konektory.

5) str. 71 – 72: k tzv. Pragmatic Reason - Result. Z čistě povrchové struktury je mnohdy obtížné tento vztah odlišit od „běžného“ Reason - Result. Srov. typicky „kauzální“ věta:

Chodníky jsou mokré, protože pršelo.

může být za různých okolností interpretována různě. Je nutné znát celý výpovědní kontext. Z této perspektivy se mi jeví pouhých 40 výskytů tohoto vztahu jako relativně málo (chápe-li tento vztah v intencích parafráze uvedené na str. 72).

6) str. 74 - 75: podle autorky je rozdíl mezi Reason – Result a Explication dán absencí kauzálního vztahu. Nebylo by na tzv. Explication výhodnější nazírat jako na diskurzní funkci ve smyslu: „tvrdím A a dále připojuji B, které podporuje můj nárok tvrdit A“. Vysvětlovat se přece dá i kauzálním vztahem. Dále: byla by anotace jiná, pokud by v příkladu typu (63) bylo místo *totiž* užito výrazu *protože*?

7) str. 81, příklad (74): nabízí se interpretace přes implicitní argument: zdá se mi, že věta s *dokonce* není v gradačním vztahu z větou předchozí, ale že je spíše třeba ji považovat za podpoření (argumentaci) ve prospěch tvrzení věty první (dosazení *nejen – ale* do obou vět cítím problematicky). Výraz *dokonce* uvádí silnější argument pro tvrzení obsažené v první větě. Onen slabší argument, ke kterému *dokonce* odkazuje, ale není uveden, což může být z hlediska diskurzní strategie výhodné (viz rovněž níže, odrážka 11).

8) str. 88: považuji za velmi přínosné, že se do diskurzních vztahů zahrnují i výčtové organizátory. Bylo by možné za výčtové konektory považovat i výrazy typu: *nejprve, dále nakonec*?

9) str. 103 – 109, stat' o víceznačnosti konektorů. Důsledné řešení tohoto problému vyžaduje úvahu o vztahu mezi významem konektoru a kontextem. V praxi bývá nejčastější přístup ten, že se definuje vztah celku (tedy textu «P, Conn Q»), a poté se tvrdí, že příslušný konektor uvádí/vyjadřuje tento vztah. To je rovněž přístup, který volí L. Poláková. Tato metoda pak nutně vede k tomu, že jeden konektor figuruje jako „operátor“ více různých vztahů, mnohdy i kontradiktorních. To zcela legitimní, ovšem problém nastává v okamžiku, kdy hovoříme o polysémii konektoru samotného. V tom případě je nutno uvést jak chápeme význam konektoru (jak jej lingvisticky modelujeme) a následně ukázat, jak v rámci onoho modelu uchopit postulovanou polysémii. Viz např. text:

Pracoval, když byl ve vězení.

Možné interpretace tohoto spojení - časové, podmínkové, kauzální (ve smyslu podpory nároky tvrdit P) - jsou dány syntaktickými, prozodickými a kontextovými parametry (včetně sémantiky samotného vztahu mezi propozičními obsahy P a Q). Je tedy podle mého názoru výhodnější tvrdit, že ten či onen konektor příslušné interpretace umožňuje (na základě kontextových parametrů), než že je sám nějakým způsobem explicituje.

Přístup L. Polákové je ovšem metodologicky konzistentní: anotátor definuje diskurzni vztah celku a následně přidělí konektoru danou hodnotu. Statistická data pak uvádějí, v kolika různých vztazích konektor figuroval, aniž by ovšem bylo zmíněno, jak se tento vztah vytváří a jaká je role jednotlivých částí (např. sémantický model, kognitivní model apod.). Prosim pro potřeby obhajoby o uvedení konkrétního textu, kde konektor *ale* artikuluje vztah Reason (tabulka na str. 181).

10) stat' 4.2. – obsahuje velmi podnětné úvahy o vztahu mezi syntaxí a diskurzem. Dovoluje ocenit výhodu anotace PDiT na syntakticky anotované texty. Stat' otvírá celou řadu velmi zajímavých problémů, které současná diskurzni lingvistika řeší, zejména otázku atribuce garancí tvrzení, ať už pomocí verba dicendi či sentiendi nebo pomocí předložkových syntagmat. Vzhledem k tomu, že diskurzni konektory jsou jedny z nejvýraznějších stop argumentace v textu, nutně se účastní i těchto složitých vztahů (polyfonních).

11) stat' 4.3. – zkušenost anotátorů ukazuje na problematičnost exaktních interpretací implicitních vztahů (tedy vztahů nevyjádřených konektorem). Zde lze očekávat, že lokální lexikalistický přístup PDTB bude narážet na větší omezení než přístup globální typu RST, který diskurzni vztahy definuje z hlediska celkové textové sémantiky (pragmatiky) a nikoli primárně na základě lokálních lexikálních signálů. Uchopení implicitnosti je tak další výzvou v anotačním procesu PDiT. V rámci této otázky bych rád upozornil ještě na jeden rozměr implicitnosti, naznačený již pod odrážkou 7). Některé konektory totiž mohou operovat s implicitním argumentem, typicky např. konektory gradační, které nutně konektují koordinované argumenty. Srov. příklad s *dokonce*:

Na dovolenou pojedeme do Itálie. Dokonce tam letos zlevnili zmrzlinu.

Levostranným argumentem konektoru *dokonce* není přímo předcházející věta, ale implicitní argument typu *Je to blízko*. Obě entity konektované výrazem *dokonce* pak z hlediska globálních diskurzni vztahů fungují jako argumenty na podporu tvrzení v první větě. Počítá anotační proces PDiT i s těmito případy?

Závěr:

Práce L. Polákové je velmi kvalitní, a to jak po stránce obsahové, tak po stránce formální. Autorka bezesbytku plní úkol, který si v úvodu stanovuje. Práce je založena na exaktním a originálním výzkumu, autorka prokazuje rozhled jak v problematice teoreticko-lingvistické, tak v otázkách technických. Považuji ji za významný příspěvek do problematiky počítačové anotace diskurzni vztahů.

Práce prokazuje předpoklady autorky k samostatné tvořivé práci. Bez výhrad ji doporučuji k obhajobě před komisí.

V Českých Budějovicích, 13. 8. 2015



doc. PhDr. Ondřej Pešek, Ph.D.