ČASOPIS PRO MODERNÍ FILOLOGII 97, 2015, Č. 1, S. 51–57

Kolokace a kolokabilita synsémantik: substituty a interjekce

František Čermák (Praha)

COLLOCATIONS AND COLLOCABILITY OF CZECH GRAMMATICAL WORDS: SUBSTITUTES AND INTERJECTIONS

While most attention is generally paid to lexical (autosemantic) words, the goal of this paper is to map grammatical (or synsemantic) words which have been largely neglected. The specific aim is to look into combinations of all types of grammatical words whose common feature, though they are widely different in most other respects, is their auxiliary role with regard to lexical words. This auxiliary role is to be viewed as having two functional interpretations: grammatical words may be seen either as linkers (serving as links between lexical words) or as substitutes or pro-forms (replacing lexical words in text to avoid their repetition in general). Since until now most authors have paid attention to various word classes in this area (see Bibliography), this paper focuses on the possibilities of their roles as substitutes and the limitations of their combinations with one another, a subject hardly ever mentioned, let alone researched. The investigation, involving pronouns and numerals, is based on a large corpus of Czech (SYN2010). Because of their special character and function, interjections are loosely added to the bulk of the grammatical words. It is shown that, contrary to simplified expectations, most grammatical words have a life of their own, i.e. they combine with each other and form collocations independently of lexical words.

KEYWORDS

grammatical words, combinations, collocations, pronouns, numerals, interjections

KLÍČOVÁ SLOVA

synsémantika, kombinace, kolokace, pronomina, numerale, interjekce

1. ÚVOD A PŘEHLED OBLASTI

Mluví-li se o kombinatorice slovních druhů, mluví se obvykle o velkých autosémantických slovních druzích. V následujícím pokusu půjde ale o kombinatoriku jinou, o přehled zcela zanedbávané oblasti synsémantik, a v širokém smyslu o pokus zjistit, nakolik se tyto slovní druhy kombinují nezávisle na autosémantikách, tj. především mezi sebou, uvnitř své třídy; role synsémantik vázaných na autosémantika je většinou dobře známá.

Do záběru se takto dostane šest slovních druhů, funkčně zcela heterogenních, které spojuje jednak to, že nejsou autosémantiky, a jednak to, že mají vůči nim v různém smyslu většinou služebnou úlohu, i když ne vždycky. Konkrétně tedy půjde o prepozice, konjunkce, partikule, pronomina, numerale i interjekce, u nichž se až dosud o syntagmatických aspektech příliš neuvažovalo (pokud vůbec).

První část stručně vymezí široké a hrubé pojetí synsémantik, jejich typy (jednoslovné i víceslovné) a kombinatoriku a především funkční aspekty, popř. i pragmatiku. Necháme-li stranou definitoricky bokem stojící interjekce, pak lze oblast synsémantik dělit na dvě velké skupiny, na substituty (pronomina a numerale) a relátory (prepozice, konjunkce, včetně odlišných relátorů pragmatických, tj. partikulí); více o pojetí viz Čermák (2010).

Protože autor pozornost některým z nich už věnoval dříve (tj. prepozicím, konjunkcím a partikulím) a protože je tato oblast velmi široká, bude se příspěvek věnovat především substitutům, tj. pronominům a numeraliím, v závěru však i oblastem dalším, především znovu interjekcím.

Druhá část, založená na korpusovém rozboru, si postupně všimne jednotlivých takto vymezených synsémantik. Základním půdorysem k pohledu na ně bude jejich obecná syntagmatická funkce, komentovaná stručně i z hlediska jejich sémantiky a pro ně specifické funkce. Důraz se bude klást, v souladu s mým opakovaným zjištěním, na potřebu lišit vedle tradičně formálně povědomých synsémantik jednoslovných i synsémantika víceslovná, v řadě případů tedy půjde o synsémantické kolokace ustálené (často frazeologické). Jedním z cílů bude naznačit kombinatorické možnosti, ale i omezení, především vůči sobě navzájem. Tato specifická otázka nebyla dosud v zásadě systematicky studovaná.

V části věnované zájmenům a číslovkám se prezentují možnosti a hranice jejich kombinatoriky a komentují některé nápadné zvláštnosti. Obdobně se nabídne pohled na, běžně nečekanou, kombinatoriku číslovek, které nejsou vždy jen textové; najdou se tu i případy ustálené.

2. ÚVOD: SYNSÉMANTIKA

Synsémantika bývají vymezována různě, především však negativně jako ty slovní druhy, které zůstávají po vydělení autosémantik, kterým různým způsobem jako pomocná slova „slouží" a vyjadřují různé vztahy mezi nimi i jejich role. Teprve relativně nedávno se ukázalo, že tento obrázek je nedokonalý, především s důrazem na jejich (**A**) **funkce**, které umožňuje nejlépe studovat až zkoumání korpusu (patří mezi ně donedávna nezmiňované aspekty (**A**1) **pragmatické**), a tedy i jejich (**A**2) **kombinatoriku**, ale také na jejich (**B**) **kombinatoriku vlastní**, a tedy mj. víceslovnou povahu a aspekty ustálenosti. Ukazuje se však taky, máme-li se podržet naznačeného prostého dělení slovních druhů na autosémantika (S, A, V a ADV) a „neautosémantika", že sem můžeme přiřadit i (**C**) **interjekce**, které jsou funkčně zcela jiné a lze je chápat i mimo rámec slovních druhů. Není nijak zvlášť nutné zdůrazňovat, že jako celku se synsémantikům velká pozornost nikdy nevěnovala, zvláště ne jejich aspektům funkčním a kombinatorickým.

Ne právě zanedbatelným faktem přitom je dosud nepříliš reflektovaná skutečnost, že víceslovných synsémantik (necháme-li stranou číslovky) je aspoň v případě předložek a partikulí mnohem víc než jednoslovných, jde tedy pochopitelně o kombinace ustálené a součást systému, jakkoliv se nikde systematicky neevidují. Je možné dokázat, že navzájem se můžou kombinovat, byť v různé míře, všechna synsémantika a vznikají tak jak kombinace neustálené, tak ustálené.

Jakkoliv je funkce takto široce pojatých synsémantik pestrá a různá, je (skrze jejich syntagmatiku) jejich stránkou zcela zásadní. Naproti tomu jejich stránka sémantická je v tradičním pohledu omezena pouze na předložky, popř. specificky ještě číslovky; další slovní druhy vlastní význam nemají (je ale třeba lišit význam od funkce, např. u konjunkcí). To, co se namísto významu u nich nabízí, je funkce, většinou syntaktická, resp. syntagmatická (spojky), jakkoliv s tradičním výlučně syntaktickým pohledem u částic nevystačíme — dominuje tu funkce pragmatická a ta většinou není pouze textová, ale zahrnuje i různé vztahy mimojazykové. S většinou čistě sémioticky (ne však nutně vždy pragmaticky) pojatou funkcí se ale setkáváme i u citoslovcí (k tomuto pojetí i klasifikaci viz blíže Čermák 2010, 2.32).

Proti v podstatě relační syntagmatické a textové funkci výše, tj. u prepozic, konjunkcí a zčásti partikulí, nacházíme především u interjekcí funkci nominativní (např. *brr* označuje přímo pocit zimy apod., a tedy nic nespojuje ani nekvalifikuje, jakkoliv je bližší větě), a tedy paradigmatickou. Jinou stránkou této paradigmatické, nominativní funkce je však také substituce, tj. schopnost kvalitativně nebo kvantitativně zastupovat, resp. v textu nahrazovat autosémantika (většinou jen nominální povahy). Tu mají pronomina a numerale.

Formálně jsou synsémantika většinou stálá a bez flexe (srov. však časování u *aby-ch/-s/-chom/-ste* aj.), frekventované však mívají varianty. Základním nereflektovaným problémem je standardní grafická povaha u některých interjekcí, kdy není jasné, jak je vlastně zapsat (mluví-li mluvčí hodně emocionálně, kolik **é** zapíšeme optimálně do interjekce údivu *jéé*??).

3. SYNSÉMANTIKA FUNKČNĚ A KOMBINATORICKY

V dalším se pozornost soustředí především na povahu funkcí vybraných typů synsémantik a jejich hlavní typy, a to na pozadí kombinatoriky zjišťované z korpusu.

Připomeňme si, že specifická pozornost věnovaná partikulím, prepozicím, konjunkcím a také interjekcím, a to na korpusových datech, se odrazila v dřívějších studiích autora (Čermák 1996, 2007, 2008a, 2008b). Tento příspěvek tak završuje a uzavírá celistvý pohled na synsémantika jako celek, jakkoliv jen stručně, a je ho třeba chápat především jako inspiraci pro systematičtější a podrobnější popis.

3.1 SUBSTITUTY: PRONOMINA A NUMERALE

Substituty jsou synsémantická slova/lexémy, která funkčně v textu obecně zastupují autosémantika, především substantiva a adjektiva, a to proto, aby se nemusela opakovat, srov. *Dostal jsem nové* ***knížky***, a možná substituční užití zastupující jako pokračování informace v dané větě substantivum (a) prostým a (b) kvantifikovaným způsobem, tj. (a) *Už jsem* ***je*** *přečetl.* ***Všechny*** *jsou zajímavé* nebo (b) ***Tři*** *už ale znám*. V prvním případě (a) zastupuje, substituuje substantivum (*knížky*) pronomen (*je*), v druhém (b) numerale (*tři*).

Takováto základní substituce má však svá omezení, mj. nemožnost další modifikace stejným způsobem. Substituty jsou korelovány s tím, co zastupují, i subkategoriálně, např. osobní zájmena a substantiva, číslovky řadové a adjektiva aj. Je třeba si

také uvědomit, že vzhledem ke své substituční funkci se můžou, analogicky k substantivům, substituty také kombinovat mezi sebou. Budou nás tedy zajímat aspoň obecně možnosti a omezení (srov. triviální případ: *Piju čaj s mlékem. Už jsem **ho** vypil.*, a nikoliv **Už jsem **ho** s **ním** vypil.*).

3.1.1 PRONOMINA A JEJICH KOMBINACE

Z 1 128 286 dvoukombinací typu **Pron-Pron** (SYN 2010) (A) některé jsou textově náhodné, a tedy nesmyslné (*co se, ho tím*), (B) jiné ilustrují možné typy kumulace (*já sám*) a některé z nich jsou dokonce (C) ustálené jako lexémy (*ten kdo*). Na náhodném vzorku 1000 z nich se podíváme na některé rysy. Srov. *těch jejich, ten můj, ten svůj, to všechno, všechny tyhle, všechny naše*, popř. hojný typ s prepozicí (zde nesledovaný) *nikdo z nich* aj.

Typ **A** je zcela *náhodný*, tj. daný slovosledem, a je také zdaleka nejčastější (cca 98 %), např. *všechno se, ten ji, ty si* aj. Typ **B** zahrnuje kumulace, které jsou věcně, logicky možné, ve kterých druhý komponent specifikuje první, např. *on sám, co všechno, nikdo z nich, ten jejich, všechny tyhle, všechny naše, ten svůj*. Vzorek naznačuje možnost různých kombinací, mezi nimiž vyniká kombinace demonstrativ a posesiv aj., v zásadě však nikdy členů stejné třídy, pokud nejde o kumulaci sumární typu *já a/i on* (užívané obvykle se spojkou).

Typ **C** je nejmenší a jasně uzavřený (tvoří uzavřené paradigma), jakkoliv všechny kombinace nejsou stejně časté. Vedle idiomatického *ten můj* jde o paradigmatické kombinace členů několika tříd dávající teoreticky minimálně 145 různě ustálených kombinací, srov.:

1) v pozici 1. komponentu: pronomina ukazovací/neurčitá (*ten, tento, tenhle, tamten, tamhleten, takový, /ten/týž, některý, kterýkoliv, jakýkoliv, každý, všechen, žádný /leckterý/)* nebo pronomina osobní/neurčitá (*já, ty, on, my, vy, oni, někdo, nikdo, leckdo, kdokoliv, něco, nic, leccos, cokoliv, všechno*);

2) v pozici 2. komponentu: *který, co1* (bezrodé, obecné), *jenž, kdo, co2* (rodové, v neutru).

Trojkombinace jsou podobné, v korpusu je jich 70 602, ustálená není žádná. Pro představu se podívejme ještě na ***čtyřkombinace* Pron-Pron-Pron-Pron**, kterých je v korpusu 1783. Žádná z nich však není ani typu B, ani C výše, všechny jsou náhodné, a tedy významově nesouvisejí, srov. *co se něco takového, já vám to všechno (vysvětlím), Já se sám tomu (divím), já mu nic takového (dávat)*. Dodejme ještě, že pětikombinací se nalezlo jen 74 a že ani tam nelze najít smysluplné kombinace, srov. *Na něco takového se jí nikdo (v životě nezeptal), kdo si je všeho tohoto (vědom)*. Je zřejmé tedy, že zájmena kombinatoricky nepřekračují u smysluplných kombinací kombinace dvoučlenné.

Zvláštní případ ovšem představuje řídký, ale výrazný typ, kdy se zájmena kombinují v podobě frazeologických binomiálů s předložkou, srov. *pro mne za mne, pro nic za nic, z ničeho nic*.

3.1.2 NUMERALE A JEJICH KOMBINACE

Číslovky (v dalším jen vlastní, bez substantivizovaných typů jako *trojka*) jsou na první pohled užívané v kombinacích, které jim „dovoluje“ matematika. Můžeme se však ptát, jaké ty kombinace vlastně jsou a je-li veškerý jejich úzus podřízený jen matematice, zvl. algebře.

Dvoukombinace číslovek jsou překvapivě hojné (90 423), jde ale hlavně o kombinace těch základních, jakkoliv překvapivě hodně je i kombinací substantiv typu *dvě třetiny*, kde se už hranice slovního druhu překračuje. Trojkombinací, tj. užití tří číslovek vedle sebe, je výrazně méně (26) a zase jde hlavně o číslovky základní typu *čtyři čtyři sedm* (což je ale spíš ad hoc záznam čtení číslic a lze je pak najít i hodně dlouhé).

Pro úplnost představy uveďme stručně i vyšší kombinace, tj. s více komponenty. Všechny mají po obecném zadání do korpusu jednotnou, nikde obvykle nezmiňovanou povahu, tj. že jsou co do obsazení komponentů vždy homogenní, stejného druhu, nejčastěji jde však o číslovky základní.

Čtyřkombinace jsou početně nízké taky (115), jinak jsou povahou stejné jako trojkombinace; vzácně se najdou i řadové číslovky typu *dvou stý osmdesátý první vězeň* nebo ty, co jsou časté při udávání letopočtu (*Devatenácti stý osmdesátý sedmý byl rokem Posledního císaře*). Smíšený, a tedy „nematematický" způsob čtení a psaní s nedodržením jinak obligatorního věcného pořadí představuje např. *dvě stě pětasedmdesát tisíc* (vedle přehození sledu srov. i aglutinační způsob psaní dohromady). Mimo stojí nehomogenní, smíšený způsob vyjadřování data *jedenáctého třetí čtyřicet tři (se měl hlásit)*.

Pětikombinace (26) jsou omezené na číslovky základní, z toho se vymyká 6 výskytů záznamu způsobu pochodu a jiného rytmu, srov. *raz dva raz dva raz dva*. ***Šestikombinace*** (7) a vyšší jsou zcela periferní, jediné běžné je čtení telefonního čísla, srov. *Tady je osm šest osm sedm dva sedm osm sedm.* Sedmikombinací (7) a osmikombinací (4) je minimum, neliší se od šestikombinací, srov. *Tady je osm šest osm sedm dva sedm osm sedm. Nechte nám vzkaz.* Zcela mimo stojí ojedinělé vícekomponentové výskyty číslovek základních jako např. *devět set devadesát devět tisíc devět set devadesát devět miliónt*in.

Nicméně objevují se tu jisté uzance, které s matematikou nemají nic společného (viz už výš), srov. např. oznamování, resp. záznam času *čtyři čtyřicet pět* nebo obecný způsob naznačování přibližnosti užitím dvou sousedních číslovek naznačujících rozmezí *čtyři pět tisíc* ap.

4. PROPOZICE: INTERJEKCE A JEJICH KOMBINACE

Citoslovce jsou mezi slovními druhy jakožto neslova, resp. mikrověty odsunovány stranou, jakkoliv ani syntaktikové se nevyznamenávají pozorností, kterou jim věnují. V příspěvku z r. 2007 (Čermák 2007) se analyzuje jejich základní úzus i problematické aspekty. Poukazuje se tu na korpusových datech na jejich využití v textu, na souvislosti s jinými slovními druhy (zvl. s partikulemi a adverbii), na jejich relativně častou funkci v kumulaci vystupovat jako propoziční frazémy (*No jéjé*), na jejich pozici ve větě, jejich valenci (*Hybaj do postele!*) i na otevřené problémy. Mezi ty patří jejich nepevná, neustálená délka, resp. forma (*jéje* i *jéjéjéjé*…, viz výše) a variabilita (*hergot / hernajs / hejšvec*). Zde se tudíž omezíme na jejich vzájemnou kombinatoriku, v článku už naznačenou. I citoslovce se navzájem kombinují a je třeba zjistit jak.

Dvoukombinací citoslovcí je pochopitelně nejvíc (1339), přitom celá (a) polovina (48 %) je založená na opakování, vyjadřujícím obvykle větší intenzitu či délku, např. *jupí jupí, ha ha, hm hm, ťuk ťuk, cha cha, uf uf* aj. Zbytek připadá většinou na víceméně ustálené víceslovné kombinace, počítané mezi (b) propoziční frazémy a zpracované

ve slovníku (SČFI), např. *ach jó, k sakru, hola hej, fuj tajbl, enyky benyky,* a jen nepočetná menšina připadá na zřejmě (c) autorské kombinace typu *kruci hergot, kruci sakra, hergot sakra, ach propána, himl laudón.* Při bližším pohledu na kontext ale i zde vidíme jasnou funkční a sémantickou souvislost a snahu zesíleně vyjádřit totéž. Žádná kombinace interjekcí tedy není zcela náhodná a bez vnitřní souvislosti.

Trojkombinace jsou opět výrazně řidší (299). Dominantní tu je prosté opakování typu *ach ach ach* či *ha ha ha,* výjimky jsou smíšené *hopla hopla hej, jupí jupí cupity* (tato jediná se zdá pro své větší opakování na hranicích ustálení). ***Čtyřkombinací*** je ještě méně (64) a akcentuje se tu zase prosté opakování, srov. *Uf uf uf uf…* Řídké smíšení s jiným citoslovcem může vyjadřovat signál změny rytmu, srov. *Ťuk ťuk ťuk prásk.* Z důvodu dominantního opakování však nemá velký smysl hledat vícekomponentové shluky interjekcí, které jsou povahou i strukturou podobné a jejichž v některých případech vysoké opakování slouží spíš jako rytmicko-stylistický prostředek než jako vyjádření postoje či zvlášť výrazného zvuku, srov. *Aj aj aj , aj aj aj aj , aj aj aj aj , aj aj aj aj, buráceli od dveří v rytmu reggae.*

5. KOMBINACE SMÍŠENÉ (NAPŘÍČ)

Kombinace zapojující členy více než jediné ze zkoumaných tříd neprozrazují ani zvláštnosti, ani tendence. Podíváme-li se na takové kombinace zájmen, zjišťujeme, že jsou kvantitativně velmi různé (od statisíců s předložkou po tisíce s interjekcí). Uveďme namátkou pár ilustrací:

— **Pron-Prep***: kdo z, každý z, ten o* (specifikace)
— **Pron-Num***: ti dva, ten druhý* (specifikace)
— **Pron-Kon:** *on a (já), já však* (koordinace, subordinace)
— **Pron-Int***: co sakra* (náhoda, vztah jen k propozici).

6. ZÁVĚRY A PROBLÉMY

Z výsledků průzkumu je zjevné, že i synsémantika se kombinují navzájem uvnitř svého slovního druhu, popř. napříč s jinými synsémantiky, a nejenom, jak se tradičně předpokládá, jakožto služebná slova pouze s autosémantiky. Nabývají v této své roli přitom poněkud jiné funkce.

Kromě případů modifikace, posílení či zpřesnění významu, který vidíme u zájmen (typ B *co všechno*) a kombinace reference a syntaktických vztahů (C *ten kdo*), je pozoruhodná tendence některých typů kombinací se ustalovat a vytvářet víceslovné lexémy (*ten můj, jeden dva, raz dva*), jakkoliv je řídká. Možnosti a hranice tohoto úzu však neznáme, nemáme-li je počítat mezi náhodné výtvory, kam patří vlastně celá frazeologie.

Závěrem je třeba znovu zdůraznit, že jakkoliv se tu dodává poslední část komplexního popisu synsémantik, je to jen popis stručný a často jen ilustrativní; snad bude inspirovat k důkladnému korpusovému popisu systematickému.

POUŽITÉ KORPUSY

Křen, M. — Bartoň, T. — Cvrček, V. — Hnátková, M. — Jelínek, T. — Kocek, J. — Novotná, R. — Petkevič, V. — Procházka, P. — Schmiedtová, V. — Skoumalová, H. (2010): *SYN2010: žánrově vyvážený korpus psané češtiny*. Praha: Ústav Českého národního korpusu FF UK. Dostupný z: <http://www.korpus.cz>

LITERATURA

Čermák, F. (1996): Systém, funkce, forma a sémantika českých předložek. *Slovo a slovesnost*, 57, s. 30–46.

Čermák, F. (2007): Povaha a úzus interjekcí: případ češtiny. In: J. Levická — R. Garabík (eds.), *Computer Treatment of Slavic and East European Languages. Slovanské a východoeurópské jazyky v počítačovom spracovaní.* Bratislava: Slovak National Corpus L. Štúr, Slovak Academy of Sciences, s. 299–307.

Čermák, F. (2008a): Syntagmatika, kombinace a kumulace konjunkcí. In: F. Štícha — M. Fried (eds.), *Grammar & Corpora, Gramatika a korpus*. Praha: Academia, s. 315–323.

Čermák, F. (2008b): Partikule, jejich syntagmatika a kumulace v mluvené češtině. In: M. Kopřivová — M. Waclawičová (eds.), *Čeština v mluveném korpusu*. Praha: NLN, s. 63–74.

Čermák, F. (2010): *Lexikon a sémantika*. Praha: NLN.

Čermák, F. — Holub, J. (1982): *Syntagmatika a paradigmatika českého slova I: Valence a kolokabilita*. Praha: SPN, kap. 2.6.

Šimandl, J. (2008): Problém doloženosti — případ tvarů číslovek. In: F. Štícha — M. Fried (eds.), *Grammar & Corpora* 2007. Praha: Academia, s. 427–435.

Štícha, F. a kol. (2013): *Akademická gramatika spisovné češtiny*. Praha: Academia.

František Čermák | Ústav Českého národního korpusu, FFUK | nám. J. Palacha 2, 116 38 Praha 1
frantisek.cermak@ff.cuni.cz