

## Kontextové vymezení kolokací: o slovech monogamních a promiskuitních

čtvrtek 4. 4. 2013, 17.30–18.45, místnost č. 104

Václav Cvrček

Mnohokrát citovaná pravda, že slovo poznáš podle toho, na jakou společnost si potrpí (Palmer 1968: 179), se ve světle korpusové lingvistiky stává poprvé v historii testovatelnou hypotézou. V situaci, kdy k mnoha slovům (a nejen slovům, ale jazykovým jevům obecně) máme v korpusech k dispozici stovky dokladů, můžeme z pohodlí svého pracovního stolu sledovat, jak se užívají, do jakých kontextů a jak často vstupují. Problém nastává v situaci, kdy máme rozhodnout, jak susedství dvou jevů kvantitativně hodnotit. Jak poznáme, že určitý typ souvřskytu je relevantní pro lingvistickou analýzu? Jak rozsáhlý kontext musíme vyhodnotit, abychom získali smysluplný a stále ještě přehledný obraz o zkoumaném jevu? Co všechno z kontextu vyčíst lze a co naopak ne?

Učinit kontext výchozím bodem lingvistické analýzy jevů se ve světle těchto otázek snadněji prohlašuje, než udělat. Obecné kvantitativní charakteristiky kontextu totiž nejsou dosud příliš prozkoumány a porovnávat tak nalezená zjištění není moc s čím. Pokud zároveň přistoupíme na to, že lingvisticky popsat jev znamená uchopit ho po syntagmatické i paradigmatické stránce, dostáváme se do paradoxní situace: na jedné straně tvrdíme, že (téměř) vše je obsaženo v kontextu, zároveň ovšem to jediné, co máme v korpusech k dispozici, je množství textů – lineárních řetězců jednotek (tedy informací o jejich syntagmatické). Cílem kontextové analýzy bude tedy vytěžit z dat, jejichž povaha je povrchově pouze parolové syntagmatická, informace o systému a to jak syntagmatické, tak paradigmatické povahy.

Kvantitativní poznání obecných zákonitostí kontextu, o které nám zde jde, je na základě výše uvedených úvah možné rozdělit na tři vzájemně propojené a související otázky. Za prvé jde o problém toho, jak vypadá úhrnný kontext jednotky, tj. s kolika slovy se spojuje a na kterých pozicích. Tento pohled na kontext by měla prezentovat veličina, kterou můžeme nazvat variabilita. Jejím zkoumáním zjistíme, jaké jsou průměrné trendy v obsazování kontextových pozic a např. to, jestli můžeme slovo s celkovou frekvencí 100 výskytů a s 50 různými doplněními na první právě pozici považovat za jednotku s malou, velkou nebo extrémní variabilitou kontextu.

Druhý okruh problémů se týká toho, jaká je průměrná vzdálenost mezi dvěma jednotkami – proximita. Tento ukazatel může za určitých okolností vypovídat o tom, do jakých syntagmat sledovaný jev typicky vstupuje. Ukazuje se totiž, že to, co je pravidelně blízko sebe „na papíře“, je v těsné souvislosti i v jazykovém systému. Případy slov, jejichž kolokabilita je zásadně omezená (např. monokolokabilní slova, kolokace, reflexiva tantum, víceslovné termíny apod.), ukazují, že obligatorní doplnění ve většině případů nestojí daleko (a to navzdory volnému slovosledu v češtině).

Poslední veličina, která umožňuje popsat kontext, je obligatornost, tj. míra, s jakou jsou jednotky na sebe vázané, jak jsou monogamní a nebo naopak promiskuitní. Tento aspekt byl zkoumán v korpusové lingvistice nejvíc, a to právě v souvislosti s asociačními mírami, které se používají k identifikaci kolokací (viz např. Pecina 2010, Evert 2004).

Kombinací těchto veličin je možné charakterizovat kontext, a to jak pro účely popisu syntagmatického (výzkum kolokací, valence apod.) tak pro účely výzkumu paradigmatického (výzkum synonymie, slovtvorných paradigmat, slovních druhů apod., srov. Cvrček 2011).

Po úvodní části věnující se vymezení základních pojmů kvantitativního popisu kontextu a příkladům jejich možného využití se přednáška zaměří na výzkum kolokací právě prizmatem kontextové analýzy.

### Literatura:

Cvrček, V. (2011): Kontextová analýza a systém slovních druhů. *Korpus – gramatika – axiologie* 3/2011, s. 6-32.

Evert, S. (2004): *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. PhD dissertation. IMS. University of Stuttgart. Published in 2005. (dostupné z <http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/>).

Palmer, F. R. (ed.) (1968): *Selected Papers of J. R. Firth, 1952–1959*. Bloomington/London.

Pecina, P. (2010): Lexical Association Measures and Collocation Extraction. *Language Resources and Evaluation* 44/1-2, s. 137–158.