



Partikule v Pražském mluveném korpusu

Ana Adamovičová

ABSTRACT:

Particles in Prague Spoken Corpus. The article deals with particles, usually considered to be a residual part of speech, and strives to come to some general conclusions on occurrence and frequency of particles as well as their function in common spoken language. The basic source is the reference Prague Spoken Corpus (PSC) which is a part of the Czech National Corpus. The fact that particles are after verbs and pronouns the third most frequent word in spoken Czech appears for the first time in the Frequency Dictionary of Spoken Czech based on the PSC. This finding demands new tasks on linguists, especially more detailed description of this so frequently used part of speech, which hasn't been so far thoroughly analyzed on the basis of true authentic data. PSC provides the unique possibility to describe functions and the meaning of particles in the direct authentic context and usage, where they naturally appear. Large contextual scope is the decisive criterion for their identification. Description of particles requires a practical approach and by analyzing their real occurrence and co-occurrence one can prove, deny or change all theoretical premises. What we haven't found in the corpus is also a positive knowledge — the prove that in the corpus with a size of almost three quarter million tokens a particular word didn't appear. The article presents all types of particles which appear in the corpus and provides both its quantitative analyses dealing with original particles as well as with those homonymous with other parts of speech. It also deals with the existing processing of particles in various linguistic manuals.

KLÍČOVÁ SLOVA / KEY WORDS:

částice, frekvence slov, homonymní slova, korpusová lingvistika, kvantitativní analýza, mluvený jazyk, partikule, slovní druhy
corpus linguistics, frequency of words, homonymous words, particles, parts of speech, quantitative analysis, spoken language

O ÚVOD

Článek pojednává o partikulích, tradičně zbytkovém slovním druhu, a pokouší se dobrat některých obecnějších závěrů o jejich výskytu a frekvenci v běžně mluveném jazyce. V první části článku se stručně poohlédneme za zpracováním partikulí v některých vybraných příručkách. V další části představíme Pražský mluvený korpus (PMK) a konečně v třetí části se zaměříme na popis druhů partikulí vyskytujících se v PMK.

1 SLOVNÍ DRUH PARTIKULE

O partikulích jako o slovním druhu se začalo uvažovat poměrně pozdě: např. první řecké gramatiky, které převzali staří Římané, uváděly jen osm slovních druhů, mezi



nimiž se partikule ještě nevyskytovaly.¹ Ve 4. století byl člen, původní slovní druh vyskytující se v starořeckých gramatikách, nahrazen interjekcemi. První zmínka o partikulích se objevuje v Číně, a to dost pozdě, až ve 14. století. Nicméně právě čínský spis o partikulích se považuje za první, alespoň částečný pokus o sepsání první čínské gramatiky — do té doby se čínští jazykovědci věnovali hlavně zvukové stránce jazyka, popisu tónů a intonace. Musíme však mít na zřeteli, že partikule v čínštině nebo v japonštině nelze srovnávat s pojetím partikulí, které je rozšířeno v Evropě. Do evropské lingvistiky se úvahy o partikulích dostávají až začátkem 20. století, a to spíše okrajově. Většinou se za ně označuje jen několik málo slovních tvarů, s jejichž popisem či zařazením do jazykového systému si jazykovědci nevěděli rady.

1.1 NĚKOLIK POZNÁMEK K SOUČASNÉMU STAVU BÁDÁNÍ

Během posledních čtyřiceti let přístup k partikulím, slovnímu druhu postiženému syndromem „odpadkového koše“ (Čermák, 2008, s. 63), bezesporu zaznamenal pozoruhodný vývoj, který však zatím ještě nebyl zcela systematicky zpracován. Průlom v nahlížení na partikule přináší až orientace lingvistiky k pragmatice a běžně mluvenému jazyku. Genezi tohoto přístupu ke zkoumání jazykových dat a jejich funkcí ve větě lze patrně hledat v již poměrně dávné minulosti, a sice v časovém horizontu zhruba první poloviny a začátku druhé poloviny 20. století. V těchto souvislostech je jistě třeba zmínit de Saussura a Wittgensteina — vzpomeňme jejich známé výroky „v mluvě se nalézá zárodek všech změn“, které se postupně dostávají „do obecného úzu“ (de Saussure, 1996, s. 125), a „význam slov je v jejich užití“ (Wittgenstein, 1993, s. 34–43) — anebo Austina a Searla s jejich teorií mluvních aktů zasazených do kontextu promluvy. Při postupném odklonu lingvistiky od gramatiky (morfologie či syntaxe) a zvýšené orientaci na sémantiku, pragmatiku a textovou lingvistiku se začalo uvažovat o komunikačních funkcích jazyka, na jejichž realizaci se ve velké míře podílejí též partikule.

Začátkem sedmdesátých let minulého století si ve světle nových teorií i čeští lingvisté uvědomili, že modální adverbia mají ve větě především funkci postojovou, a tak se z nich staly modální partikule (Bauer — Grepl, 1972). V osmdesátých letech vychází 2. díl tzv. akademické *Mluvnice češtiny* (MČ2), v němž se řady partikulí rozšířily o další desítky příkladů (Uličný, 1986). Grepl (2011) v devadesátých letech vydává sérii přednášek *Co děláme, když mluvíme*, v nichž také pojednává o mluvních aktech a partikulích. V polovině devadesátých let a začátkem nového tisíciletí vycházejí další práce — *Příruční mluvnice češtiny* (PMČ), partikulím se zde věnuje Nekula, a *Encyklopedický slovník češtiny* (EŠČ), kde hesla o partikulích zpracovávají Grepl a Nekula. Takto mohutný vývoj se koncem sedmdesátých let do jisté míry odrazil i v lexikografické rovině: v SSČ (1978) se řada adverbíí přehodnocuje na partikule.

Partikule jsou však v různých příručkách definovány často nejednotně a někdy i nahodile. EŠČ (2002, s. 62) to přímo připouští, když je označuje za „jeden ze slov-

1 Ačkoli se nevyskytovaly v gramatikách, dnes už máme důkazy, že se o nich vědělo, srov. např. Wiktionary, seznam *Ancient Greek Particles*, v němž se uvádí, že je jich alespoň 20, a pětisvazkový sborník *Particles in Ancient Greek Discourse* (Bonifazi — Drummen — de Kreij, 2016).



ních druhů, který je vymezován velmi různě“. PMČ (1995, s. 358) uvádí, že částice jsou „slova se synsémantickým významem nezahrnující předložky a spojky“. Tato poslední definice jednak obsahuje jistý protimluv, a sice „synsémantický význam“, jednak vyznívá tak, že se, na rozdíl od předložek a spojek, o kterých nejspíš víme, jaký mají význam a funkci, o partikulích totéž říci nedá. Dalším problémem při určování partikulí je splývavá, nejasná hranice mezi partikulemi a dalším „zbytkovým“ slovním druhem, interjekcemi, na což trefně ve své studii upozornil M. Vondráček (1998). Kromě nejednotného přístupu je mnohem vážnějším problémem skutečnost, že příklady uváděné pro ilustraci jednotlivých tezí vycházejí z teorie, která však nebyla empiricky ověřena v úzu, ovšem také skutečnost, že se k pragmatice přistupuje z různých východisek.

Zlom, který do lingvistiky vneslo korpusové zpracování jazykových dat, je tak obrovský, že právem lze mluvit o mezníku rozdělujícím lingvistická bádání na éru před vznikem korpusové lingvistiky a na éru zahájenou po jejím vzniku. Korpusový materiál je také nepřeborným zdrojem dat poskytujícím cenné informace právě při studiu partikulí.

1.2 VYMEZENÍ PARTIKULÍ

Klíčem k lepší či přehlednější strukturaci partikulí je hledisko, v němž se prolíná funkčně-formální přístup: funkce v promluvě je prezentována jistou formou. Máme-li už zjištěnou formu a vyjdeme-li z ní, dostaneme se znovu k funkci ozřejmené o způsob užití, kontext, význam, valenci a začlenění do textu/promluvy. Tam se ukáže i onen sémantický posun, ke kterému dochází při slovnědruhové transpozici, která vznik partikulí často provází. Jde především o dva procesy: procesy lexikalizace gramatických slov (*co když, jen aby, ještě že, už aby* apod.) a procesy delexikalizace (pragmatizace) autosémantik (*čistě, prostě, klidně, pomalu, rozhodně* apod.). Partikule jsou jistými pragmatickými komentáři, referujícími formálně i sémanticky k rozličným aspektům promluvy/textu, k postojům mluvčího, často však i k širší mimojazykové situaci (Čermák, 2008, s. 66). Obecně vzato by se dalo říct, že je partikule specifickým „kořením“ jazyka a má, na rozdíl od adverbia, především komunikační význam, který se realizuje v kontextu. Zároveň je třeba zdůraznit, že se partikule, na rozdíl od interjekce, nikdy nevyskytuje samostatně, nýbrž v kombinaci s dalším slovem (adlexémové partikule) či větou (adpropoziční partikule) a přitom může stát na různých místech v promluvě.

2 PRAŽSKÝ MLUVENÝ KORPUS

Pražský mluvený korpus (PMK), který začal vznikat koncem 80. let minulého století, je prvním korpusem mluvené češtiny zachycujícím autentický a spontánní mluvený jazyk. Budování PMK začalo na popud F. Čermáka, který již delší dobu pocítoval potřebu systematicky, důsledně a taxativně popsat mluvený jazyk jakožto „primární zdroj komunikace a zdroj jazykových změn zakládajících jeho vývoj“ (Adamovičová — Čermák — Pešička, 2008, s. 7).



2.1 PRVNÍ FÁZE PROJEKTU

Na začátku projektu bylo pořízeno 304 nahrávek² obsahujících přes 15 015 promluv, čímž vzniklo 7600 minut nahraných monologů a dialogů, tj. 316 hodin mluveného projevu. Při nahrávkách byly pořízeny jak formální rozhovory čili řízené podle předem připraveného scénáře (193 nahrávek), tak rozhovory neformální čili neřízené, spontánní (103 nahrávek). Formální nahrávky neboli monology vycházely z dotazníku zaměřeného na zachycení názorů respondentů na otázky týkající se běžného života: škola, vzdělávání, práce, postavení žen na pracovišti, manželství, rodina, výchova dětí, mezilidské vztahy, život ve městě vs. život na venkově, šťastný a úspěšný život apod., v druhé vlně nahrávání po změně politického klimatu v r. 1989 přibýly otázky o podnikání a cestovním ruchu. Neformální nahrávky neboli dialogy jsou volným rozhovorem dvou respondentů, kteří se dobře znají, na jakékoli téma. Jde tedy o prototypickou mluvní situaci — neformální, nepřipravenou a dialogickou.

Respondentům (v úhrnu jich bylo 504: 287 žen a 217 mužů) byla zaručena anonymita a nebyl jim předem sdělen účel (tj. ryze jazykový výzkum), za jakým se nahrávky pořizují. Mnozí se tak domnívali, že jde o jakési sociologické či sociolingvistické šetření zaměřené na názor dotazovaných na běžná témata z osobního, pracovního a společenského života. Přestože se většina mluvčích dokázala uvolnit a rozpovídat, přítomnost nahrávacího zařízení pro některé z respondentů byla nicméně stresujícím faktorem, mnohem častěji však v začátcích natáčení, kdy roli hrála i všeobecně panující nedůvěra (před rokem 1989), což se nejednou odrazilo v tom, že respondenti mluvili více spisovně a strojeně, více se hlídali, u některých byl dokonce patrný i strach mluvit otevřeně. V druhé vlně nahrávání v polovině devadesátých let se tyto obavy respondentů celkem rozplynuly, a tak nahrávky působí poněkud přirozeněji. Pro úplnost je ještě potřeba dodat, že se při kombinaci dalších sociolingvistických parametrů, jako jsou věk a vzdělání respondentů, včetně typu promluvy, dbalo o jejich maximální vyváženost.

2.2 DRUHÁ FÁZE PROJEKTU

Ve druhé fázi projektu byly všechny pořízené nahrávky přepsány do počítače maximálně věrně, což bylo časově velmi náročné. Při přepisu se hlavně dbalo na to, přenést do psané podoby každý zvuk, včetně hezitačních, různé odchylky ve výslovnosti, delší či kratší odmlky mluvčího, zánik kvantity apod. Rozhodující slovo však měl přepisovač, a tak to, zda polodélku zachytil jako délku, či nikoli, záleželo pouze na jeho posouzení. Asimilace souhlásek se však nezaznamenávala (mělo se zapisovat výlučně např. *radši*; v korpusu se však ojediněle vyskytují i tvary *rači*). Počítačový přepis má tedy spíše fonologicko-morfologický charakter, a je tudíž snadno čitelný i pro méně zasvěceného uživatele. Hlavním důvodem takto zvolené přepisovací strategie bylo výhledové zařazení PMK do databáze Českého národního korpusu (o kterém se v té

2 Původně bylo nahrávek mnohem více, avšak řada z nich musela být z další fáze zpracovávání materiálu vyřazena pro špatnou kvalitu zvuku způsobenou nedokonalým nahrávacím zařízením.



době teprve uvažovalo). Přepis je takto kompatibilní s psanými korpusy, na druhou stranu však obsahuje méně informací suprasegmentálního charakteru, než je pro některé, především novější mluvené korpusy běžné. Takto vlastně vznikl Pražský mluvený korpus, který můžeme označit za referenční, tj. stálý, uzavřený a neměnicí se (detailněji viz FSMČ, 2007, s. 11–35).

2.3 TŘETÍ FÁZE PROJEKTU

Ve třetí fázi projektu se přistoupilo k anotaci takto získaného materiálu na základě kódovníku, který pro tyto účely sestavil F. Čermák. Šlo o ruční označení každého slova až jedenáctimístnými kódy poskytujícími především úplnou morfologickou informaci, včetně široce pojaté valence a funkce, částečně také syntaktickou informaci³ a údaj o stylové charakteristice promluvy. Celý materiál obsahuje více než 7 milionů indexů.

Pražský mluvený korpus je tedy první český korpus, který nabízí plné, systematické a pečlivě ověřované tagování všech svých forem. Pro časovou náročnost celé práce byla část korpusu okódována jen redukovane, tj. označena jen tří- až čtyřmístnými kódy. Vedle deseti slovních druhů byly také označeny a blíže zařazeny i citátové výrazy jednoslovné i víceslovné (CV), neslovní zkratky (ZKRAT), propria (PROP) a frazémy (IF). Byly také poprvé označeny všechny víceslovné analytické gramatické formy (*abych šel, kdyby přišlo* atd.), jakož i další víceslovné výrazy včetně partikulí (PAR), víceslovných spojek (CONJ) a frazémů. Identifikace a označení frazémů vycházejí ze zpracování frazeologie ve *Slovníku české frazeologie a idiomatiky*.

Velmi pracná ruční anotace oproti mnohem běžnějšímu automatickému tagování má bezesporu výhody právě v případě zkoumání partikulí: vzhledem k vysokému procentu homonymních tvarů řady partikulí s dalšími slovními druhy je jejich automatická identifikace jinde dosti nepřesná. Autoři modifikovaného analyzátoru *Ajka* pro anotaci mluvených korpusů (Hlaváčková — Osolobě, 2008, s. 111) uvádějí, že např. k slovnímu tvaru *tak* analyzátor přiřadí až čtyři možné interpretace — adverbium, spojka, citoslovce nebo částice — [tag="k6"] [tag="k8"] [tag="k9"] [tag="ko"], což bez časově náročné kontextové a manuální analýzy nelze úspěšně disambiguovat, tj. zjednoznačnit. Otázkou zůstává, zda vůbec, anebo do jaké míry, lze pro trénovací korpus jednoznačně postihnout všechny kontextové výskyty řady partikulí, což by

3 K syntaktické informaci nutno podotknout, že ačkoli se s její analýzou při sestavování kódovníku původně počítalo, v praxi se od toho z větší části upustilo, neboť se rozdíly v syntaxi mluveného a psaného jazyka ukázaly být tak velké, že se syntax psaného jazyka, obzvláště na úrovni větněčlenské kategorizace, pro analýzu jazyka mluveného stala jen stěží použitelnou. Vznikaly mj. různé problémy související nejen s označením hranice věty v souvislém proudu vyprávění, ale také s jejím vnitřním členěním a odlišnou větnou perspektivou mající často eliptický ráz. Není to nic překvapivého: každý korpus, obzvláště však mluvený, je v zásadě pestřejší a bohatší než sebelépe navržený systém značek. Proto také u řady kategorií přibýly značky — *jiné*, příp. *jiné/nejasné*, nebo *nelze určit*, které byly poměrně často obsazovány, ačkoli se ještě, pokud je nám známo, nikdo popisu těchto problémových kategorií nevěnoval.

mohlo značně zvýšit úspěšnost automatické disambiguace. Při další analýze materiálu bychom se však měli dobrat některých konkrétních návrhů na zlepšení automatického rozpoznání partikulí.

PMK se stal podkladem pro vytvoření *Frekvenčního slovníku mluvené češtiny* (FSMČ), vydaného nakladatelstvím Karolinum v roce 2007. Slovník obsahuje 780 322 pozic včetně interpunkcí, po jejichž odečtení máme přesně 644 909, respektive 548 091 slov (první číslo udává počet všech tvarů včetně tvarů vázaných ve víceslovných spojeních; komponenty vázané na výskyt ve víceslovných jednotkách mají 96 818 výskytů). V PMK najdeme 46 798 různých slov, tj. lemmat, respektive 30 570 slov (opět po odečtení vázaných slov).

Cílem projektu PMK–FSMČ bylo vůbec poprvé souvisle a vyčerpávajícím způsobem zmapovat autentickou mluvenou češtinu a nezkráceně představit její jádro — obecnou češtinu. Korpus má přídomek „pražský“, protože se vycházelo z představy, že Praha, vzhledem ke svému centrálnímu postavení, svým způsobem reprezentuje běžnou českou mluvu v Čechách a zčásti i na Moravě, mj. i vzhledem k množství původem mimopražských Pražanů. PMK díky této strategii — systematickému pokrytí jazykové oblasti obohacenému o intenzivní anotaci — skýtá zcela ojedinělý obraz české mluvené situace z konce 20. století, přičemž odráží jak jazyk předchozího společenského uspořádání, tak začátek nového. Přestože lze namítat, že je daný rozsah pokrytí omezený, slovník představuje cennou materiálovou bázi pro studium běžné mluvené češtiny.

Pokud jde o lexikon, je zde ale nutno poznamenat, že četnost výskytů některých slov, obzvláště substantiv a některých frází, může být poněkud zkrácena, neboť respondenti automaticky opakovali slova obsažená v otázkách dotazníku a navíc se vyjadřovali k předem určeným tématům. Týká se to hlavně slov a obrátů jako *škola*, *učitelka*, *rodiče*, *děti*, *trestat*, *vzít pásek*, *postavení žen na pracovišti*, *zaměstnání*, *doplňovat si kvalifikaci*, *život ve městě vs. život na venkově*, *vidět si do talíře*, *šťastný a úspěšný život* atp.

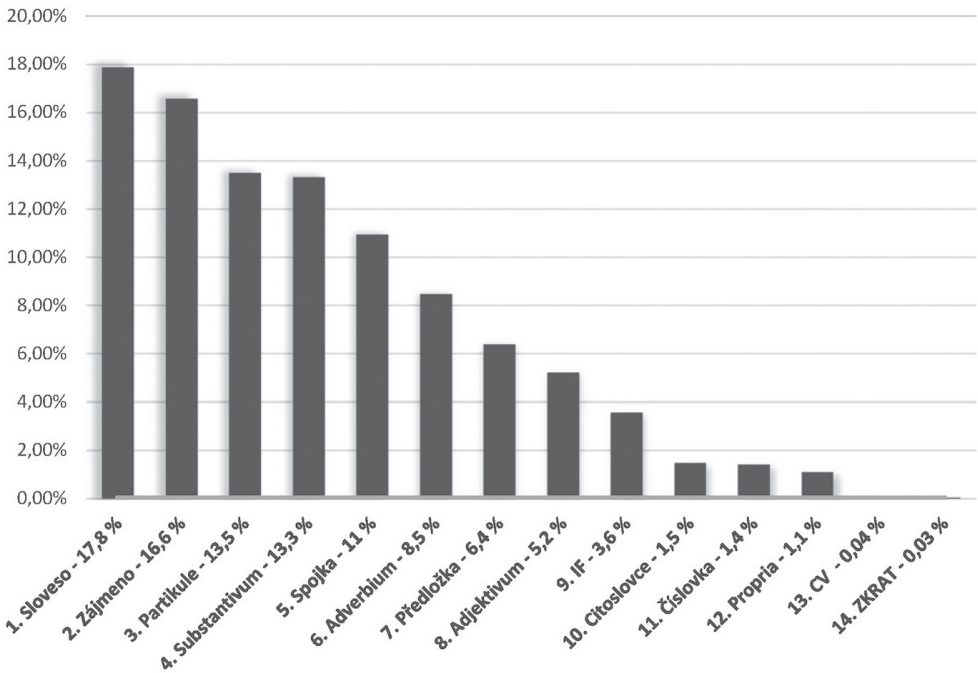
Výskyt partikulí, na rozdíl od zmíněných autosémantik a některých frazémů, není však ovlivněn žádnými vnějšími faktory, a jejich frekvenci lze tudíž považovat za plně vypovídající a nosnou. FSMČ poprvé uveřejnil údaj, že jsou partikule třetím nejfrekventovanějším slovním druhem v mluvené češtině, jinými slovy — v průměru každým osmým slovem v promluvě bude partikule.

3 PARTIKULE V PRAŽSKÉM MLUVENÉM KORPUSU

Podívejme se nyní blíže na posledně zmíněný údaj o frekvenci partikulí, ovšem také dalších slovních druhů v běžně mluveném projevu, který znázorňuje graf 1.

Frekvenční pořadí slovních druhů v psaném korpusu je pochopitelně značně jiné. Uvedme pro srovnání, že distribuce slovních druhů v žánrově vyváženém stomilionovém korpusu SYN2000, v němž převažují texty z let 1990–1999, tedy z podobného období jako nahrávky tvořící PMK, vypadá takto: 1. substantiva (tvoří zhruba třetinu velikosti korpusu), 2. slovesa (jejich výskyt je však na první pohled překvapivě o polovinu nižší), 3. adjektiva (12,5 mil.), 4. předložky, 5. zájmena, 6. spojky, 7. adverbia, 8. číslovky, 9. partikule a na posledním, 10. místě citoslovce. Je zajímavé, že stejné

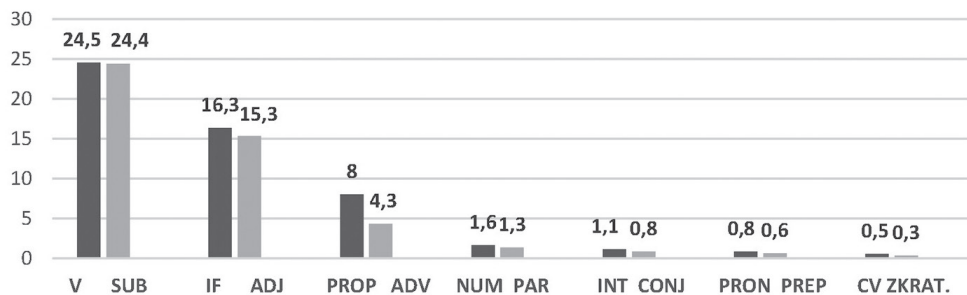




GRAF 1: Frekvence slovních druhů, IF, proprií, citátových výrazů a zkratk v PMK.

frekvenční pořadí, pokud jde o první tři nejfrekventovanější slovní druhy (S, V a ADJ), uvádí i první česká publikace věnovaná frekvenci slov (Jelínek — Bečka — Těšitelová, 1961). Jejich korpus tvoří 1 623 527 slov, je převážně sestaven z beletrie, literatury pro děti a mládež a dramát (cca 57 % hesel), odborné a vědecké literatury (cca 25 % hesel) a v poslední necelé čtvrtině se převážně vyskytuje žurnalistika, okrajově poezie a mluvený projev (nutno podotknout, že většinou šlo o projevy čelných komunistických představitelů té doby). Pro nás je obzvláště cenný poznatek, že se v roce 1961 partikule jako slovní druh vůbec neuvažovala.

V porovnání s mluveným korpusem v psaném korpusu nejnvýraznější propad zaznamenaly právě partikule (z 3. na 9. místo), ovšem také zájmena (z 2. na 5. místo), přičemž jsou slovesa procentuálně mnohem méně zastoupena v psaném než v mluveném korpusu, na rozdíl od substantiv, která jsou výrazně víc zastoupena v korpusu psaném. Tyto výsledky nejsou však nikterak překvapivé. Partikule (jak už jsme na to upozornili) jsou při automatické anotaci těžko rozpoznatelné (v automaticky tagovaném korpusu je rozlišení partikulů značně problematické vzhledem k vysoké míře homonymie, kterou partikule vykazují s dalšími slovními druhy: řada partikulů v takovém, především psaném korpusu zůstává proto skryta zejména za adverbii, spojkami, tvary sloves, citoslovci apod.), deiktika se v mluvených projevech vyskytují mnohem více. Pokud jde o nízký procentní podíl sloves oproti substantivům, zde se nabízí vysvětlení, že v psaném projevu tíhnuoucím ke kondenzaci mají nominální konstrukce jednoznačnou převahu. S tím přímo souvisí i vysoká zastoupenost předložek



GRAF 2: Frekvence lemmat v PMK.

v psaném korpusu (v mluveném korpusu je najdeme na 7. místě, v psaném na 4.). Na velkém zastoupení předložek v psaném korpusu se podílejí především víceslovné předložky vyskytující se hojně právě v nominálních konstrukcích, které nejsou charakteristické pro mluvený projev.

Překvapivě vysoko se však v porovnání s mluveným korpusem umístila adjektiva (v mluveném korpusu jsou na 8. místě, v psaném na 3.). V souvislosti s tímto zjištěním se nutně nabízí možné vysvětlení: v psaném jazyce se často při popisu uchylujeme k metaforickému vyjadřování, pro něž jsou právě charakteristické synonymické řady adjektiv. Při redakční práci text dále třídíme, obohacujeme ho o další řady synonym. Při mluveném projevu na to není čas. Stačí se podívat na nejfrekventovanější adjektiva v psaném a mluveném korpusu. V první stovce nejfrekventovanějších autosémantik se v mluveném korpusu vyskytlo pouze 8 adjektiv: *dobrej* (36), *jinej* (37), *velkej* (41), *mladej* (57), *malej* (74), *celej* (78), *stejne* (96), *špatnej* (100). M. Kopřivová (2006, s. 17) ve své práci uvádí seznam 100 nejfrekventovanějších adjektivních lemmat v korpusu SYN2000. Stejně, anebo podobné frekvenční pořadí se vyskytlo pouze u slov, která můžeme označit za jakési neutrální konstanty popisování kvality, případně relací. Partikule jsou oproti adjektivům univerzálnějšími komentátory promluvy, jejichž význam se konkretizuje až v kontextu — podstatné je, že se na rozdíl od adjektiv často vztahují k celé promluvě a jsou jednoznačnější z hlediska postojů mluvčího.

Podíváme-li se ovšem na zastoupení slovních druhů v PMK z hlediska počtu lemmat ve FSMČ znázorněného v grafu 2, situace se jeví značně jinak.

Zde dle očekávání u vrchu tabulky najdeme slovesa, substantiva, po nich frazémy, adjektiva, propria a adverbia. U těchto slovních druhů se vyskytlo nejvíce různých slov, respektive jejich repertoár je nejpestřejší. Počet různých lemmat u partikulí je 400, a takto se partikule umístily v dolní části tabulky mezi číslovkami (NUM) a citoslovci (INT), předčí však zájmena (PRON; 249 lemmat) nebo předložky (PREP; 199 lemmat) a hodně zaostávají za adverbii, u nichž najdeme až 1324 různých lemmat. Těchto 400 různých lemmat partikulí představuje jen 1,32 % slovníku. Tato slova však, jak už víme, vykazují tak vysokou frekvenci, že partikule jsou celkově třetím nejfrekventovanějším slovním druhem v mluveném jazyce, respektive každým 8. slovem v promluvě (13,51 % slovníku).

Porovnáme-li partikule z obou zmíněných hledisek — frekvence tvarů a frekvence lemmat — například s číslovkami, zjistíme, že 480 lemmat číslovek předsta-



vuje 1,57 % slovníku, z hlediska frekvence jejich výskytu je však zastoupení číslovek podobné (1,42 %), a jde tedy na rozdíl od partikulí o slovní druh s poměrně dosti nízkou frekvencí. Porovnejme ještě z tohoto hlediska adverbia a partikule: 1324 všech lemmat adverbii tvoří 4,33 % slovníku (je jich tedy 3× více než partikulí), frekvence jejich tvarů je pak 8,47 % (více než o třetinu nižší než u partikulí).

Uvedme nyní ještě prvních 20 nejfrekventovanějších slov / stylových lemmat doložených v PMK.

1. ten	40 271	11. tak	6 224
2. být	20 255	12. ale	6 022
3. a	19 590	13. tam	5 704
4. že	13 542	14. teda	4 276
5. já	8 743	15. prostě	3 924
6. mít	7 871	16. myslet si	3 263
7. no	7 785	17. nebo	3 262
8. v	7 716	18. třeba	3 186
9. na	7 067	19. s	3 151
10. jako	6 717	20. protože	3 086

TABULKA 1: 20 nejfrekventovanějších slov v PMK.

Z autosémantik na seznamu najdeme pouze tři verba (*být*, *mít* a *myslet si*) a jedno adverbium (*tam*). Z jednoznačných synsémantik se zde nejvíce vyskytují partikule, a sice čtyři (*no*, *teda*, *třeba*⁴ a *prostě*), dále tři prepozice (*v*, *na* a *s*), dvě zájmena (*ten* a *já*) a dvě spojky (*nebo* a *protože*). Na frekvenci všech dalších pěti slov (*a*, *že*, *jako*, *tak*, *ale*) se však v různé, někdy i ve větší míře (např. u *jako*), podílejí též partikule.

Je třeba si uvědomit, že tato čísla jsou ve skutečnosti ještě vyšší, neboť se do nich promítají jen počty všech tvarů těchto slov izolovaně čili slovníkové lemma, ne však další kombinace lemmat. A tak např. sloveso *být* nezahrnuje celou řadu propozičních frazémů typu *to je bezvadný*, *to je blbost*, *to je dobrý* a spousty dalších, do frekvence spojky a partikule *a* se též nepromítají propoziční frazémy *a dost*, *a hotovo*, *a je to*, *a tak dále*, *a šmytec* atp. či víceslovné spojky *a proto*, *a přitom*, *a tak*, do frekvence spojky a partikule *že* nejsou zahrnuty výrazy *že jo*, *že ano*, *že ne*, do počtu výskytů zájmena *já* se nepromítají propoziční frazémy *já ne/vím*, *(co všechno)* atp., partikule a interjekce *no* nezahrnuje propoziční víceslovné partikule či frazémy *no tak*, *no jo*, *no vidíš*, *no nazdar* apod., sloveso *mít* je též základním komponentem spousty frazémů, např. *mít ambice*, *mít co dělat*, *mít chuť*, *mít logiku*, *mít rád* atd., prepozice *v* je součástí řady frazémů či víceslovných prepozic, např. *v každém případě*, *v podstatě*, *v pohodě*, *v průměru*, *v tomhle směru*, *v zásadě*, *v žádném případě* či *v důsledku*, *v porovnání s*, *v průběhu*, *v případě*, stejně jako prepozice a interjekce *na*, např. *na co*, *na druhou stranu*, *na druhý straně*, *na první pohled*, *na úrovni* či *na konci*, *na počátku*, *na základě*, partikule a spojka *jako* je též součástí víceslovných spojek *jako dyby*, *jako když* či frazému *jako takovej*,

4 Výraz *třeba* byl jen 58× anotován jako adverbium.

partikule a adverbium *tak* se hojně vyskytuje jako první komponent frazémů *tak jako tak, tak řík, tak vidíš* atd.

Zdaleka nejvyšší frekvenci má zájmeno *ten*. Pro dokreslení významu *deixe* a anafory v češtině si uvědomme, že do lemmatu *ten* nebyla zahrnuta samostatná lemmata *tendle, tendleten, tento* ani samostatné lemma partikule *to*, víceslovná vztažná zájmena *ten — co, ten — kdo, ten — kterej* či víceslovné zájmeno *ten samej*. Poznamenejme ještě, že k vysoké frekvenci verba *myslet si* nejspíš přispěly, jak už jsme na to upozornili výše, znění dotazníku a typ průzkumu vyžadující po respondentech zaujetí vlastního stanoviska. Stejně je tomu u druhého nejfrekventovanějšího zájmena *já*.

Vidíme tedy, že na frekvenci nejfrekventovanějších slov v běžně mluveném jazyce mají partikule jednoznačně dominantní podíl.

3.1 LEMMATA PARTIKULÍ

V PMK se, jak už bylo zmíněno, vyskytlo 400 lemmat partikulí. Všechna lemmata zahrnují jak slovníkové (stylové) lemma, tak všechny textové varianty, tj. doložené tvary toho kterého lemmatu. Najdeme tedy mezi nimi jak *dyť, taky* nebo *teda*, tak také *vždyť, také* i *tedy*. Do těchto lemmat se pochopitelně nepromítají víceslovné partikule, a tak se tvary vázaných lemmat zde vyskytují izolovaně (diskrétně), např. *dejme* je součástí víceslovné partikule *dejme tomu*, frekvence *no* zahrnuje jak samotné *no*, tak *no* vázané ve víceslovné partikuli *no tak, dyž* se vyskytuje jak samostatně, tak jako vázaný tvar ve víceslovných lemmatech *dyž už, dyž tak* apod.

V níže uvedené tabulce 2 zobrazíme prvních 120 nejfrekventovanějších tvarů partikulí, tedy zhruba první čtvrtinu všech jejich doložených tvarů (všech doložených forem je ve skutečnosti 480 — kromě hapaxů se mezi nimi vyskytují i různé okrajové tvary — individuální formy (*aj, asima, zasej*), ojediněle zaznamenaná nedbalá výslovnost (*rostě, kórt, upně*), překlapy (*zdaleko*) atp. Kurzivou jsou v tabulce označeny variantní tvary (též tvar *také*, přestože byl při sestavování slovníku považován za samostatné lemma, tvar *alespoň* s 11 výskyty se do tohoto seznamu nedostal).

Podíváme-li se na čtyři sloupce ve výše uvedené tabulce, hned na první pohled nás zaujme délka jednotlivých slov. V prvním sloupci, tedy v první třicítce nejfrekventovanějších tvarů, se převážně vyskytují velmi krátká slova s jedinou výjimkou slova *samozřejmě*. Počet jednoslabičných slov s klesající frekvencí pozvolna klesá (1.–4. sloupec: 12, 9, 7, 6). V každém sloupci je nejvíce dvouslabičných slov (1.–4. sloupec: 15, 17, 11, 12). Tříslabičných slov je nejvíce ve 3. sloupci (10), čtyřslabičných ve 4. sloupci (4), jde o slova *absolutně, víceméně, každopádně a náhodou*. Ve 3. sloupci je jen jedno čtyřslabičné slovo — *vyložně*. Pětislabičná slova jsou spíše výjimkou — vyskytla se pouze dvě: *pochopitelně* ve 3. sloupci a *pravděpodobně* ve 4. sloupci. Dohromady tříslabičná (23 tvarů, 19,1%), čtyřslabičná (6 tvarů, 5%) a pětislabičná slova (2 tvary, 1,6%) mají tedy jen 31 výskytů čili tvoří 25,7% všech tvarů. Další téměř 75% všech tvarů tvoří slova jednoslabičná (34 tvarů, 28,3%) a dvouslabičná (55 tvarů, 46%). Preference krátkých slov z hlediska jejich četnosti je zcela v souladu s jednou ze základních jazykových zákonitostí, a sice jazykovou ekonomii, a jen potvrzuje pravidlo popsané v Zipfových zákonech.

Pro další výzkum je zajímavé zjistit, v jakém kontextu se vyskytuje v prvním sloupci vyčnívající co do délky slovo *samozřejmě* (frekvence 760, 21. ve frekvenčním





1. no	8263	31. ne	471	61. co	161	91. absolutně	70
2. jako	5448	32. takže	469	62. nebo	155	92. víceméně	70
3. tak	5132	33. možná	421	63. prej	147	93. teprve	69
4. teda	4250	34. jó	418	64. totiž	141	94. také	63
5. prostě	3906	35. tedy	406	65. naprosto	139	95. klidně	62
6. třeba	3202	36. hlavně	395	66. normálně	134	96. každopádně	60
7. že	2855	37. až	375	67. fakt	131	97. takhle	59
8. a	2854	38. dyť	375	68. například	125	98. holt	59
9. asi	2440	39. aspoň	313	69. ať	124	99. jakoby	57
10. jo	2362	40. skutečně	301	70. celkem	124	100. přitom	56
11. taky	2248	41. přece	300	71. dokonce	122	101. když	56
12. to	1574	42. víc	298	72. akorát	122	102. radši	54
13. i	1439	43. ně	289	73. vyloženě	118	103. konkrétně	53
14. ale	1385	44. zrovna	282	74. dejme	110	104. von	52
15. vlastně	1266	45. jinak	281	75. třeba	109	105. jen	50
16. zase	1143	46. pak	278	76. opravdu	108	106. zvlášť	49
17. už	1018	47. zas	273	77. ještě	106	107. teďka	47
18. eště	963	48. rozhodně	248	78. naopak	104	108. navíc	47
19. vůbec	906	49. jaksi	229	79. přímo	104	109. vona	46
20. nó	856	50. potom	229	80. nakonec	101	110. pravděpodobně	45
21. samozřejmě	760	51. teď	219	81. zřejmě	98	111. ono	45
22. ani	676	52. stejně	203	82. ňák	93	112. pomalu	45
23. určitě	646	53. bohužel	175	83. jenom	90	113. jistě	43
24. vopravdu	608	54. docela	173	84. řekněme	90	114. především	42
25. spíš	595	55. úplně	171	85. přeci	86	115. především	42
26. právě	565	56. ovšem	169	86. zkrátka	86	116. třebaš	41
27. vlasně	515	57. myslim	167	87. pochopitelně	82	117. zejména	39
28. vono	478	58. úplně	166	88. prakticky	79	118. á	39
29. vid'	475	59. takle	163	89. dyž	75	119. náhodou	38
30. snad	474	60. vůbec	161	90. dost	71	120. voni	36

TABULKA 2: Prvních 120 tvarů partikulí podle frekvence.

pořadí) a jak jsou na tom jeho kratší synonyma *určitě*, *jasně* a *jistě*. Zatímco *určitě* s 646 výskyty a 23. místem v pořadí najdeme těsně za *samozřejmě*, *jistě* je v pořadí až 113. a má pouze 43 výskytů. *Jasně* se jako partikule mezi prvními nejfrekventovanějšími 120 tvary vůbec nevyskytlo. Najdeme ho až v druhé půlce seznamu s 6 výskyty, 196. v pořadí, čili někde v půlce seznamu všech tvarů partikulí. Vysvětlení pro to je jednoduché: *samozřejmě* a *určitě* vykazují syntagmaticky mnohem výraznější tendenci vázat se k dalším slovům ve větě, a tedy plnit funkci partikule, zatímco *jistě* a *jasně* (mimochodem obě slova v korpusu mají stejný počet výskytů — 43) vykazují mnohem výraznější tendenci stát samostatně čili ve funkci interjekce (*jistě* 28×, *jasně* 25×).

Prvních 11 partikulí má velmi vysokou frekvenci, která pozvolna začíná klesat. Více než 1000 výskytů najdeme u prvních 17 partikulí, v pořadí 80. partikule *nakonec*



se vyskytla už jen 101×. Přes 300 všech doložených tvarů má frekvenci nižší než 100. Jinými slovy — pouze 20 % partikulí má frekvenci vyšší než 100 výskytů. Pokud se zaměříme na podíl jednotlivých frekvencí v rámci těchto prvních nejčetnějších 20 % všech tvarů partikulí, získáme následující poměry.

Frekvenční podíl partikulí s frekvencí > 100		
Frekvence	Procentní zastoupení	Počet tvarů
> 1000	21,25 %	17
> 500	12,50 %	10
> 300	17,50 %	14
> 200	12,75 %	11
> 100	35 %	28
	99,00 %	80

TABULKA 3: Frekvenční podíl prvních 80 partikulí tvořících 20 % všech doložených tvarů.

Vidíme, že je co do počtu tvarů daleko nejpočetnější skupina s nejnižší frekvencí, druhou nejpočetnější skupinou jsou partikule s nejvyšší frekvencí. Ve středním pásmu — frekvence vyšší než 100 výskytů a nižší než 1000 — je počet tvarů více méně vyrovnaný. Tato tendence pokračuje — u dalších 80 % všech doložených tvarů majících frekvenci nižší než 100 výskytů je frekvenční distribuce doložených lemmat ve středním pásmu srovnatelná, avšak daleko nejvíce zastoupenou frekvencí je frekvence 1. Nejvíce partikulí (116) má frekvenci 1, což je téměř třetina (29 %) všech vyskytujících se tvarů partikulí a 66 % všech tvarů majících frekvenci nižší než 10.

Celkové proporční zastoupení lemmat z hlediska frekvence zobrazuje tabulka 4:

Frekvenční podíl partikulí s frekvencí vyšší než 100, vyšší než 10 a nižší než 10 výskytů	
> 100	20 %
> 10	23,75 %
< 10	56,25 %
	100,00 %

TABULKA 4: Procentuální podíl všech tvarů podle jednotlivých frekvencí.

V rámci frekvence 1 se vyskytují jak překlepy, je jich 15, což činí 3,75 % celkového materiálu. Mezi překlepy najdeme např. jak čtyřikrát velké počáteční písmeno⁵ či zřetelné překlepy typu *řeva*, *tyd'*, *přostě*, *žo*,⁶ tak různě individuálně nezřetelně vyslovené či zkomolené tvary (např. *oak*, *otak*). Mnohem důležitější však je, že se mezi

5 Dle zásad přepisu se velká písmena měla užívat jen pro propria a některé zkratky.

6 Při sestavování FSMČ jsme řadu evidentních překlepů odstranili: např. *řekněmě*, které je doloženo dokonce pětkrát, jsme opravili a opravený tvar přidali k lemmatu *řekněme*, nicméně některé překlepy zůstaly nepovšimnuty.



tvary majícími frekvenci 1 skutečně nacházejí lemmata, která se v celém korpusu vyskytla pouze jednou, čili hapax legomena. Je jich celkem 22 — 13 jednoslovných: *bodejť*,⁷ *bohudíky*, *cožpak*, *div*, *houby*, *koukám*, *nechtě*, *neřkuli*, *pranic*, *přinejlepším*, *vopravdicky*, *zaboha*, *zhola* a 9 víceslovných: *div že*, *hlavně aby*, *chceš nechceš* (a varianty *chca nechca*, *chtěj*, *nechtěj*), *nebo tak něco*, *než tak to radši*, *podle toho co*, *tak ať*.

Nyní uvedeme prvních 20 nejfrekventovanějších partikulí včetně víceslovných, které v tabulce 1 nebyly zobrazeny:

no	7986	taky	2257
jako	5479	vlastně	1784
tak	5033	to	1593
teda	4678	zase	1423
prostě	3924	i	1415
třeba	3242	ale	1396
a	2581	no tak	1144
že jo	2542	vůbec	1068
asi	2438	eště	1055
jo	2366	už	1016

TABULKA 5: 20 nejfrekventovanějších partikulí.

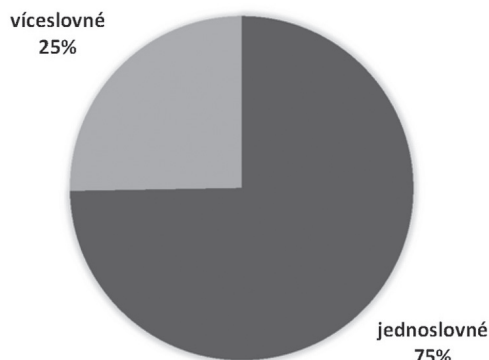
V seznamu najdeme pouze dvě víceslovné partikule, jejichž výskyt a frekvence jsou obecně mnohem nižší než výskyt a frekvence jednoslovných partikulí, jak ukážeme dále.

3.2 JEDNOSLOVNÉ A VÍCESLOVNÉ PARTIKULE

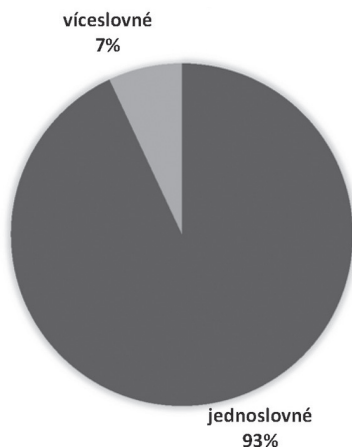
V materiálu PMK byly poprvé označeny též víceslovné partikule. Vůbec poprvé se tomuto gramaticko-lexikálnímu jevu soustavně věnoval F. Čermák ve IV. svazku *Slovníku české frazeologie a idiomatiky* pojmenovaném *Výrazy neslovesné*. Tento svazek obsahuje několik typů neslovesných frazémů, kromě jiného frazémy relační, které mají za základ složené výrazy gramatické povahy, jejichž funkce ve výpovědi však odpovídá jednoslovným výrazům synsémantickým. Tyto výrazy, kterými jsou obvykle ustálené kombinace komponentů, modifikují různým způsobem řídicí substantivum, verbum, popřípadě adjektivum či adverbium anebo také celou výpověď. Jen zčásti jde o výrazy čistě nociónální, většinou má jejich význam a celkové vyznění hodnotící anebo intenzifikační ráz. Z hlediska stylu je pro ně příznačná příznakovost a expresivita.

V našem materiálu se vlastních víceslovných partikulí vyskytlo 54, což je 13,5 % výskytu všech lemmat partikulí. Je však nutné zdůraznit, že se v materiálu kromě těchto výrazů vyskytuje ještě dalších 14 slovních spojení označených jako víceslovné partikule, jež však ve výpovědi mohou mít i jiné funkce — převážně se jedná o fra-

⁷ *Bodejť* se ještě vyskytlo jako součást frazému *bodejť by jo*, *houby* pak jako substantivum, interjekce (5×) nebo část frazému *to je na houby* (2×), *koukám* (17) jako verbum.



GRAF 3: Jednoslovné a víceslovné partikule — lemmata.



GRAF 4: Frekvence jednoslovných a víceslovných partikulí — tvary.

zémy, víceslovné/korelované spojky a adverbia. Poměr všech jednoslovných a víceslovných lemmat partikulí v PMK je tedy 201 : 68, jak znázorňují grafy 3 a 4.

Řada těchto výrazů se však vyznačuje velmi nízkou frekvencí, stejně jako zdaleka ne tak pestrou variantností tvarů, jakou jsme zaznamenali u jednoslovných partikulí. Pokud jde o kombinace, je zřejmé, že se většinou kombinují krátké spojky: *a, i, aby, ani, ať, dyž, dyby, jak, nebo, že* a partikule, případně adverbia: *eště, jen, možná, něk, přece, spíš, takle, už*. Je zajímavé, že některá slova (převážně jde o spojky a partikule — jedinou výjimkou je imperativ *dejme*) mohou stát jen na začátku spojení: *a* — 14× jen první pozice, *nebo* 4×, *ať, čím, ne* 3×, *ani, přece, že* 2×, *eště, hlavně, dejme, no, už, kdyby* jednou, některá zase jen na konci spojení (zde se kromě spojek a partikulí vyskytují také adverbia): *aby, kór, potom, přitom, takle, možná, jesli, spíš*. Část výrazů může však stát jak na začátku, tak i na konci spojení. Nejvyšší počet kombinací vykazují spojka *a* a partikule/adverbium *tak*. Je to bezesporu otevřená třída a nejspíš zde můžeme očekávat nárůst různých kombinací.



3.3 VARIANTY LEMMAT

Z celkového počtu 400 lemmat, respektive 385 po vyloučení překlepů (15), je doložených 269 původních lemmat (67,25 %) a 116 variant (29 %).

Nyní se blíže podíváme na varianty, jejichž podíl je skutečně nesmírně vysoký a tvoří téměř třetinu všech lemmat. Ve výše uvedené tabulce 2 jsou varianty pro lepší přehled označeny kurzivou. První variantu *nó* najdeme na 20. pozici, poslední *á* na pozici 188. V této tabulce, zobrazující zhruba čtvrtinu všech doložených tvarů, je variant poměrně málo (jen šestina všech tvarů). Podíváme-li se však na druhou půlku seznamu, počet variant pochopitelně rapidně stoupá, a tak můžeme říct, že čím vyšší pořadí seznamu, tím nižší frekvence a vyšší podíl variant. Některé tvary však za varianty považovat rozhodně nelze — může se jednat buď o zcela ojedinělý tvar v idiolektu konkrétního mluvčího, anebo spíše o různá přehnutí a nedbalou výslovnost. Znamená to, že „malým slovům“ při výslovnosti nevěnujeme přílišnou pozornost? Že jen o ně jaksi zavádíme a rychle se ve svém myšlenkovém pochodu ubíráme dál? Není-li pak jejich funkce pouze v jakémsi podvědomém zapojení opory (berličky) při hledání adekvátního slova, přesnějšího vyjádření a pevnějšího uchopení myšlenky?

Varianty lze dále členit na základě dvou hledisek: ustálenosti, uzuálnosti (určené frekvencí a území; toto hledisko je rozhodující) na ustálené vs. neustálené (nahodilé, individuální) a z hlediska druhu na fonetické vs. lexikální. Obě hlediska se vzájemně prolínají. Za ustálené varianty, někdy s mírným stylovým rozdílem, můžeme považovat *přece* vs. *přeci*, *nikoli* vs. *nikoliv*, *taky* vs. *také*. Ustálená varianta se může natolik osamostatnit, že ji lze také považovat za samostatné lemma: ve FSMČ jsou jako samostatná lemmata označeny *aspoň* a *alespoň*, *taky* a *také*, *jen* a *jenom*, *jo* a *jó* (další analýza jejich samostatnosti potvrdí či vyvrátí), jako varianty lemmatu se posuzovaly např. *nikoli* a *nikoliv*, *dnes* a *dneska*, *teprve* a *teprv*, *tedy* a *teda*, *no* a *nó*. Domnívám se ale, že jako o samostatných lemmatech lze uvažovat především o tvarech *teda* a *tedy*, které, na základě mých dosavadních pozorování, nejsou ve všech kontextech zaměnitelné. Jako hlavní kritérium při určování samostatnosti lemmatu je dle mého názoru nutné uplatnit hledisko funkčně-sémantické čili kontextové. Druhé hledisko — druh varianty — bere v potaz její původ: buď vychází z obecněčeského hláskosloví a běžné obecněčeské výslovnosti typické pro konsonantické shluky (varianta čistě fonetická), anebo jde o charakteristickou variantu pro to které slovo. V prvním případě mám na mysli *dyž* vs. *když*, *dyť* vs. *vždyť*, *vopravdu* vs. *opravdu*, v druhém jde o řadu ustálených, uzualizovaných variant, např. *teď*, *teďka*, *teďko*, *teďkon*, *teďkonc* nebo *hned*, *hnedka*, *hnedle*, kde varianty z hlediska obecněčeských fonetických zákonitostí nejsou nikterak odvoditelné a předvídatelné.

Uvedme nyní frekvenci některých dublet (zvýrazněné slovo je frekventovanější variantou).

aspoň — *alespoň*: 313 : 11, *dyť* — *vždyť*: 387 : 7, *eventuelně* — *eventuálně*: 12 : 1, *nikoliv* — *nikoli*: 20 : 9, *obzvlášť* — *vobzvlášť*: 17 : 2, *prej* — *prý*: 158 : 6, *přece* — *přeci*: 174 : 34, *přece jenom* — *přeci jenom*: 158 : 51, *spíš* — *spíše*: 615 : 14, *taky* — *také*: 2257 : 65, *teda* — *tedy*: 4678 : 406, *teprve* — *teprv*: 108 : 25.



3.3.1 LEMMATIZACE SLOVNÍCH TVARŮ

Při značkování korpusu anotátoři používali značkovací program EDITIN (Savický, 1994), který při lemmatizaci korpusu nabízel tři kroky: 1. upravený textový tvar umožňoval opravit případné evidentní překlepy, 2. zavést stylové lemma, což je pak převedení tvaru na základní, standardizovanou normu (infinitiv u slovesa, 1. pád u jmen, pozitiv u adjektiv a adverbíí, nevolalizovaný tvar u předložek), a 3. dodat pro srovnání spisovné lemma, které reprezentuje základní tvar ve spisovné podobě, což je důležité zvláště jako společný jmenovatel umožňující souhrnné statistiky všech stylů dohromady. V případě partikulí se spisovná podoba uváděla jen u některých variant (např. *obzvlášť*, *tedy*, *teď* apod.). Některé tvary jako například *furt* spisovnou podobu nemají (spisovná synonyma se zásadně nepoužívala), a tak se jako spisovné lemma uvádí tento tvar.

1. upravený textový tvar	především	furt	tedy	tea	řikali bysme	abysme
2. stylové lemma	především	furt	tedy	teda	řikat	abysme
3. spisovné lemma	především	furt	tedy	tedy	říkat	abychom

TABULKA 6: Příklady lemmatizace tvarů.

Slovníkové lemma představuje pak souhrn více frekventovaných tvarů lemmat. Například tvar bez kvantit *především* se vyskytuje stejně často jako *především*, avšak po zohlednění 6. pozice kódovníku — stylová charakteristika slova (viz níže), která se v poslední fázi určování výsledné podoby lemmatu stala rozhodujícím třídícím kritériem odrážejícím celkový styl promluvy, se varianta bez kvantit stala slovníkovým lemmatem. U řady dalších tvarů byla situace mnohem jednodušší: *teda* se vyskytuje 10× častěji než *tedy*, *prej* 26× častěji než *prý*, a tak se jako slovníkové lemma — standardní tvar mluveného jazyka — v těchto a dalších podobných případech uvádějí lemmata *především*, *teda*, *prej*.

3.4 PARTIKULE PODLE KÓDOVNÍKU PMK

Kódovník, na jehož základě proběhla ruční anotace celého materiálu, pro slovní druh částice obsahuje 6 pozic: na 1. pozici se značí slovní druh, tedy partikule, na 2. pozici *druh partikulí* (bližší určení jejich původu), na 3. pozici *třída* (informace o jejich funkcích, případně jejich sémantických vlastnostech), na 4. pozici *valence* (respektive místo výskytu částice v promluvě), na 5. pozici *modus věty* (oznamovací, tázací, zvolací), v níž se partikule objevuje, a na 6. pozici *stylová charakteristika* partikulí (odráží stylovou charakteristiku toho kterého slova s přihlédnutím k celkovému vyznění promluvy). První a šestá pozice jsou pro všechny slovní druhy stejné. Nyní se ještě blíže podíváme na druhou pozici v kódovníku.



3.4.1 DRUH PARTIKULÍ

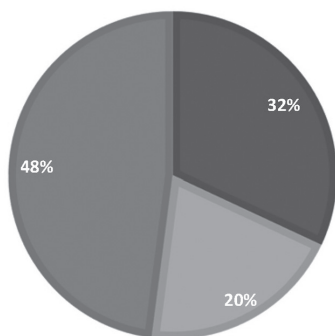
Na druhé pozici, jak je patrné z níže uvedeného seznamu, rozlišujeme 5 podtříd: 1. vlastní, nehomonymní partikule: *ať, asi, prej*, 2. partikule homonymní s adverbem: *jistě, klidně, stejně, dovopravdy, zvlášť, jen*, 3. partikule homonymní se spojkou: *však, tedy*, 4. jiné partikule, a to a) homonymní se substantivem nebo jeho předložkovým pádem: *fakt, celkem, například, dokonce*, b) homonymní se zájmenem nebo jeho předložkovým pádem: *což, copak, vono (von, voni), to, ovšem*, c) homonymní se slovesným tvarem: *myslím, prosím*, d) mající formu adverbia, ale nehomonymní s adverbem užívaným jako deadjektivní adverbiale: *hlavně, vlastně*, e) homonymní s citoslovce nebo citoslovečným frazémem: *vid', víš, bohužel*, f) jinojazyčného původu: *holt, kór*, 5. partikule víceslovné: *že jo, dejme tomu, no tak, nějak tak, (k)dyž tak, když už, přece jenom, hlavně že*.

Vyjdeme-li z tohoto popisu, z celkového počtu 269 stylových lemmat se vyskytlo jen 20 vlastních partikulí (např. *asi, taky, vůbec, spíš, právě, snad, dyť, aspoň, totiž, přece, prej, teprve*), což je jen o něco málo přes 7 % zkoumaného materiálu. Uplatnili-li však kontextové kritérium, nehomonymních jednoslovných partikulí se vyskytlo mnohem více — 85. Všechných dalších 65 lemmat partikulí, které jsou svým původem sice obecně homonymní s dalšími slovními druhy, se však v kontextu pořízených promluv vyskytlo pouze ve funkci partikule, a tudíž v našem materiálu nejsou homonymní co do své funkce s dalšími slovními druhy, a tak bychom z hlediska úzu, o který nám hlavně jde, na ně měli nazírat jako na partikule vlastní. Jsou to např. partikule *bohužel, bohudík, celkem, dokonce, dovopravdy, hlavně, jakoby, jaksi, každopádně, konec-konců, kór, kupodivu, mimochodem, obzvlášť, popřípadě* a další. V tomto bodě se tedy rozcházíme s kódovníkem.

Do této klasifikace se promítá hned několik rovin: sémantika, etymologie, forma, funkce a úzus (pragmatika), přičemž se v některých bodech některé roviny prolínají. Funkčně-sémantické hledisko se vztahuje k bodu 2, funkční k bodu 3, bod 4 je z tohoto hlediska nejpestřejší: na sprežky z bodů a) a b) se uplatňuje hledisko etymologické, v bodě d) dominuje formální hledisko, v bodě e) zase funkční, zatímco v bodě f) etymologie. Pragmatika se samozřejmě týká všech bodů a je rozhodujícím kritériem, na jehož základě můžeme s jistotou určit, zda se jedná o partikuli, anebo o některý z dalších homonymních slovních druhů. Pro naši analýzu je právě toto hledisko klíčové.

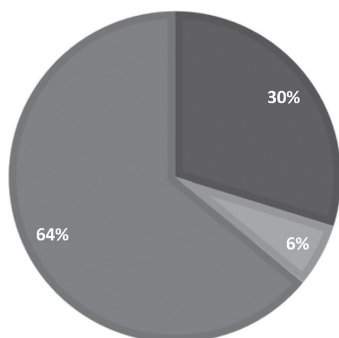
K přeražení zhruba dalších dvou třetin partikulí do kategorie „vlastní“ nás tedy vede skutečnost, že vycházíme pouze z funkcí a užití partikulí skutečně doložených v PMK (např. *bohužel se* ve funkci interjekce vůbec nevyskytuje). Dále se domníváme, že ze synchronního hlediska homonymie některých z výše zmíněných slov má ryze formální charakter (např. homonymie se substantivy u *dokonce, naštěstí, například*), a tudíž není pro naše účely prvoplánová. Tímto synchronním přístupem oprostěným od etymologického anebo diachronního hlediska značně rozšiřujeme seznam vlastních partikulí. Tento posun, který především závisí na kontextu a pragmatice, je bezesporu důkazem toho, že se třída partikulí pozvolna rozrůstá. Dalo by se zde právem namítnout, že „nehomonymnost“ některých z výše zmíněných partikulí „platí“ jen pro materiál doložený v PMK. Ano, lze připustit, že je to pravda. Budeme-li se však držet materiálu, musíme zároveň uznat, že jde o velmi výraznou tendenci, kterou je

■ Vlastní jednoslovné ■ Vlastní víceslovné ■ Homonymní



GRAF 5: Frekvence lemmat vlastních a homonymních partikulí.

■ Vlastní jednoslovné ■ Vlastní víceslovné ■ Homonymní

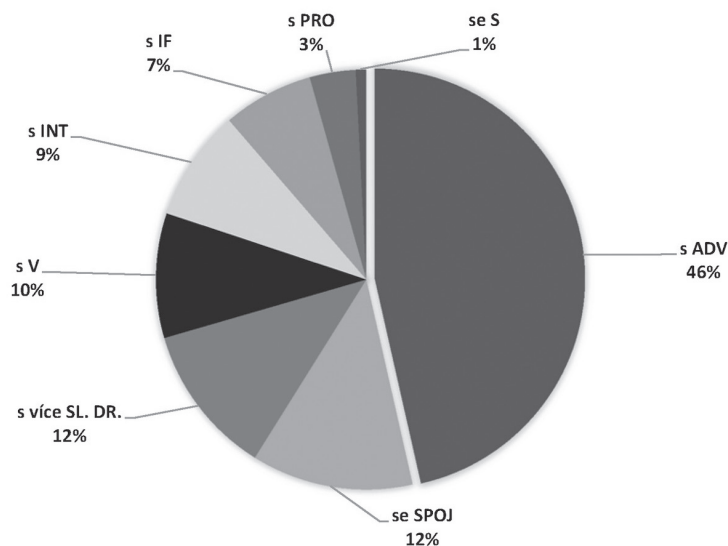


GRAF 6: Frekvence tvarů vlastních a homonymních partikulí.

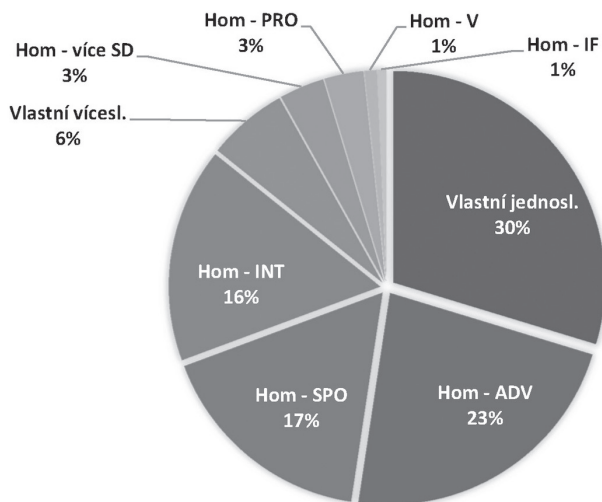
třeba jasně pojmenovat. Kontextové hledisko je rozhodující a mělo by mít poslední slovo. Zastáváme-li toto hledisko, pak mezi nehomonymní partikule musíme zařadit ještě dalších 9 hapaxů: *bohudíky*, *cožpak*, *div*, *neřkuli*, *pranic*, *přinejlepším*, *vopravdicky*, *zaboha* a *zhola* a několik dalších víceslovných hapaxů zmíněných výše.

Dále bych se ještě chtěla zastavit u termínu „vlastní“ ve smyslu „původní“, který zde asociuje, či spíše přímo koresponduje s termínem vlastní předložky, tedy předložky původní. Pokud jsme třídu vlastních partikulí rozšířili o řadu příkladů doložených materiálem, v tomto rozšířeném přístupu k této třídě bych pro ni spíše navrhla termín primární, tj. partikule nehomonymní z hlediska užití. Pak se přímo nabízí i termín sekundární, tj. partikule homonymní s dalšími slovními druhy. Mezi vlastními (či primárními) partikulemi najdeme tedy i partikule *bohužel*, *dovopravdy*, *naštěstí*, *dokonce*, *ovšem*, *vlastně*, *halt*, *kór*, *teda/tedy*, *však* apod.

Grafy 5 a 6 znázorňují frekvenci lemmat a tvarů partikulí v kontextu promluvy. Zhruba polovina lemmat a dvě třetiny tvarů partikulí doložených v PMK jsou homonymní s lemmaty a tvary dalších slovních druhů.



GRAF 7: Frekvence lemmat partikulí homonymních s dalšími slovními druhy.



GRAF 8: Frekvence tvarů jednotlivých druhů partikulí.

Dále jsme zkoumali podíl partikulí homonymních s dalšími slovními druhy. Ze všech typů partikulí homonymních s dalšími slovními druhy jsou na prvním místě partikule homonymní s adverbii.⁸ Je to zároveň po vlastních jednoslovných partikulích druhá nejpočetnější skupina, vyjdeme-li z celého materiálu (23 %).

⁸ K tomu detailněji Adamovičová (2014).

Na závěr se ještě podíváme na frekvenci tvarů jednotlivých druhů partikulí. I zde se v promluvě nejčastěji vyskytují vlastní jednoslovné partikule. Z partikulí homonymních s dalšími slovními druhy jsou nejvíce zastoupeny partikule homonymní s adverbii, spojkami a citoslovci. Součet tvarů těchto čtyř skupin dohromady představuje 86,5 % výskytu všech typů partikulí v promluvě. Za výraznější skupinu lze ještě považovat vlastní víceslovné partikule (6,16 %), zatímco poslední čtyři typy — partikule homonymní zároveň s více slovními druhy, dále se zájmeny, verby a frazémy, představují co do frekvence svých tvarů spíše marginální skupinu.

4 ZÁVĚR

Z tohoto krátkého nástinu výskytu partikulí v PMK plyne hned několik závěrů. Partikule jsou nesmírně frekventovaným slovním druhem, kterému doposud nebyla věnována patřičná a systematická pozornost ve světle nových možností, jež poskytuje korpusové zpracování dat. Nyní se poprvé na partikule můžeme dívat v autentickém kontextu, tam, kde se přirozeně vyskytují a kde často i vznikají. Detailní popis výskytu a souvýskytu partikulí, jejich funkce a pozice v promluvě pomůže však při jejich lepší klasifikaci, která by se měla odrazit jak v lépe zpracovaných trénovacích souborech pro automatickou anotaci, tak i v lexikografické praxi a v neposlední řadě také při výuce češtiny jako cizího jazyka.

Všechny tyto tři oblasti jsou mimořádně důležité. Automatická anotace partikulí má zatím velmi nízkou úspěšnost, zpracování tohoto slovního druhu v příručkách a slovnících je doposud nedostačující, s čímž přímo souvisí i jeho nedostatečné a nysystematické zapojování do učebnic češtiny pro nerodilé mluvčí a obecně zřejmě i do výukového procesu. Nemám přitom na mysli jen jazyk mluvený, nýbrž i jazyk psaný. Vyčerpávající popis kontextového výskytu partikulí a jejich klasifikace by měly zaplnit všechny výše uvedené mezery.

LITERATURA:

- ADAMOVIČOVÁ, Ana — ČERMÁK, František — PEŠIČKA, Jiří (2008): *Čítanka mluvené češtiny*. Praha: Karolinum.
- ADAMOVIČOVÁ, Ana (2014): Partikule vs. adverbia — rozpoznání slovního druhu. In: Vladimír Petkevič — Ana Adamovičová — Václav Cvrček (eds.), *Radost z jazyků: Sborník příspěvků k 75. narozeninám prof. Františka Čermáka*. Praha: Nakladatelství Lidové noviny, s. 188–200.
- BAUER, Jaroslav — GREPL, Miroslav (1972): *Skladba spisovné češtiny*. Praha: Státní pedagogické nakladatelství.
- BONIFAZI, Anna — DRUMMEN, Annemieke — DE KREIJ, Mark (2016): *Particles in Ancient Greek Discourse* [online]. Boston, MA: Center for Hellenic Studies, Harvard University. Cit. 12. 6. 2017. Dostupné z WWW: <<https://kleos.chs.harvard.edu/?p=5982>>.
- ČERMÁK, František (2008): Partikule, jejich syntagmatika a kumulace v mluvené češtině. In: Marie Kopřivová — Martina Waclawičová (eds.), *Čeština v mluveném korpusu*. Praha: Nakladatelství Lidové noviny, s. 63–74.
- ČERMÁK, František (2011): *Jazyk a jazykověda: Přehled a slovníky*. Praha: Karolinum.



- DE SAUSSURE, Ferdinand (1996): *Kurs obecné lingvistiky*. Praha: Academia.
- ESČ: KARLÍK, Petr — NEKULA, Marek — PLESKALOVÁ, Jana (eds.) (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- FSMČ: ČERMÁK, František et al. (2007): *Frekvenční slovník mluvené češtiny*. Praha: Karolinum.
- GREPL, Miroslav (2011): Co děláme, když mluvíme. In: Miroslav Grepl, *Jak dál v syntaxi*. Brno: Host, s. 61–106.
- HLAVÁČKOVÁ, Dana — OSOLSOBĚ, Klára (2008): Morfologické značkování mluvených korpusů, zkušenosti a otevřené otázky. In: Marie Kopřivová — Martina Waclawičová (eds.), *Čeština v mluveném korpusu*. Praha: Nakladatelství Lidové noviny, s. 105–114.
- JELÍNEK, Jaroslav — BEČKA, Josef Václav — TĚŠITELOVÁ, Marie (1961): *Frekvence slov, slovních druhů a tvarů v českém jazyce*. Praha: Státní pedagogické nakladatelství.
- KOPŘIVOVÁ, Marie (2006): *Valence českých adjektiv*. Praha: Nakladatelství Lidové noviny.
- MČ2: KOMÁREK, Miroslav — KOŘENSKÝ, Jan — PETR, Jan — VESELKOVÁ, Jarmila (eds.) (1986): *Mluvnice češtiny: 2, Tvarosloví*. Praha: Academia.
- PMČ: KARLÍK, Petr — NEKULA, Marek — RUSÍNOVÁ, Zdenka (eds.) (1995): *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.
- ULIČNÝ, Oldřich (1986): Částice. In: Miroslav Komárek — Jan Kořenský — Jan Petr — Jarmila Veselková (eds.), *Mluvnice češtiny: 2, Tvarosloví*. Praha: Academia, s. 228–238.
- VONDRÁČEK, Miloslav (1998): Citoslovce a částice — hranice slovního druhu. *Naše řeč*, 81(1), s. 29–37.
- WITTGENSTEIN, Ludwig (1993): *Filosofická zkoumání*. Praha: Filosofia.

Ana Adamovičová | Ústav bohemistických studií FF UK
 <ana.adamovicova@ff.cuni.cz>