

Univerzita Karlova v Praze

Filozofická fakulta

Ústav Českého národního korpusu

Matematická lingvistika

Disertační práce

Zkoumání hierarchizace pádů českého substantiva

v sémantických (kolokačních) třídách

Vedoucí práce: Prof. PhDr. Karel Kučera, CSc.

2015

Karin Schöne

Poděkování

Nejprve bych ráda poděkovala vedoucímu disertační práce profesoru Karlu Kučerovi za vedení a konzultace, poskytnutí cenných připomínek a za čas, který mé práci věnoval.

Velké poděkování patří rovněž mé kolegyni Mgr. Janě Škodové za jazykové korektury a připomínky, ale také za její morální podporu.

Děkuji v neposlední řadě spolupracovníkům Ústavu Českého národního korpusu a rovněž své rodině za projevenou důvěru a trpělivost.

Karin Schöne

Prohlášení:

Prohlašuji, že jsem disertační práci napsala samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne

[vlastnoruční podpis]

.....

Karin Schöne

Abstrakt

Disertace si klade za cíl zjišťovat a analyzovat frekvenční distribuci substantivních pádových tvarů v současné češtině. Práce vychází z analýzy textů nerodilých mluvčích češtiny a jejich frekventovaných chyb při produkci substantivních tvarů. Na pozadí poznatků badání v oblasti osvojování jazyka na jedné straně a zásad komunikační metody na druhé straně jsou analyzovány zavádění a explikace pádů, jejich funkcí a znázornění pomocí typických konstrukcí v devíti učebnicích češtiny jako cizího jazyka. Práce dochází v tomto ohledu k závěru, že je potřeba, aby se funkce pádu vyjasnila na typických konstrukcích a na dostatku konkrétních substantiv, pro něž jsou tyto funkce typické. Celkově si syntagmatická stránka substantiva zaslouží větší pozornost oproti pohledu na substantivum primárně jako na součást deklinačního paradigmatu. S tím se spojuje jedna z výchozích otázek, které slovní tvary substantivního lexému jsou závažná pro jeho přiměřenou prezentaci v didakticky orientovaném popisu. Pilotní studie ukazuje na datech Českého národního korpusu, že se frekvence pádových tvarů pro konkrétní lexémy významně liší od průměrného rozložení pádových tvarů v korpusu. V omezeném rámci pilotního vzorku se analyzuje frekvence pádových tvarů jednotlivých lemmat i na pozadí její závislosti na funkčním stylu a na sféře komunikace. Vlastní studie na náhodném vzorku o rozsahu 203 lexémů, jež považujeme za relevantní pro úroveň B1 Společného evropského referenčního rámce pro jazyky, poznatky pilotní studie potvrzuje. Lexémy se třídí do skupin podle dvou nejfrekventovanějších morfologických tvarů lemmat. Pomocí těchto skupin (tedy morfologických profilů substantiv) lze zachytit substantiva v sedmi vysoko frekventovaných, osmi středně a devíti nízko frekventovaných typech. Ukazuje se, že jednotlivé morfologické profily mohou být typické pro konkrétní sémantickou třídu substantiv. Pro některé zástupce vzniklých typů jsou vyhodnoceny frekventované kolokáty za účelem zjistit typické kolokační vzorce. Zároveň se zkoumá, do jaké míry tyto profily zrcadlí charakteristické rysy sémantických kategorií substantiv, jako abstrakta, konkréta, názvy osob, živočichů a míst. Analýza se provádí ve vyváženém stamilionovém korpusu současné psané češtiny SYN2010 a je založená na metodologii korpusové lingvistiky.

Klíčová slova: analýza chyb, český jazyk, čeština jako cizí jazyk, kolokace, korpusová lingvistika, morfologie substantiva, pád, sémantická klasifikace substantiv

Abstract

This dissertation aims to identify and analyse the frequency distribution of morphological forms of nouns in contemporary Czech. The work is based on an error analysis of texts of non-native speakers revealing and describing frequent types of errors in noun inflection. Against the background of recent research in the field of second language acquisition on the one hand and the principles of communicative approaches to language teaching on the other hand nine textbooks of Czech as a foreign language are analysed for approaches to the introduction and presentation of grammatical cases, their functions and typical constructions. It is concluded that it is necessary to exemplify case functions by prototypical constructions and sufficient sample nouns for which these functions are typical. With regard to the often observed emphasis on the noun as a part of the inflectional paradigm and thus on the nominative singular as its basic form, it is eligible for textbooks and language classes to shift the focus from the dominant paradigmatic to the syntagmatic aspect of nouns and their word forms. One of the initial questions of these theses therefore is which forms of a noun are essential to the adequate (didactically oriented) presentation of the noun and its use. A pilot study conducted on the data of the Czech National Corpus reveals that the frequency of case forms for specific lexemes differs significantly from the average distribution of case forms in the corpus. The small pilot sample is also analysed for the frequency distribution of the lemmas' case forms across different domains and types of communications. The actual study carried through on a random sample of 203 lexemes considered to be relevant to the level B1 of the *Common European Framework of Reference for Languages* confirms the findings of the pilot study. Lexemes can be classified according to their two most frequent morphological forms (their morphological profiles) and thus be quantified in seven highly frequent, eight medium frequent and nine low frequent types. It turns out that different morphological profiles may be specific to a particular semantic noun class (e.g. nouns denoting languages, nationalities, towns, or subject matters). For some exponents of the incurred morphological profiles their frequent collocations are evaluated in order to explore typical collocational patterns. Finally, we examine to what extent the profiles reflect characteristics of semantic categories of nouns, such as abstract and concrete nouns, names of human beings, animals, and places. The Analysis is

conducted on the data of a synchronic representative corpus of written Czech comprising 100 million tokens.

Key words: case, collocation, corpus linguistics, Czech language, Czech as a foreign language, error analysis, noun morphology, semantic classification of nouns

Obsah

1 Úvod.....	8
2 Osvojování si morfologických forem a jejich prezentace v učebnicích	11
2.1 Osvojování si morfologických forem: faktory a překážky	11
2.1.1 Popis rozvoje jazykového systému založeného na úzu	12
2.1.2 Vliv frekvence tvarů a konstrukcí na jejich kategorizaci	13
2.1.3 Vliv frekvence na rychlost zpracování forem.....	15
2.1.4 Další faktory	16
2.1.5 Shrnutí se zřetelem k češtině	19
2.2 Typy chyb jinojazyčných studentů českého jazyka ve skloňování	20
2.2.1 Cíle analýzy chyb.....	20
2.2.2 Výběr textů	20
2.2.3 Rozsah analýzy	21
2.2.4 Postup anotace chyb.....	22
2.2.5 Vyhodnocení	26
2.2.6 Shrnutí.....	31
2.3 Prezentace substantivních tvarů v učebnicích	32
2.3.1 Role gramatiky v komunikační metodě	32
2.3.2 Prezentace pádů a pádových forem v učebnicích	37
2.3.3 Výběr substantiv a jejich forem	46
2.3.4 Shrnutí.....	48
3 Korpusový přístup k popisu substantiv.....	49
3.1 Korpusový přístup ke gramatice	49
3.2 Forma a význam, kontext a význam	52
3.3 Paradigmatické a syntagmatické vztahy substantiva	55
3.4 Shrnutí.....	58
4 Frekvence pádů v současné češtině	60
4.1 Frekvence substantivních morfologických kategorií v češtině.....	60
4.1.1 Mluvená a psaná čeština	61
4.1.2 Funkční styly psané češtiny	65
4.2 Frekvence syntaktických funkcí	69
4.3 Shrnutí.....	71
5 Metodický postup a data	73
5.1 Metodický postup a pilotní studie.....	73
5.1.1 Úvod.....	73
5.1.3 Pilotní studie	74
5.1.4 Závěry z pilotní studie	81
5.2 Užité korpusy	82

5.3 Výběr substantiv a vytvoření náhodného vzorku	86
5.3.1 Zjišťování vyjadřovacích potřeb studentů češtiny.....	86
5.3.2 Výběr substantivních výrazů	88
5.3.3 Úprava výběru substantiv na bázi frekvenční distribuce	91
5.3.4 Charakteristika základního souboru.....	96
5.3.5 Vytvoření náhodného vzorku.....	104
5.4 Problémy kvantifikace morfologických tvarů	105
5.4.1 Víceznačnost tvarů.....	105
5.4.2 Víceznačnost lexémů a víceslovné lexémy	106
5.4.3 Vytvoření morfologických profilů.....	107
6 Vyhodnocení	109
6.1 Typy profilů - obecně	109
6.2 Roztřídění substantiv	110
6.3 Frekventované typy morfologických profilů	112
6.3.1 Typ <S4-S2>.....	112
6.3.2 Typ <S4-S1>.....	119
6.3.3 Typ <S2-S6>.....	120
6.3.4 Typ <S1-S2>.....	123
6.3.5 Typ <S6-S4>.....	124
6.3.6 Typ <S1-S6>.....	125
6.3.7 Typ <S1-P1>.....	126
6.3.8 Okrajové jevy.....	127
6.4 Věci abstraktní a konkrétní	129
6.5 Místa, osoby a živočichové.....	131
6.5.1 Kategorie „člověk“ (Hum).....	131
6.5.2 Kategorie „živočich“ (ANIM)	132
6.5.3 Kategorie „místo“ (LOC).....	132
6.6 Morfologický profil substantiv v mluveném jazyce	134
7 Závěr	136
7.1 Souhrn výsledků korpusové analýzy a metodické aspekty.....	136
7.2 Didaktické aspekty prezentace substantivních tvarů	142
Literatura.....	146
Příloha.....	157
1. Výběr substantiv	157
2. Základní soubor a náhodný vzorek	170
3. Vyhodnocení dat	180

1 Úvod

Předkládaná práce si klade za cíl zjišťovat a analyzovat frekvenční distribuci substantivních pádových tvarů v současné češtině. Záměrem studie je identifikovat typické užití jednotlivých lemmat pomocí dat z Českého národního korpusu. Analýzu provádíme na náhodném vzorku substantiv, jež byla vybírána z didaktických hledisek pomocí popisu referenční úrovně B1 *Společného evropského referenčního rámce pro jazyky*, a rovněž z hlediska pozice substantiv ve frekvenčním slovníku¹. Na základě nejfrekventovanějších tvarů jednotlivých substantiv jsou vytvářeny profily užití, tzv. morfologické profily, jež nám umožňují srovnávat a třídit substantiva na základě společné podobné frekvenční distribuce pádových tvarů. Pro některé zástupce vzniklých tříd jsou vyhodnoceny jejich frekventované kolokáty za účelem zjistit typické kolokační vzorce.

Námět práce vychází z analýzy textů nerodilých mluvčích češtiny a jejich frekventovaných chyb při produkci substantivních tvarů, a má tudíž především didaktické pozadí. Chybová analýza ukazuje, že v 80 procentech nesprávného užití morfologické substantivní koncovky studenti volí tvar patřící sice ke stejnému paradigmatu, ale k jinému pádu. Nejčastěji se v takových případech vybírá nominativ singuláru², což je signálem jeho dominantního postavení v učebnicích a ve výuce. Navzdor tomu mohou však být i tvary akuzativu nebo lokálu upřednostňovány v určitých syntaktických konstrukcích, zejména v případě volného doplnění k určení směru nebo místa.

Český deklinační systém připravuje pro cizince učícího se češtině jako cizímu jazyku řadu překážek, zejména homonymie pádových formantů a rozmanitost deklinačních vzorů. Poznatky bádání v oblasti osvojování jazyka ukazují, že učení je pevně spjato s vyhodnocováním informací o frekvenci a distribuci forem v textu. Velký význam pro zapamatování a zobecňování konstrukcí mají jejich prototypičtí zástupci (srov. např. Goldberg a Casenhiser, 2008). Přitom je jedním z předpokladů úspěšného učení se (a osvojování) nových konstrukcí, k nimž připočítáme i užití substantiv v konkrétních syntaktických funkcích, dostatek inputu a konkrétních příkladů, jež by měly být k dispozici již i v rané fázi osvojovacího

¹ Referencí je *Frequency dictionary of Czech. Core vocabulary for learners* (Čermák a Křen, 2011).

² Přičemž u slov vzorů *hrad* a *město* je totožný s akuzativem.

procesu a kterými by se studenti měli vědomě zabývat i v pozdější fázi procesu učení se (srov. k tomu Goldberg a Casenhiser, 2008 a Ellis, 2014).

Frekvence slovních tvarů jako kritérium výběru slov a jejich tvarů se proto přikládá v oblasti SLA³ velkou závažnost. Není ovšem jednoduchým úkolem vyhovět při výběru substantiv pro učebnice a kurikula jak funkčnímu hledisku, tak smysluplnému uspořádání pádů a zároveň frekvenci substantivních lexémů a tvarů. V učebnicích, jež se v této práci analyzují vzhledem k prezentaci pádů obecně a substantivních tvarů konkrétně, narážíme často na dominantní postavení paradigmatického pohledu na substantiva. Při vysvětlování substantivních morfologických paradigmát pádů a jejich funkcí v příslušných cvičeních je nápadný právě nedostatek příkladových konstrukcí. Považujeme proto za důležité, aby se syntagmatické stránce substantiva a jeho frekventovaným tvarům v učebních materiálech přidělilo více prostoru. Diskutujeme proto v této práci roli frekvence substantivních tvarů a jejich kolokací při výběru a prezentaci substantivních lexémů.

Zvolený přístup k popisu substantivních lexémů a jejich tvarů se záměrně opírá o velké množství dat, jež mohou dokládat chování lexémů v textu, a tudíž jejich užití. Vyhodnocování distribuce tvarů lexémů nám umožňuje činit závěry nejen o deklinačním systému v jeho úplnosti, ale především o chování individuálních lexémů a o podobnostech mezi nimi. Analýza je v tomto smyslu založená na metodologii korpusové lingvistiky a snaží se o postup řízený korpusem.

Substantivní tvar se tradičně vnímá jako součást systému, jako nositel morfologické koncovky vyjadřující gramatický rod, pád a číslo. Korpusovým přístupem k substantivu ho však zachycujeme nejen jakou součást paradigmatu lexému, ale také jako část textu, zasazený do konkrétního kontextu - tudíž v jeho pozici na rozhraní systému a textu, lexikonu a syntaxe. Na tomto pozadí se tato práce pokouší o klasifikaci substantivních lexémů na základě jejich nejfrekventovanějších morfologických tvarů. Diskutuje se přitom, jaké frekvenční spektrum tvarů nejlépe reprezentuje typické užití lexému a může být považováno za součást jeho morfologického profilu. V neposlední řadě se práce zabývá otázkou, do jaké míry odrážejí jednotlivé typy morfologických profilů sémantické vlastnosti substantiv nebo vlastnosti známých sémantických tříd. Při vyhodnocování se proto přihlíží

³ Second language acquisition = osvojování (prvního) cizího jazyka.

ke klasifikaci substantiv podle povahy denotátu navržené Čermákem (2010, str. 165n.)⁴.

Frekvenční analýza substantivní morfologie byla provedena již v osmdesátých letech (Těšitelová 1985) a nyní jsou k dispozici data o četnosti jednotlivých pádů, jež se opírají o stamilionový korpus SYN2005 (srov. Bartoň et al., 2009). Tyto kvantitativní charakteristiky češtiny sice nepřihlížejí k jednotlivým lemmatům, ale zviditelňují rozložení pádů v různých funkčních stylech češtiny. Číselné přehledy zřetelně objasňují souvislost mezi funkčním stylem a frekvencí pádů. V omezeném rámci pilotního vzorku se proto v této práci analyzuje frekvence pádových tvarů jednotlivých lemmat i na pozadí její závislosti na funkčním stylu a na sféře komunikace.

⁴ Na věci abstraktní, věci konkrétní, názvy osob, zvířat a míst.

2 Osvojování si morfologických forem a jejich prezentace v učebnicích

2.1 Osvojování si morfologických forem: faktory a překážky

Otázka reprezentace gramatických tvarů v mozku, zapamatování a rychlost aktivizace flektivních forem při produkci jazykových projevů zaměstnává psycholinguisty a kognitivní lingvisty už několik desetiletí. Jejich studie zkoumají především zpracovávání jazyků rodilými mluvčími – jak dospělými tak dětmi – a osvojování struktur a forem dětmi, věnují se ale rovněž osvojovacím procesům dospělých studentů cizích jazyků.

V této části se chceme soustředit na faktory, jež ovlivňují osvojování gramatických forem. Zvláště nás zajímá, jaký dopad má frekvence tvarů (frekvence tokenů) v inputu na jejich zapamatování a co ovlivňuje rychlost zpracování forem. Uvedeme především poznatky kognitivní lingvistiky a bádání z oboru SLA⁵. V popředí stojí modely, které popisují systém žákovského jazyka⁶ jako výsledek vyhodnocování statistické informace z jazykového úzu. Popisu morfologických tvarů a jejich kolokací z hlediska syntagmatického i paradigmatického se budeme věnovat ve třetí kapitole.

Student cizího jazyka bezesporu není jazykově nepopsaným listem. Na rozdíl od dítěte dokáže dospělý mluvčí jazyk do jisté míry analyzovat a rozpoznat z jazykového proudu jednotlivá slova a určovat morfémy a jejich funkce. Na vysvětlení rozvoje systému druhého jazyka (L2) na základě či bez vlivu mateřského jazyka se rozvíjely různé koncepce (přijímány a hojně diskutovány byly zejména tzv. *contrastive hypothesis* a *identity hypothesis*⁷). Nejvlivnější koncepcí pro teorii SLA se stala Selinkerova koncepce *interlanguage* (*mezijazyk*). Tato koncepce předpokládá, že si student v procesu učení se dalšího jazyka vytvoří samostatný jazykový systém, přičemž mu pomůže geneticky daná struktura (*latent psychological structure*), která je však odlišná od struktury vznikající při přirozeném osvojování jazyka (Selinker, 1972). Selinker tak přesunul pozornost na prozatímní

⁵ Second Language Acquisition

⁶ Podle úzu v aktuální lingvodidaktické a akviziční literatuře používáme i my v této práci termíny *mezijazyk* a *žákovský jazyk* jako české ekvivalenty pojmu *interlanguage*. Srov. definice žákovského jazyka např. u Štindlové: „Žákovský jazyk je standardně v kontextu akvizice druhého/cizího jazyka nahlížen jako jazykový systém sám o sobě, tzv. mezijazyk, a měl by být analyzován včetně nekorektních struktur.“ (Štindlová, 2012: 36)

⁷ viz Klein (1990, s. 23 nn.)

system cílového jazyka, na pravidla a zásady, podle kterých ho student rozvíjí, a tedy na studentskou či žakovskou „představu“ o systému cílového jazyka⁸.

2.1.1 Popis rozvoje jazykového systému založeného na úzu

Psycholinguvistické bádání uplynulých padesáti let ukázalo, že zpracování jazyka je v mnoha hlediskách vázáno na zažitou či vnímanou frekvenci a distribuci jazykových jednotek. Vliv častosti výskytu lexikálních jednotek, jejich forem a kolokací na tvoření jazykových kategorií se chápe v mnoha modelech jazykového popisu jako nesporný.

Konstruktivní gramatika (*construction grammar*) zkoumá rozvoj systému mateřského jazyka (L1) už dlouho pomocí korpusů mapujících jazykové výstupy dětí po delší dobu. Studie založené na konstrukčních přístupech prokazují, že si děti i dospělí studenti dokážou osvojit slovesné struktury (*verb argument structure*) na bázi minimálních vstupních dat. Na základě konkrétně zažitých příkladů, ucelených konstrukcí, si tvoří dítě, resp. student obecnější, generalizovaný pojem konstrukce (*generalizations*). Konstrukční gramatika tím oponuje hypotéze univerzální gramatiky Chomského, že základní aspekty jazyka jsou do jisté míry vrozené, předspecifikované a nedají se učit (Goldberg a Casenhiser, 2008, str. 197). Gramatické fenomény v pojetí konstrukční gramatiky vznikají generalizací a kategorizací konkrétních případů. Všechny gramatické jevy je možné pochopit jako naučené kombinace forem s jejich sémantickou funkcí (Goldberg, 2006, str. 18).

Goldberg a Casenhiser citují různé studie založené na koncepci konstrukční gramatiky, podle nichž vysoká frekvence tokenů⁹, tedy častý výskyt konkrétních instancí určitých konstrukcí, může napomáhat tvoření generalizací. Studie prokazují, že cvičení, jenž prezentuje prioritně příklady podobných případů, podporuje učení se konstrukčním kategoriím. Prezentuje-li se však zároveň případy, které do příslušné kategorie nepatří, proces učení či osvojování se zpomaluje nebo se mu dokonce zabraňuje (Goldberg a Casenhiser, 2008, str. 210). Doporučení konstrukční gramatiky pro SLA se zakládají na domněnce, že vědomí o jednotlivých

⁸ Při dalších úvahách o odchylném užití morfologických forem nebo o jejich prezentaci jak v hlavě studenta, tak v učebnicích se o tuto koncepci opíráme i my.

⁹ Frekvence tokenů se podle Bybeeové vztahuje na výskyt konkrétní jednotky v textu. Za jednotku se považují jazykové prvky na různých úrovních komplexnosti: souhláska, slabika, slovo, kolokace, věty. (Bybee, 2008, s. 218)

konstrukcích či jednotkách (*item-specific knowledge*) koexistuje s vědomím o generalizacích (Goldberg, 2006, str. 63). Oboje se získává prostřednictvím analýzy inputu, pomocí učení se. Z četných studií zabývajících se osvojováním nových konstrukcí jak dětmi, tak dospělými Goldberg a Casenhiser vyvozují: „Collectively, they suggest that it is advantageous to supply targeted input that includes ample prototypical instances early in training.“ (Goldberg a Casenhiser, 2008, str. 210)

Je tedy potřeba konfrontovat žáka/studenta již v rané fázi osvojovacího procesu s frekventovanými a „prototypickými“ jednotkami a konstrukcemi z roviny *parole*, na jejichž základě si může jak zapamatovat konkrétní tvary a konstrukce, tak utvořit představu o jejich struktuře.

2.1.2 Vliv frekvence tvarů a konstrukcí na jejich kategorizaci

Konkrétní vliv frekvence tvarů a konstrukcí v jazykovém úzu na procesy gramatikalizace, vytváření gramatických kategorií a na ustálenost morfologických tvarů zkoumají přístupy ke gramatice založené na úzu (*usage-based*) (Bybee, 2002; Bybee, 2006). Konstrukcí se v těchto přístupech rozumí kombinace formy s významem uložené v paměti¹⁰. Konstrukce zahrnuje prvky všech jazykových rovin a zařazuje je na ose konkrétnost – abstraktnost či ustálenost - neustálenost. Na takové ose můžeme pochopit gramatické tvary buď jako část víceslovného pojmenování (např. v *souvislosti s ...*), jako část konstrukce připouštějící zaplnění mezer různými (sémanticky vymezenými) jednotkami (např. <*nastavit bratrovi / nepříteli / kolemjdoucímu nohu*>) nebo jako plně abstraktní konstrukce (např. <sloveso + přímý předmět v akuzativu>: <*mít strach | hlad | čas*>; nebo <sloveso + adverbialní výraz>: <*letět/jet + do*> vs. <*letět/jet + na*>)¹¹.

Osvojování gramatických forem a kategorizace jejich vzorů závisí podle Bybeeové nejen na vysoké frekvenci tokenů, ale také na frekvenci *typů* (*type*). Pojem

¹⁰ “Constructions are stored pairings of form and function that range over units at the level of word up to and including complex sentences.” (Bybee, 2008, str. 217)

¹¹ Příklady byly tvořeny podle modelu konstrukcí v Bybee, 2008, str. 217. Z hlediska lexikální syntagmatiky, kterou lze popsat metodami korpusové lingvistiky, lze tutéž problematiku popsat jako omezenou kolokabilitu lexémů. Podle Čermáka (2010, str. 295) je kolokabilita lexémů trojího typu: vysoká, střední a periferní. Kombinace lexémů gramatických se zdá být neomezená, ačkoliv ve skutečnosti neomezená není. Kolokabilitě lexémů se blíže věnujeme ve třetí kapitole.

type frequency se vztahuje na počet odlišných jednotek, jež reprezentují konstrukční vzor (Bybee, 2008, str. 218).

Diachronní studie prokázaly, že nepravidelné tvary sloves mají sklon připodobnit se pravidelným vzorům (Bybee 2006; Bybee 2008). Vysoká frekvence tokenů však může mít konzervační efekt: nepravidelné tvary mohou odolat vlivu produktivních tříd, vyskytují-li se často. Bybeeová uvádí příklad nepravidelných tvarů minulého času anglických sloves: *keep* -> *kept*; *sleep* -> *slept*, která zachovala iregulární tvary oproti slovesům stejného typu, u nichž se postupně vyvíjely či vyvíjejí tvary podle pravidelného modelu <kmen + *ed*>: *weep* -> *weaped*; *leap* -> *leaped* (Bybee, 2008, str. 219). V případě složených tvarů může jejich případná vysoká frekvence (*token frequency*) způsobit i to, že se časem přestávají vnímat jako konstrukce složené z jednotlivých částí, takže ztrácejí svoji transparentnost a nabývají autonomie (srov. příklad tvarů anglického složeného futura *be going to* Bybee, 2008, str. 219).

Dále platí, že vzory a konstrukce s vysokou frekvencí *typů* se s vyšší pravděpodobností přenášejí i na další nové jednotky, u nichž se daný vzor zatím neuplatňoval (Bybee, 2008, str. 221).

Zkušenost s frekvencí a distribucí jazykových jednotek a jejich struktur tedy řídí jejich kognitivní reprezentaci. Student, který chce úspěšně aplikovat určitý vzor (*pattern*), se musí setkat s určitým množstvím odlišných příkladů (Bybee, 2008, str. 221). To platí i u nepravidelných tvarů: studenti musí být vystaveni dostatečným množstvím příkladů, dostatečnému množství příkladů, umožňujícím zakotvení těchto tvarů v paměti.

Deklinační třídy mají charakter paradigmatický, nemají přímo funkci ve větě. Samostatně je proto zřejmě nelze považovat za konstrukce v daném smyslu. Tvoří však rámec pro jemnější třídění abstraktních konstrukcí, srov. např. <*jet do -y / -a / -e / -u ...*>. Podle výše popsaného principu lze očekávat, že čím více substantiv daný vzor reprezentuje, tím větší je pravděpodobnost, že se přenáší i na substantiva, s kterými se student v daném vzoru ještě nesetkal a která jsou z tohoto hlediska nová.

Při všech zmíněných principech však zůstává nedořešeno, kudy přesně probíhá hranice mezi vysokou, střední a nízkou frekvencí.

2.1.3 Vliv frekvence na rychlost zpracování forem

Další centrální otázkou kognitivní lingvistiky je, jaký vliv má frekvence morfologických tvarů v řeči na způsob jejich reprezentace a na aktivizaci tvarů při vyjadřování.

Model dvojitého přístupu (*Dual access model*) navrhuje, aby morfologické tvary byly zpřístupněny buď jako ucelené reprezentace (whole-word presentations), nebo po cestě dekompozice tvarů. O tom, jaký přístup se volí, rozhoduje frekvence. Alegreová a Gordon zkoumali v sérii experimentů, jaký dopad má frekvence pravidelných flektivních tvarů sloves na jejich reprezentaci. Výsledky výzkumu je vedly k předpokladu, že flektivní tvary s vysokou frekvencí jsou reprezentovány jako celek (Alegre a Gordon, 1999, str. 56); k podobným výsledkům došli před nimi i Burani, Salmaso a Caramazza (1984). Jakmile ovšem frekvence pravidelných tvarů klesá, přístup k tvarům cestou kompozice je spolehlivější (Alegre a Gordon, 1999, str. 56).

Častý výskyt tokenů, v tomto případě pravidelných morfologických tvarů, tedy napomáhá tomu, že tvary jsou k dispozici přímo jako celek. Naproti tomu méně frekventované pravidelné tvary jsou i u rodilých mluvčích častěji tvořeny kompozicí.

Můžeme shrnout, že vědomí o pravidelných a nepravidelných jazykových jevech jazyka získá student (více či méně vědomým) vyhodnocováním a zobecňováním kvantitativních informací o konstrukcích, které plní určitou funkci v diskurzu, a o jejich distribuci. Frekvence tokenů značně ovlivňuje rychlost aktivizace tvarů při produkci, frekvence typů je zase důležitým indikátorem pro platnost kategorie nebo abstraktní konstrukce. Ve svém přehledu studií zkoumajících vliv frekvence na vědomí o jazyce Ellis (2014, str. 196) shrnuje: „Language knowledge involves statistical knowledge, so humans learn more easily and process more fluently high frequency forms and ‘regular’ patterns which are exemplified by many types and which have few competitors.” Pokud jde o osvojování morfologie jazyka, dodává, že existuje mechanická asociace mezi základním a flektivním tvarem.

2.1.4 Další faktory

Vedle frekvence ovlivňuje tvoření systému žakovského jazyka, zejména hypotéz o morfologických pravidelnostech, řada dalších faktorů, které jsou doloženy psycholingvistickými experimenty.

Význačnost gramatických markerů (salience)

Pojem *salience* se vztahuje na pozornost člověka věnovanou stimulům v řeči, přesněji řečeno na sílu či stupeň aktivizace stimulu odkazujícího na určitou funkci. Ellis (2006) zkoumal konkurenci lexikálních markerů času s morfémy tvaroslovnými (-s ve 3. osobě singuláru anglických sloves). Došel k závěru, že zvlášť flektivní morfémy mají nižší salienci než konkurující lexikální prostředky (*today*), a proto činí studentovi větší potíže (tamtéž). V případě deklinačních morfémů, které jsou v jazyce bez pevného slovosledu, jako je čeština, často jediným poukazem funkce daného slovního tvaru ve větě (až na implikace dané sémantikou zúčastněných aktantů), nemůžeme toto tvrzení přejmout bez dalších zkoumání.

Prototypičnost významu

Pro mnoho kategorií v jazyce hraje klíčovou roli její prototypický zástupce. Prototyp je jejím nejlepším reprezentantem, vzorově shrnuje nejvíce podstatných vlastností. Ellis (2014) postuluje, že čím větší je frekvence konkrétního exempláře, tím více přispívá k vymezení kategorie, a pokud jde o výuku cizích jazyků, dodává: „The best way to teach a concept is to show an example of it.“ (tamtéž, str. 201). Tuto zásadu můžeme zvážit pro prezentaci jak deklinačních vzorů, tak kolokačních tříd (či abstraktních konstrukcí v pojmech konstrukční gramatiky).

Spolehlivost shody mezi formou a funkcí

Teorie *natural morphology* se zabývá otázkou, jak univerzální preference omezují nebo podporují interference mezi morfologickými systémy cílového a mateřského jazyka. Stephan Schmid (1997) zkoumal žakovský jazyk španělských studentů italštiny a pokusil se o funkčně orientovaný popis problémů, které při osvojování morfologických forem vznikaly. Ve svém modelu *Naturalness Differential Hypothesis* (NDH) ověřil a dále upřesnil čtyři univerzální preference¹² uživatele přiro-

¹² Teorie *natural morphology* a předchozí *natural phonology* tvrdí, že přirozenost a lidská schopnost hraje podstatnou roli při explikaci jazykových procesů na různých rovinách systému samotného nebo i na rovině typologické. „This theory shares the basic assumptions of Natural Phonology, in that it views linguistic universals as reflecting the substantial capacities and

zeného jazyka předložené teorii přirozené morfologie. Podle jeho zjištění platí následující zásady:

- *biuniqueness*: v ideálním případě odpovídá jednomu významu jedna forma; alomorfie není uživatelem jazyka preferovaným případem
- *morphotactic transparency*: jazyk by měl mít jasné hranice mezi morfémy; v ideálním případě si odpovídají hranice morfémů a slabik
- *morphosemantic transparency*: význam komplexního¹³ tvaru je funkcí významu jeho komponentů; synkretismus není uživatelem jazyka preferovaným případem
- *constructional iconicity/diagrammaticity*: uvnitř gramatické kategorie sémioticky bezpříznaková hodnota zůstane i formálně bez příznaku. (Schmid, 1997, str. 336n)

Schmid pojmenovává dva další aspekty preference, které se uplatňují v případě konkurence tříd nebo morfémů. Dle zásady *system-congruity* (systémová shodnost) při konkurenci morfologických příznaků vítězí více frekventovaný příznak (uvádí příklad tvoření německého plurálu pomocí přehlásek vs. přípon). Pro češtinu je zvláště zajímavá zásada *dominant paradigm condition* (zásada dominantního paradigmatu), podle níž lexikální jednotky přecházejí z menších deklinačních skupin do skupin, jež jsou platné pro větší počet jednotek (Schmid, 1997, str. 337).

Z hlediska NDH jsou tedy problémy v procesu osvojování si morfologického systému do jisté míry předvídatelné. Se zvláštním zřetelem na ohýbání slov postuluje Schmid:

“Morphological items of the target language which differ from the native language will be difficult, if they are not uniformly encoded, not morphotactically and/or morphosemantically transparent, not iconic/diagrammatic, not coherent with system defining structural properties, or members of a less stable and less numerous inflectional class.” (Schmid, 1997, str. 340)

Psycholingvistka Bordagová prováděla analýzu chyb studentů českého jazyka, které jsou spjaté s alternací hlásek uvnitř kmene a ověřovala ve své studii výše zmíněné zásady pro osvojování češtiny. Vzhledem k morfologii podstatných

difficulties of human beings in language performance; in morphology, the extralinguistic' foundation of linguistic naturalness predominantly depends on restrictions on perception, processing and memorisation” (Schmid 1997, str. 336).

¹³ Ve smyslu: skládající se z různých morfémů.

jmen prokázala. Četné případy, ve kterých studenti zachovali kmen nominativu i při ohýbání, dokládají platnost principu *biuniqueness*. Studenti zkusí zachovat jednoznačný vztah mezi významem a tvarem (Bordag, 2006 S. 207). Na příkladu chybách v substantivní morfologii dále zkoumala princip systémové shodnosti (*system-congruity*). Chyby dokládají, že studenti v případě souběhu dvou příznaků – morfologické zakončení a alternace kmenu – preferují volbu jen jednoho, silnějšího příznaku. Přitom dochází jen k chybnému vynechání alternace (srov. např. chybně utvořený tvar genitivu singuláru: **stůlu* místo *stolu* nebo chybný nominativní tvar **žáci* místo *žáci*) (Bordag, 2006, str. 216n).

Princip dominantního paradigmatu, přenášíme-li ho na proces osvojování si české deklinační soustavy, anticipuje, že substantiva příslušející k méně četným, méně typickým vzorům by studenti skloňovali spíše podle vzoru dominantního – ve smyslu frekventovaného – pro daný rod. Bordagová přezkoumala princip na deklinačním vzoru *žena* a došla k závěru, že v případě češtiny lze spíše mluvit o dominantní rodové nebo plurálové koncovce (Bordag, 2006, str. 221). Poznatky ověříme v další části na vlastním menším žakovském korpusu.

Na závěr uvádíme studii MacWhinneyho a Kempeové (1998), která zkoumala osvojování pádové morfologie ruštiny a němčiny dospělými studenty za účelem zjistit, které faktory ovlivňují tento proces, a to bez přihlídnutí k mateřskému jazyku. Diskuse o jednoznačnosti vztahu mezi morfologickým formantem a jeho funkcí je v této studii rozšířena o zřetel ke spolehlivosti tohoto vztahu. Na anglicky mluvících studentech autoři testovali pomocí krátkých transitivních konstrukcí s jedním agentem a jedním pacientem¹⁴ vliv signálů (cues), jako je pádový formant, slovosled nebo životnost, na rychlost osvojování morfologického systému (Kempe a MacWhinney, 1998).

Ruský pádový systém se vyznačuje tím, že jeho pádové koncovky poměrně spolehlivě poukazují na funkci ve větě, ačkoliv se v paradigmatu některé funkce neutralizují (životnost ve všech pádech až na akuzativ mužských životných; akuzativ a nominativ u neživotných). Oproti ruštině je německý systém méně komplexní a systém morfologického označení pádů poskytuje posluchači méně spolehlivou

¹⁴ Patiens i agens se v konstrukcích, které studenti uslyšely, objevily jednou na poslední a jednou na první pozici.

informaci o tom, který člen nebo která kombinace členu s pádem vyjadřuje kterou větnou funkci¹⁵.

Autoři vyhodnotili výsledky na pozadí tzv. *competition model*. Oproti přístupům, jež předpokládají, že procesu učení podléhá opakovanému testování hypotéz řízenému pravidly, dává *competition model* větší váhu rozvoji žákovského morfologického systému na základě asociací, a tudíž učení se, které čerpá ze vstupných dat (inputu) a řídí se platností signálů v inputu. Tento model zaměřuje pozornost na deklinační formanty jako signál či poukaz (cue) na příslušné tematické role (Kempe a MacWhinney, 1998, str. 545).

Výsledky ukazují, že ruský systém značkování pádů si studenti osvojili rychleji než německý (Kempe a MacWhinney, 1998, str. 582): spolehlivost, s kterou pádový formant poukazuje na větnou funkci, je v ruštině větší, zatímco větší ambiguita v německém systému spíše zpomaluje jeho osvojování.

Komplexnost deklinačního systému tedy nutně nepředpokládá složitější proces učení se. Závaznější je podle studie platnost „pádového signálu“ a spolehlivost, s kterou poukazuje na funkci ve větě či tematickou roli.

2.1.5 Shrnutí se zřetelem k češtině

Z hlediska poznatků SLA připravuje český deklinační systém pro cizince učícího se češtině jako cizímu jazyku následující překážky:

- homonymie některých pádových formantů způsobuje porušení spolehlivosti vztahu mezi tvarem a jeho funkcí ve větě, a tudíž znejednoznačení platnosti morfologického signálu
- alternace kmene, která se může týkat více tvarů v paradigmatu, je vzhledem k vyjádření funkce pádu nespécifická a zatěžuje proces osvojování
- rozmanitost deklinačních vzorů a skutečnost, že každému rodu je přiřazeno více vzorů, zvyšuje konkurenci při vyjadřování kategoriálního významu a komplikuje tvoření kategorií
- neutralizace pádové opozice u některých vzorů (např. u vzoru *růže* v akuzativu, lokálu a dativu singuláru) zpomaluje proces osvojování

¹⁵ Pád se vyjadřuje v singuláru pomocí členu, který např. u ženského vzoru má jen dvě podoby: *die* pro nominativ a akuzativ, *der* pro genitiv a rovněž pro dativ singuláru; v plurálu je přes plurálový morfém *-(e)n* tvarová homonymie daná v nominativu a akuzativu.

Citované studie ukazují, že prezentace a výuka deklinační morfologie musí přihlížet ke konkrétním tvarům a k jejich distribuci v jazykovém úzu, a svědčí o tom, že učení je pevně spjato s vyhodnocováním informací o frekvenci a distribuci forem v textu. Zažité frekvence tvarů v inputu určují rychlost pozdějšího přístupu a zapamatování pravidelných i nepravidelných tvarů. Se vzrůstající pravidelností nějaké konstrukce roste její produktivita a vzniká abstraktní vzorec. Zásadní roli pro učení se těchto vzorců sehrávají prototypičtí zástupci.

Pro další výzkum vyplývají z diskuse následující úkoly. Je potřeba:

- identifikovat morfologické konstrukce, které činí zvláštní potíže studentům češtiny jako cizího jazyka
- zkoumat distribuci substantivních tvarů a jejich preferovaných konstrukcí (kolokací) v současném jazykovém úzu
- identifikovat prototypické zástupce těchto konstrukcí a pokud možno vymezit jevy periferní a jádrové
- analyzovat v tomto rámci roli deklinačních tříd a identifikovat jejich prototypické zástupce

2.2 Typy chyb jinojazyčných studentů českého jazyka ve skloňování

2.2.1 Cíle analýzy chyb

Tato část práce si klade za cíl ověřovat na menším žákovském korpusu výše popsané překážky při osvojování si substantivních tvarů a blíže specifikovat místa či konstrukce, ve kterých studenti češtiny často chybují. Analýza typů chyb má ukázat, o které vzory či prototypické exempláře se studenti při tvoření substantivních tvarů opírají.

2.2.2 Výběr textů

Pro danou analýzu je nám k dispozici malý ručně anotovaný korpus písemných studentských projevů. Studenti se zúčastnili zkoušky UNI[®]cert I, která uzavírá třísemestrální studium českého jazyka a testuje komunikativní schopnosti a dovednosti na úrovni A2+ (produkce) a B1 (recepce). Test UNI[®]cert I klade zvláštní důraz na schopnost studentů vyjadřovat se ke všedním studijním záležitostem v rámci popisu dané úrovně. Zkouška se skládá ze čtyř částí: poslech a čtení s porozuměním, psaní dopisu a část ústní. Do analýzy byly zahrnuty výsledky písemného projevu, celkem 119 dopisů. Úlohy spočívaly v tom, že studenti měli

reagovat na osobní dopis, který obsahoval otázky jak ze studijního, tak osobního života¹⁶.

Autoři textů byli studenti ve věku od 20 do 30 let, kteří studovali na drážďanské Technické univerzitě nefilologické obory. Pro převážnou většinu byla mateřským jazykem němčina. Ojedinělí autoři byli rodilými mluvčími ruštiny nebo polštiny.

Jsme si vědomi, že takto elicitované texty odrážejí jen do jisté míry přirozená, volná a neřízená vyjadřování, a lze je tudíž jen se rezervou považovat za plně autentické, tedy projevy vzniklé v reálné komunikační situaci.¹⁷

V současné době jsou k dispozici dva rozsáhlé žákovské korpusy češtiny, CzeSL¹⁸ a MERLIN¹⁹. Oba korpusy mají ruční chybovou anotaci a automaticky provedenou lemmatizaci i anotaci slovních druhů. Bohužel byly zpřístupněny teprve v závěrečné fázi této práce. Poněvadž má rozbor chyb v rámci práce jen pilotní charakter a není jejím těžištěm, neopakovali jsme časově náročnou analýzu morfologických chyb v obou korpusech, přihlížíme však k jejich metodologickým přístupům. Připouštíme, že by bylo potřeba ověřit výsledky dané analýzy v rámci rozsáhlejší samostatné studie.

2.2.3 Rozsah analýzy

V naší analýze nezkoumáme všechny chyby studentů. Omezujeme se na jevy, které se projevují v morfologické rovině. Na odchylky lexikálně-sémantické povahy a rovněž chyby v rovině pragmatické nebereme ohled. V rámci průzkumu frekvence morfologických tvarů podstatných jmen a znázornění substantivních paradigmat nás zajímá hlavně nesprávné²⁰ užití tvarů podstatných jmen a v omezené míře i adjektiv, pokud jsou součástí chybné nominální skupiny.

¹⁶ Zadání tedy již obsahovalo český text, z kterého bylo v podstatě možné přebírat některé formulace.

¹⁷ Problematiku omezené autentičnosti diskutuje Šebesta (2012). Podle něj je vznik žákovských textů ovlivněn různými podmínkami, a proto žádá jemnější členění než na data autentická a neautentická. Texty, které se zařazují do žákovských korpusů, se pohybují na ose autentičnosti mezi póly přirozeného, volného, „neřízeného“ vyjadřování v autentických situacích reálného života a vyjadřování, které je jasně řízeno jinou osobou než mluvčím s primárním zřetelem k (jazykové) formě sdělení (Šebesta, 2012, str. 18). Dochází k závěru, že většina dnešních žákovských korpusů jsou budované na textech elicitovaných badatelem s pozorností zaměřenou na obsah či funkci sdělení, nikoli na jeho jazykovou formu (Šebesta, 2012, str. 20).

¹⁸ Šebesta et al., 2014, CZESL-PLAIN, <http://akces.ff.cuni.cz/> a <http://www.korpus.cz>

¹⁹ Project Merlin, 2004, <http://merlin-platform.eu>

²⁰ Pokud v této práci užíváme slovo *nesprávný* v kombinacích jako *nesprávné užití*, *nesprávná volba*, *nesprávný tvar*, máme na mysli užití popř. tvary, které odchylují od noremního užití, tj. od

Nezkoumají se chybné tvary zájmen ani případy, v nichž substantivum bylo nahrazeno osobním zájmenem, poněvadž se jim studenti na této úrovni nemohli důkladně věnovat a často jejich tvary nedostatečně zautomatizovali.

2.2.4 Postup anotace chyb

Žákovské texty byly analyzovány s důrazem na syntagmata s defektním (z hlediska cílového jazyka) substantivním tvarem. Byly elicitovány všechny chybné pádové tvary substantiv bez ohledu na to, zda byly výsledkem nesprávné volby pádové koncovky, porušení shody nebo porušení valence uvnitř verbální, předložkové nebo nominální skupiny. Chyby byly popisovány na více rovinách:

- 2.1. Rovina 1 – specifikace větných konstituentů, v nichž došlo k chybě
- 2.2. Rovina 2 – specifikace druhu chyby (lingvistická kategorizace)
- 2.3. Rovina 3 – jemnější specifikace chyby podle lingvistických subkategorií a způsobu modifikace cílového tvaru
- 2.4. Rovina 4 – specifikace zdroje transferu

2.2.4.1 Rovina 1 – Specifikace větných konstituentů, ve kterých došlo k chybě

Rovina 1 určuje, ke které syntaktické složce se chyby vztahují. Pro popis se užívá terminologie složkové strukturální analýzy, tj. jednotkami syntaktické struktury věty jsou složky či fráze (angl. *phrase*). Jsme si vědomi, že takový postup v české lingvistice není běžný. Není však pro nás možné jednoznačně vyjasnit závislostní vztahy mezi chybným tvarem a nadřizenými větnými členy pomocí komplexní závislostní analýzy, poněvadž je často těžko určit, zda je příčinou chyby neznámá vazba, nesprávně realizovaná vazba či nesprávná volba koncovky²¹. Pro formální vymezení syntaktické složky, ve které k chybě došlo, se nám postup složkové strukturální analýzy zdá dostačující. Přikláníme se tak k postupu anotace chyb v

toho, co je českým jazykovým společenstvím přijato a co je konvencionalizováno. V tomto smyslu rozumíme zde i pojmem *chyba* a *chybný* především odklon od jazykové normy.

²¹ Usouvztažňování chyby k jejímu kontextu umožňuje víceúrovňová chybová anotace, která byla provedena např. v korpusu CzeSL. Anotační schéma dovoluje anotaci chybného tvaru nezávisle na kontextu při současném zachování jeho syntaktických relací. Tvary jsou emendované na rovině 1, tedy opravené jen pravopisně a morfologicky, připojeny k jejich opraveným reprezentantům na vyšší rovině a opatřeny navíc značkou, která ukazuje na typ chyby. Tím způsobem je možné sledovat syntaktické závislosti skrz různé úrovně chybové emendace (Štindlová et al., 2014, str. 67).

korpusu MERLIN, který volí jako rozsah tagů (tag span) pro morfologicky ne-
správně realizované tvary obvykle jednotku fráze²².

Chyby byly usouvztažňovány k nejbližší vyšší (řídící) složce. Podle toho se dělí
na chyby v NP²³ (nominální fráze), VP (verbální fráze), PP (předložková fráze) a
AdP (adverbiální fráze). NP se považují za část verbální složky (VP) tehdy, jsou-li
ve funkci obligatorního či fakultativního aktantu (*Budu psát diplomovou práci*).²⁴
Z funkčního hlediska se NP a PP dále člení na složky, které nahrazují adverbiální
výraz, tj. mají ve větě funkci příslovečného určení směru, místa, způsobu, příčiny
a času. Podle funkčně generativního popisu se jedná o volná doplnění k řídícímu
slovu, která ale mohou být pro některá slovesa obligatorním, konstitutivním čle-
nem. Dále se vyčleňují NP, jež nahrazují kvantifikující výraz. Následující tabulka
uvádí typy složek s příkladem.

Fráze		Příklad
NP	Složka podstatného jména (nominální skupina)	
NP	<ul style="list-style-type: none"> ▪ samotné jádrové substantivum, vlastní nebo obecné jméno v pozici subjektu nebo přísudku (predikativum) ▪ samotné jádrové substantivum v N²⁵/A/G/I ▪ jádrové substantivum s neshodným substantivním atributem ▪ jádrové substantivum + substantivum v předložkovém pádě 	<i>moje zkoušky</i> <i>Praha je krásné město</i> <i>jako slavík, vlakem, je povolá- ním</i> <i>gramatika češtiny, cesta vlakem</i> <i>zkouška z historie, strach ze zkoušky</i>
NP- kvant	<ul style="list-style-type: none"> ▪ jádrové substantivum + kvantitativní substantivum ▪ jádrové substantivum + základní číslovka 	<i>milion+NP, řada+NP, kou- sek+NP</i> <i>2-4+NP; 5+NP</i>
NP-adv	<ul style="list-style-type: none"> ▪ složka podstatného jména, která nahrazuje adverbiální výraz (volná doplnění k řídícímu slovu, která mohou být fakultativní). ▪ jádrové substantivum v A ▪ jádrové substantivum v G 	<i>Jedu <u>autem</u> do Švýcarska.</i> <i>příští středu</i> <i>dvacátého února</i>

²² „A phrase is considered a unit of one or more words that build a functional unit, that are adjacent and that can be permuted/moved as a whole within a sentence. Each phrase has a head that determines the formal features of the other elements of the unit and that determines the name of the unit.” (MERLIN project, 2014)

²³ V našem přehledu jsou použity nejběžnější anglické zkratky: NP = nominal phrase, AdvP = adverbial phrase, VP = verbal phrase, PP = prepositional phrase.

²⁴ Srov. dělení slovesných doplnění ve funkčně generativním popisu na doplnění vnitřní (aktanty obligatorní a fakultativní) a doplnění volná (ESČ, Karlík et al., 2002, str. 518).

²⁵ Pro konkrétní pád a číslo používáme v této práci následující zkratky: N = nominativ, G = genitiv, D = dativ, A = akuzativ, V = vokativ, L = lokál, I = instrumentál, sg. = singulár, pl. = plurál.

VP	složka slovesná	
	<ul style="list-style-type: none"> ▪ jádrové sloveso a jeho valenční doplnění s výjimkou pozice podmětu (viz NP): <ul style="list-style-type: none"> ○ jádrové sloveso + nominální složka v A, G, D a I ○ jádrové sloveso + nominální složka v předložkovém A/L/I ▪ konstrukce s nulovým podmětem / neosobním slovesem anebo s vedlejší subjektovou větou ▪ plnovýznamové BÝT + časová orientace 	<i>znát někoho, ptát se někoho zabývat se něčím, blahopřát někomu, rozumět někomu děkovat za, mluvit o, mluvit s</i> <i>bolet někoho někde, líbit se někomu, těšit někoho, že ..., zdát se někomu, že ...</i> <i>včera bylo šestého června</i>
PP	složka předložková	
PP-adv	<ul style="list-style-type: none"> ▪ předložková složka ve funkci volného doplnění, která nahrazuje adverbialní výraz: směru, pohybu, místa, příčiny, účelu 	<i>do práce, na dovolenou, k moři jít <u>pro peníze</u>, kvůli tomu, pro nemoc, ze strachu</i>
AdP- kvant	<ul style="list-style-type: none"> ▪ adverbialní kvantifikátor + substantivum 	<i>několik+N, mnoho+NP, moc+NP</i>

Tab. 2.1 Přehled o složkách užitých při anotaci v rovině 1

2.2.4.2 Rovina 2 a 3 – určení druhu chyby

Interpretaci chyby v jejím kontextu musí předcházet vytvoření tzv. cílové hypotézy, tedy varianty žakovského výrazu, která se v cílovém jazyce považuje za správnou či akceptovatelnou (Reznicek, Lüdeling a Hirschmann, 2013, str. 103). Cílová hypotéza může přihlížet k různým úrovním lingvistické analýzy (od vytvoření ortograficky správného textu, přes vytvoření morfologicky správného textu až po vytvoření textu, který je nejen lexikálně ale i pragmaticky hodnocen rodilým mluvčím jako přiměřený). Za účelem vyhodnocení morfologických chyb jsme formulovali ke každé větě minimální cílovou hypotézu, jež koriguje ortografické a morfologické odchylky s minimálními zásahy a změnami původního textu tak, aby vznikla gramaticky správná věta (srov. student: **Pojedu na Itálie*. → cílová hypotéza: *Pojedu do Itálie*). Správnost věty se přitom posuzovala jen v kontextu dotyčné věty²⁶. Řídili jsme se přísnými pokyny pro vytvoření minimální cílové

²⁶ Pokud nebylo např. v tomto kontextu možné jednoznačně rozhodnout, co měl student na mysli, nebyl tvar označen za chybný. Srov. např. *budu na moři* se nepovažuje za chybu, protože taková interpretace v úzkém kontextu jedné věty přípustná je, ačkoliv student měl zřejmě na mysli: *u moře*.

hypotézy, které byly použity v žákovském korpusu FALKO²⁷ a MERLIN (Boyd et al., 2014, str. 1284).

Na základě cílové hypotézy byly chyby podle lingvistické povahy odchylky od spisovné normy rozříděny na chybu v předložce, neexistující tvar, morfologicky nesprávný tvar (tvar, jenž existuje, ale není v daném kontextu správný)²⁸ a pravopisnou chybu. Takové třídění posuzuje chyby výhradně v úzkém morfologickém kontextu bez předpokladů příčiny chyby, jako je např. nesprávná shoda nebo nesprávně realizovaná valence. Ku prospěchu studenta byly chybné délky v morfologické koncovce posuzovány za chyby pravopisné (např. *seminární práce*) stejně jako záměnu *-e s -ě* a *-i s -y* (např. *do kavárni, v jídelne*).

Až na další, třetí rovině byl blíže určován typ odchylky od správného tvaru. Morfologicky nesprávné či neexistující tvary byly jemněji rozříděny podle morfologické kategorie (chybný pád, chybné číslo, chybný rod). Chyby v předložce a pravopisné chyby se dále specifikovaly podle způsobu odchylky od cílového tvaru podle spíše mechanického modelu²⁹ (viz tab. 2.2).

2.2.4.3 Rovina 4 – specifikace zdroje transferu

Účelem pilotní sondy žákovských textů bylo zjistit, o které deklinační vzory nebo pádové tvary se studenti při produkci textů opírají. Proto byla zavedena další rovina, která určuje pro každou morfologickou chybu, zda pádová koncovka existuje v jiném deklinačním vzoru, zda zvolená koncovka patří k jinému pádu stejného vzoru nebo zda byla vybrána variantní koncovka, která odpovídá příslušnému pádu a užívanému vzoru, avšak je nepřipustná pro daný lexém nebo v daném syntaktickém kontextu (např. **v srpne, *pozdravy k Petru*).

Následující tabulka znázorňuje souvislosti mezi rovinou druhou, třetí a čtvrtou.

Rovina 2	Rovina 3	Rovina 4
Chyba v předložce	nesprávná volba předložky (wrong choice)	-
	nadbytečná předložka (addition)	
	chybějící předložka (omission)	

²⁷ Srov. koncepci *Minimal target hypothesis (TH1)* (Reznicek, Lüdeling a Hirschmann, 2013, str. 109)

²⁸ Rozlišení morfologických chyb na morfologicky nesprávný tvar a neexistující tvar jsme převzali z chybové koncepce v projektu MERLIN (MERLIN project, 2014).

²⁹ Taxonomie chyb, která se opírá o formální popis odchylky od cílového jazyka jako např. *omission, addition, wrong choice* etc., je v mnoha žákovských korpusech běžná a známá jako tzv. *target modification taxonomy* (Díaz-Negrillo, 2006, str. 92).

Neexistující tvar	chybný rod nelze určit	jiný vzor jiný vzor anebo jiný pád
Morfologicky chybný tvar	chybný pád	jiný vzor
	chybné číslo	jiný pád
	chybný rod	jiný vzor anebo jiný pád
	nelze určit (ambiguous)	
	nesprávná varianta koncovky	nesprávná varianta koncovky
Pravopisná chyba	nesprávná volba grafému	-
	chybný grafém (omission)	
	nadbytečný grafém (addition)	

Tab. 2.2 Typologie chyb, rovina 2, 3 a 4

2.2.5 Vyhodnocení

V analyzovaných písemných projevech studentů se vyskytlo celkem 175 morfologicky chybných nebo neshodných substantivních tvarů. 72 % chyb přitom tvořily tvary s nesprávnou morfologickou koncovkou (např. *letím do Polsky; jedu do Švýcarsky*), 3,4 % pravopisné chyby. 24,6 % substantivních tvarů vybočovalo ze shody v rámci syntagmatu v důsledku nesprávné volby předložky (např. *jedeme do Baltské moře ve dvě týdnu*).

	absolutní počet chyb	%
morfologicky chybný tvar	105	60,3 %
chyba v předložce	43	24,7 %
neexistující tvar	20	11,5 %
pravopisná chyba	6	3,4 %
CELKEM	174	100 %

Tab. 2.3 Počet chyb podle druhu chyby

V další analýze se soustředíme na 125 chybně utvořených substantivních tvarů (morfologicky chybných tvarů a neexistujících tvarů).

2.2.5.1 Specifikace větných konstituentů, ve kterých došlo k chybě

V 38,1 % případů došlo k chybě v předložkových skupinách, které fungují jako příslovečné určení místa nebo směru a jsou buď fakultativním, nebo konstitutivním aktantem ve větě, např. *pojedu na dovolené, když setkáme v šest hodin před menze*. Většina nesprávných syntagmat (45,8 %) připadá na doplnění směru (popř. účelu) u sloves pohybu, např. *Pojedu do Francii na dovolene; Pojedu vlakem na Slovensku; Pojedu k mořem; Kdy přijížděš do Drážďanech?; Jedu do Švýcarsky*.

Víc než čtvrtiny morfologických chyb (29,4 %) se dopustili studenti v přímých a nepřímých předmětech slovesné složky (např. *Pracuju už na diplomku.*; *Měj se hezky a zdravím Petr.*; *Dovolená plánuju mezi studijní prací a diplomoví prací.*; *Minulý týden jsem napsala diplomce*).

Chyby v nominálních frázích čítají celkem 41 případů, tedy 30,9 %. Potíže s tvořením správného tvaru měli studenti v nominálních složkách s neshodným atributem, např. *Moje zkoušky z historie byly výborně.*; *hledám dobrou gramatiku českého jazyku*; *Přeji ti štěstí za tvůj pohovor v Pirně.*; *Pozdravy k Petru*; *zajímavý inzerát o kurzu češtiny* (57,4 % všech chyb v NP). Zbylé problémy vznikly při užití časového výrazu v akuzativu, např. *Příští semestru budu psát diplomovou práci.*; *Příští středa budu mít čas.* (22,2 % chyb v NP), a při užití kvantifikujícího výrazu, např. *protože napíšu pět zkoušky* (20,4 % chyb v NP).

Fráze	absolutní počet chyb	%
NP	22	17,6
NP-adv	7	5,6
NP-kvant	11	8,8
PP-adv	48	38,4
VP	37	29,6
CELKEM	125	100%

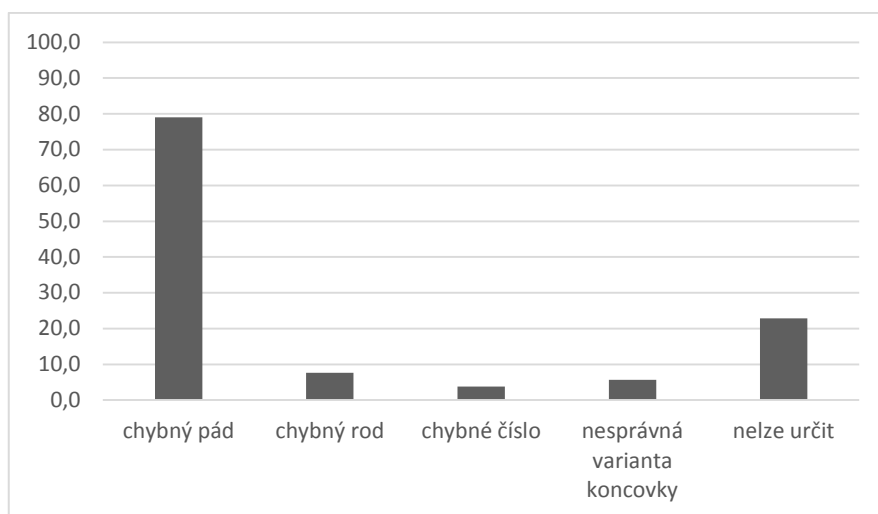
Tab. 2.4 Rozložení chyb podle syntaktických složek

Můžeme shrnout, že se sledované morfologické chyby rozkládají rovnoměrně do všech složek ve větě. Nerovnoměrnost distribuce chyb je nápadná teprve z funkčního hlediska: problémy mívali studenti nápadně často s tvořením substantiv v předložkových nebo substantivních skupinách plnících funkci příslovečného určení místa, směru, účelu a času (44 % všech případů).

2.2.5.2 Specifikace druhu chyby

Ve 105 ze 125 chybně utvořených substantivních tvarů volili studenti existující tvar, který je ale v daném kontextu chybný. Pouze ve 20 případech (15,8 %) utvořili tvar, který neexistuje, a s kterým se tedy před tím s nejvyšší pravděpodobností nesesetkali. Další roztřídění chyb podle morfologické kategorie odhaluje, že téměř v 80 % případů vznikla chyba volbou tvaru patřícího k jinému pádu. Poměrně zříd-

ka se stalo, že studenti volili tvar, který patří do paradigmatu jiného čísla (3,8 %) nebo jiného rodu (7,5 %). V 5,7 % případů je pád a deklinační vzor správný, byla však vybrána variantní koncovka, kterou s daným lexémem nelze spojit. Tvary neexistující není vždy možné dále specifikovat. Nacházejí se mezi nimi tvary slov, jež byly tvořeny podle jiného vzoru, a navíc byly užity s chybnou pádovou koncovkou³⁰. Ovšem nebylo ani u všech nesprávných ale jinak existujících tvarů možné jednoznačně určit, co vedlo ke vzniku chyby. Celkem 22,9 % užitých tvarů tak nelze vztáhnout k nesprávné volbě pádu nebo rodu.



Obr. 2.1 Frekvence typů chybných substantivních tvarů (rovina 3) v %

V neposlední řadě nás zajímá, o které vzory nebo tvary se studenti opírají při produkci substantivních tvarů. Do jaké míry používají dominantní tvary, a pokud tvoří nové formy, z kterých předloh nebo vzorů čerpají? V 20 případech byla užitá koncovka jiného vzoru, v 85 případech koncovka jiného pádu.

Zkoumali jsme nejprve tvary, jejichž koncovka odpovídá syntaktickým požadavkům v tom smyslu, že naznačuje správný pád, ale patří k jinému deklinačnímu vzoru. V 60 % sloužil studentům jako předloha vzor *žena* a to zejména u substantiv typu *dovolená*, ale také u substantiv patřících ke vzoru *město*. V 25 % případů volili studenti koncovka vzoru *hrad* místo vzoru *pán* (týká se to zejména užití akuzativního tvaru, srov. *Pozdravuj Petr.*). Vzácně, a to v následujících případech, došlo k výpůjčce koncovky od vzoru *hrad* u substantiva ženského rodu (*zítra mám zkoušku*), k výpůjčce koncovky od vzoru *stroj* u slova *kurz* (*posileš mně inzerát o*

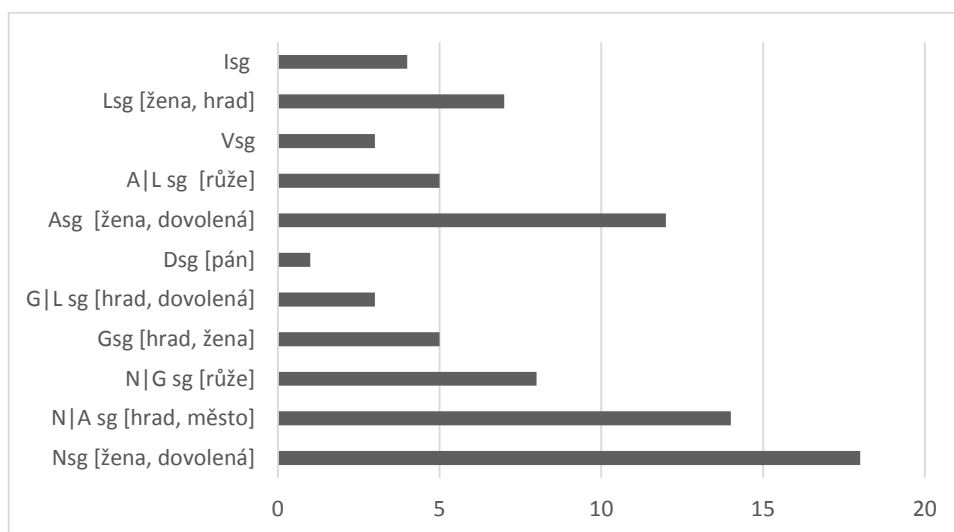
³⁰ Problém se opakuje např. u slova *dovolená*, např.: *Ne, neplánuju o dovolenu.*; *Jsem o dovolenu.*

prázdninám kurzi češtiny) a od vzoru *město* u slova *dovolená* (*budu brzo mít dovoleno*).

Srovnávání frekvencí deklinačních vzorů *žena* se vzorem *dovolená* a vzoru *hrad* se vzorem *pán* v SYN2005³¹ ukazuje, že vzory *žena* s 22,79 % a *hrad* s 25,96 % všech výskytů mají oproti vzoru *pán* (9,79 %) a vzoru *dovolená* (0,55 %) výrazně dominantní postavení. Studenti tedy v případě nejistoty skloňovali substantiva přináležející k méně často zastoupeným vzorům spíše podle vzoru dominantního pro daný rod.

Stojí za to zmínit, že koncovky vzorů *žena* a *hrad* byly uplatněny také u lexémů jiného rodu, např. *letím do Polsky; jedu do Švýcarsky*. To může být indikátorem nejen dominance deklinačního vzoru, ale i celé konstrukce typu <*sloveso pohybu do ...-y*>; naše datová báze však neposkytuje dostatečnou základnu pro jednoznačný závěr.

V dalším kole byly podrobeny důkladnější analýze chybné formy, jež v paradigmatu daného substantiva existují, ale vyjadřují jiný než požadovaný pád. Účelem bylo zjistit, které pádové tvary jsou upřednostňovány v rámci deklinačního vzoru. Pro značnou homonymii pádových tvarů ve všech deklinačních paradigmatech, zejména však u vzorů *růže*, *moře* a *stavění*, nebylo vždy možné jednotlivé pády jednoznačně určit. Tudíž byl pro lemmata uváděn pádový tvar, jenž studenti zvolili, a obě možné funkce (srov. obrázek 2.1, např. *Francii*: vzor *růže* → A | L sg).



Obr. 2.1 Transfer jiného pádové koncovky (absolutní počet)

³¹ Srov. Statistiku češtiny (Bartoň et al., 2009, str. 139)

Z přehledu (v obrázku 2.1) je patrné, že tvary nominativu singuláru jsou nejvíce vybírány, celkem u 40 % slov, přičemž u 17,5 % slov je tvar nominativu totožný s akuzativem (*hrad, město*). Toto zjištění možná nepřekvapuje, neboť nominativ je tvarem substantiva, který je v učebnicích, ve výkladech o gramatice i v lekcích uváděn jako ten tvar, podle kterého se přiřazují slova k deklinačním vzorům. U některých lemmat je to zřejmě i slovní tvar, s kterým se studenti ve výuce nejčastěji setkali. Jen částečně sem spadá užití tvaru nominativu vzoru *růže*, který je formálně totožný i s jinými tvary, zejména s genitivem singuláru či nominativem plurálu.

Vedle nominativu mohou mít roli dominantního tvaru i akuzativní tvary a tvary lokálu singuláru (celkem 30 %), u slov vzoru *žena, moře a hrad* pozorujeme také nesprávné užití tvaru instrumentálu nebo genitivu.

Chybně užití tvary v sestupném frekvenčním pořadí

vzor	1.	2.	3.	4.	5.
<i>žena</i>	N sg (12)	A sg (9)	L sg (5)	G sg (3)	I sg (3)
<i>růže</i>	N G sg (8)	A L sg (3)			
<i>dovolená</i>	N sg (4)	A sg (2)	L G sg (2)		
<i>hrad</i>	N A sg (10)	G sg (2)	G L sg (2)		
<i>pán</i>	N sg (2)	D sg (1)	V sg (1)		
<i>město</i>	N A sg (1)				
<i>moře</i>	N A sg (1)	G L sg (1)			

Tab. 2.5 Nahrazující (přenášené) morfologické tvary a frekvence jejich užití dle vzorů

Po nominativu je druhým nejčastěji přenášeným pádovým tvarem akuzativ, pokud se liší od nominativu. Sonda do našich dat ukazuje, že akuzativní tvar nahrazuje např. tvar lokálu v předložkových skupinách s funkcí určení místa (*budu na dovolenou, budu na naši svatbu, setkáme se na univerzitu*) nebo že se užívá u předmětového substantiva v rámci konstrukcí s tranzitivním slovesem, kde se vedle náležitých případů (např. *mám práci, píšu diplomku*) vyskytuje i tam, kde by pád předmětového substantiva měl být řízen např. kvantifikátorem nebo (nenáležitě) užitou předložkou, např. *mám moc práci; píšu na diplomku; mám moc práce a málo čas; nemám moc peníze*. Také nesprávné užití lokálu pozorujeme u substantiv odlišných vzorů. Studenti mu dávali přednost zejména u podstatných jmen označujících místa (v devíti z deseti případů), srov. *Chci jet do Praze.; Kdy přijížděš do Drážďanek?; Chtěla bych jet do Florencii.; když setkáme v šest hodin před menze; Můžeme jít do kina nebo do hospodě.; Pojedu na dovolené. (2x); Přiju na hradě Křivoklat.*

2.2.6 Shrnutí

Ve většině zkoumaných morfologických chyb byl preferován jiný tvar stejného lemmatu (63,2 % případů). Bez zjevných rozdílů u jednotlivých deklinačních vzorů byl ve většině případů zvolen nominativní tvar. To však neznamená, že nominativní tvar převládal u jednotlivých slov ve všech kontextech. Dominantním tvarem byl v některých kontextech (např. v NP u tranzitivních sloves nebo v PP plněním funkce příslovečného určení místa nebo směru) i tvar akuzativu nebo lokálu. Data naznačují, že určitý pádový tvar je preferován jen určitou syntaktickou konstrukcí s konkrétní funkcí, v tomto případě volným doplněním k určení směru nebo místa. Užití lokálních tvarů je zjevně upřednostňováno u substantiv s lokální sémantikou, jako jsou názvy měst. U těchto slov je lokální tvar zřejmě lépe přístupný a zakotvený, a dá se tedy předpokládat, že se s ním studenti setkali v učebnici nebo ve výuce častěji než s tvarem nominativním.

V 16,8 % morfologických chyb byla užitá koncovka jiného deklinačního paradigmatu. Zjevné dominantní postavení má vzor *žena* pro ženský rod (ale i pro některá slova jiného rodu) a vzor *hrad* zejména pro maskulina životná a neživotná. Data tedy potvrzují zjištění, že v případě substantiv náležejících k méně typickým vzorům mají studenti tendenci skloňovat podle vzoru dominantního pro daný rod. Výsledky mohou tedy podporovat předpoklad principu dominantního paradigmatu. Data ale dávají podnět ke zvážení, zda pádová koncovka dotyčného deklinačního vzoru má dominantní postavení jen v rámci určité syntaktické konstrukce. Prototypické se zdá být např. pro slovesa pohybu užití předložky *do* + genitivní koncovky vzoru *žena*. Pro ověření této hypotézy by však bylo potřeba provést sondu do většího žákovského korpusu.

Můžeme shrnout, že nominativní tvar má sice obecně velmi dominantní postavení pro lemmata, ale není a priori pro každé lemma nejdůležitějším reprezentantem. Takovou roli může převzít i tvar lokálu nebo akuzativu. Naše data ukazují, že se tyto tvary upřednostňují v určitých syntaktických konstrukcích.

2.3 Prezentace substantivních tvarů v učebnicích

2.3.1 Role gramatiky v komunikační metodě

Ke gramatice se v lingvodidaktických metodách předminulého a minulého století přistupovalo velmi odlišně. V metodě gramaticko-překladové, která se dlouhá léta aplikovala i v Čechách, se podřizovalo zprostředkování komunikačních schopností a často i výběr učebních textů systematickému učení gramatických paradigmat. V jiných metodách se gramatika prezentovala induktivním způsobem (např. ve strukturně globální audiovizuální metodě) nebo se uvažovalo o přirozené posloupnosti osvojování jinojazyčného kódu bez přímého zprostředkování znalostí o jazykovém systému (např. v přirozené metodě)³². Zmíněným i dalším vyučovacím koncepcím, jako např. metodě přímé nebo audioorální, se mj. vytýkalo jednostranné zaměření na jednu z řečových dovedností, omezená škála trénovacích metod a neuspokojivá aplikovatelnost v běžné cizojazyčné komunikaci (srov. Hrdlička, 2009, str. 52n.). Od sedmdesátých let minulého století posléze doplňovala a střídala předchozí koncepcie *komunikační metoda*³³. Komunikační metoda uznává svébytnost všech řečových dovedností a zahrnuje komplexní přístup k cíli připravit studenta na úspěšnou komunikaci, přiměřenou v daných situacích. Jde tudíž o kompetenci všestrannou, jež zahrnuje nejen jazykovou či gramatickou složku, ale i složku sociolingvální, diskurzní a strategickou.

V tradičních tzv. strukturních kurikulech zaujala systematická prezentace jazykových struktur přední pozici. Současně s komunikačně-pragmatickým obratem a zavedením komunikační metody jim od sedmdesátých let minulého století konkuruje tzv. pojmově-funkční kurikulum (*notional/functional syllabus*). Zakládá se na jazykových funkcích (*pozdravit někoho, představit se*) a sémanticko-gramatických pojmech jako např. častost, prostor, čas³⁴ jako na hlavních organizujících prostředcích učebních plánů. Tato koncepce je úzce spjatá s prací Evropské rady na

³² Pro souhrnný přehled o roli gramatiky ve vybraných metodách výuky cizích jazyků v devatenáctém a dvacátém století viz Hrdlička (2009).

³³ V české lingvodidaktice se setkáváme s dvojím pojmenováním. Užívá se jak *komunikační metoda*, tak *komunikativní metoda*. Hrdlička zdůrazňuje, že je důležité použít přívlástek komunikační, pokud se vztahujeme k autonomní vyučovací koncepci, a výstižně rozlišuje: „Komunikační metoda je tedy jistý přesně definovaný specifický způsob prezentace (výchozího a cílového) jazykového kódu; komunikativních (a také nekomunikativních metod) je vícero.“ (Hrdlička, 2009, str. 60)

³⁴ V anglickojazyčné literatuře se paralelně používají pojmy *notions* a *semantico-grammatical categories* (srov. Johnson, 2009, str. 318). V české verzi Threshold level, Prahová úroveň (Bischofová, J. et al., 2002), se pojem zachycuje termínem *obecné pojmy*.

celoevropsky platném rámci pro výuku, kurikula a testování cizích jazyků - Společný evropský referenční rámec pro jazyky (SERR). Obracení se ke studentovi jako činiteli, který se učí přiměřeně užívat jazyk ke zvládnutí úloh (jak jazykových, tak společenských), je rovněž jedním z metodických pilířů SERR³⁵. SERR uplatňuje pro svůj účel – poskytnout obecný základ pro vypracování jazykových sylabů, směrnic pro vývoj kurikul, zkoušek, učebnic atd. v celé Evropě - velmi obecný, funkční pohled na učení se jazyků a užívání jazyka. Pokud jde o přípravu kurikul kurzů a hodin, klade důraz na komunikační úlohy a doporučuje uživatelům rámce zvážit, ve kterém kontextu studenti budou cizí jazyk užívat, a podle využitelnosti vybírat komunikační úlohy, témata, jazykové struktury, slovní zásobu apod. (SERR, 2002, str. 47nn. a 153n.)

Vypracování jednotlivých kompetenčních úrovní SERR, jako anglický *Threshold level* a jeho česká obdoba *Prahová úroveň*, je typickou ukázkou funkčního typu kurikul. V posledních letech se však ozývaly kritické hlasy, mezi nimi i ze strany učitelů. Vytýkají výuce orientované primárně na vypěstování komunikační dovednosti bez systematického učení jazykových struktur, že zanechá studenta s neúplným seznamem naučených frází, ale bez schopnosti udělat jasné zobecnění o jazykovém úzu a užívat jazyk i mimo tyto fráze (srov. Johnson, 2009, str. 320).

V současnosti se proto vracejí do kurikul prvky strukturních kurikul a vznikají hybridy, které organizují výuku podle užití jazykových prostředků v komunikačních situacích, ale dovolují zaměřit se ve vhodném okamžiku na systematickou výuku určité jazykové struktury. Označují se jako multidimenzionální kurikula (srov. Hrdlička, 2009, str. 51; Holá, 2014, str. 1; Johnson, 2009, str. 331nn.)

V následující části budeme diskutovat, jakou roli hraje explicitní instrukce a zamýšlení se nad strukturami a pravidelnostmi jazyka v komunikačně orientované výuce a v metodách „post-komunikačních“ (*post-communicative approaches*) a jak se metody odrážejí v příslušných kurikulech či osnovách učebnic se zvláštním zřetelem na českou situaci.

³⁵ Metodicko-didaktická koncepce, která podléhá referenčnímu rámci, sleduje *akčně zaměřený přístup*, a to v tom smyslu, „že chápe uživatele a studenty jazyka v první řadě jako společenské činitele, tj. členy společnosti, kteří mají své úlohy (nejen jazykové) a ty musí za daných okolností ve specifickém prostředí a v určitém poli působnosti splnit.“ (SERR, 2002, str. 9)

2.3.1.1 Komunikační kompetence a gramatická kompetence

Nejen Společný referenční rámec ale i moderní vyučovací koncepce se opírají o pojetí gramatiky vycházející z Hallidayovy koncepce funkční gramatiky. Funkční gramatika se systematicky dívá na gramatiku ‚shora‘ jako na zdroj smyslu (Halliday a Matthiessen, 2014, str. 49). Podle tohoto pojetí je předmětem gramatiky výzkum jazykových forem, jež vyjadřují funkce nebo význam. Hallidayova koncepce systematické funkční gramatiky sahá ještě mnohem dále. Systematicky postihnout gramatiku jazyka, určité struktury, neznamena jen postihnout tyto struktury, ale vzájemné vztahy k jiným strukturám (Halliday a Matthiessen, 2014, str. 49).

Zabývat se gramatikou znamená v tomto funkčním pojetí zabývat se funkcemi, textem a jeho významem. Nejde o jednodimenzionální výtažek celého systému, ale o vzájemně související struktury. Taková koncepce nepojednává o sémantice a gramatice jako o oddělených (zvláštních) stránkách jednoho jevu³⁶ (Halliday a Matthiessen, 2014, str. 48).

Pojem komunikační kompetence, která je cílem komunikačně orientované výuky, má více dimenzí či složek. Van Ek (1986), jenž je také autorem anglického *Threshold level* (Ek a Trim, 1980), rozlišuje kompetenci strategickou, sociolingvální, diskurzivní, sociokulturní a společenskou a zařazuje je na stejnou úroveň jako kompetenci jazykovou či lingvistickou.

Referenční rámec považuje právě rozvoj lingvistických kompetencí studenta za ústřední a nezastupitelný aspekt učení se jazyku (SERR, 2002, str. 152). Gramatická kompetence je neoddelitelnou součástí takovéto jazykové kompetence, rozumí se jí „znalost jazykových prostředků určitého jazyka a schopnost je používat“ (SERR, 2002, str. 114). V kapitole 6 ji autoři Referenčního rámce vymezují jako „schopnost uspořádat věty tak, aby vyjadřovaly význam“.

Znalost jazykového systému a zároveň schopnost ji (přiměřeně) používat je tedy neoddelitelnou součástí komunikační kompetence (srov. také Hrdlička, 2009, str. 56). Komunikační vyučovací koncepce proto nevyklučuje v podstatě obrácení se ke

³⁶ Halliday dovává, že popis gramatiky či popis jazyka jako systému je nutně kompromisem, protože se na ni díváme z odlišných vrstev, tedy z vrstvy sémantiky („shora“) a z vrstvy fonologie („zdola“) a zároveň na systém samotný (Halliday a Matthiessen, 2014, str. 48).

gramatice, alespoň u méně radikálních příznivců komunikační metody³⁷. Podle nich se student učí jazyk, aby ho mohl používat, a součástí výuky může být i učení se pravidel jazyka, která se pak ve skutečných situacích aplikují³⁸ (Thornbury, 1999, str. 18).

2.3.1.2 Česká situace

Řada soudobých učebnic češtiny jako cizího jazyka se explicitně vztahuje ke stupnici dovedností SERR, např. *Česky krok za krokem 2* (Holá a Bořilová, 2009a), *Čeština Expres 1 - 3* (Holá a Bořilová, 2010; 2011; 2013), *Tschechisch kommunikativ I a II* (Maidlová a Nekula, 2007; 2013), *Cvičebnice češtiny, čeština jako cizí jazyk* (Sýkorová a Davidová, 2013), *Čeština pro život* (Nekovářová, 2010).

Hrdlička (2009) analyzuje a diskutuje pojetí a aplikaci komunikační metody v oblasti češtiny jako cizího jazyka. K tomu účelu provedl rozbor šestnácti učebnic češtiny pro jinojazyčné mluvčí, jehož výsledky publikoval mezi roky 1962 a 2003. Ve své studii kritizuje, že se gramatika a její role v komunikační metodě ve sféře češtiny jako cizího jazyka pojímá nevhodně. Zaznamenává rozšíření domněnky, že se komunikační metoda nutně vyznačuje minimálním zastoupením gramatiky a že je potřeba ji proto gramaticky obohacovat (srov. Hrdlička 2009, str. 60n.). Některým autorům vytýká nevhodnou a neuváženou simplifikaci gramatiky češtiny a nekritické přenášení některých postupů z výuky analytických jazyků (Hrdlička, 2009, str. 62n.).

Hrdlička opakovaně zdůrazňuje, že ke komunikativnosti nepřispívá obcházení gramatiky nebo dokonce agramatičnost (Hrdlička, 2009, str. 61nn.). Podobně se vyjadřují i Bozděchová a Šašková-Pierce (1995, str. 37): „[...] pokud komunikativní metoda bere v úvahu potřebu a skutečnou nutnost výuky explicitní gramatiky, může být její teorie úspěšně využita ve výuce obtížných syntetických jazyků, jako jsou slovanské jazyky.“

³⁷ Srov. Thornburyho rozlišení dvou myšlenkových škol: přístup zmírněný (*shallow-end approach*) a přístup radikální (*deep-end approach*) (Thornbury, 1999, str. 18).

³⁸ Oproti tomu purističtí zástupci nepovažují přímé zprostředkování gramatiky za nutné, poněvadž předpokládají, že si student osvojí strukturu jazyka, když se bude zabývat komunikačními situacemi, čili přímo během užívání jazyka (Thornbury, 1999, str. 18).

Hlavní – a zdá se – nejtěžší úlohou je smysluplné zařazení výuky gramatiky do komunikačně orientované výuky, zejména u jazyků s bohatou flexí. Vzhledem k tomu, že gramatika sama o sobě není cílem, ale prostředkem k dosažení komunikační kompetence, je zřejmou a metodiky uznávanou skutečností, že se gramatika nelze učit vně kontextu³⁹. Výběr, posloupnost a parcelace gramatické látky se proto řídí v komunikačních kurikulech právě pohledem funkčním, pragmatickým, tedy i tím, s kterými komunikačními situacemi se student pravděpodobně setká, a které by tudíž měl jazykově zvládnout.

Holá a Hanková (2014, str. 1) funkčnímu pohledu dávají přednost před výhradně systémovým úhlem pohledu, jenž zdůrazňuje např. frekvence výskytu gramatických jevů či jejich komplexnost. Navrhují multidimenzionální syllabus umožňující probírat pod jednotlivým tématem slovní zásobu, gramatiku, výslovnost, jazykové funkce a řečové dovednosti vztahující se k danému tématu (tamtéž)⁴⁰. Pokud jde o odstupňování gramatického jevu při jeho zavádění, Holá a Hanková navrhuje, aby se neprobíraly všechny významy a užití jednoho jevu najednou, ale aby se gramatika vyučovala funkčně a ve smysluplném kontextu (Holá a Hanková 2014, str. 2). Proti hrozícímu tříštění pohledu doporučují odkázat na komplexnější gramatickou informaci (tamtéž).

Komunikační metoda má v dnešní době několik odvětví, které pod vlivem konstruktivistických modelů přesunují ohnisko pozornosti na studenta a aktivní vypořádání se s novými jevy různým způsobem. Zaměřují se např. na zabývání se autentickými a smysluplnými úlohami (*task-based language learning and teaching*) nebo na kooperaci při řešení takových úloh (*cooperative language learning*). Na podporu osvojování gramatických struktur se proto nyní často aplikuje metoda indukce. Vyučovací prostředí má nabízet dostatek struktur a příkladů a tak navádět studenty k tomu, aby objevili pravidla užití dané struktury sami. Příkladem takového přístupu je metoda SOS⁴¹ – shromáždit – organizovat – systematizovat (srov. Komárek, 2006, str. 49n.), jež se aplikuje úspěšně i v prostředí češtiny jako cizího jazyka (Holá & Hanková, 2014; Komárek, 2006). Kritické hlasy však při-

³⁹ Srov. např. Bozděchová a Šašková-Pierce (1995, str. 37): „Gramatiku přitom nelze učit formou drilů vytržených z kontextu a memorováním dlouhých přehledů a tabulek. Domníváme se, že gramatika by se měla učit spíše významuplným demonstrováním její funkce v širším a zajímavějším kontextu.“

⁴⁰ Tzv. multidimenzionální kurikula povolují přesunout ohnisko ve vyučovacím procesu uvnitř jedné lekce (jedné vyučovací jednotky) buď na určitou jazykovou funkci, na určitou strukturu nebo na vyučovací strategii (srov. Johnson, 2009, str. 331).

⁴¹ Podle německého: Sammeln – Ordnen – Sortieren.

pouštějí, že přístup osvojování znalostí pozorováním není vhodný, jakmile cílový a mateřský jazyk jsou od sebe dost vzdálené (jako např. čeština pro mluvící analytických jazyků) (Bozděchová & Šašková-Pierce, 1995, str. 38).

Shrnutí

Explicitní výklad či poučení o jazykové soustavě má i v komunikačně orientované výuce češtiny jako cizího jazyka stále silné postavení. To nemusí být v rozporu s ideami komunikačního vyučování, pokud je takový výklad zasazen do komunikačního kontextu. Tzv. multidimenzionální kurikula dovolují zaměřit se na možné realizace daných jazykových funkcí, obecných pojmů a zároveň na lexikální jednotky i gramatické struktury, které jsou s nimi spojené pod jednotlivými tématy nebo komunikačními situacemi.

Klíčové jsou smysluplné odstupňování a parcelace uvádění gramatických jevů. V jazycích s bohatou morfologií se osnovy neobejdou bez určitého zjednodušení nebo dočasného soustředění na jeden aspekt tohoto jevu. To však lze vyvážit explicitním odkazem na ucelenou strukturu nebo ucelené paradigma.

V následující části kapitoly nás zajímá, jak se didakticky přistupuje k odstupňování substantivní deklinace a která kritéria slouží k výběru poslovnosti forem nebo funkcí.

2.3.2 Prezentace pádů a pádových forem v učebnicích

V následujícím rozboru prezentací deklinačních paradigmat v učebnicích češtiny se zaměříme na to, jak přistupují jednotlivé učebnice k prezentaci vzorů, která slova vybírají jako vzorová a jak se komplexní látka parceluje, zjednodušuje a uspořádává. Zároveň se podíváme na to, jak zvládají učebnice přechod od prezentace deklinačního systému pomocí vzorů k zasazení substantiv do situačního kontextu. V neposlední řadě nás zajímá, ve kterých konstrukcích se substantiva procvičují.

Podívali jsme se na 9 učebních sad⁴², které se v současnosti užívají zejména v Německu a v Česku při počáteční výuce češtiny jinojazyčných mluvčích. Zároveň se opíráme o výsledky Hrdličkova rozboru (Hrdlička 2009).

Zvlášť zajímavá z daného úhlu pohledu by byla kvantifikace tvarů, v nichž substantiva v učebnicích vystupují, a analýza jejich distribuce. K tomuto účelu by bylo vhodné sestavit učebnicový korpus, což v rámci naší práce není možné. V současnosti se připravuje projekt, jenž buduje specializovaný korpus sedmnácti učebnic češtiny jako cizího jazyka (Vališová, 2013). Jakmile bude k dispozici, bude možné zkoumání distribuce substantivních tvarů prohloubit.

2.3.2.1 *Prezentace deklinačních paradigmat v učebnicích češtiny*

Komplexnost české deklinační soustavy autoři učebnic češtiny tradičně řeší roztríděním substantiv do (jmenných) deklinačních vzorů. Substantiva se přiřazují ke vzorům na základě zakončení kmene a koncovky v nominativu singuláru⁴³ (Nsg.)⁴⁴.

Postupy jednotlivých učebnic se liší:

1. počtem prezentovaných vzorů
2. výběrem vzorových slov
3. znázorněním vzorů v učebním textu a způsobem přenášení vzorů na další slova.

Ad 1)

Tradičně se uvádí 14 vzorů: *hrad, stroj, pán, muž, předseda, soudce, žena, růže, píseň, kost, město, moře, stavení, kuře*⁴⁵. Podle zaměření jednotlivých učebnic se uvádějí paradigmata vzorů buď uceleně na začátku, nebo na konci v gramatických přehledech. Některé publikace, nejen komunikačně orientované, však zužují výběr ve vysvětlujících textech jednotlivých lekcí na 13, 12, 10 či 9 tříd. Zjednodušení spočívá např. v tom, že se redukuje počet vzorů mužských životných na *pán* a

⁴² Amorová, 1998; Confortiová et al. 2003; Hasil et al., 2007; Hirschmann, 2011; Holá, 2006; Holá & Bořilová, 2010, 2011, 2013; Maidlová & Nekula, 2005, 2013; Remediosová et al., 2002; Sodeyfi & Newerkla, 2002.

⁴³ Koncovka nominativu singuláru může být v případě většiny maskulin a několik feminin i nulová.

⁴⁴ Pro konkrétní pád a číslo používáme v předložené práci následující zkratky: N = nominativ, G = genitiv, D = dativ, A = akuzativ, V = vokativ, L = lokál, I = instrumentál, sg. = singulár, pl. = plurál.

⁴⁵ Např. Sodeyfi a Newerkla, 2002 a Hasil et al., 2007.

*muž*⁴⁶ nebo že se jako zvláštní vzor neuvádí *kuře* anebo *píseň*⁴⁷ či *kost*. Remediosová et al. (2002) nabízejí navíc vzorové slovo *centrum*.

V postupu ke vzorům poněkud vyčnívají publikace *Chcete mluvit česky* (Remediosová et al., 2002) a *Vítáme vás* (Amorová, 1998). Ačkoliv v přehledových tabulkách na konci nabízejí jako referenci paradigmatu klasických vzorů, znázorňují jednotlivé pády na příkladu substantiv, jež vystupují v daném situačním nebo vyučovacím kontextu. V přehledové tabulce o lokálu, který se zavádí nejprve jako předložkový pád s funkcí určení místa nebo polohy, nenarazíme proto na jediný příklad pro maskulina životná. Funkce se exemplifikuje s využitím slov *obchod* [a *stůl*], *pokoj*, *třída* [a *škola*, *fakulta*], *skříň* [a *nemocnice*, *práce*], *kino* [a *divadlo*, *město*] (Remediosová et al., 2002, str. 41). Podobně je tomu u genitivu, jenž se zavádí ve funkci vyjádření směru v předložkových skupinách.

Ad 2)

U tří z devíti učebnicových sad se jako vzorová slova v čele paradigmat ocitají odlišná substantiva (srov. také Hrdlička, 2009, str. 97). Pozorujeme tendence nahradit klasická vzorová slova slovy, s kterými se student setkává v učebních textech, např.:

stroj (1212⁴⁸) → *čaj* (1729), *počítač* (1081); *pán* (mezi 5000 nejčastějšími se nachází *pan* s pozicí 199) → *student* (1175); *žena* (204) → *káva* (2205); *růže* (3655) → *židle* (1684), *restaurace* (1591); *stavení* (není mezi 5000 nejfrekventovanějšími slovy) → *nádraží* (2228).

Volba se zdá být libovolná⁴⁹, respektive se nikde explicitně nezduvodňuje. K obdobnému úsudku dochází Hrdlička ve svém rozboru (tamtéž). Jak ukazují frekvence, často ale (nikoliv vždy) jsou zvolené reprezentanty jednotlivých vzorů frekventovanější než tradiční vzorová slova.

Ad 3)

Zjistíme, že všechny zkoumané učebnice podávají měkké a tvrdé vzory současně, s výjimkou učebnice *Tschechisch. Faszination der Vielfalt*⁵⁰. Přehledy konco-

⁴⁶ Např. Maidlová a Nekula, 2005 a Confortiová et al., 2003.

⁴⁷ Remediosová et al., 2002.

⁴⁸ Rank slova podle *A Frequency Dictionary of Czech. Core Vocabulary for Learners* (Čermák & Křen 2011).

⁴⁹ Jsme si vědomi, že i klasické třídění substantiv a zároveň výběr reprezentanta třídy – vzorového substantiva – je do jisté míry libovolné (srov. MSC 2010, str. 143).

⁵⁰ Autoři nejdříve uvádějí pády na příkladu substantiv, která se skloňují podle tvrdých vzorů a postupně je doplňují o ostatní vzory (Sodeyfi a Newerkla, 2002).

vek jednotlivých vzorů pro probraný pád se strukturují nejprve podle rodu a v druhé řadě podle povahy kmenové finály (tzv. měkké vzory a tvrdé vzory). Na rozdíl narazíme teprve při skloubení vzorů a učebního textu. Učebnice, které se snaží znázornit určitý pád pro všechny vzory, nevyhnutelně podávají celé funkční spektrum pádu, aby se našel vhodný příklad pro každý vzor. Přehledová tabulka je doprovázena explikací funkcí za pomoci dvou až tří modelových vět nebo konstrukcí, které se vyplní slovy typickými pro danou funkci (srov. Maidlová a Nekula, 2005; Confortiová et al., 2003; Hirschmann, 2011). Hirschmannová (2011, str. 30nn) k tomu účelu nabízí zvláštní tabulku s modelovými větami omezenými na dvě až tři konstrukce (např.: **Znám** toho mladého inteligentního pána. **Vidím** ten velký moderní stroj.) Konstrukce se explikuje většinou na příkladu vzorového slova plus jednoho dalšího, např. vlastního jména nebo jiného, často frekventovaného slova (*taška, kolo*). Přenášení vzorů na další slova se uskutečňuje procvičováním slov v daných konstrukcích, slova se podávají typicky v Nsg. V komunikačně orientovaných učebnicích *Tschechisch kommunikativ 1* a *2* předchází prezentaci gramatické látky řízené pozorování struktur, jež dominují v předem podaném textu. Následující cvičení se už se vrací k tematické slovní zásobě a modelovým konstrukcím sloužícím plnění komunikačního záměru⁵¹.

Pro všechny tři publikace musíme konstatovat, že explikují funkce pádu jen na omezeném počtu příkladů, a tudíž se podává studentům k analýze jen velmi málo příkladů substantivních tvarů vyjadřujících daný pád (často po tabulce a funkčním přehledu následují cvičení jen s jedním modelovým řešením úkolu). Tvary musí studenti ve cvičení sami utvořit podle vzorů. Podobný postup k prezentaci vzorů a jejich procvičení nacházíme v učebnici Hasila et al. (2007) *Brána jazyka českého otevřená*. Poněvadž učebnice není určená úplným začátečníkům, je substantivní skloňování shrnuté v jedné lekci. V lekci věnované tématu *výlet* jsme se ale setkali i s prezentací dvou pádů, a to spíš z lexikálního hlediska: dvě sémanticky blízké funkce předložkového G a L se znázorňují a následně procvičují na příkladu slov s významem místa (města, hory a další místa, které mohou sloužit jako cíl výletu). Nehledě k přináležitosti slov k nějakému vzoru se podávají jejich tvary

⁵¹ Maidlová a Nekula (2005) uvádí substantivní skloňování na základě modelových konstrukcí, které prototypicky realizují daný komunikační záměr nebo jsou typické pro danou situaci. V předmluvě autoři detailně poukazují na to, že se důraz klade v první řadě na sloveso jako centrální složku věty. Kategorie pádu se postupně uvádí na základě modelových konstrukcí (Maidlová a Nekula 2005, str. V).

v G a L v „hotových“ konstrukcích s předložkou (Hasil et al., 2007, str. 28). Autoři vybrali velká a známá místa či města, a tím i běžně užívaná.

Výše jsme se již zmínili o postupu Remediosové et al. (2002). Autorky předkládají vždy jen jednu funkci pádu a soustředují se v tabulkových přehledech na vzory, které tuto funkci typicky vyjadřují. Cvičením, která pobízejí studenty k samostatnému tvoření tvarů, vždy předchází prezentace konstrukcí, které vyžadují daný pád, v minimálním kontextu (často krátký rozhovor) a cvičení, ve kterém má student přímo užívat již skloňované tvary, např. při reakcích na otázku (*Máte dceru? Ano, mám ...; Díváš se na učitelku? ...*) nebo při pouhém imitování (*Piju kafe s: mlékem | cukrem | koňakem. Do kina jdu s: manželkou | dcerou a synem | svým kolegou*)⁵². Výsledkem takového postupu je, že se vzory maskulin životných z přehledů o Asg. nebo o Npl. vyjímají a podávají zvlášť.

Střední cestu volí Holá (2006) tím, že úvodem do lekce prezentuje nejdříve celé funkční spektrum jednoho pádu a příslušné morfologické tvary vzorů. Samotná lekce se soustřeďuje na komunikační situaci a její realizaci. Jakmile dochází k objevování pravidel za pomoci příkladových konstrukcí, autorka obrací pozornost studenta už na slova, jež se v daném kontextu skutečně užívají (např. akuzativ se v lekci věnované návštěvě restaurace uvádí na příkladu slov *káva, rýže, jogurt, mléko* apod., srov. Holá 2006, str. 57nn.). Navíc podporuje spojení slov s jejich vzory (alespoň na úrovni rodu a životnosti) pomocí barevného označení (např. maskulina životná = tmavomodrá, maskulina neživotná = světlomodrá, apod.), které dodržuje v gramatických vysvětleních a ve slovníkové části.

Ve své nejnovější publikaci dávají Holá a Bořilová (2011) studentům více prostoru na objevování struktur na základě jazykového materiálu z textu. Tabulky obsahují mezery, do kterých si studenti zanášejí koncovky sami po systematizování za pomoci příkladů. Na konci explikace je odkaz na gramatická vysvětlení a tabulkové přehledy v příloze. Pádové tvary vidí a slyší studenti na několika místech v lekci (v dialozích a cvičeních) před tím, než jsou vyzváni k tvoření vět na základě tvarů v Nsg.

⁵² Tato cvičení často sledují princip drilu. Příklady podané v krátkých rozhovorech jsou pro studenty užitečné, někdy však schází zasazení do kontextu lekce.

2.3.2.2 Posloupnost uvádění pádů a čísla substantiv

V naprosté většině současně užívaných učebnic se aplikuje při prezentaci pádů tzv. *horizontální postup*, tj. skloňování substantiv a adjektiv se předkládá postupně po jednotlivých pádech⁵³. Hrdlička se zmiňuje o pěti učebních sadách, jež pracují se souhrnným zavedením všech pádových forem najednou (srov. Hrdlička 2009, str. 91), jedná se vesměs o publikace vydané před rokem 1999. V našem vzorku se vyskytuje jen jedna učebnice, která takový přístup volí⁵⁴. Hrdlička soudí i ze svých dlouholetých zkušeností z vyučovací praxe, že oba postupy mohou vést k dobrým výsledkům (Hrdlička 2009, str. 93). Musíme konstatovat, že nemáme oporu pro jeden ani pro druhý postup v empiricky získaných datech. Hrdlička si vysvětluje dominanci postupu horizontálního tím, že autoři učebnic považují tento přístup za méně náročný hlavně pro Neslovaný (tamtéž).

Zároveň nemáme informace o tom, na kterých principech autoři stavějí posloupnosti zavedení pádů. Hrdlička ve svém rozboru učebnic nezjistil žádnou výraznou tendenci upřednostňování určité posloupnosti, a to ani pokud jde o pořadí prezentace singuláru a plurálu (Hrdlička, 2009, str. 90). Konstatuje, že při oddělené prezentaci jednotlivých pádů se často podávají paradigmatu jednotného a množného čísla zvlášť, a plurálová méně detailně (tamtéž). Obě zjištění můžeme na svém vzorku potvrdit. Zdá se, že řazení pádů podléhá celé řadě kritérií, z nichž jedním, ne však řídicím, je frekvence.

Když porovnááme posloupnost uvedení pádů (s přihlédnutím ke gramatickému číslu) v učebnicích s frekvenční distribucí v korpusu psané češtiny (SYN2005, viz tab. 2.6) zjišťujeme následující diskrepance:

- Genitiv, který má pozici druhou v řazení v psané češtině, se ocitá jen ve dvou učebnicích na druhé nebo třetí pozici, ve zbylých je odsunutý na čtvrtou nebo i další pozici ve prospěch akuzativu nebo lokálu.
- Akuzativ singuláru často předchází genitivu singuláru, což by bylo možné připsat a) skutečnosti hojnějšího výskytu akuzativu v mluvené češtině a b) skutečnosti, že z didaktického hlediska akuzativ lépe navazuje na nominativ, od kterého se liší jen u dvou vzorů.

⁵³ Termín *horizontální přístup* přebíráme od Hrdličky (2009). Proti horizontálnímu postupu staví *vertikální postup*, podle kterého se deklinační paradigmatu probírají komplexně, podle jednotlivých vzorů (Hrdlička 2009, str. 91).

⁵⁴ Hasil et al. (2007). Publikace se liší se od ostatních tím, že je určena mírně pokročilým studentům. Byla koncipována především pro potřeby pražské Letní školy slovanských studií jako materiál pro intenzivní kurz.

- Lokálu singuláru se ve čtyřech z osmi učebnic dává přednost před častějším akuzativem a genitivem a probírá se už jako druhý v pořadí. Předpokládáme, že je to dáno tím, že se poměrně brzo klade velký důraz na vyjádření prostorových vztahů.
- Genitiv plurálu je často upozaděný, v 6 z 8 sad se vůbec nevyskytuje nebo až v druhém dílu, pokud je tento díl k dispozici.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
Nsg.	Gsg.	Asg.	Lsg.	Gpl.	Isg.	Npl.	Apl.	Dsg.	Lpl.	Ipl.	Dpl.	Vsg.	Vpl.
23,0	18,5	14,2	9,7	8,9	6,6	6,0	5,0	2,7	2,1	1,9	0,92	,035	0,001

Tab. 2.6 Rozložení gramatického čísla a pádu u substantiv v procentech výskytů v korpusu SYN2005 podle Statistik češtiny (Bartoň et al. 2009, str. 134), řazené podle frekvence

2.3.2.3 Možnosti redukce komplexnosti a alternativní přístupy k prezentaci deklinačních paradigmat

I přes to, že se substantivní skloňování zprostředkovává po jednotlivých pádech, jsou studenti vystaveni velkému množství forem, které mohou vyjadřovat daný pádový význam. Autoři učebnic v našem vzorku uplatnili různé metody, aby tomuto problému čelili a vhodně odstupňovali komplexní tvarový systém české deklinace:

1. uvádějí jen jednu funkci pádu (zejména u lokálu a genitivu) a spirálově se vracejí k pádu v pozdějších lekcích, popř. uvádějí další vzory (Holá, 2006, Holá a Bořilová, 2011, Confortiová et al., 2003; Remediosová et al., 2002; Sodeyfi a Newerkla, 2002)
2. probírají funkce pádu jen v daném kontextu, a tudíž jen v rámci omezeného počtu relevantních nebo častých slov (Holá 2006; Holá a Bořilová, 2011)
3. v prvních lekcích procvičují méně časté vzory a frekventovaná substantiva, která se podle těchto vzorů skloňují, jen v kolokacích (Holá 2006; Holá a Bořilová, 2011)

Jako příklad ad 2 a 3 uvádíme Holou (2006). Lokál zavádí poprvé ve 4. lekci, a to výlučně v předložkových výrazech jako *ve škole, v práci, v České republice, v hotelu*. Odkazuje na gramatický přehled a podrobné vysvětlení ve 14. lekci. Tvary akuzativu jsou v 5. lekci zasazené do kontextu návštěvy restaurace. V téže lekci se procvičují také hojně kolokace slovesa *mít* a tím i frekventovaná substantiva méně častých vzorů: *mám dovolenou, mám žízeň, mám rande*.

Holá a Bořilová se ve svých novějších publikacích *Čeština expres 1-3* a *Česky krok za krokem 2* navíc pokusily o odstupňování systému deklinačních tříd. Navr-

hují systém 3 obecnějších deklinačních skupin, které se podle koncovky v nominativu singuláru a podle rodu dělí opět do 12 menších podskupin. Tradiční dělení na měkká a tvrdá substantiva se v podstatě přejímá – substantiva končící na tvrdou souhlásku, na *-a* a *-o* jsou zachycená v první skupině, substantiva zakončená na měkké souhlásky (*-š, -č, -ř, -ž, -ň, -d', -t', -c, -tel*) či na *-e/ě* a feminina na *-del* a *-ev* tvoří druhou deklinační skupinu. Třetí je vyhrazená neutřím vzoru *kuře, nádraží*, maskulinům životným na *-a* a femininům zakončeným na *-st*. Jako výhodu autorky uvádějí zařazení do deklinačních skupin podle koncovek v N sg., které v přehledech pro studenty přímo pojmenovávají (nemluví se o tvrdých či měkkých vzorech). Vydělení druhé deklinační skupiny má napomáhat studentům zejména při osvojování koncovek genitivu, dativu a lokálu singuláru a také akuzativu a lokálu plurálu (Holá a Bořilová, 2009b). Za největší přínos však považují právě možnost parcelace výuky pádů i z hlediska fázování výuky a látky podle úrovně SERR⁵⁵: „Např. na úrovni A1 můžou studenti věnovat pozornost osvojování substantiv z I. a II. deklinační skupiny, na úrovni A2 se lze zaměřit i na III. deklinační skupinu, zatímco úroveň B1 již může akcentovat nácvik všech tří deklinačních skupin.“ (tamtéž)

Navržené dělení zrcadlí rovněž zjištění o frekvenci koncovek substantiv (v nominativu) v současné češtině⁵⁶. Podle Holé a Bořilové (2009a, str. 10) zahrnuje první deklinační skupina substantiva s typickými koncovkami. Představují 66 % všech českých substantiv. 34 % substantiv má netypické koncovky a naplňuje zbylé dvě skupiny. Autorky v této souvislosti mluví o koncovkách majoritních a minoritních (netypických). I neutra na *-í* se řadí k minoritním koncovkám, lemma ta s touto koncovkou se však vyskytují mnohem častěji než neutra zakončená na „typické“ *-o*⁵⁷. Až na tuto výjimku má předložené pravidlo oporu v datech, a rozdělení může, dle našeho názoru, pomoci studentům soustředit se na typické jevy a snadno nalézt paralely mezi pádovými koncovkami i přes rozdílné rody a vzory.

⁵⁵ Na rozdíl od Hrdličky, který považuje právě dlouhé časové rozpětí, v němž se probírá český deklinační systém, za nevýhodu horizontálního přístupu (Hrdlička, 2009, str. 92).

⁵⁶ Autorky postavily koncepci na průzkumu Českého národního korpusu 2005.

⁵⁷ Podle *Statistik češtiny* tvoří substantiva zakončená na *-í* největší skupinu v rámci rodu. Představují 75 % všech neutrálních lemmat, kdežto substantiva vzoru *město* mají podíl jen 20,9 % (Bartoň et al. 2009, str. 139).

Nakonec se chceme ještě zmínit o svérázném přístupu k prezentaci a popisu deklinačního systému v pedagogické gramatice určené jinojazyčným mluvčím. Zdá se nám být zajímavý proto, že obdobně bere ohled na typické a frekventované typy koncovek. Poldauf a Šprunk (1968) navrhnou systém, jenž se opírá o distinkci základních tzv. „bázových“ tvarů vs. jejich „zvláštní“ podoby. Jejich přístup se pokouší nalézt typy tvarů na základě jejich tvarové podobnosti a vysvětlovat zvláštní koncovky odlišnými funkcemi nebo příslušností substantiv k různým (sémantickým) třídám. Student musí ovšem při vědomém tvoření gramatických tvarů podle předloženého systému přihlížet k celé řadě parametrů. Navíc z tohoto přístupu vyplývá celá řada dalších vzorů, které jsou podle navrženého systému výjimkami, jako např. neobvyklé podoby bázového tvaru, smíšené a adjektivní skloňování (k nimž patří mimo jiné slova *týden, kámen, zvíře, den, téma, muzeum, řeč, tvář, správce, kolega, láhev, centrum*) a řada nepravidelných a nesklonných substantiv, mezi nimiž se nalézají slova s častým výskytem: *dcera, dveře, bratr, člověk, dítě, host, nebe, oblak, rodič, ruka, oko, noha, paní, úterý*.

Shrnutí

Představené přístupy se pokoušejí o třídění komplexní soustavy českého skloňování z didaktického hlediska a z hlediska lepší parcelace výuky. Rozmanitost morfémů, které mohou vyjadřovat pádové funkce, řeší tím, že zaměřují pozornost vždy jen na jednu funkci pádu a tu probírají v daném kontextu anebo v rámci jen omezeného počtu relevantních nebo častých slov. Několik málo učebnic zařazuje z hlediska systému méně typické jevy, jako zástupce méně frekventovaných deklinačních tříd, do počáteční výuky tím, že nabízí a procvičuje jejich tvary v kolokacích či předložkových skupinách. Jinde se s nimi zachází v podobě lexikální poznámky.

Zvolená vzorová slova, jako zástupci deklinační třídy, slouží studentům hlavně jako reference a bez zasazení do konkrétní konstrukce nemohou plnit funkci prototypu, o které byla řeč v části 2.1. Explikace funkce pádu na konkrétních příkladech a v konkrétních konstrukcích často, nikoliv vždy, postrádá dostatek různých substantiv, a tudíž pádových forem. Studenti si tak těžko mohou odvodit prototypického zástupce celé konstrukce nebo zapamatovat správný substantivní tvar v kontextu. Na druhé straně jsme příležitostně pozorovali, že učebnice obracejí pozornost studentů záměrně na substantivní tvary (jiné než nominativní) v určitých

tematických kontextech nebo v konstrukcích vázaných na určitou jazykovou funkci, jako např. kolokace slovesa *mít* a *mít rád* nebo předložkové konstrukce k vyjádření data nebo cíle cesty či výletu. Systematicky kolokacím věnují prostor v lekcích jen Holá (2006) a Holá a Bořilová (2011).

Zdá se, že vodítkem pro volbu nebo upřednostňování vzorového slova nebo pádu je především frekvence, ale důležitá je i řada dalších faktorů. Můžeme konstatovat, že v rámci učebnice, která chce současně systematicky zprostředkovávat deklinace češtiny a brát ohled na smysluplné uspořádání a zasazení pádů do komunikačního kontextu, není možné činit frekvenci prvním kritériem volby vzorů, konstrukcí relevantních pro daný pád a posloupnosti zavedení pádu.

Je však možné brát ohled na substantiva, která sdílejí podobné konstrukce a zároveň podobné sémantické rysy. K tomuto účelu je nutné zjistit rozložení jejich pádových tvarů. Informace o frekvenci tvarů konkrétních substantiv lze zohlednit například při prezentaci forem ve slovníkové části, ale zároveň ve cvičeních věnovaných nácviku pádových tvarů v konstrukcích s daným pádem.

2.3.3 Výběr substantiv a jejich forem

Výběr lexikálních jednotek pro učebnice se řídí ve funkčním kurikulu mj. jejich významem pro komunikační situace a funkce. Je tedy možné přiblížit se tématu výběru pádů a pádových forem i ze strany lexikální. Sinclair a Renouf (1988) zkoumají rozložení lexikálních jednotek a jejich forem v učebnicích angličtiny jako cizího jazyka. Kritizují, že přístup autorů kurikul a učebnic k lexikonu je často nesystematický, a že lexémy slouží převážně ilustraci jiných jazykových jevů. Zjistili, že počet uvedených slovních tvarů jednoho lexému se ve zkoumaných učebnicích značně liší a obdobně se liší i průměrný počet jejich opětovného výskytu (Sinclair a Renouf 1988, str. 141).

Kognitivní a aplikovaná lingvistika jsou zajedno v tom, že se vyučování nemůže soustředit jen na jednotlivá slova, ale že musí přihlížet i k víceslovným lexémům, ustáleným kolokacím a konstrukcím (*chunks* či *formulaic sequences*⁵⁸), jichž jsou

⁵⁸ V posledních letech je mezi aplikovanými lingvisty rostoucí zájem o roli tzv. *formulaic sequences*. Podle Wraye je formulaic sequence: “[...] a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray, 2002, str. 9). Synonymně se používá pojem *chunks*, který preferují i

součástí (např. Robinson a Ellis, 2008; Kennedy, 2008; Lewis, 2002; Boers et al., 2010; Wray, 2002). Je akceptovanou skutečností, že frekvence zkušeností se slovy, s jejich slovními tvary v kolokacích a konstrukcích ovlivňuje proces učení. Zároveň se předpokládá, že jevy vysoce frekventované jsou i užitečné pro „budoucího“ uživatele jazyka. Vzhledem k omezenému času ve třídě se tak zdá rozumné, aby se vyučování koncentrovalo právě na frekventovaná slova a jejich kolokace (srov. např. Kennedy, 2008). Sinclair a Renouf (1988) navrhují pro (tehdejší) absenci jiných zásad, že by se student angličtiny měl soustředit:

- a) na nejběžnější slovní tvary
- b) na jejich centrální vzory užívání
- c) a na jejich typické kombinace

(Sinclair a Renouf, 1988, str. 147)

Soustředit se v řízeném vyučování na nejčastější jevy implikuje, že se osvojování středně až málo frekventovaných jevů víceméně ponechává náhodnému setkání s nimi mimo lekce a učební materiály. Pokud jde o výběr kolokací ve smyslu *chunks* čili *formulaic sequences* pro vyučování, argumentují Boers et al. (2010), že pravděpodobnost, že se student setká s novými kolokacemi mimo třídu, je právě u vysoce frekventovaných konstrukcí větší, a proto by se mělo vyučování soustředit také na středně frekventované jevy (Boers et al., 2010, str. 242).

Ačkoliv vysoká frekvence v inputu může kladně ovlivnit to, že si studenti nějaký jev, ať gramatickou konstrukci či určitou lexikální kombinaci (kolokaci), zapamatují, není univerzálním lékem. Studie, jež se věnovaly vlivu zadání na zpracování nových kolokací a konstrukcí při práci s textem, prokázaly, že si čtenář při četbě zaměřené především na porozumění textu obtížně uvědomuje neznámé kolokace a gramatické konstrukce (pro přehled srov. Schmidt 1990). Podle Schmidta (1990) je však uvědomění a všímání si jevu předpokladem pro to, aby ho studenti později aktivně zvládli. Rovněž Robinson a Ellis (2008) potvrzují pozitivní vlivy, jež má obrácení pozornosti studentů na formální jevy při zpracovávání a vyhodnocování textů.

Boers et al. Vymezuji ho jako: “[...] (semi-)fixed expressions of many different sorts, including strong collocations (e.g. *commit a crime*), social routine formulae (e.g. *Have a nice day*), discourse markers [...], compounds (e.g. *peer pressure*), idioms (e.g. *take a backseat*), standardised similes [...], proverbs [...], genre-typical clichés [...], exclamations [...], and more. (Boers et al., 2010, str. 239n.)

2.3.4 Shrnutí

Jak jsme viděli, není – zvláště pro jazyk s bohatou flexí – jednoduchým úkolem vyhovět při výběru substantiv pro učebnice jak funkčnímu hledisku, tak smysluplnému uspořádání pádů a jejich parcelaci. Frekvence lemmatu samotného, ale i frekvence kolokací jeho tvarů by měla být jedním z předních kritérií výběru, ne však jediným⁵⁹.

Závažné je, že vyučovací koncepce a osnovy učebnic přistupují k zavedení nejběžnějších lemmat a jejich tvarů s přihlédnutím k různým vztahům lexikální jednotky jak v rovině systému, tak v rovině textu. Podle našeho názoru by se osnovy učebních materiálů měly pokusit odrážet zejména následující:

- vztahy ke komunikačnímu kontextu: kontexty, ve kterých se substantiva a jejich tvary typicky vyskytují (pohled funkční)
- vztahy k textu: pádové funkce, které substantivum typicky vyjadřuje (pohled syntaktický)
- vztahy k jiným lexikálním jednotkám:
 - její typické kolokace (pohled syntagmatický)
 - vztahy ke slovům stejného kolokačního paradigmatu (pohled paradigmatický)
- vztahy uvnitř deklinačního systému: její příslušnost k některé z deklinačních tříd (pohled systémový)

Tím se předchází nebezpečí, že učební materiály obracejí pozornost studentů na tvary substantiva přednostně v paradigmatických tabulkách a že tvary slouží hlavně k explikaci gramatických kategorií.

Závažným úkolem učebnic je ukázat studentovi užití podstatných jmen v různých kontextech. Kurikula musí látku nutně parcelovat a vybírat. V počáteční výuce se nabízejí situace nejběžnější pro cílovou skupinu. Podstatná jména se zde mohou prezentovat v komunikativních kontextech, které jsou pro ně typické, a to rovněž v příslušných tvarech a kolokacích. Výběr textů a formulace úloh by měly studenty podporovat v tom, aby se soustředili jak na funkci tvaru v kontextu, tak na tvar samotný a na jeho roli v systému. Na základě výše řečeného považujeme za důležité, aby se funkce pádu vyjasnila na typických konstrukcích a na dostatku konkrétních příkladů, tedy pádových forem v kontextu. K tomuto účelu může být užitečné identifikovat typické chování substantiv v textu, rozložení jejich tvarů v korpusu, a najít podobnosti mezi lexémy na základě dat.

⁵⁹ Vrátime se k této otázce v kapitole 5.

3 Korpusový přístup k popisu substantiv

Tato kapitola se věnuje teoretickým předpokladům předkládaného postupu. Ve své práci se pokoušíme zachycovat a analyzovat užití substantivních forem na základě korpusových dat. Volíme korpusový přístup, protože nám dovoluje prozkoumat velké množství substantivních lexémů⁶⁰ a jejich užití v textu, tedy distribuci jejich slovních tvarů. Vycházíme v této souvislosti z toho, že frekvence výskytu nějakého tvaru je indikátorem jeho typičnosti. Jsme si vědomi, že se typičnost nerovná automaticky relevantnosti, frekvence nám však v rámci této didakticky orientované práce může pomoci identifikovat vzorce užití, které by se měly zrcadlit v učebních materiálech určených širšímu publiku. Dále předpokládáme, že kontext⁶¹ slovního tvaru, který můžeme prostřednictvím korpusu prozkoumat, přispívá k poznání jeho významu a funkce a že je potřeba analyzovat jednotlivé tvary jednoho lexému v této souvislosti zvlášť. Úzus sledovaný v korpusu nám dovoluje porovnávat chování jednotlivých substantiv v textu. Budeme zjišťovat distribuci jejich tvarů a zkoumat, zda se na základě sdílené preference určitých pádových tvarů (a pokud možno preference obdobných kontextů) dají rozřadit do nových paradigmat. Pokládáme si otázku, zda obdobné chování lexémů v textu, tedy jejich realizace konkrétními slovními tvary, alespoň částečně odráží obdobné sémantické rysy substantiv.

3.1 Korpusový přístup ke gramatice

Tradičně se v lingvistice pohlíželo na lexikon a gramatiku (zejména na syntax) jako na dva rozdílné aspekty jazyka. V rámci gramatiky flexivních jazyků je slovo vnímáno jako součást různých formálně nebo funkčně vymezených tříd, a tudíž zejména jako nositel gramatických morfémů, které vyjadřují např. slovo tvorný nebo deklinační typ. Nezávisle na tom se jeho obsahovou stránkou zabývá lexikon, jenž se věnuje lexému jako celku bez ohledu na případná rozdílná užití jeho jednotlivých slovních tvarů.

⁶⁰ Lexémem rozumíme abstraktní jednotku lexikonu, která zahrnuje všechny své konkrétní textové tvary, v případě substantiva celé paradigma. Lemma je jeho konkrétní reprezentativní podoba, jež je charakterizovaná jistou gramatickou formou, v případě substantiv zpravidla nominativu singuláru. (srov. EŠC, Karlík, 2012, str. 242). Přikláníme se k definici pojmu *lexém* u Čermáka (2010, str. 17), podle níž lexém může být formálně jednoduchý nebo složený (víceslovný), a to na rovině textu (např. jít → *šel bych*) nebo i systémově (např. *kyselina siřičitá, vzít do zaječích*).

⁶¹ Pod pojmem *kontext* zde chápeme přímé textové okolí hledaného slova oproti širšímu Firthovu pojmu *situačního kontextu* (*context of situation*), srov. (Firth, 1969).

Učební materiály angličtiny, které se zvláště zabývají tzv. *phrasal verbs*, už brzy pronikaly do přechodného pásma mezi gramatikou a lexikonem, kdežto v učebních programech a učebnicích flexivních jazyků gramatika přetrvává stále jako nezávislá část a příruček i cvičebnic vycházejících prvořadě z lexikonu je málo. Mimo jiné se jako důsledek tohoto dělení jeví gramatika jako sada (omezených) pravidel, jež se aplikují na (naučenou) slovní zásobu za cílem vytvořit gramaticky správné věty. Je pochopitelné, že se výuka uchyluje k jasně formulovaným pravidlům, která pomáhají redukovat komplexní systém jazyka a „zajišťují“ produkci gramaticky správných vět, do nichž se vsazují slova dle aktuální potřeby. V didakticky orientovaném popisu jazykových jevů se tak, zdá se, zrcadlí Chomského základní distinkce mezi kompetencí (*competence*) generovat na základě znalosti jazykových pravidel neomezenou sadu gramaticky správných vět, popř. rozhodovat o gramatičnosti vět, a produkcí i interpretací individuálních promluv v konkrétních situacích (*performance*). V nejrůznějších lingvistických kruzích bylo diskutováno, zda Chomského distinkce poskytuje dostatečnou základnu pro to, aby bylo možné přiměřeně objasnit a popsat užití jazyka a „gramatičnost“. Lingvisté, kteří pracují s jazykovými korpusy, mohou na jejich základě odhadnout, které z potenciálně *možných* struktur a vět se skutečně „generují“, popř. užívají. Narážejí na dost nevyvážený poměr *competence* a *performance*. Obrovské množství dat o jazykovém úzu jim povoluje ověřovat a vyvracet hypotézy o platnosti jazykových pravidel, poznávat tendence i typické užití a vyvozovat závěry o pravděpodobnosti, s níž se konkrétní lexikální jednotka objevuje v určitém kontextu. Ohnisko při vytváření jazykových modelů se tak za vlivu korpusů posunulo od *možného* (toho, co je dáno v podstatě systémem) k *pravděpodobnému* (srov. Hanks, 2000, str. 4).

Jazykové korpusy poskytují velké množství informací o textu (*parole*). Disciplínou, která data přirozeného jazyka systematicky vytěžuje a analyzuje, je korpusová lingvistika. Čermák (2006, str. 9) tvrdí, že povaha poznání, která získáme o jazyce pomocí korpusového výzkumu, je hlubší a přesvědčivější než cokoliv, co bylo k dispozici v době předkorpusové. Korpusová lingvistika je schopná lépe vymezovat jevy syntagmatické povahy a studovat problematiku hranice mezi lexikonem a gramatikou. Čermák zároveň naznačuje, že souhrn výskytů jednoho jevu (slova) v korpusu sice může odrážet, jaké kolokace preferuje a které významy jsou typické – tedy jádro – a co lze naopak pro velice omezený nebo individu-

ální výskyt označit za periferii, žádné ostré hranice však mezi jednotlivými typy kolokací a významy nejsou (Čermák, 2006, str. 14). Studované jevy a výsledky je třeba vidět na sestupné, popř. vzestupné škále (tamtéž).

Korpus odráží reálné užívání jazyka k nejrůznějším cílům, jazykové jednání lidí, přirozený jazyk. Korpusová lingvistika tak přispívá k poznání funkce jazyka jako nástroje komunikace. Vzhledem k tomu je veškerý korpusový výzkum nutně funkční, a tudíž přínosný pro lingvistickou analýzu, která je založená na skutečném jazykovém úzu (Meyer, 2002, str. 11). Bereme-li v potaz, že je cílem komunikační výuky rozvíjet způsobilost jazykově zvládat a řešit úkoly a situace v cizojazyčném prostředí, měl by být korpusový přístup k popisu jazykového úzu přínosný i pro didaktické účely. Při posuzování modelů, které vznikají na bázi korpusových dat, je ovšem nutno mít na zřeteli, že mohou popsat jen to, co je v korpusu obsaženo a dokumentováno (srov. Mahlberg, 2005, str. 32).

Z pohledu korpusové lingvistiky je mnohem zřejmější, že popis jazykového systému jako sady pravidel na jedné straně a inventáře významů na druhé straně není přiměřený. Hallidayova systémová funkční gramatika vychází z předpokladu, že gramatika je reprezentována jako komplexní síť systémů, nikoliv jen jako inventář struktur (Halliday, 2007, str. 22). Halliday pohlíží na gramatiku a lexikon jako na dva konce kontinua, dva pohledy na tentýž komplexní jev. Gramatika přitom tvoří „hlubší“ konec, tj. výběr z velice obecných významů jednoho paradigmatu je velice omezený (Halliday, 1991, 32n., uvádí jako příklad polaritu, slovesný rod a slovesný způsob). Tvrdí proto, že stejně jako u jevů lexikálního rázu nám povolí i frekvence abstraktních gramatických významů činit závěry o jejich pravděpodobnostech, jež mohou být signifikantní pro celý systém (tamtéž)⁶². Pravděpodobnost nacházíme tedy i v gramatice, i povaha jevů gramatických paradigmat je škálovité povahy. Není tudíž vždy možné jasně a ostře oddělit gramaticky správné užití od nesprávného⁶³. Právě korpusový přístup ke gramatice pomůže odhalit typickou realizaci gramatických významů, ale i přechodná pásma. Z hlediska přístupu řízeného korpusovými daty (*corpus-driven*) oproti přístupu na korpusu založenému (*corpus-based*) lze však proti Hallidayovu paradigmaticky orientovanému přístupu k popisu jazykových jevů namítat, že přes přihlídnutí k jeho škálovité charak-

⁶² Halliday takto ztotožňuje frekvenci výskytu nějakého jevu s jeho pravděpodobností.

⁶³ Srov. k tomu i Hanks (2000), podle kterého je i gramatičnost škálovité povahy (Hanks, 2000, str. 4).

teristice zanedbává právě souvztažnost mezi lexikonem a gramatikou (srov. Tognini-Bonelli, 2001, str. 76).

Za korpusový a zároveň lexikální přístup ke gramatice se přimlouvají i Francisová a Hunstonová (Francis, 1993 a Hunston a Francis, 2000). Francis vytýká tradičním deskriptivním gramatikám, že se málokdy pokusí spojit danou strukturu s konkrétními lexikálními jednotkami, tj. že se zřídka uvádějí lexikální exponenty strukturních prvků (s výjimkou slovesných doplňů) (Francis, 1993, str. 141). Přitom korpusová data jasně ukazují, že se určité syntaktické struktury jako např. v angličtině <sloveso + *it* + ADJ> vyskytují jen s omezeným počtem lexikálních jednotek, že jsou tedy zvoleny jen určitými slovesy (Francis, 1993, str. 141). Francis proto zastává názor, že se syntaktické struktury a lexikální jednotky vybírají souběžně a že je nelze chápat jako navzájem nezávislé (tamtéž). Vzájemnou podmíněnost syntaktických struktur a lexikálních jednotek se pokouší Hunstonová a Francisová (Hunston a Francis, 2000) prozkoumat a zachytit právě pomocí přístupu řízeného korpusem, tj. jen s minimálními teoretickými předpoklady.

Za účelem sestavit deskriptivní, pedagogickou gramatiku angličtiny s využitím korpusu si kladou za cíl „zachytit všechny význačné lexikální jednotky z hlediska jejich syntaktických preferencí a všechny gramatické struktury z hlediska jejich klíčového lexikálního obsazení a frazeologie“ (Francis, 1993, str. 153). Výsledkem je tzv. *pattern grammar*, tj. gramatika založená na vzorcích jednotlivých slov, přičemž vzorcem slova rozumějí „všechna slova a struktury, které se s ním běžně asociují a které přispívají k jeho významu“ (Hunston a Francis, 2000, str. 37). Identifikace a vymezení takového vzorce (*pattern*) však nejsou přímočaré. Autorky narážejí na některé problémy při identifikaci vzorců, např. na problematiku určení, zda předložková skupina následující po zkoumaném substantivu je součástí jeho vzorce (Hunston a Francis, 2000, str. 72). Dokazují, že není ostrá hranice mezi spojením ustáleným a značně neurčitým vzorcem.

3.2 Forma a význam, kontext a význam

V předchozí podkapitole jsme viděli, že lexikon a syntax od sebe nelze odpojit pro jejich vzájemnou podmíněnost. V následující podkapitole se blíže věnujeme právě úzké souvislosti mezi významem lexikální jednotky a jejím přímým okolím v textu.

Z didaktického hlediska nás zajímá, které kontexty užití jsou pro konkrétní slovo typické a jak se odráží preference v užití na sémantice lexému. Je významová jednotka dostatečně prezentována, když se její význam explikuje jen pomocí základního tvaru, lemmatu, zpravidla NSg.?

V tradici strukturalismu je uznávanou skutečností, že forma a význam jsou úzce spojené, respektive vzájemně propojené. Centrální otázkou nejenom lexikografie je – a v době korpusové se znovu stalo – vymezení rozsahu lexikální jednotky, která není a priori totožná s grafickým slovem, a určení jejího významu za pomoci analýzy jejích vztahů k jiným jednotkám. Saussure (2007) odlišuje od termínu *význam* termín *hodnota*, který vystihuje vzájemnou souvislost znaků v jazykovém systému, poněvadž není možné zjistit význam bez přihlédnutí k dalším jednotkám v systému. Jazyk je „systém, jehož všechny termíny jsou vzájemně solidární a kde hodnota jednoho vyplývá ze současné přítomnosti druhých [...]“ (tamtéž, str. 142). Jeho pojem *hodnota* zahrnuje všechny vztahy slova jak na ose paradigmatické, tak na ose syntagmatické.

V době korpusové je na bázi velkého množství dat možné identifikovat jednoslovné jednotky, ale i zvlášť jejich smysluplné kombinace, které se blíží oné základní významové jednotce a hledané úzké korespondenci významu a formy. Nejdále jde Sinclair (2004) v pokusu podchytit lexikální jednotku (*lexical item*). Potvrzuje, že forma a význam jsou dvě perspektivy téhož jevu (Sinclair, 2004, str. 139). Zároveň dodává: „It is asserted that form and meaning cannot be separated because they are the same thing“ (tamtéž). Následkem takového pohledu je, že se ambiguita prakticky rovná nule a že specifikace, ale i forma lexikální jednotky musí zahrnout všechny její komponenty, a to tak, že na ně lze nahlížet už jako na téměř nezávislé na ko-textu⁶⁴ (tamtéž).

Je zřejmé, že při posouzení slova vně kontextu se ztrácejí informace závažné k určení jeho významu, že teprve tak vzniká víceznačnost. Zároveň musíme vycházet z toho, že kontext jazykového jevu významně přispívá k utváření jeho vlastností (jeho gramatických kategorií nebo i pragmatických vlastností); srov. Cvrček, 2013, str. 23. Cvrček proto staví systematický, kvantitativní výzkum kontextu slova na jeho *úhrnný kontext*, jenž „představuje všechna textová okolí, do nichž hledaný prvek vstupuje“ (Cvrček, 2013, str. 24). Podle Cvrčka zrcadlí úhrn-

⁶⁴ Termín *ko-text* zde užívá ve smyslu přímého textového okolí a vystihuje tak rozdíl k širšímu pojetí termínu kontext, který zahrnuje i situační kontext.

ný kontext relevantní vlastnosti lingvistických jevů (tamtéž). Dalo by se tedy zároveň tvrdit, že slovní tvar, jenž prokazuje konkrétní gramatické vlastnosti a zahrnuje v sobě možné funkce v textu, leccos napovídá o svém kontextu.

Cvrček na konkrétním příkladu distribuce predikátové a atributivní funkce adjektiv ukazuje, že úhrnný kontext může odrazit i distribuci odlišných funkcí jazykových jednotek⁶⁵. Pokud je slovní tvar nositelem konkrétní gramatické informace o vztazích k jiným slovům v textu, vyvstává otázka, zda se lemmata, která upřednostňují stejné tvary, také shodují v kontextu. Cvrčkovo porovnání signifikantního kontextu lemmat (konkrétně apelativ) zatím ukázalo, že většina lemmat - soudíc z kontextu - je funkčně a sémanticky jedinečná. Ve své pilotní studii, jejímž cílem bylo identifikovat v synchronním korpusu češtiny *SYN2010* takové dvojice lemmat, jejichž seznam kolokátů se shoduje alespoň z 50 %, Cvrček dospěl k základnímu zjištění: „jen velmi málo lemmat se shoduje v tom, do jakých kolokací vstupuje“ (Cvrček, 2013, str. 57)⁶⁶. Avšak ta lemmata, která identifikoval, bylo možné přiřadit následujícím sémantickým skupinám: názvy měsíců, dny v týdnu, číslovky, pojmenování rodinných příslušníků, světové strany, jednotky a řady (tamtéž). Bylo by zajímavé zjistit, zda preference obdobných kontextů koreluje s upřednostňováním stejných slovních tvarů.

Studium kolokace se stalo centrálním prostředkem k prozkoumání syntagmatické stránky jazyka jak na textové rovině, tak na rovině systému. Sehrává podstatnou roli při vymezení víceslovných pojmů.

Kolokace jako vztah mezi dvěma nebo i více slovy se vymezuje různě, podle závažnosti, která vyděluje buď jejich sousedství, frekvenci souvýskytu, anebo kohezi mezi nimi (srov. k tomu i Čermák, 2006, str. 11). Přikláníme se tu k pojmu *kolokace*, jak ho razil Čermák (2010, str. 291), jenž ho charakterizuje jako termín lexikologický, omezený na smysluplnou kombinaci⁶⁷ lexémů. Důsledně tak oddě-

⁶⁵ Konkrétně se odlišná variabilita kontextu adjektiv v predikátovém užití a v atributivním užití ukazuje na výskytu příslušných kontextových typů na odlišných frekvenčních hladinách. Cvrček (2013, str. 54) postuluje: „V úhrnném kontextu, jako v reflexi relevantních vlastností lingvistických jevů, se může tento jev [primární vs. sekundární funkce jednotky, pozn. autorky] projevit tím, že různé funkce pozorované jednotky mají kontextový korelát na různých frekvenčních hladinách.“

⁶⁶ Cvrček ale upozorňuje, že asociační míry užití k extrakci kolokací prozkoumaných lemmat nejsou optimálním ukazatelem prototypického kontextového zapojení jednotek (Cvrček, 2013, str. 58). Zároveň se přimlouvá za přístup k extrakci kolokací, který přihlíží i k uspořádání jednotek v kontextu zkoumané jednotky, k jejich prototypické pozici a proximitě v korpusu (srov. Cvrček, 2013, str. 197).

⁶⁷ Kombinace vidí přitom jako preteoretický, obecný termín (tamtéž).

luje pojem *kolokace* závislý na pravidlech sémantické slučitelnosti⁶⁸ od pojmu *syntagma*, který se už neomezuje pouze na kombinaci lexémů.

V korpusové lingvistice se navíc vyděluje pojem koligace, čímž se podchycuje vzájemná kolokabilita gramatických kategorií, tedy kombinace jazykových jednotek na abstraktnější rovině, jako např. slovních druhů (srov. např. Čermák, 2010, str. 291). Do popisu syntagmatiky se tak dostávají slova i pojmy, jež byly doposud centrálními jednotkami morfologie a syntaxe. Sinclair řadí kolokaci, koligaci a sémantickou preferenci (*semantic preference*) do nepovinných součástí lexikální jednotky (*lexical item*), za jejichž pomoci se uskutečňují vztahy k jiným jednotkám, a vyděluje je tak od jejího jádra (Sinclair, 2004, str. 141). Sémantická pravidla, jež určují, jak se slova smysluplně kombinují, a zároveň pravidla, která řídí kombinace s gramatickými třídami, jsou tedy součástí jejich významů nebo jejich *hodnot*, řečeno v pojmech saussurovských.

Vzhledem k závažnosti gramatických a sémantických pravidel při tvoření syntaktických celků se má proto v korpusové lingvistice za to, že sémantická pravidla mají přednost před gramatickými (srov. Čermák, 2006, str. 10; Čermák, 2010, str. 126). I v pedagogické gramatice by proto měla připadnout větší důležitost popisu sémantické, a tedy i kombinatorické stránce substantiva. Ptáme se v této souvislosti, zda je substantivum jako významová jednotka dostatečně prezentováno, když se jeho význam explikuje jen pomocí lemmatu, jímž bývá zpravidla NSg. Z výše řečeného spíš plyne, že právě slovní tvar je tou jednotkou, jež v sobě nese parametry, které řídí souvztažnosti s jinými jednotkami v textu.

Výchozím bodem dané práce je proto otázka, které slovní tvary jsou v didakticky orientovaném popisu substantiva závažné pro přiměřenou prezentaci lexému jako abstraktní významové jednotky.

3.3 Paradigmatické a syntagmatické vztahy substantiva

Slovo čili lexikální jednotka podléhá omezení výběru jiných jednotek v textu a zároveň i ono samo vnáší do textu takové omezení. Vlastnosti, které mu umožňují vstupovat do určité struktury a které zároveň můžeme chápat jako výsledek všech

⁶⁸ Srov. definici pojmu kolokace u Čermáka (2006, str. 11): „Kolokaci je možné vymezit zhruba tak, že je to *smysluplné spojení slov/lexémů, resp. lexikální syntagma, zvl. v podobě víceslovných pojmenování, jehož vznik je podmíněn jejich vzájemnou kolokabilitou, a tedy i kompatibilitou.*“

jeho užití v různých kontextech, sdílí s jinými jednotkami. Utváří s nimi na základě nějaké vlastnosti třídu, kterou lze souhrnně označit jako paradigmatickou informaci slova. Informaci o jeho kompatibilitě s jinými jednotkami lze pak chápat jako jeho syntagmatickou stránku. Při popisu (lexikální) jednotky, jejího významu a její funkce se musí přihlídnout k oběma stránkám, resp. obě perspektivy by se měly doplňovat (srov. Halliday, 1991). Obě stránky jazykového jevu je možné pozorovat v konkrétním textu a zároveň jako abstrakci různých užití. I spíše gramatické souvislosti mezi slovy by se měly popsat jak z hlediska paradigmatického, tak syntagmatického⁶⁹.

Substantiva jako slovní druh tvoří v jazyce abstraktní třídu (paradigma), a to zejména na základě společných vlastností relevantních na rovině syntaxe. Skupinu substantiv lze však také vymezovat morfologickými kritérii. Vyznačují se jednak ohebností (svou flexí vyjadřují morfologické kategorie čísla, pádu, rodu i životnosti), jednak umožňují různými slovtvornými prostředky tvoření dalších substantiv a tvoří tak slovtvorné třídy a typy (např. názvy činitelské). V neposlední řadě lze na třídu substantiv nahlížet také z hlediska sémantického (zejména z lexikologického pohledu, srov. Čermák, 2010, str. 160). Jako slovní druh jsou substantiva vymezována jen intenzionálně, tj. na základě kategoriální vlastnosti, která je společná všem prvkům, jedná se o paradigma virtuální (srov. Čermák a Holub, 2005; Čermák, 2010).

Substantiva patří k autosémantikům, jež jsou v textu nejčastěji zastoupená (plnovýznamové slovní druhy tvoří 70,6 % všech výskytů v korpusu SYN2005⁷⁰). S podílem 30,2 % jsou substantiva přímo nejpočetnějším slovním druhem v současné češtině⁷¹. Jejich velký podíl je důkazem skutečnosti, že se jedná o tu třídu, která nejvíce reaguje na změny pojmenovávacích potřeb komunikující společnosti. Je to třída otevřená⁷². Substantiva jsou velmi často součástí víceslovných pojmenování (jako např. *závěrečná zpráva*, *dnešní zprávy* nebo *filozofická fakul-*

⁶⁹ Na závažnost a potřebu „reorientovat a zvracet nepříznivou a jednostrannou rovnováhu při jeho [dostupného jazyka, pozn. autorky] popisu od paradigmatického pohledu k pohledu syntagmatickému“ upozornil Čermák (2006, str. 9).

⁷⁰ Dle Statistiky češtiny (Bartoň et al., 2009, str. 130).

⁷¹ Dle Statistiky češtiny (tamtéž).

⁷² Srov. Čermák 2010. Čermák (2010, str. 258) však ukazuje na příkladu třídy číslovek, která se s přihlídnutím k číslovkám složeným nedá na základě kritéria „základní jednoslovný lexém“ jasně vymezit, že je pojetí ne/uzavřenosti paradigmatu poněkud relativní a představuje určité zjednodušení.

ta), která mohou být ustálená, a tudíž součásti lexikonu (lexémy) nebo i neustálená, tj. momentální kombinace slov v textu. Rozsah jak ustálených, tak neustálených pojmenování lze v praxi jen těžce číselně vymezit (srov. Čermák, 210, str. 164).

Oproti verbu, které se především ocitá ve funkci vyjadřovat ve větě predikát, plní substantivum různé syntaktické funkce, a má tudíž souvztažnosti s různými větnými členy (srov. Čermák 2010, str. 23). Převážně je argumentem tj. v pozici přímého nebo nepřímého předmětu, nebo podmětem, ale může také být součástí predikátu nebo fungovat jako jiný větný člen (např. jako atribut). Jako část složené prepozice je schopno vyjadřovat např. i časoprostorové, kvalifikační nebo příčinné relace (např.: *v blízkosti něčeho, s přihlédnutím k něčemu, z důvodu něčeho*).

Jednotlivá substantiva se dále řadí do řady dalších paradigmat (tříd), která se vymezují buď na základě morfologického kritéria, např. příslušnosti k určitému slovo-
tvornému typu nebo k jedné z deklinačních tříd, nebo na základě sémantického kritéria (např. názvy měsíců, názvy rodinných příslušníků ap.). Z onomaziologického hlediska se dělí na konkréta a abstrakta (viz k tomu podrobněji kapitola 5). Vzhledem k tomu, že substantivum je v jazyce centrálním prostředkem vyjadřujícím pojmenování⁷³ a k tomu, že je tak hojně v textu zastoupené, není adekvátní, že se v učebnicích věnuje substantivům pozornost převážně jako součástí abstraktních morfologických tříd a proč lexikální a syntagmatické hledisko často ustupuje zcela do pozadí.

Substantiva jsou v textu součástmi syntagmat. Vstupují do nich s vlastnostmi danými svou příslušností k virtuálním paradigmátům. Zároveň však s sebou nesou sémantické rysy, které jim povolují spojovat se s dalšími jazykovými jednotkami v textu, anebo jim naopak spojení s nimi znemožňují, resp. omezují. Na základě jejich kombinovatelnosti s dalšími lexémy se vymezuje *kolokační paradigma* konkrétního substantiva⁷⁴.

⁷³ Srov. k tomu Čermák (2010, str. 155): „Zatímco se v syntaxi jednostranně věnuje pozornost slovesu a se substantivem se zachází jako s jakýmkoliv jiným slovním druhem, v oblasti nominace je situace opačná. Základním slovním druhem z tohoto hlediska v jazyce není, jak nám namlouvají syntaxe, verbum, ale substantivum.“

⁷⁴ „**Kolokační paradigma** je naproti virtuálnímu takový úhrn všech lexémů, které se obvykle kombinují s daným lexémem (ten je však externí a do něj nepatří) a které ho svým složením (to je pro každý lexém unikátní), i jednoznačně identifikují.“ (Čermák, 2010, str. 252). Podíváme-li se na kolokační třídu zevnitř neboli na konkrétní substantivum jako na součást tohoto kolokačního paradigmatu, tak se nachází k dalším členům třídy ve vztahu vzájemně funkční ekvivalence (srov.

Kombinatoriku substantiva je možné popsat pro jednotlivý lexém, zároveň lingvisty zajímá rozřídění substantiv a jejich kolokací na základě obdobné spojitelnosti. Na rovině slovních druhů vyděлил Čermák (2010) devět podtypů kombinací substantiv: S-V⁷⁵, V-S, S-Inf, A-S, S-S, S-Konj-S, ADV-S, S-ADV, S-KL. Považha substantiv (konkréta, abstrakta) opět umožňuje rozčlenit typ S-V do funkčně-sémantických tříd možných kolokací (Čermák, 2010, str. 301nn.; viz tamtéž pro příklady). Velmi dobře a podrobně lze podle Čermáka vymezit a vystihnout kombinace substantiv kvantitativních (např. *spousta, kapka, stéblo, sbírka*) typu S-S. Považha kvantifikovaného substantiva (zda se např. řadí do třídy singulativ, numerativ nebo substantiv počítatelných) zde do značné míry určuje, s kterým kvantifikátorem nebo s kterou třídou kvantifikátorů se spojuje (viz tamtéž).

Úzce spjat se schopností lexémů vázat na sebe určité množství dalších jednotek a ukládat na výběr následujícího tvaru omezení zejména gramatického rázu je pojem valence. V české lingvistice se tradičně přistupuje k valenci spíš syntakticky a s hlavním zřetelem na sloveso. Korpusově orientované přístupy si kladou otázku, do jaké míry je valence vázána i na sémantické rysy samotného slova (např. Čermáková, 2009, str. 16). Takové přístupy se opírají o pojetí valence, které se vztahuje k zobecněné kolokabilitě celé sémanticky vymezené třídy. Čermák vystihuje součinnost kolokability lexémů a valence následovně:

„Vzájemný vztah kolokability a valence je tedy v zásadě prostý: kolokabilita daného lexému, pro něj specifická a individuální, představovaná jeho kolokačním paradigmatem, je v textu sémantickou a lexikální konkretizací jeho obecné a formální valence a jedno bez druhého neexistuje. [...] V tomto smyslu má valence obecně sémantický základ a vznikla na základě zobecněného a obdobného chování tříd lexémů se společnou sémantikou.“ (Čermák, 2010, str. 284)

3.4 Shrnutí

Soustavný přístup k popisu substantiv pro didaktické účely by měl vzít v úvahu jeho přináležitost k třídám gramatickým, ale také sémantickým a zároveň jeho syntagmatickou stránku. Poněvadž se kategoriální schopnosti, schopnost vyjadřo-

definici kolokačního paradigmatu v Čermák a Holub, 2005, str. 11n). V textu se tedy mohou na tomtéž místě potenciálně střídát, např. *Od 15 h mi nepřicházejí [zprávy / smsky / e-maily]*.

⁷⁵ zde S = substantivum, A = adjektivum, V = verbum, Inf = infinitiv, konj = konjunkce/spojka, ADV = adverbium, KL = klauze

vat nějakou gramatickou kategorií, neboli gramatické vlastnosti opírají do nemalé míry o sémantickou stránku lexému, je potřeba tuto souvztažnost mezi zobecněnými sémantickými rysy a výrazem gramatických vlastností, pokud možno, reflektovat. V této práci se budeme pokoušet roztřídit substantiva na základě společné preference realizovat se s určitým slovním tvarem a zjemňovat obecnější sémantické třídy vzhledem ke vztahům k okolním jednotkám v textu, tj. ke vztahům, které se odrážejí na slovním tvaru.

Pro přiměřené znázornění syntagmatické stránky substantiv v učebních pomůckách by bylo dále důležité:

- (a) zjišťovat substantivní lexémy (jednoslovné i víceslovné) tvořící kolokační paradigma konkrétního verba nebo substantiva ve struktuře, která je centrální pro plnění procvičované jazykové funkce (např. dát si <*kafe / kávu / čaj | likér / pivo / sklenku / šálek kávy / panáka / zmrzlinu / pivo a párek*>)⁷⁶.
- (b) zjišťovat kolokační paradigma konkrétního substantivního lexému, a tudíž kontexty, ve kterých se často a preferovaně ocitá.

Touto cestou by bylo možné ukázat substantivum jako součást různých kolokačních paradigmat relevantních sloves, substantiv nebo i adjektiv a zároveň ho prezentovat v preferovaných syntagmatech. Struktura současných učebnic často neposkytuje dostatečný prostor pro specifikaci gramatické struktury na příkladu sémanticky určených tříd slov, která je typicky vyjadřují. Jsme si vědomi, že v omezeném rámci tištěné publikace často není možné poskytnout dostatečně velké množství příkladů. Rozmanité možnosti, jak zapojit elektronická média v rámci e-learningu, by však už nyní mohly problém vyřešit.

⁷⁶ Jde o kolokace lemmatu *dát si* v korpusu SYN2000; prozkoumána byla frekvenční distribuce kolokátů na pozicích 1-3 vpravo od KWIC.

4 Frekvence pádů v současné češtině

Již jsme se výše zmínili o tom, že substantiva jsou nejčastěji zastoupeným slovním druhem v češtině. Jejich frekvence však podléhá nemalé variaci, přihlížíme-li k jejich zastoupení v různých subkorpusech. Podle Mluvnice současné češtiny (MSC, 2010) jsou nejpočetnější v publicistice a v odborných textech⁷⁷. Nejnápadnější je velký rozdíl mezi frekvencí substantiv právě v těchto sférách užívání a užitím v korpusu přátelské konverzace, kde substantiva pokrývají jen něco přes 10 % všech slovních druhů (MSC, 2010, str. 34). Závislosti frekvence výskytu slovních druhů na funkčním stylu si všimla již Těšitelová (1985, str. 56). Zatímco v uměleckém stylu podíl substantiv a adjektiv klesá, roste zároveň podíl sloves a s nimi i podíl zájmen. Oproti tomu se podíl předložek a spojek celkem nemění (tamtéž). Rozložení substantiv, adjektiv a sloves je tedy poměrně citlivé na různé funkční styly a dalo by se předpokládat, že totéž platí i pro syntaktické funkce, které substantivum plní.

V této kapitole proto přihlížíme nejen k obecné frekvenci morfologických kategorií v korpusech současné češtiny, ale rovněž se věnujeme faktorům, na něž je jejich frekvence citlivá. Povahou jednotlivých kategorií a jejich vztahy k sémantické složce lexémů se zabývá kapitola 5.

4.1 Frekvence substantivních morfologických kategorií v češtině

Morfologické kategorie substantiva lze poměrně dobře postihnout a kvantifikovat, poněvadž jejich počet je omezen a protože se v češtině vyjadřují synteticky, tedy hlavně prostřednictvím koncovky resp. kmene⁷⁸. Při analýze konkrétní morfologické kategorie jako např. pádu je však třeba mít na zřeteli, že se vždy vyjadřuje v kombinaci s ostatními kategoriemi, tedy zároveň s číslem a rodem (popř. životností a neživotností). Tyto kategorie se mohou vzájemně doplňovat a ovlivňovat, jak uvidíme níže.

⁷⁷ Jejich frekvence se uvádí se zhruba 350 000 ppm (parts per million), tedy počtem výskytů substantiv v korpusu textů s jedním milionem slov (MSC, 2010). Předpokládáme, že se kvantifikovala lemmata, nikoli slovní tvary, nemáme však o tom z textu přesné informace. Statistiky češtiny udávají podíl substantiv ze všech výskytů v korpusu SYN2005 s 30,2 %, oproti tomu mají adjektiva 11,7 % a slovesa 17,5 % (Bartoň et al., 2009, str. 130).

⁷⁸ Např. význam ‚nominativ singuláru‘ se u maskulin vzoru *hrad* vyjadřuje pomocí kmene. Je však možné mluvit v tomto případě o *nulové koncovce*.

4.1.1 Mluvená a psaná čeština

Těšitelová a kol. (1985) provedli první rozsáhlejší analýzu číselného zastoupení jednotlivých pádů v užívání českého jazyka. Jejich *Kvantitativní charakteristika současné češtiny*, jež představuje dodatek k akademické mluvnici, se opírá o korpus, na kterém je založený i první velký frekvenční slovník současné češtiny z roku 1961⁷⁹. Korpus má část uměleckou a část věcnou, což umožňuje zkoumání závislostí na funkčním stylu. Těšitelová et al. zapojují do své analýzy korpus věcného stylu (VS), jež se skládá z textů publicistické, odborné a administrativní povahy, a to z 75 % z projevů psaných a z 25 % z projevů mluvených⁸⁰. Od vzniku velkých automaticky lemmatizovaných a morfologicky značkových korpusů psané češtiny (např. korpusy SYN2005 a SYN2010 z řady SYN Českého národního korpusu) je nám nyní k dispozici informace o kvantitativní charakteristice morfologických kategorií, která má oporu v mnohem větší datové bázi. V mnohem větším rozsahu se mapuje i situace v mluvené češtině, zejména s důrazem na živý jazyk a skutečný, typický úzus. Rozsáhlou analýzu morfologické stránky mluvené češtiny poskytuje Šonková (2008) na bázi dat Pražského mluveného korpusu (PMK, Čermák et al., 2001). Od roku 2006 vznikly tři další korpusy mluvené češtiny, které dovolují zkoumat mluvenou podobu jazyka, zachycenou v rozhovorech v neformálních komunikačních situacích (řada ORAL). Tyto korpusy však neobsahují lemmatizaci ani morfologické značky, což ztěžuje kvantifikaci morfologických kategorií.

Pro substantivum se z kvantitativního hlediska charakterizují tři kategorie, a to rod (kombinuje se s kategorií životnost/neživotnost), číslo a pád. O rozložení kategorií rodu a čísla se zde zmíníme také, poněvadž obě mají vliv na četnost a rozrůzněnost deklinačních vzorů a samozřejmě i morfologických koncovek.

Následující dvě tabulky shrnují frekvenční rozložení jmenného rodu a čísla v korpusech SYN2005 a v Pražském mluveném korpusu (PMK) v procentech výskytů. Údaje jsou převzaty ze *Statistik češtiny* (Bartoň et al., 2009, str. 132) a *Morfologie mluvené češtiny: Frekvenční analýza* (Šonková, 2008, str. 53nn.). Poněvadž se podíly jednotlivých významů v korpusu FSC a SYN2010 výrazně neliší od poměrů v SYN2005, neuvádíme je zde zvlášť.

⁷⁹ *Frekvence slov, slovních druhů a tvarů v českém jazyce* (Jelínek et al., 1961) dále jen FSC

⁸⁰ Nahrávky přednášek a rozhlasových příspěvků, z nichž analýza vychází, však nelze dost dobře považovat za ukázkou autentického nepřipraveného mluveného projevu.

rod	SYN2005			
	SYN2005	BEL	ODB	PUB
mask.anim ⁸¹	16,3	18,5	10,1	19,6
mask.inan	29,7	28,3	32,6	28,4
fem	38,5	38,9	39,4	37,4
neut	15,6	14,4	17,9	14,6
CELKEM	100	100	100	100

rod	PMK
mask.anim	15,1
mask.inan	25,4
fem	44,5
neut	14,9
nelze určit	0,0
	100

Tab. 4.1 Rozložení jmenného rodu u substantiv v korpusu SYN2005, v jeho subkorpusech a v PMK

Maskulina jsou s frekvencí 46 % nejpočetnějším jmenným rodem v psané češtině. V PMK je jejich podíl o něco menší (jen 40,5 %) zejména ve prospěch feminin. Mužská jména budou zřejmě vždy mít nerovnoměrně větší podíl oproti ženským jménům, poněvadž většina slov, jako jsou označení profesí, sociálních rolí apod., má mužský rod⁸². Pohled na distribuce rodu v subkorpusech SYN2005 naznačuje odlišné užití substantiv zejména v odborné literatuře.

Singulárové tvary v jazyce jednoznačně převažují. Když však porovnááme jejich užití v psané češtině s užitím v mluvené podobě, zjišťujeme nápadný posun v podílech singuláru a plurálu ve prospěch plurálových tvarů (celkem 32,5 % v PMK vs. 24,9 % v SYN2005).

číslo ⁸³	SYN2005			
	SYN2005	BEL	ODB	PUB
sg	75,1	77,4	72,8	74,9
pl	24,9	22,6	27,2	25,1

číslo	PMK
sg	66,7
pl	30,6
pluralia tantum	1,9
kolektiva	0,7
nelze určit	0,1

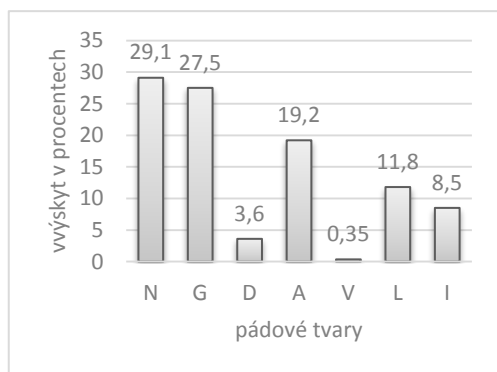
Tab. 4.2 Rozložení čísla u substantiv v korpusu SYN2005 a jeho subkorpusech a v PMK

Rozložení pádových tvarů je opět podobné pro korpusy SYN2005 a SYN2010. Uvádíme vzorově jejich rozložení v SYN2005 a PMK.

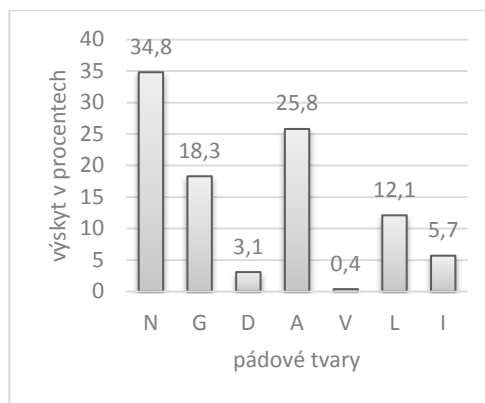
⁸¹ Shodně s užitím ve *Statistikách češtiny* (Bartoň et al., 2009) užíváme v této práci k označení gramatického rodu následující zkratky: mask = mužský rod (maskulina), anim = životná (animata), inan = neživotná (inanimata), fem = ženský rod (feminina), neut = střední rod (neutra).

⁸² Srov. MŠČ (2010, str. 136).

⁸³ Zbytky duálu jsou zahrnuty pod plurál.



Obr. 4.1 Rozložení pádových tvarů v korpusu SYN2005



Obr. 4.2 Rozložení pádových tvarů v PMK⁸⁴

Nejpočetnějšími pády jsou jak ve psané, tak v mluvené podobě češtiny N, G a A. Pokrývají celkem 75,8 % všech výskytů pádových tvarů substantiv v psané a dokonce 78,9 % všech výskytů v mluvené češtině. Zatímco se rozložení tvarů instrumentálu, lokálu, dativu a vokativu v SYN2005 jen málo liší od rozložení v PMK (nejvíce u instrumentálu), jsou právě u kvantitativně nejzávažnějších pádů poměrně velké rozdíly mezi psanou a mluvenou podobou. Obrázky 4.1 a 4.2 názorně ukazují, že genitiv je mnohem četněji zastoupený v psané češtině a že se četností výskytu téměř rovná nominativu, kdežto v mluvené podobě jednoznačně převládá nominativ (s téměř dvojnásobnou frekvencí forem než genitiv). I pozice akuzativu je v psané češtině oproti mluvené češtině oslabená. Rozdíly v četnosti těchto pádů vzhledem k mluvené a psané podobě češtiny ukazují na určitou variabilitu ve frekvenci syntaktických funkcí, jež tyto pády převážně vyjadřují, zejména funkcí objektu, subjektu a neshodného atributu (viz k tomu podrobněji 4.2).

Frekvenční distribuce pádových tvarů se nápadně liší v singuláru a plurálu. Vzájemné vztahy čísla a pádu ukazuje tabulka 4.3 (pády jsou v ní uspořádány podle sestupné frekvence). Frekvenční údaje ze SYN2005 a PMK jsme záměrně doplnili o data z korpusu SYN2010. Chceme tím ukázat, že distribuce pádů s číslem je poměrně stabilní v textech současné češtiny alespoň v období přibližně 10 let⁸⁵. Údaje se nikde neliší o víc než jeden procentní bod.

⁸⁴ Podle Šonkové (2008, str. 55).

⁸⁵ V korpusu SYN2005 převažují texty z let 2000-2004 a v korpusu SYN2010 z let 2005-2010.

Korpus SYN2005 ⁸⁶				Korpus SYN2010 ⁸⁷				Pražský mluvený korpus (PMK) ⁸⁸			
pořadí v sg		pořadí v pl		pořadí v sg		pořadí v pl		pořadí v sg		pořadí v pl	
	%		%		%		%		%		%
N	23,0	G	8,9	N	22,0	G	8,5	N	23,0	N	11,0
G	18,5	N	6,0	G	17,6	N	5,6	A	17,7	G	7,3
A	14,2	A	5,0	A	15,0	A	5,4	G	10,2	A	7,2
L	9,7	L	2,1	L	9,6	L	2,1	L	9,4	L	2,3
I	6,6	I	1,9	I	6,3	I	1,8	I	4,0	I	1,5
D	2,7	D	0,92	D	2,8	D	0,9	D	1,9	D	1,0
V	0,35	V	0,001	V	0,3	V	0,002	V	0,2	V	0,1
cel.	75,2	cel.	24,8	cel.	73,6	cel.	24,3	cel.	66,4	cel.	30,4
				2,8% případů nelze určit včetně podstatných jmen pomnožných				0,8% případů nelze určit 2,4% případů připadá na pluralia tantum a kolektiva			

Tab. 4.3 Rozložení jmenného pádu a čísla u substantiv v korpusech SYN2005, SYN2010 a PMK

Je nápadné, že na rozdíl od singulárových tvarů převažuje u plurálových tvarů genitiv. To však platí jen pro užití substantivních tvarů v psaných textech. V mluvené češtině převládá nominativ jak v singuláru, tak v plurálu. Porovnáváme-li užití genitivu s užitím akuzativu, zjistíme, že genitiv je v psané češtině vždy četnější než akuzativ, výjimku tvoří jen singulárové tvary v mluvené češtině. Pořadí lokálu, instrumentálu, dativu a vokativu je v plurálu stejné jako v singuláru, podobně jako v korpusech SYN2005, SYN2010 a PMK.

Můžeme shrnout, že složení pásma tří nejfrekventovanějších pádů je v psané a mluvené češtině stabilní, tj. že se skládá vždy z pádů N, G, A, i když v odlišných poměrech. Tyto nejfrekventovanější pády pokrývají mezi 75 % a 79 % užití všech substantivních tvarů⁸⁹. Obdobně je i složení pásma na spodním kraji frekvenční distribuce pádů stálé. Nenapovídá však nic o rozložení frekvence pádů pro jednotlivé lexémy. Největší variabilitu vzhledem ke kombinaci se jmenným rodem, ale zejména v závislosti na psanosti či mluvenosti projevů, ukazují nominativ, genitiv a akuzativ.

⁸⁶ Podle *Statistik češtiny* (Bartoň et al., 2009, str. 134).

⁸⁷ Udává relativní frekvenci pádových forem; přitom platí, že součet frekvencí všech substantivních forem by měl dávat 100 %. Uváděné výsledky však nelze považovat za stoprocentně správné: je třeba počítat s chybovostí alespoň čtyř procent, protože úspěšnost morfologické analýzy celého korpusu se pohybuje okolo 96 %.

⁸⁸ Frekvence pádů v Pražském mluveném korpusu podle Šonkové (Šonková, 2008, str. 55).

⁸⁹ Tvrzení, že zmíněné tři pády mohou představovat jádro zkoumaného jevu, opíráme o zjištění Těšitelové (1990). Ve své stati (1990, str. 10) dospěla k závěru, že pokud se seřazují jazykové jevy v sestupném pořadí, mají obvykle tři (nebo maximálně čtyři) jevy nápadně vyšší frekvenci (s určitou gradací), zatímco frekvence ostatních jevů je výrazně nižší.

4.1.2 Funkční styly psané češtiny

Závislost výskytu morfologických kategorií na funkčním stylu pozorujeme srovnáním subkorpusů vytvořených na základě typu textů, a tudíž na základě určitých stylistických charakteristik textů. Korpusy SYN2005 a SYN2010 se skládají z textů beletristických (subkorpus BEL), odborných (subkorpus ODB) a publicistických (subkorpus PUB) a jejich složení lze považovat za vyvážené z hlediska užívání jazyka (dle principu recepce)⁹⁰. Na průměrném poměru type / token na vzorcích o velikosti 1 000 a 10 000 tokenů, vidíme, že se subkorpusy liší i vzhledem k pestrosti obsažených textů. Podle *Statistik češtiny* má subkorpus publicistiky nejvyšší poměr type/token, „protože je sestaven z pestřejších, kratších textů, s větším množstvím rozmanitých vlastních jmen“ (Bartoň et al., 2009, str. 131).

4.1.2.1 Rod

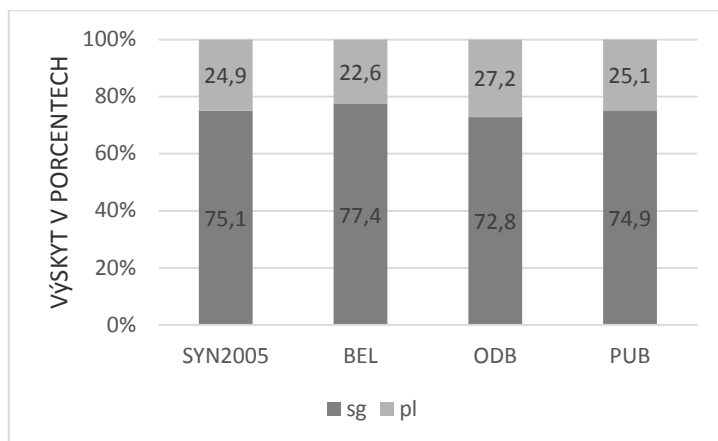
Rozložení rodu substantiv v korpusu SYN2005 se poměrně výrazně liší v závislosti na funkčním stylu⁹¹ (srov. také Těšitelová et al. 1985). Zatímco v subkorpusech beletrie a publicistiky mají maskulina životná i neživotná, feminina a neutra podobné rozložení, jejich chování se značně liší v odborné literatuře. Nápadný je zejména řídký výskyt maskulin neživotných (10,1 % všech substantiv v ODB oproti např. 19,6 % všech substantiv v PUB). Jejich nižší zastoupení se vyrovnává častějším výskytem neuter (srov. 17,9 % v ODB vs. 14,6 % v PUB) a maskulin neživotných (srov. 32,6 % v ODB vs. 28,4 % v PUB). Obdobně zaznamenala už Těšitelová et al. (1985, str. 60), že počet neuter je v subkorpusu stylu věcného o 5 % vyšší než ve FSC a je rovněž vyšší než v korpusu stylu uměleckého. Autoři to vysvětlují zejména přibýváním podstatných jmen slovesných.

4.1.2.2 Číslo

Výkyvy v poměru singulárových a plurálových forem ve srovnání s celkovým rozložením jsou nejvýznamnější v beletrických textech a v odborné literatuře (viz obr. 4.3). O něco větší podíl mají plurálové tvary v odborných textech (27,2 %), jejich podíl v krásné literatuře je naproti tomu nižší (22,6 %).

⁹⁰ O konkrétním složení viz Příručku Českého národního korpusu (Cvrček a Richterová, 2015) a dále také v kapitole 5.

⁹¹ Opíráme se opět o údaje ve *Statistikách češtiny* (Bartoň et al., 2009).

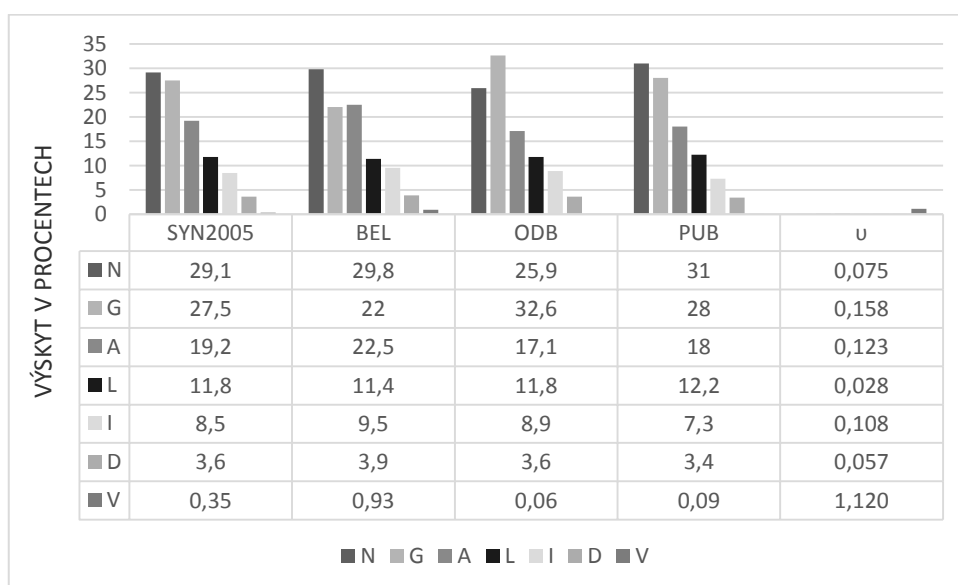


Obr. 4.3 Rozložení jmenného čísla v korpusu SYN2005 a v jeho subkorpusech

Možnými faktory vedoucími k tomuto stavu jsou rozdílné četnosti vlastních, zejména místních jmen a pluralií tantum. Významnější vliv mají však podle Těšitelové et al. (1985, str. 60nn.) tematika a styl autora. Zároveň autoři poukazují na souvislost čísla s rodem. Zejména u neuter, a zvláště u jmen slovesných, je ve FSC poměr singuláru k plurálu výrazně vyšší (srov. poměr sg. a pl. ve FSC u neuter 85,98 % : 14,02 % vs. poměr sg. a pl. ve FSC bez přihlédnutí k rodu 75,81 % : 24,1 %, tamtéž). Totéž pozorujeme i v korpusu SYN2005, v němž mají singulárové tvary podíl 83,7 % u všech neutrálních tvarů (srov. poměr sg. : pl. v celém korpusu 75,1 % : 24,9 %). Poměr sg. : pl. se výrazně odchyľuje od průměru už jen u maskulin životných, a to v opačném směru: podíl plurálových forem u nich činí 29,3 %. Důvodem může být, že se mezi nimi nacházejí názvy činitelů, národností apod., které se častěji vyskytují v množném čísle. Ve prospěch tohoto vysvětlení by svědčilo i skutečnost poměrně vysoké procento plurálových tvarů maskulin životných v publicistice a v odborné literatuře, kde nacházíme poměr tvarů plurálových k singulárovým 31,2 % : 68,8 % a 38,0 % : 62,0 %, kdežto v beletristice jsou v poměru 23,0 % : 77,6 %. Bez podrobnější manuální analýzy nemůžeme souvislosti mezi frekvencí kategorie čísla a slovtvornými třídami jednoznačně určit, ale můžeme říct, že funkční styl ovlivňuje frekvenci množného či jednotného čísla, a to v závislosti na rodu, jenž do jisté míry zrcadlí onomaziologické kategorie, a tudíž slovtvorné typy.

4.1.2.3 Pád

Když porovnááme frekvenční distribuci pádů v jednotlivých subkorpusech SYN2005 jen pohledem na graf v obrázku 4.4, mohli bychom říct, že největší výkyvy lze pozorovat u genitivu, nominativu a akuzativu (viz obr. 4.4). Pohled na variační koeficient (υ) však odhaluje, že se nejvíce od průměru odchyluje frekvence vokativu. Výrazně také vybočuje frekvence genitivu, a to v odborné literatuře (o 5,1 procentních bodů), kde genitiv je nejužívanějším pádem. Zároveň s přibýváním genitivních tvarů klesá užití nominativu (o 3,2 procentní body) a akuzativu (o 2,1 procentní body). Beletristické texty se vyznačují nízkou frekvencí genitivu a nejsilnějším zastoupením akuzativu, jenž se zde užívá častěji než genitiv. V publicistických textech je nápadná opět poměrně vysoká frekvence nominativu. Zatímco se zde genitiv ani dativ výrazně neliší od průměru, mírně klesá frekvence akuzativu, instrumentálu (v obou případech o 1,2 procentního bodu) a také vokativu.



Obr. 4.4 Rozložení pádů v subkorpusech SYN2005, v sestupném řazení podle frekvence v korpusu SYN2005⁹²

Můžeme shrnout, že v závislosti na stylu textů se nejméně mění frekvence dativu a lokálu, naopak nejvýrazněji na tento faktor reagují vokativ a genitiv. V beletristice a v odborné literatuře změnu v užití genitivu doprovází změna frekvence akuzativu. Četnost nominativu se zdá být postihnuta výrazně jen v odborné

⁹² υ = variační koeficient; Udáváme variační koeficient, tedy relativní velikost rozptýlenosti frekvence pádů v subkorpusech vzhledem k průměru, aby bylo možné srovnávat variace jednotlivých pádů.

literatuře. Celkově však jeho variabilita vzhledem k funkčnímu stylu zaostává za variabilitou nejen genitivu, akuzativu, ale i instrumentálu. Podobně tomu, co jsme zjistili při porovnání mluvené češtiny s psanou, nemění se ani zde pořadí pádů na frekvenčním žebříčku ve středním ani v nízkém frekvenčním spektru rozložení pádových tvarů (v sestupném pořadí: L – I – D – V).

Interpretace těchto faktů je obtížná, poněvadž musí vzít v úvahu mnoho aspektů odlišných typů textů. Typy textů se liší velkým počtem faktorů jako např. tematikou, užitím terminologie, výskytem proprií, způsobem realizace valenčních pozic (*někdo zvažuje problém – zvážení problému*), poměrem ustálených (anomálních) lexikálních kombinací (frazémy) a pravidelných jazykových kombinací, preference aktivních či pasivních syntaktických konstrukcí, ale také mírou výraznosti autorského stylu apod. Souhra těchto faktorů ovlivňuje frekvenci výskytu pádů a do jisté míry i jejich charakteristiku. Podle závislosti selekce pádu buď na sémantické složce nositele valence (obvykle slovesa) nebo na gramatickém okolí se rozlišují v *teorii pádů* strukturní a nestrukturní pády. Toto rozlišení přejímá pro češtinu i Karlík (2000). Strukturní pád je takový pád, jehož realizace se v závislosti na gramatickém okolí mění (tamtéž, str. 179). Právě nominativ, který není vybrán na základě sémantické struktury lexie⁹³ (Karlík, 2000, 179), je vysoce strukturním pádem. Potvrzuje to skutečnost, že se jeho frekvence mění v subkorpusech jen nevýznamně. I pády, jež jsou selektovány významem lexie⁹⁴, jsou v různé míře strukturními pády. Nejvíce strukturním pádem je pro Karlíka A, druhým nejvíce strukturním je potom D (Karlík, tamtéž). Akuzativ stojí v hierarchii mezi interními valencemi nejvyš, a představuje tzv. *default-pád*⁹⁵. Při bližším pohledu na data zjišťujeme, že ani dativ není příliš citlivý na změnu funkčního stylu. Totéž však nelze tvrdit o akuzativu. Všechny ostatní pády, tedy G, I, L, V, jsou podle Karlíka v různé míře lexikální/ idiosynkratické. Vzhledem k variaci jejich frekvencí v různých subkorpusech se však chovají dost odlišně. Přímé a obecnější souvislosti mezi charakteristikou pádů a jejich kolísající frekvencí v subkorpusech můžeme

⁹³ Postavení subjektového nominativu a jeho vztahů k valenci slovesa, tj. zda je skutečně řízen predikativním verbem finitem, je mezi lingvisty spornou otázkou (srov. k tomu EŠČ [Karlík, 2002, str. 306]). Karlík navrhuje distinkci externí valenční pozice a interní valenční pozice. Nominativ je tedy jediná možnost vyjadřovat externí valenční pozici. Je selektován slovesným flexémem jakožto složka nositele valence (Karlík, 2000, str. 179).

⁹⁴ Naplňují tedy podle Karlíka interní syntaktickou valenci (Karlík, 2000, str. 179).

⁹⁵ Srov. Karlík (2000, str. 179): „Je to pád mezi interními valencemi nejvíce strukturní, tj. nejsou-li nějaké sémantické důvody kódovat A nějakým jiným pádem, a tím ho označit jako aktant s jinou sémantickou rolí než *patiens* (v širokém smyslu), dá se předvídat, že A [aktant, pozn. autorky] bude vyjádřen právě v pozici s formou Ak.“

jen těžko popsat. Ale mohli bychom předpovídat, že např. v textech s vyšším podílem pasivních konstrukcí je méně čteně zastoupený A a častěji v nich nalzáme I, nebo že řidšímu výskytu akuzativu může např. nahrávat častější užití deverbálních substantiv, jež vznikla transpozicí (*vyčerpát – vyčerpání* apod.). Jejich vyšší výskyt by zároveň měl být doprovázen vyšší frekvencí genitivu a případně i vyšší frekvencí instrumentálu (srov. např.: *Petr kritizuje Janu – kritizování Jany Petrem*; příklad z Karlíka 2000, str. 178).

4.2 Frekvence syntaktických funkcí

Souhru mezi frekvencí pádových tvarů a funkcí pádů vyjasňuje analýza syntaktických funkcí slovních druhů u Těšitelové et al. (1985) a u Těšitelové 1990).

Z analýzy četnosti syntaktických funkcí v již zmíněném korpusu FSČ i korpusu věcného stylu vyplývá, že substantivum má v 35 % případů funkci subjektovo-objektovou⁹⁶, ale v 43 % svých výskytů vystupuje ve funkci atributivní⁹⁷ (Těšitelová et al. 1985, str. 103), což může vysvětlovat vysokou pozici genitivu na frekvenčním žebříčku pádů, zejména jeho silnou roli v odborných textech. Autoři tvrdí, že „syntaktické uplatnění slovních druhů je ve své podstatě nezávislé na funkčním stylu.“ (tamtéž). Dokládají to tím, že se frekvence tzv. primárních funkcí (tedy např. funkce predikátní jako primární funkce sloves a funkce subjektovo-objektové jako primární funkce substantiv, viz tamtéž) ani v různých funkčních stylech výrazně nemění. Připouštějí však, že se významné rozdíly mohou projevit v uplatnění slovních druhů v jejich funkcích sekundárních. Autoři vyčleňují ze substantiv různé podtřídy a ukazují, že např. substantiva číselná a deadjektivní jsou nápadná nadprůměrně častým výskytem atributivní funkce (53 % u číselných a 50 % u deadjektivních). Je tedy zřejmé, že se projevuje ve frekvenci i sémantické vlastnosti jednotlivých podtříd (Těšitelová et al., 1985, str. 104).

Na tomto místě se chceme zmínit ještě o pokusu Těšitelové (1990) roztřídit substantiva na základě frekvence syntaktických funkcí, poněvadž tento postup přihlíží k individuálním substantivním lexémům a třídí je na základě jejich syntaktického chování. Analýza syntaktických profilů substantivních lexémů byla provedena v korpusu věcného stylu o rozsahu 540 000 textových slov na příkladu 50 nej-

⁹⁶ Tato funkce se považuje pro substantivum za primární (viz Těšitelová et al. 1985, str. 103).

⁹⁷ 19 % připadá na funkci adverbialní a 3 % na funkci predikátu.

frekventovanějších substantiv. Záměrem analýzy bylo: a) popsat substantiva vzhledem k tomu, v jakých sémanticko-syntaktických jednotkách vystupují a které kombinace těchto jednotek se u jednotlivých lexémů projevují, a b) popsat distribuce tří nejfrekventovanějších jednotek.

Těšitelová (1990, str. 10) identifikuje čtyři typy kombinací syntakticko-sémantických funkcí:

- typ 1 - kombinace se syntaktickou funkcí atributu (38)⁹⁸
- typ 2 - kombinace se syntaktickou funkcí objektu (7)
- typ 3 - kombinace se syntaktickou funkcí subjektu (3)
- typ 4 - kombinace se syntaktickou funkcí adverbiale (3)

Je tedy vidět, že 38 zkoumaných substantiv se vyskytuje nejčastěji ve funkci atributu, a zároveň, že se substantiva liší vzhledem k frekvenci, s kterou nějakou funkcí plní. První skupina (typ 1) se vyznačuje i největším počtem podtypů (celkem 11). Ty se tvoří na základě druhé a třetí nejčastější funkce, např. prvním podtypem typu 1 (podtyp 11) je „attr. + obj. + subj.“ (tamtéž, str. 10n.). Spadají do něj mj. lexémy *činnost, síla, voda a výrobek*. V podtypu 13 („attr. + adv. + obj.“) se nacházejí společně např.: *země, čas, forma a systém*. Do třetí skupiny (typ 3) náleží substantiva *doba, období a případ*, což nepřekvapuje, přihlížíme-li k jejich časovým užitím ve spojení s předložkou: *v době, v období, v případě*.

Podrobnější analýza frekvenční distribuce funkcí uvnitř podtypů odhaluje, že se substantiva mohou významně lišit vzhledem k frekvenci, s kterou plní jednotlivé funkce. Lexémy podtypu 11 vyjadřují funkci atributivní např. v rozsahu od 60 % (*energie*) do 26 % (*pojem*) a funkci objektu od 26 % (*pojem*) do 16 % (*jednotka*). Těšitelová upozorňuje při analýze funkce objektové u slov tohoto podtypu, že se substantiva spojují se významově dost odlišnými slovesy v souladu s jejich odlišnou lexikální sémantikou (Těšitelová, 1990, str. 14).

Na základě poznatků Těšitelové můžeme shrnout, že substantiva lze roztřídit podle preferovaných syntaktických funkcí a podobně je možné jejich jemnější roztřídění na subtypy. Členové jednotlivých skupin (typů) a podtypů se částečně vyznačují společnými sémantickými rysy. Analýza individuálních substantiv ale také ukázala, že lze v podtypech pozorovat poměrně velkou variaci ve frekvencích, s nimiž jednotlivá substantiva tuto funkci vyjadřují. Substantiva s obdobnou pre-

⁹⁸ Číslo v závorce udává absolutní frekvenci tohoto typu kombinace funkcí.

ferencí vyjadřovat syntakticko-sémantickou funkci se mohou navíc rozlišovat podle toho, s kterými slovesy se pojí nebo které role upřednostňují.

Myslíme si, že tento přístup k analýze syntaktického uplatnění substantiv může nastínit souvislosti mezi frekvencí pádových tvarů substantiv a jejich sémantickou povahou. Bylo by zajímavé opakovat analýzu ve větším korpusu, který pokrývá i větší spektrum stylu⁹⁹. Takovou analýzu by ovšem bylo nutno provádět manuálně, což není v rámci této práce možné.

4.3 Shrnutí

Viděli jsme, že obecná frekvence pádů je citlivá na sféru užívání (mluvená / psaná čeština) a na funkční styl, a to zejména pokud jde o frekvenci genitivu, akuzativu, ale i vokativu. Zatímco se pořadí ve středním a nízkém pásmu frekvenčního rozložení pádů (tedy u instrumentálu, lokálu, dativu a vokativu) jen málo liší jak v subkorpusech různých funkčních stylů, tak v závislosti na mluvené versus psané formě textů, jsou zde právě u kvantitativně závažnějších pádů G a A poměrně velké frekvenční rozdíly. Z dat není patrné, do jaké míry odlišná frekvenční distribuce pádů v subkorpusech odráží kolísání ve výskytu syntaktických funkcí. Bylo zatím zjištěno jen to, že primární funkce substantiva významně nezávisí na funkčním stylu a že je v subkorpusech poměrně stálá (Těšitelová et al.). Autoři ale zároveň ukázali, že četnost výskytu syntaktických funkcí se může významně lišit v různých podtřídách substantiv, jako např. u deadjektivních nebo číselných substantiv.

Zkoumání chování substantiv v různých funkčních stylech nám pomohlo nastínit, které další faktory (doprovázející změnu funkčního stylu) ovlivňují obecnou frekvenci pádů. Mimo těžko postihnutele aspekty, jako jsou tematický okruh a individuální styl autora, sehrávají roli např. podíl určitých slovtvorných typů, užití deverbálních substantiv a pasivních konstrukcí. Změna v jejich podílech se může odrážet ve variaci frekvence nominativu a akuzativu jako strukturních pádů, ale rovněž ve variaci frekvence genitivu.

⁹⁹ Je potřeba se zmínit o tom, že se analýza opírá o korpus věcného stylu, nezkoumá užití substantiv v beletrii, a nezrcadlí tedy plné spektrum užití slov. Tak je např. nápadné, že se slovo *energie* ocitá v 94 % jeho výskytů jako genitivní tvar ve spojení s přídavným jménem *elektrická*. (srov. Těšitelová, 1990, str. 13).

Pád je navíc vzájemně propojený i s kategoriemi jmenného čísla a rodu. Analýza čísla a rodu naznačuje, že slova různých onomaziologických kategorií, např. *propria*, a slovotvorných typů, např. deverbativa na *-í*, ovlivňují frekvenci jednotného čísla nebo neuter v subkorpusech. Pokud se posuzuje frekvence pádů u určitých vzorů¹⁰⁰ nebo slovotvorných typů, mělo by se k této souvztažnosti přihlížet.

Až dosud jsme ukázali, že obecná frekvenční distribuce pádů je závislá na funkčním stylu a sféře užívání. Souvislost frekvence morfologické kategorie pádu nebo čísla se sémantickou stránkou lexémů se brala v úvahu zatím jen na obecnější úrovni, např. u deverbativ, substantiv číselných nebo *proprií*. Na konkrétních substantivních lexémech ji zatím zkoumala jen Těšitelová (1990), která poukázala na souhru mezi preferencí syntaktických funkcí (a tudíž realizace lexémů v odlišných pádech) a individuální sémantikou lexémů.

Chceme proto v této práci systematicky zkoumat, do jaké míry je frekvenční distribuce pádů závislá na sémantické složce individuálních lexémů. Pokusíme se k tomuto účelu o roztřídění substantivních lexémů na základě profilů, jež vzniknou hierarchizací jejich pádových tvarů podle frekvence v korpusu.

¹⁰⁰ Myslíme zde vzorová slova deklinačních tříd.

5 Metodický postup a data

5.1 Metodický postup a pilotní studie

5.1.1 Úvod

V dané práci se snažíme o přístup k analýze, jež lze spíš charakterizovat jako řízený korpusem (*corpus-driven*) nežli založený na korpusu (*corpus-based*). Touto základní distinkcí podchycuje Tognini-Bonelli (2001) vztah mezi teorií a daty a zároveň vystihuje rozdíl mezi přístupy, jež využívají korpus k ověřování a opodstatnění hypotézy, která je odvozená od jednotlivých příkladů a vznikla hlavně ze zkušeností či intuice lingvistů, a přístupy, jež vycházejí přímo z korpusových dat s minimálními teoretickými předpoklady a přihlížejí při vytváření hypotéz k celku dat (srov. Tognini-Bonelli, 2001, str. 65nn).

Ve smyslu analýzy *corpus-driven*¹⁰¹ se pokoušíme nejprve pozorovat frekvenční distribuci zkoumaného jevu v korpusu na menším pilotním vzorku. Na základě výsledků a zjištěných podobností v chování postupujeme k hypotéze o možném roztřídění. Tuto hypotézu potom ověřujeme na větším náhodném vzorku, který vytváříme ze základního souboru substantiv sestavovaného na základě didaktických a zároveň frekvenčních hledisek. Závěry a tvrzení, které o tvarových profilech substantivních lexémů činíme, se tudíž vztahují především na zkoumaný základní výběr slov, a nemohou být tedy chápány jako platné pro celý korpus, popř. určitou frekvenční hladinu¹⁰².

Analýzu jak pilotního, tak vlastního náhodného vzorku provádíme v lemmatizovaném a morfologicky značkováném korpusu SYN2010. V korpusu je každému slovnímu tvaru přiřazené lemma, tedy tvar reprezentující celé paradigma slovníkové jednotky¹⁰³, a zároveň morfologická značka, která určuje slovní druh a hodnoty příslušných morfologických kategorií. Jednoznačná anotace slovního tvaru je výsledkem morfologické disambiguace, která je zvláště pro češtinu složitý proces vzhledem k náhodné homonymii a morfologickému (zejména pá-

¹⁰¹ Dle Tognini-Bonelli (2001) se postup *corpus-driven* na rozdíl od postupu založeného na korpusu vyznačuje právě empiričností: „The corpus-driven approach, on the other hand, fits well the definition in that it aims to derive linguistic categories systematically from recurrent patterns and the frequency distribution that emerges from language in context“ (tamtéž, str. 87).

¹⁰² Jsme si vědomi, že náš postup není ryze *corpus-driven* v původním smyslu. Neobejdeme se totiž úplně bez teoretických předpokladů, když využíváme morfologicky anotovaný korpus; srov. k tomu níže.

¹⁰³ Na rozdíl např. od pojetí lemmatu jako kořenu sémantického hnízda.

dovému) synkretismu (srov. Jelínek a Petkevič, 2011, str. 160). Disambiguace vybírá podle kontextu nejvhodnější interpretaci slovního tvaru, a to hlavně, avšak nejenom, na základě ručně vytvořených pravidel, která zachycují zejména negramatické strukturní a slovosledné kombinace, a vylučují tudíž z možných interpretací nevhodné. Tato metoda je v kombinaci s modulem, který identifikuje kolokace a frazémy, poměrně úspěšná a resultuje v úspěšnosti pokrytí 98,82 %. V závěrečné fázi disambiguace se aplikuje ještě strojově řízený, statistický tagger *MorČe*. Na češtině se tento hybridní postup osvědčil (srov. Jelínek a Petkevič, 2011). Chybovost automatického značkování dosahuje už jen cca 4 %.

Při vyhledávání tvarů nutně využíváme danou lemmatizaci, jež se opírá o morfologický slovník o rozsahu 350 000 lemmat včetně lemmat propriálních (tamtéž, str. 159). Jsme si vědomi toho, že taková anotace korpusu představuje jistou interpretaci textů, je však nutným předpokladem pro analýzu flexivních tvarů ve stamilionovém korpusu, která by na manuální vyhodnocení kladla příliš velké časové nároky. Jak ale již poznamenali Jelínek a Petkevič (2011, str. 154), shoduje se většina lingvistů v interpretaci jevů na úrovni lemmatizace a morfologických a syntaktických údajů. Zároveň je právě pádový tvar předmětem daného výzkumu, a automatické značkování nám umožňuje získat vůbec poznatky o jeho distribuci v tak velkém množství textů a učinit závěry o jejich užití. Jsme si vědomi problému, na nějž již upozornil Sinclair (Sinclair 1991, str. 8¹⁰⁴), že tvary nereflektují automaticky chování celého lemmatu. Podle něj jsou tedy svědky hlavně a jenom samy pro sebe (tamtéž). Postupujeme proto tak, že na základě pozorování slovních tvarů, které nám jsou nutně přístupné jen přes jejich lemma, činíme závěry o tom, zda odlišné užití neukazuje na zvýraznění určité sémantické složky uvnitř jednoho lemmatu.

5.1.3 Pilotní studie

Pro zjištění základní charakteristiky distribuce substantivních tvarů s přihlédnutím k individuálním substantivním lexémům jsme provedli nejprve pilotní průzkum v korpusu psané češtiny.

Pilotní vzorek byl vytvořen na základě katalogu témat a jazykových funkcí *Prahouvé úrovně* (Bischofová, 2001), jenž byl upraven podle potřeb zjištěných

¹⁰⁴ „There is a good case for arguing that each distinct form is potentially a unique lexical unit, and that forms should only be conflated into lemmas when their environments show a certain amount and type of similarity.”

v dotazníkovém šetření pro projekt *Euromobil*¹⁰⁵. Ze seznamu navržených exponentů relevantních tematických okruhů¹⁰⁶ a jazykových funkcí byla vybrána typická slovesa (*dát si* a *mít chuť*; *studovat*; *bydlet*) a vypočítány jejich frekventované kolokáty¹⁰⁷. Ty byly doplněny o substantiva tematických oblastí *zaměstnání* a *povolání*, jež se ocitají mezi 5000 nejfrekventovanějšími slovy (dle Čermák a Křen, 2011, str. 155). Postup výběru byl tudíž založen hlavně na odhadovaných potřebách a doporučeních učebních pomůcek. K frekvenci sice bylo přihlíženo při výběru kolokátů a substantiv z některých tematických okruhů, nikoli však systematicky. Výsledky není proto možné vztahovat na určité frekvenční hladiny ve slovní zásobě současné češtiny. Výsledný pilotní vzorek omezený na celkem 50 substantiv (apelativ i proprií) zahrnuje následující lexémy:

akademie, alkohol, architektura, biologie, Brno, byt, čaj, centrum, čeština, doktor, dům, ekonomie, fakulta, fyzika, gymnázium, herec, hotel, chemie, chléb (včetně chleba), chování, klavír, knihy, malířství, maso, ministr, mluvčí (tvar mužský), novinář, panelák, párek, patro, pití, podnikatel, policista, politik, Praha, přízemí, ředitel, režisér, řízek, šálek, sklenička, sousedství, trenér, ubytovna, učitel, ulice, venkov, vesnice, víno, zmrzlina

Uvedená substantiva pilotního vzorku jsme podrobili analýze distribuce jejich slovních tvarů v žánrově vyváženém korpusu synchronní češtiny SYN2010, a to jak v celém korpusu, tak v jeho subkorpusech: beletrie, odborná literatura a publicistika¹⁰⁸. Korpus je lemmatizovaný a označovaný. Lemmatizace bylo využito při vyhledávání a zároveň při analýze výskytů pádových tvarů (viz 5.1.1). Pro jednotlivá lemmata byla tudíž vypočítána frekvenční distribuce jejich morfologických forem pomocí funkce „Frekvenční distribuce“ pro KWIC s přihlédnutím k atributu „tag“. Výsledkem analýzy byl seznam všech slovních tvarů jednoho lemmatu v podobě konkrétní značky (např.: NNNS6-----A-----), s údajem o jeho absolutní četnosti. Značku tvoří vždy řetězec 16 znaků¹⁰⁹, ve kterém každá pozice vyjadřuje hodnotu z určité morfologické třídy (v uvedeném příkladu tedy: sub-

¹⁰⁵ Euromobil je jazykový vzdělávací a informační program na podporu mobility studentů, který vznikl v rámci evropské spolupráce a za podpory Evropské komise. Anketa prováděná v rámci projektu zachycuje potřeby zahraničních studentů, stážistů, v oblasti jazykové a sociokulturní (Benešová-Grundová, 2006, str. 114).

¹⁰⁶ Konkrétně se jedná o témata: jídlo a pití / stravování mimo domov, vzdělání a studium, informace o osobě (adresa a povolání).

¹⁰⁷ Pro podrobný popis postupu viz Schöne (2011) a dále také část 5.3.

¹⁰⁸ Korpus SYN2010 se skládá ze 40 % z beletristických, z 33 % z publicistických textů a z 27 % z odborné literatury (viz Cvrček & Richterová, 'cnk:syn2010', 2015).

¹⁰⁹ Ucelený soubor pozičních atributů (tagů) je popsán v *Průručce ČNK* (Cvrček a Richterová, 2015).

stantivum | obyčejné | neutrum | singulár | šestý pád | afirmativ)¹¹⁰. Morfologické značkování korpusu SYN2010 přihlíží i k variantním tvarům a stylovým příznakům (např. se značkuje na patnácté pozici variantní koncovka lokálu pro střední vzor *město*

-u nebo koncovka standardní obecné češtiny pro instrumentál plurálu -ama). Aby bylo možné srovnávat lexémy vzhledem k frekvenci jejich pádových tvarů (s přihlédnutím k číslu), zahrnuli jsme všechny variantní tvary pod příslušný pádový tvar. Stejně jsme postupovali i při analýze vlastního, většího vzorku.

Výsledkem byl výčet absolutní a relativní četnosti pádových tvarů pro jednotlivá lemmata v celém korpusu a zvlášť v subkorpusech. V pilotní studii jsme zvlášť nepřihlíželi k víceslovným lexémům ani jsme se u víceznačných lexémů nesnažili rozčlenit konkordanční řádky podle významu (např. u lexémů *párek* jsme ponechali jak význam a) ‚dvojice lidí, zvířat n. věcí nějak k sobě patřících‘, tak význam b) ‚uzenky spojené po dvou‘).

Následovně byly pro relativní frekvence pádových tvarů vypočítány míry jejich rozptylu, konkrétně směrodatné odchylky, a to:

- a) zvlášť pro frekvence všech pádů u všech lemmat → za účelem zjistit, jak velká je variace¹¹¹ mezi jednotlivými substantivy v četnosti výskytu určitého pádu a
- b) zvlášť pro frekvence pádů u všech lemmat v subkorpusech → za účelem zjistit variabilitu výskytů pádů v závislosti od funkčního stylu, a to rovněž pro jednotlivá lemmata.

Protože nelze směrodatné odchylky pro jednotlivá lemmata přímo porovnávat, vypočetli jsme variační koeficienty (v).

Vyhodnocení pilotní studie

Tabulka 5.1 ukazuje jednak frekvenční distribuci pádů v celém korpusu SYN2010, jednak podíl jednotlivých pádů (N-I) na všech pádových tvarech v pilotním vzorku. Pravá část tabulky znázorňuje, do jaké míry se odchylojí výskyty pádových tvarů u jednotlivých lemmat.

¹¹⁰ Pozice 6 – 10, 12 – 14 a 16 se pro substantiva neurčují.

¹¹¹ Zde: v obecnějším smyslu.

pád	relativní frekvence v %		směrodatná od- chylka	variační koefi- cient
	SYN2010	lemm. pilotní studie		
N	22,0	32,9	18,46	0,662
G	17,6	28,5	11,05	0,374
D	2,8	2,6	3,57	1,142
A	15,0	13,2	13,41	0,726
V	0,3	0,3	1,36	5,541
L	9,6	16,9	16,70	1,146
I	6,3	5,7	3,12	0,509

Tab. 5.1 Distribuce pádových tvarů v pilotním vzorku a jejich variace vzhledem k jednotlivým lemmatům

Vidíme, že variace pádové frekvence je poměrně velká, tj. variační koeficient přesahuje pro všechny pády až na genitiv hodnotu 0,5 (popř. 50 %) - hranice, která je ve statistice obecně chápána jako ukazatel, že střední hodnota už není vhodná jako reprezentant jednotlivých dat pro jejich vysoký rozptyl. Pro názornost ukazujeme pro vybrané pády

a) variace pádů dle funkčního stylu: N $v = 0,075$ | G $v = 0,158$ | V $v = 1,12$ a

b) variace dle jednotlivých lemmat: N $v = 0,662$ | G $v = 0,374$ | V $v = 5,541$.

Poněvadž se při znázornění variace pádové frekvence podle jednotlivých lemmat nemůžeme opírat o přímé srovnání jejího rozptylu s rozptylem pádů dle typů textů (subkorpora beletristika, odborná literatura a publicistika¹¹²), uvádíme zde alespoň maxima a minima relativní pádové frekvence v pilotním vzorku:

Pád	Extrémní hodnoty
N	max: <i>trenér</i> 67,2 %, min: <i>sousedství</i> 4 %
G	max: <i>chemie</i> 54,8 %, min: <i>párek</i> 8,8 %
D	max: <i>novinář</i> 18,6 %, min: <i>dům</i> 0,3 %
A	max: <i>sklenička</i> 55,6 %, min: <i>politik</i> 2,6 %
V	max: <i>doktor</i> 9,8 %, min: 0 % (u 32 lemmat)
L	max: <i>sousedství</i> 71 %, min: <i>mluvčí</i> 0,1 %
I	max: <i>pití</i> 11,9 %, min: <i>sousedství</i> 1,3 %

Tab. 5.2 Maxima a minima v distribuci pádových tvarů v pilotním vzorku

Můžeme shrnout, že nejvíce kolísá frekvence vokativu, dativu a lokálu. Tyto pády se ukazují jako více závislé na sémantice nebo individuálním užití substantiv než ostatní pády. Naopak nejméně se odchyľují od průměrné hodnoty substantiva pilotního vzorku, když porovnáme frekvenci v genitivních tvarech.

¹¹² V této práci užíváme dále jen zkratky BEL, ODB a PUB.

Příklady toho, do jaké míry může být četnost určitého pádového tvaru ovlivněna funkčním stylem, naznačují tabulky 5.3a a 5.3b, které ukazují pro vybrané lexémy výkyvy pádové frekvence v subkorpusech SYN2010 BEL, ODB a PUB pomocí varičního koeficientu. Zvýrazněny jsou hodnoty *u*, které jsou větší než příslušné hodnoty pádové variace v přehledu všech substantiv v korpusu.

Pád	akademie		čeština		čaj		sklenička	
	F rel. v %	<i>u</i>	F rel. v %	<i>u</i>	F rel. v %	<i>u</i>	F rel. v %	<i>u</i>
N	17,3	0,15	19,7	0,11	24,2	0,33	10,0	0,29
G	51,6	0,10	33,4	0,13	28,0	0,03	20,1	0,20
D	1,8	0,04	2,7	0,40	1,9	0,47	0,7	0,12
A	13,8	0,21	11,7	0,28	36,4	0,26	55,6	0,25
V	0,0	-	0,0	-	0,0	-	0,0	-
L	12,8	0,28	26,0	0,20	3,0	0,18	6,8	0,30
I	2,7	0,16	6,4	0,40	6,4	0,20	6,7	0,18

Tab. 5.3a Relativní frekvence pádových tvarů pro vybraná lemmata v SYN2010 a její variabilita v subkorpusech

Pád	ulice		byt		ředitel		doktor	
	F rel. v %	<i>u</i>	F rel. v %	<i>u</i>	F rel. v %	<i>u</i>	F rel. v %	<i>u</i>
N	8,5	0,11	10,1	0,05	63,8	0,09	56,6	0,03
G	24,1	0,13	43,9	0,12	19,5	0,09	12,2	0,22
D	1,6	0,18	0,8	0,25	3,0	0,57	6,0	0,14
A	8,6	0,39	23,7	0,08	4,4	0,22	9,4	0,08
V	0,002	1,41	0,0	-	0,4	1,17	9,8	0,65
L	48,9	0,14	18,6	0,31	0,3	0,51	0,9	0,23
I	8,4	0,29	2,8	0,26	8,7	0,23	5,1	0,04

Tab. 5.3b Relativní frekvence pádových tvarů pro vybraná lemmata v SYN2010 a její variabilita v subkorpusech

Jak lze očekávat, je vokativ pro svou funkci sloužit jako oslovení¹¹³ pádem s velice individualizovaným užitím. Jeho výskyt je nejvíc závislý na individuálním slovu (vyskytuje se jen u 16 z 50 slov), a pokud se vyskytuje, tak nejčastěji (alespoň v 60 % výskytů) v beletristických textech. Také můžeme pozorovat tendenci, že pro polovinu substantiv připadá nejvíce výskytů dativních forem na beletristiku, např. u slov *hotel*, *dům*, *ředitel*, *mluvčí*, *patro*, *čaj*, *pítí*, *biologie*, *češti-*

¹¹³ Vokativ se vyjímá z řady ostatních pádů, poněvadž svým významem ‚oslovení‘ ukazuje ven z jazyka. Stojí vně syntaktické struktury věty (srov. EŠČ Karlík, 2002, str. 532).

na, chemie. Jinak nelze pozorovat, že by určitý pád zvláště vynikal, výkyvy frekvence jednotlivých pádů v tom či onom subkorpusu jsou významné jen pro konkrétní slova. Například *Brno* a *mluvčí* se vyskytují v publicistice výrazně častěji v N než v beletrii nebo v odborné literatuře. Lemma *sklenička* je v naprosté většině případů svého výskytu v beletrii ve tvaru akuzativním, kdežto rozložení pádů v ODB a PUB je rovnoměrnější. S instrumentálním tvarem *češtinou* se setkáváme převážně v beletrii. U slova *podnikatel* je neobvyklá vysoká hodnota u pro lokál. Důvodem je, že se lokálové tvary tohoto slova vůbec nevyskytují v publicistice a v beletrii jen řídce. Podobně tomu je u lemmat *řízek* a *párek*.

Celkově lze říci, že výkyvy pádů u jednotlivých slov v závislosti na funkčním stylu jsou výrazné jen pro jednotlivá slova, a to pro konkrétní pády. Vyšší potenciál lišit se v užití dle funkčního stylu mají málo frekventované pády jako dativ a vokativ – tuto tendenci jsme již pozorovali při souhrnném pohledu na všechna lemmata pilotního vzorku.

V dalším kroku jsme seřadili pádové tvary lemmat podle jejich frekvence. Následně byly lexémy tříděny podle slovních tvarů, ve kterých vystupují nejčastěji. Procedura byla prováděna a) jen na pádových tvarech bez přihlídnutí ke kategorii čísla a b) na morfologických tvarech i s přihlídnutím ke kategorii čísla. 50 slov pilotního vzorku lze podle toho přiřadit a) čtyřem typům a b) pěti typům, které jsou znázorněny v tabulkách 5.4a a 5.4b. Podle toho, který pádový tvar (popř. morfologický tvar) je druhým nejčastějším, byly vytvořeny příslušné podtypy.

TYPY DLE NEJFREKVENTOVANĚJŠÍHO PÁDU	POČET LEMMAT	POČET PODTYPŮ	RELATIVNÍ FREKVENCE
TYP1a Genitiv	23	3	46 %
TYP2a Nominativ	14	2	28 %
TYP3a Akuzativ	8	3	16 %
TYP4a Lokál	5	1	10 %
celkem	50	9	100 %

Tab. 5.4a Typy morfologických profilů substantiv podle první pozice ve frekvenčním pořadí (jen pádové tvary)

TYPY DLE NEJFREKVENTOVANĚJŠÍHO MORFOLOGICKÉHO TVARU (PÁD S ČÍSLEM)	POČET LEMMAT	POČET PODTYPŮ	RELATIVNÍ FREKVENCE
TYP1b Genitiv singuláru	21	4	42 %
TYP2b Nominativ singuláru	11	3	22 %
TYP3b Lokál singuláru	8	2	16 %
TYP4b Akuzativ singuláru	7	4	14 %
TYP5b Nominativ plurálu	3	2	6 %
celkem	50	15	100,0 %

Tab. 5.4b Typy morfologických profilů podle první pozice ve frekvenčním pořadí substantiv (pád s číslem)

Je patrné, že téměř polovina všech zkoumaných slov se vyskytuje nejčastěji ve tvaru Gsg. Tato skupina má zároveň nejvíce podtypů. Nejtypičtějším pádem pro většinu slov v daném vzorku tedy není nominativ, ale genitiv. Nominativ následuje až na druhé pozici v žebříčku. Pro 10 % substantiv pilotního vzorku je lokál nejfrekventovanějším pádem. Typ 3a, popř. 4b (nejfrekventovanější výskyt akuzativních tvarů) je nejvíce rozčleněn podle druhé pozice v pořadí pádů. Na výsledcích seřazení morfologických tvarů je zajímavé, že tři z 50 slov se dokonce vyskytují nejčastěji ve tvaru Npl. (*policista, politik, novinář*).

Když rozřídíme substantiva na základě frekvenční distribuce jejich pádových forem podle prvních dvou pozic, rýsují se následující podtypy:

Podtyp	1. pozice	2. pozice	počet	lemmata tohoto typu
1a1	N	G	12	<i>ředitel, ministr, mluvčí (M), doktor, trenér, režisér, policista, učitel, novinář, podnikatel, herec, politik</i>
1a2	N	L	2	<i>Brno, Praha</i>
2a1	G	N	6	<i>akademie, architektura, biologie, fyzika, chemie, kniha</i>
2a2	G	A	8	<i>alkohol, byt, ekonomie, chléb, chování, malířství, maso, víno</i>
2a3	G	L	9	<i>centrum, čeština, dům, fakulta, gymnázium, hotel, panelák, ubytovna, vesnice</i>
3a1	A	N	2	<i>klavír, řízek</i>
3a2	A	G	5	<i>čaj, šálek, sklenička, zmrzlina</i>
3a3	A	D	1	<i>pítí</i>
4a1	L	G	5	<i>patro, přízemí, sousedství, ulice, venkov</i>
celkem			50	

Tab. 5.5 Typy morfologických profilů podle prvních dvou pozic ve frekvenčním pořadí

Konkrétní příklady zřetelně ukazují na sémantickou podobnost jednotlivých členů. Veškeré názvy povolání se ocitají ve skupině 1a1, názvy dvou největších českých měst mají rovněž stejný profil a podtyp 2a2 zahrnuje látková jména

s výjimkou slov *čaj* a *zmrzlina*, která se ocitají ve skupině 3a3, a mají tedy rovněž akuzativní a genitivní tvary v čele frekvenčního žebříčku, jen v obráceném pořadí. Substantiva, která slouží k označení míst v širokém smyslu, se nacházejí v podtypech skupiny slov s genitivem jako nejfrekventovanějším pádem – jsou tu nejsilnější skupinou v podtypu 2a3 („G-L“) nebo tvoří vlastní podtyp 4a1 („L-G“). Názvy oborů lidské činnosti se nápadně nashromažďují v podtypu 2a1 (*architektura, ekonomie, malířství, fyzika, biologie*), přičemž *ekonomie* a *malířství* se dělí se jmény látkovými o podtyp „G-A“, a substantivum *čeština* se přichycuje k substantivům označujícím místa. Podobná frekvence slovních tvarů může ukazovat na užití v obdobných kontextech a na obdobné sémantické složky. Tak je např. 90,9 % všech kolokátů, s kterými se tvar *češtině* ve svém přímém levém kontextu kombinuje, předložka *v*.

5.1.4 Závěry z pilotní studie

První sondy do korpusu na malém vzorku 50 slov ukázaly, že se frekvence pádových tvarů pro konkrétní lexémy významně liší od průměrného rozložení pádových tvarů v korpusu. Četnost, s níž se určitý lexém vyjadřuje v konkrétním pádu, je do značné míry individuální jev, a to jak v celém korpusu tak v jeho subkorpusích. Obecnou tendenci, že by výskyt pádu závisel významně i na funkčním stylu, jsme pozorovali jen u vokativu a dativu. Vzhledem k postavení nominativu lze předběžně shrnout, že je sice v celkovém rozložení pádových tvarů všech substantiv, a to zejména v beletristických a publicistických textech, nejčastějším tvarem, ale jen pro zhruba třetinu substantiv tvarem nejběžnějším. Oproti tomu se vyskytuje téměř polovina slov nejčastěji v genitivu.

Jak je zřejmé, substantiva je možné roztrždit do smysluplných skupin na základě frekvenční distribuce jejich morfologických tvarů. Počet, povaha a rozsah jednotlivých typů jsou však nepochybně závislé na velikosti a složení vzorku, a výsledky je proto potřeba ověřit na větším vzorku substantiv. K tomuto účelu je nutné řešit několik metodických otázek, kterým se tato kapitola věnuje:

- Poněvadž nemůžeme zkoumat všechna substantiva obsažená v korpusu, je třeba řešit, jak vypadá základní soubor a na základě čeho vytvoříme náhodný vzorek. V této souvislosti se musíme zabývat i otázkou, zda zapojíme do studie také víceslovná pojmenování, jež jsme ve své pilotní studii opomíjeli.

- Ústřední otázkou je dále, který korpus je vhodný pro daný výzkumný záměr.
- Pro možné závěry o propojení výskytu pádových tvarů se sémantickou stránkou jednotlivých slov je důležité vyjasnit, které souvislosti mezi existujícími kategoriemi popisujícími substantiva podle typu nebo povahy denotátu a výrazem kategorie čísla a pádů jsou známe (např. distinkce abstrakta a konkréta, propria a apelativa, substantiva počítatelná i nepočítatelná) a jak se chovají vzhledem ke zkoumanému jevu.
- Vzhledem k závěrům o morfologických profilech jednotlivých lemmat je třeba si klást otázku, jak lze vymezit jádro tohoto jevu pro konkrétní lemma, a obecně, jaké pásmo frekvenční distribuce slovních tvarů resp. jaký počet tvarů můžeme považovat za typické užití lemmatu.
- V neposlední řadě je potřebné najít přístup k případné víceznačnosti lexémů, a tedy k otázce omezení konkordančních řádků, které vyhodnotíme.

5.2 Užití korpusy

Analýzu distribuce substantivních tvarů provádíme v subkorpusu stamilionového korpusu současné psané češtiny SYN2010¹¹⁴ a menší vzorek podrobíme i manuální analýze v korpusu mluvené češtiny ORAL2008.

Naše studie se snaží být reprezentativní pro češtinu jako cílový jazyk. Omezujeme se navíc na jevy relevantní pro studenty, již chtějí dosáhnout určité komunikativní kompetence, kterou vymezujeme pomocí celoevropsky uznávaného nástroje ke stanovení úrovně jazykových znalostí *Společného evropského referenčního rámce pro jazyky* (SERR, 2002) na úroveň B1. Jsme si vědomi nedostatků nástroje SERR¹¹⁵, ale spolu s popisem referenční úrovně B1 *Prahová úroveň* (Bischofová et al., 2001) nám poskytuje funkčně orientovaný a velice podrobný popis kompetenční úrovně s přihlédnutím k rozdílným rovinám jazyka. Vzhledem k tomu, že se zahraniční student nepodílí na komunikaci v nejrůznějších komunikačních situ-

¹¹⁴ Křen, et al. (2010) *SYN2010: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Dostupný z WWW <<http://www.korpus.cz>>.

¹¹⁵ Srov. kapitulu 2. Z hlediska jeho využitelnosti v pedagogické praxi se tomuto nástroji dále vytýká, že výstupní cíle jsou formulovány příliš obecně a že nedokáže propojit výuku a testování s Rámcem (viz o kritice pedagogů Šindelářová, 2010). Zcela zásadním problémem je, že referenční úrovně nemají dostatečnou oporu v autentických datech žákovského jazyka (viz o tom např. Wisniewski et al., 2013).

acích ve stejné míře jako rodilý mluvčí, poněvadž má na začátku svého kontaktu s cílovým jazykem jiné komunikační potřeby a samozřejmě protože mu/jí chybějí jazykové dovednosti v potřebné míře, pokusili jsme se podchytit omezenou účast na komunikaci vytvořením subkorpusu. Za vhodné východisko pro naši analýzu považujeme korpus SYN2010, poněvadž je korpusem reprezentativním. Zrcadlí tedy v dostatečné míře variabilitu v jazyce, popř. obsahuje v dostatečném množství texty všech variet¹¹⁶. Složení korpusu je založené na průzkumu recepce psaných textů běžnými čtenáři¹¹⁷. Lze ho proto považovat za vyvážený z hlediska užívání jazyka. Texty zahrnuté v korpusu SYN2010 odrážejí tudíž jazyk, kterým byl obklopen běžný čtenář v České republice v letech 2005–2010¹¹⁸. Z tohoto hlediska ho považujeme i za reprezentativní pro jinojazyčného mluvčího, který buď pobývá v České republice na určitou (delší) dobu, nebo takový pobyt zamýšlí, a chce se proto učit češtině jako cílovému jazyku¹¹⁹. Na základě typů textů, s kterými se student musí umět vypořádat, je ovšem možné přizpůsobit korpus cílové skupině a vymezit subkorpus.

Od studenta, který chce dosáhnout znalostí a dovedností na úrovni B1 SERR, se očekává, že:

„rozumí hlavním myšlenkám srozumitelné spisovné vstupní informace týkající se běžných témat, se kterými se pravidelně setkává v práci, ve škole, ve volném čase atd. Umí si poradit s většinou situací, jež mohou nastat při cestování v oblasti, kde se tímto jazykem mluví. Umí napsat jednoduchý souvislý text na témata, která dobře zná nebo která ho/ji osobně zajímají. Uživatel jazyka dokáže popsat své zážitky a události, sny, naděje a cíle a umí stručně vysvětlit a odůvodnit své názory a plány.“ (SERR, 2002, str. 24; zvýraznění autorka)

Prahová úroveň (Bischofová, 2001, str. 3-9) uvádí další specifika vzhledem k situacím, ve kterých si student musí umět poradit:

1. Student si bude umět poradit v praktickém, každodenním životním „provozu“, s výrazným podílem předem předvídatelného způsobu jazykového vyjadřování.

Např.: styk s úředními místy, ubytování, stravování, nákupy (spotřební zboží), veřejná doprava, soukromá doprava, využití informačních služeb, návštěva veřejných zařízení

¹¹⁶ Viz Cvrček & Richterová, 'cnk:syn2010', 2015.

¹¹⁷ Viz Cvrček & Richterová, 'pojmy:reprezentativnost', 2013.

¹¹⁸ To nevyklučuje, že korpus obsahuje i texty starší, zejména v oblasti beletrie. Přesto mají největší podíl na korpusu novější texty (tamtéž).

¹¹⁹ Cílovým jazykem rozumíme dle Šindelářové a Škodové (2013, str. 4) jazyk, „který jedinec-cizinec nabyvá a jehož osvojení je cílem učebních aktivit, v němž chce být mluvčí schopen komunikovat“.

(museí, kin, apod.), použití veřejných služeb, školská a vzdělávací zařízení (v roli studenta nebo v roli rodiče), hledání cesty, zaměstnání, soukromá pohostinnost.

2. V oblasti společenského styku je student schopen: vyměňovat si informace; vyjádřit názory a porozumět vyjádření názoru (...); dohodnout se na spolupráci a provádět ji, (...);

To vše s ohledem na náměty osobního i obecného zájmu¹²⁰.

Předvídaná činnost mluvčího na úrovni B1 se tedy vymezuje hlavně podle domény užívání jazyka, ale i tematicky a subjektivní relevantností. Texty, s kterými by se studenti měli umět poradit, jsou vymezené právě vztahem k situaci nebo k určitému tématu. Rozdíly mezi jednotlivými kompetenčními úrovněmi se činí nikoliv na základě rozsahu či povahy lexikální náplně, ale hlavně na bázi úloh, které studenti plní, a relevance pro studenta (Bischofová et al., 2001, str. 214). Autoři *Prahové úrovně* jmenují mj. např. pokyny, instrukce, oznámení, plakáty, reklamy, inzeráty, brožury, letáky, formuláře, telefonní seznamy a informační tabule, ale také osobní korespondenci, informativní články z novin a časopisů a předpověď počasí (tamtéž, str. 215). Explicitně vylučují jako nevhodné pro danou úroveň takové typy textů, které kladou na studenta větší požadavky, pokud jde o znalosti cizího jazyka, než které se právě od této úrovně mohou očekávat, jako např. „pravidla a nařízení“, která vyžadují znalost vysoce formálního úředního vyjadřování“ (tamtéž, str. 214n).

Shodně s požadavkem Bibera¹²¹ a podle specifikace úrovně B1 by bylo tedy nejlepší řešení sestavit subkorpus prioritně na základě situačních parametrů. Avšak víme, že vnější jazyková kritéria, jako situace, instituce a téma, a vnitřní charakteristiku textu, která je ovlivněná i užitím a distribucí jazykových prostředků, nelze od sebe při klasifikaci textů úplně odpojit¹²². Zmenšili jsme tudíž korpus *SYN2010* na základě strukturního atributu *textový typ* (txttype), který odráží základní třídění textů z hlediska registru a charakterizuje text po stránce stylové (Cvrček a Richterová, 2013). Další specifikace subkorpusu na základě atributu *genre* (žánr), jenž dále specifikuje zejména tematické zařazení textu, se nám nezdá být užitečná, poněvadž zájmy studentů mohou být různé.

¹²⁰ Pro podrobnější výklad viz kapitolu 3 „Rozšířená charakteristika“ (Bischofová, 2001, str. 3-9).

¹²¹ Ve své stati o reprezentativnosti v korpusu a o stratifikaci kritérií, jež se uplatňují při budování korpusu, zastává Biber (1993, str. 245) názor, že by rozdíly na rovině *register/genre* měly mít přednost před rozdíly na rovině *text type*, a tím kritéria vnější před kritérii vnitřními.

¹²² Srov. typologie textů, EŠČ (Karlík, 2002).

Subkorpus, který podléhá dané analýze (nazvali jsme ho SYNEK2010), zahrnuje tedy následující typy textů:

Beletrie
NOV - román
COL - soubor povídek, jednotlivá povídka
FAC - literatura faktu
IMA - jiné imaginativní texty
SCR - dramatický text, scénář
Odborná literatura
POP - populárně-naučná literatura (včetně profesních a zájmových časopisů)
TXB - učebnice
ENC - abecedně, systematicky a jinak uspořádaná díla
Publicistika
MIS - rozmanité (efemera)
PUB - publicistika (noviny a neoborné časopisy)

Tab. 5.6 Složení subkorpusu SYNEK2010 podle typů textů¹²³

Subkorpus tedy neobsahuje texty typu: básně a písně (VER, SON), vědeckonaučnou literaturu (SCI) ani administrativu (ADM). Básně a písně byly vyloučeny pro svoji výraznou estetickou funkci, a protože k jejich porozumění se předpokládá rozsáhlá slovní zásoba, znalost četných idiomů, a důkladná obeznámenost s kulturou. Podle požadavků SERR a *Prahové úrovně* rovněž se neočekává, aby si student uměl poradit s vědecko-naučnou literaturou. Naproti tomu považujeme za možné, že se setkává v rámci zájmového čtení nebo čtení pro profesionální potřeby s populárně-naučnou literaturou, která zahrnuje též profesní a zájmové časopisy. Vyloučením zmíněných typů textů se částečně změní původní poměr textových skupin beletrie (40 %), publicistika (33 %) a odborná literatura (27 %) v *SYN2010* ve prospěch publicistiky.

Jsme si vědomi, že některé z typů textů (zde v širokém pojetí *Prahové úrovně*), jež mohou být relevantní pro studenty na úrovni B1, se v korpusu psané češtiny neobjevují, jako např. informační brožury, letáky, informační tabule a reklamní texty.

Cizinec se v České republice setkává samozřejmě také, a v některých kontextech téměř výlučně, s mluvenou češtinou. Proto není sporu o tom, že je náležité prozkoumat syntaktické chování substantiv i tam. Navíc souhrnný přehled o frekvenční distribuci substantiv a všech pádových tvarů v mluveném korpusu současné češtině ukázal (srov. kapitolu 4.1), že substantiva jako slovní druh jsou vý-

¹²³ Cvrček a Richterová, 'seznamy:txtype', 2013

znamně méně četná v mluvené než ve psané podobě češtiny. Značné rozdíly jsou navíc patrné i v distribuci tří nejfrekventovanějších pádů (N, G, A). Nejvhodněji se nám pro výzkum distribuce pádových tvarů v mluvené češtině jeví referenční sociolingvisticky vyvážený korpus neformální mluvené češtiny ORAL2008 (Waclawičová et al., 2008)¹²⁴, a to rovněž z hlediska didaktického využití výsledků.

Korpusy řady ORAL však nejsou morfologicky označované, a analýza distribuce pádových tvarů je proto manuálním úkolem, který je časově velmi náročný. Aby bylo možné tento úkol prakticky i časově zvládnout, byla analyzována jen polovina substantiv pilotního vzorku.

5.3 Výběr substantiv a vytvoření náhodného vzorku

Chceme tuto studii opřít o základní soubor dat, který by měl zajistit, aby byly její poznatky (výstupy) závažné pro danou cílovou skupinu a na druhé straně pro výuku na vybrané úrovni jazykové kompetence. Vzhledem k didaktickému zaměření celého projektu je tedy vhodné omezit jazykový materiál na základě tří kritérií:

- a) vyjadřovací potřeby studentů češtiny
- b) popsání standard, tedy to, co by měli být studenti schopni vyjadřovat a ovládat na určité úrovni jazykového ovládnutí
- c) frekvence výrazů v soudobém úzu rodilých mluvčích, tedy jazykový svět, kterým mohou být studenti obklopeni

Základní soubor podstatných jmen byl proto vytvářen ve třech krocích:

1. zjišťování vyjadřovacích potřeb studentů češtiny
2. vymezení vyjadřovacích prostředků, které jsou potřebné pro řešení komunikačních úloh v identifikovaných komunikačních situacích na určité úrovni osvojovacího procesu za pomoci *Prahové úrovně* a učebnic *Tschechisch Schritt für Schritt* (Holá, 2006) a *Česky krok za krokem 2* (Holá a Bořilová, 2009)
3. úprava výběru substantiv na bázi frekvence výrazů v soudobém úzu rodilých mluvčích

5.3.1 Zjišťování vyjadřovacích potřeb studentů češtiny

Při výběru materiálu se chceme zaměřit především na potřeby univerzitních studentů, již plánují krátkodobý pobyt v České republice za studijními účely nebo přicházejí do styku s českými studenty doma. Tito studenti jsou na jazykových

¹²⁴ V době analýzy mluveného korpusu nebyl ještě k dispozici rozsáhlejší korpus ORAL2013.

kurzech připravování na to, aby se orientovali v českém jazykovém prostředí, aby mohli využívat veřejné služby, aby se orientovali v univerzitním prostředí a aby se mohli dorozumívat s českými studenty v běžných situacích. Cílem kurzů (ani úrovně B1) není připravit studenty na studium jejich oboru v českém jazyce.

Pro identifikaci vyjadřovacích potřeb se můžeme opřít o informace získané v rámci dotazníkového šetření pro již zmíněný projekt Euromobil, který byl realizován od roku 2005 do 2007 mezi studenty-stážisty zúčastňujícími se výměnného programu Erasmus v České republice. Šetření mělo za cíl zachytit potřeby zahraničních studentů dorozumívat se v jazykové a sociokulturní oblasti a zjistit oblasti komunikace a užívání jazyka, jež jsou závažné pro studenty pobývajících omezenou dobu v České republice. Ankety se zúčastnilo 65 studentů nejrůznějších oborů studia (Benešová-Grundová, 2006, str. 115). Respondenti viděli potřebu znalosti češtiny v následujících komunikačních sférách:

oblasti užívání češtiny	počet respondentů (z celkem 65)
obstarání jídla	20
domluva s místními lidmi	11
orientace ve městě	10
nakupování všeobecně	8
cestování	5
pochopení místního způsobu života	5
studium	4
zdraví	2
hledání ubytování	2
získávání přátel	2
sledování sdělovacích prostředků	1

Tab. 5.7 Potřeby znalosti češtiny dle komunikačních situací a oblastí (převzato z Benešová-Grundová, 2006, str. 119)

Z uvedených detailů vyplývá, že největší potřeba náležitě užívat češtinu je při využití různých služeb zejména s místními lidmi a s pracovníky univerzitních a mimoškolních služeb, kde studenti nemůžou spoléhat na znalost angličtiny. Jak ukazuje tabulka 5.7, je mezi dotazovanými studenty poměrně málo potřebností dorozumívat se česky v oblastech zdraví, hledání ubytování, získávání přátel a sledování sdělovacích prostředků. Potřebu znalosti češtiny pro studium uvedlo překvapivě málo studentů. Tuto skutečnost lze vysvětlit tím, že jen 8 dotazovaných studovalo výhradně v češtině a pouze 14 jak v anglickém, tak v českém jazyce. Nicméně se respondenti s komunikačními obtížemi

v akademickém prostředí přesto setkávali, a to tehdy, když mluvili s neakademickými pracovníky vysokých škol, jako např. s pracovníky studijních oddělení, správci počítačových pracoven, knihovníky apod. (Benešová-Grundová, 2006, str. 117). Tento průzkum zachycuje potřeby poměrně úzkého okruhu zájemců o češtinu. Nicméně v této práci využíváme jeho výsledky k vymezení komunikačních oblastí, na něž se výběr substantiv bude soustřeďovat.

Substantiva, jež podrobíme frekvenční a morfologické analýze, jsme proto vybrali z okruhu jazykových prostředků, které jsou potřebné pro řešení následujících komunikačních úloh:

- obstarávání jídla: domluva s pracovníky menz, kaváren, restaurací a obchodů s potravinami
- domluva s místními lidmi, zejména pracovníky v univerzitním prostředí, na ulici, s poskytovateli různých služeb jako např. na poště, v bance, v internetové kavárně, s poskytovateli kopírovacích služeb apod.
- orientace ve městě: dotazy na cestu, spoje, dopravní prostředky, směry a porozumění příslušným informacím
- nakupování všeobecně: domluva s pracovníky různých obchodů a služeb
- cestování: získávání informací o veřejné dopravě, podávání informací o cestách v České republice a z/do země původu, popř. do jiných zemí za účelem dovolené nebo absolvování pracovní stáže, získání informací spojených s pobytem a domluva s českými hostiteli
- pochopení místního způsobu života
- studium: podávání informací o studovaném oboru, délce a způsobu studia, získání informací spojených se studijními záležitostmi

Jsme si vědomi toho, že některé situace vyžadují aktivní účast v ústní komunikaci (např. v menším obchodě, na nádraží, překládání nějaké žádosti apod.), kdežto v jiných budou studenti převážně posluchači či čtenáři. Na rozdílnosti frekvencí slov v aktivní a pasivní slovní zásobě však nemůžeme brát ohled při výběru slov, poněvadž není těžištěm naší práce.

5.3.2 Výběr substantivních výrazů

Vytváření základního souboru substantiv proběhlo ve dvou krocích. Pro vymezení závažných vyjadřovacích prostředků bylo nejdříve třeba vztahovat identifikované potřeby na platný standard, konkrétně na popis jazykových způsobilostí pro úroveň B1 (podle SRR). Poněvadž soupis slovní zásoby vybrané *Prahové úrovně B1* není vzhledem k náplni některých tematických okruhů úplný, byl v druhém kroku doplňován pomocí slovníkových částí dvou učebnice vedoucích k úrovni B1.

5.3.2.1 Určování referenční úrovně a slovní zásoby závažné pro tuto úroveň

Pro vymezení závažných vyjadřovacích prostředků byly identifikované potřeby vztahovány na platný standard. Opřeli jsme se o šestičlennou stupnici SERR, která pro každou úroveň jazykového ovládnutí – od A1 do C2 – vykazuje, v jaké míře je učící se schopen v cizí řeči komunikovat a to vzhledem ke třem metakategoriím: *komunikativní*¹²⁵ *činnost*, *strategie* a *komunikativní jazykové kompetence* (SERR, 2002, str. 25).

Oblasti komunikace a příslušné úlohy (srov. např. *Veřejná doprava – Student by si měl umět opatřit jízdenky a žádat o vysvětlení, co se týče cílové stanice* apod.), které jednotlivé úrovně udávají, jsou podobné. Ale úlohy je možné řešit na různých úrovních jazykové způsobilosti a úrovně se tudíž liší rozsahem potřebné slovní zásoby a mírou jazykové obratnosti při řešení komunikačních úloh¹²⁶. V některých situacích se navíc očekávají i jiné komunikační činnosti.

Za cílem vhodně omezit skupinu zkoumaných substantiv a sestavit základní soubor, který je rovněž závažný pro naši cílovou skupinu, jsme si vybrali jako referenční stupeň opět úroveň B1, tj. *Prahovou úroveň*. Ta popisuje jazykové způsobilosti zásadně na dvou rovinách: a) rovina *jazykových funkcí*¹²⁷, které popisují, co lidé pomocí jazyka dělají (např. děkování, kladení otázek, omluvy apod.) a b) rovina *pojmu*, které pojmenovávají určité představy a situace, kterými se *jazykové funkce* naplňují (Bischofová, 2001, str. 11). *Pojmy* se dále rozdělují na *pojmy obecné* (např. existence, prostor, čas) a *pojmy specifické* (např. jídlo a pití, osobní informace apod.). Výčet pojmů obsahuje konkrétní jazykové složky, tzv. *exponenty*, např. „2.13 postoje morální → 2.13.1 omluva → 2.13.1.1 +prominout v imperativu; (já nerad/a, já nechtěl/a)“ (Bischofová, 2001, str. 48). Jazykové

¹²⁵ Česká verze SERR užívá pojem *komunikativní činnost* a *komunikativní úloha*. Ve vlastním textu však v tomto kontextu budeme dále užívat adjektivum *komunikační* (viz diskuzi v kap. 2.3).

¹²⁶ Srov. např. komunikační úloha „zajišťování stravování“. Zatímco se od studenta na úrovni A2 očekává, že umí „požádat o jídelní lístek, popř. o různé ingredience; objednat si jídlo a nápoje; požádat o účet a formulovat jednoduchou reklamaci na pomalou obsluhu, na nekvalitní jídlo / pití či na předražování“ (Čadská, 2005, str. 6), měl by si student na úrovni B1 umět navíc „zajistit rezervaci (vyhlédnutého) stolu; informovat se, zda je v ceně zahrnuta daň s obsluhou [a] vyjadřovat se kriticky, příp. si stěžovat na pomalou obsluhu, nekvalitní jídlo, předražování apod.“ (Bischofová, 2001, str. 4).

¹²⁷ Pojem jazykové funkce, jak ho užívá Prahová úroveň, není totožný s českým jazykovědným termínem komunikační funkce výpovědi. Výčet jazykových funkcí v Prahové úrovni přijímá hlavní seskupení jazykových funkcí anglického vzoru Threshold level 1990: 1 – Sdělování a hledání faktických informací; 2 – Vyjadřování postojů; 3 – Ovlivňování průběhu věcí, děje; 4 – Společenský styk; 5 – Strukturace projevu; 6 – Korekce při komunikaci. Jemnější roztřídění však už přihlíží k rozvinutému popisu komunikačních funkcí výpovědi v české jazykovědné literatuře (Bischofová, 2001, str. 13).

složky jsou řazené podle závažnosti, již stanoví autoři na základě svých zkušeností. Ty, jejichž znalost se považuje za nezbytnou, byly označeny znaménkem (+) a dostaly se do slovního rejstříku *Prahové úrovně*. Z tohoto slovního rejstříku jsme vybrali všechna podstatná jména, jednoslovné i víceslovné pojmy, včetně zpodstatněných přídavných jmen (např. *dovolená, cestující*).

Seznam *specifických pojmů* či tematických okruhů¹²⁸ doplňuje výčet tzv. *názevů otevřených položek*. Jedná se o souborné názvy jako např. *druhy ovoce, názvy obchodů, historické lokality a místa turistického zájmu, názvy jazyků* apod., tedy třídy vymezené jednotným sémantickým kritériem. Jak lze ze zde citovaných názvů poznat, jsou to položky dost různorodé. Jejich konkrétní náplň se v *Prahové úrovni* příkladově uvádí, nelze ji však brát jako vyčerpávající a je podle autorů *Prahové úrovně* na tvůrčích učebnic a učitelích, aby ji přizpůsobili jazykovým potřebám svých studentů (srov. Bischofová, 2001, str. 418).

Tabulka 1.1 v příloze 1 podává souborný přehled o identifikovaných oblastech užívání jazyka, v nichž dotazovaní studenti vidí potřebu dorozumívat se a náležitě se vyjadřovat, a o přiřazených komunikačních úlohách a cílech, které formuluje *Prahová úroveň*. Tabulka navíc obsahuje výčet všech tematických okruhů (čili *specifických pojmů*) a *názevů otevřených položek*, které se shodují s uvedenými komunikačními úlohami.

5.3.2.2 Doplnování vzniklého soupisu slovní zásoby za pomoci slovníkových částí učebnic vedoucích k úrovni B1

Za účelem specifikace souborných názvů čili *názevů otevřených položek* (jako např. ovoce, zelenina, názvy povolání, názvy národností apod.) byly vybrána příslušná podstatná jména z učebnice *Tschechisch Schritt für Schritt* (Holá, 2006) a z učebnice na ni navazující *Česky krok za krokem 2* (Holá; Bořilová, 2009). Oba díly vedou podle údajů autorek k úrovni B1. Náplň *názevů otevřených položek* ukazuje tabulka 1.2 v příloze¹²⁹.

¹²⁸ Pojem *specifické pojmy* budeme v dalším textu nahrazovat pro jeho běžnost pojmem *tematické okruhy*.

¹²⁹ Tabulka je již upravená podle frekvenčních úvah (viz k tomu diskusi v další podkapitole).

I po odečítání substantiv patřících k souborným názvům *otevřených* položek byl zjevný velký číselný rozdíl mezi výčtem substantiv podle *Prahové úrovně* a seznamem, jenž vychází ze slovníkové části učebnic. Pod písmenem *B* se např. v *Prahové úrovni* uvádí 15 podstatných jmen včetně víceslovných pojmenování, v dvoudílné učebnici oproti tomu 40¹³⁰. Do základního souboru byla proto doplněna podstatná jména v *Prahové úrovni* neobsažená, jež se vztahují k uvedeným komunikačním úlohám a tematickým okruhům.

5.3.3 Úprava výběru substantiv na bázi frekvenční distribuce

Základní soubor by měl být reprezentativní nejen z hlediska potřeb studentů, ale také z hlediska frekvence výrazů v současné češtině. Ve třetím kroku jsme proto posoudili vzniklý soupis substantivních výrazů z frekvenčního hlediska a doplňovali substantiva, která jsou v soudobé češtině hojně zastoupena. K tomuto postupu nás vedly dva fakty.

1. Výběr slov v *Prahové úrovni*, závažné pro úroveň B1, se neopírá o žádný sociolingvistický výzkum ani nevyplývá z přesné analýzy potřeb, jak připouští Jitka Cvejnová, jedna z jejich spoluautorů. Autoři „spíše vycházejí [...] ze zobecněné zkušenosti nashromážděné kolektivem spolutvůrců, ale i ze zkušeností vycházejících z předchozího lingvistického výzkumu“ (Cvejnová, 2006, str. 30). Slova a struktury ve slovníkém rejstříku *Prahové úrovně* ani ve slovníkové části učebnic nemají oporu v korpusových datech. Také musíme mít na zřeteli, že *Prahová úroveň* je podle autorů „pouze formulací cílů obecných, a navíc formulací takovou, která by nikdy neměla být považována za něco definitivního a uzavřeného, nýbrž by naopak měla být aplikovaná pružně a tvořivě.“ (Bischofová, 2001, str. 10)

Porovnání slovního rejstříku *Prahové úrovně* s Frekvenčním slovníkem pro studenty, *Frequency dictionary of Czech. Core vocabulary for learners* (Čermák a Křen, 2011)¹³¹ ukazuje, že slovní rejstřík *Prahové úrovně* neobsahuje řadu častých slov, např.:

¹³⁰ Výčet slov neobsahuje (stejně jako slovní rejstřík *Prahové úrovně*) slova, která se považují za součást *Názvů otevřených položek*, např. *banán* (druh ovoce), *babička* (příbuzenské vztahy)

¹³¹ Dále jen FDCL.

- *možnost* (254¹³²), *program* (431), *případ* (149), *situace* (279), *stav* (358), *systém* (274), *vztah* (339), *změna* (305)
- zvláště neologismy, jež jsou hojně doložené v korpusech současné češtiny, jako např.: *internet* (1302), *obrazovka* (3062), *sluchátko* (4765)
- slova nutná při hovoru o samotném jazyce: *věta* (1482), *výraz* (1055)
- některá frekventovaná substantivizovaná adjektiva nebo adjektiva s převážně substantivním použitím, např. *vedoucí* (adj/noun rank: 788), *mluvčí* (rank: 892), *vrchní* (adj/noun rank: 2796)

Je tedy třeba zvážit, jestli a popřípadě v jaké míře by měla vstoupit do našeho základního souboru zatím opomíjená, vysoce frekventovaná slova.

2. Náplň slovníkových částí, kterými jsou doplněny učebnice, závisí na výběru učebních textů. Proto se ve slovníčcích učebnic vyskytují i velice okrajová slova (viz např. slovo „kastelánka“, které se vyskytuje v textu o neobvyklých bydlištích v učebnici *Česky krok za krokem 2.*). Mimo to se autorky zde zvolených učebnic snažily vyhovět vyjadřovacím potřebám jejich „čtenářů“ a uváděly dlouhé seznamy názvů zemí a národností (mezi národnostmi najdeme např. *Alžírán, Libanonec, Novozéland'an*). Vyloučení takových okrajových jevů pomůže vhodně omezit počet substantiv, jež vcházejí do základního souboru.

V posledním kroku byl proto předběžný soubor substantiv upraven následujícím způsobem:

a) Z frekvenčního slovníku FDCL byla vybrána všechna substantiva, jež se ocitají mezi 1000 nejfrekventovanějšími slovy v soudobé češtině. Podle Čermáka a Křena (2005, str. 6) pokrývá nejčastějších 1000 slov 64,9 % všech slov ve psaných českých textech¹³³. FDCL je založený na třech obsáhlých korpusech – a to na stomilionovém korpusu psané češtiny (SYN2005) a na korpusech mluvené češtiny ORAL2006 a 2008, každý z nich o rozsahu jednoho milionu slov. Za pomoci FDCL bylo zjištěno a doplněno 77 podstatných jmen s rankem od 200 do

¹³² Číslo v závorce ukazuje pořadí slova podle Frekvenčního slovníku češtiny pro studenty: *A Frequency Dictionary of Czech. Core Vocabulary for Learners* (Čermák a Křen, 2011).

¹³³ Pro češtinu se vztahují tyto údaje na *Frekvenční slovník češtiny* z roku 2004. Opírají se tudíž o korpus psané češtiny, který však není úplně totožný s korpusem na kterém je založený *Frequency dictionary of Czech*. Avšak hodnoty lexikálního pokrytí tohoto frekvenčního pásma i pro jiné jazyky popř. mnohem menší korpusy se liší jen minimálně (srov. Čermák a Křen, 2005, str. 6).

993, jež nebyla obsažená ani ve slovním rejstříku *Prahové úrovně*, ani ve slovníkové části použité učebnice. Jedná se o následující¹³⁴:

*způsob, systém, situace, změna, základ, podmínka, pocit, skutečnost, typ, tělo, smysl, činnost, prostředí, vývoj, úroveň, role, akce, míra, krok, funkce, představa, rámec, zdroj, postup, dílo, rada, množství, podoba, bod, vedení, materiál, úspěch, podstata, autor, škoda, věk, počátek, organizace, smlouva, rozhodnutí, průběh, volba, řízení, podpora, zařízení, šéf, souvislost, zbytek, řeč, boj, požadavek, hledisko, stavba, obyvatel, snaha, ministerstvo, síť, šance, čelo, pokus, zástupce, metoda, mluvčí, příprava, krev, nabídka, ztráta, text, studium, ohled, občan, rozhovor, kultura, aktivita, spolupráce, odborník*¹³⁵

Nejevilo se nám jako smysluplné doplnit do našeho souboru všechna substantiva s rankem do 5000, protože častost některých slov je vázaná na určité žánry (např. *brankář, bojovník, bitva* se nejvíce vyskytují v novinách), které jsou málo relevantní pro naši cílovou skupinu. Na druhé straně jsme zjistili, že seznam 5000 nejfrekventovanějších slov postrádá výrazy, které mohou být potřebné pro řešení identifikovaných komunikačních úloh, např.: *batoh, bufet, bazén, bydliště*. Ty ovšem pokrývaly seznamy slov učebnic. V této souvislosti chceme dát ke zvážení, zda se vyjadřovací potřeby cizince, který se dotyčný jazyk učí, neliší od potřeb rodilého mluvčího. A to jednak, protože obvykle nevidí potřebu sledovat média v cílovém jazyce v té míře, jak je zvyklý z domova, jednak proto, že omezené vyjadřovací prostředky zpětně limitují jeho komunikační potřeby a možnosti, např. mu mohou bránit v četbě krásné literatury. Důkladně se zabývají otázkou složení závazné či relevantní slovní zásoby pro výukové účely Sinclair a Renoufová (Sinclair a Renouf, 1988). Mají za to, že by se nejfrekventovanější tvary slov, které se denně užívá, měly do nějaké míry reflektovat v učebnicích a kurzech (tamtéž, str. 150). Na druhé straně připouštějí, že by seznam slovní zásoby měl obsahovat i výrazy ze středních a nižších frekvenčních pásem, aby autoři učebnic měli možnost nabízet zajímavé a pestré texty. Proto doporučují tzv. doplňující seznam (*additional list*). „These additions should be monitored carefully, so that the final word list contains items of maximal utility and power. A balance has

¹³⁴ Slova jsou řazená v sestupném pořadí podle ranku.

¹³⁵ Jsme si vědomi toho, že některá ze zvlášť častých substantiv vděčí za vysokou frekvenci skutečnosti, že se společně s předložkou uplatňují ve funkci adverbialie nebo víceslovní předložky (srov.: Čermák F. J., 2005, str. 85-86). Některá z takových substantiv ve zkoumaných slovníkových seznamech postrádáme, např. *základ* a *způsob*. Jiná však ve slovním rejstříku nebo ve slovníkové části učebnice Česky krok za krokem obsažená jsou, např. *období, pořádek, příležitost, případ*.

to be achieved, however, between natural usage and utility.“ (Sinclair a Renouf, 1988, str. 151).

b) Byly vymezeny seznamy substantiv, která jsou náplní názvů otevřených položek (jako např. povolání a jména národností a zemí), aby příliš netrpěla vyváženost výběru. Zkrácení se týkalo zvláště těch substantiv, které si vyplňuje každý „uživatel“ podle potřeby a jejichž celek není relevantní pro celou cílovou skupinu, zejména *jména národností* (o rozsahu 59 substantiv)¹³⁶, *jména zemí* (70 substantiv), *názvy povolání* (61 substantiv), *oblečení a obuv* (46 substantiv).

U názvů zemí a národností jsme se omezili na nejčastěji užitá názvy národností a příslušných zemí v česky mluvícím prostředí. Byly vybrány ty názvy, které odpovídají nejčastějším přídavným jménům označujícím národnosti a jazyky¹³⁷. Příslušné přechýlené ženské názvy byly v seznamu ponechány, protože výběr rodu zde podléhá individuálním potřebám studentů a ženské tvary pro ně mohou být ve stejné míře závažné jako mužské protějšky.

V oblasti povolání, oblečení i obuvi je řešení poněkud komplikovanější. Samotný seznam nejfrekventovanějších profesí ve slovníku FDCL zahrnuje již 57 (jednoslovných) pojmů. Rozhodli jsme se zde pro průnik obou množin (tedy seznamu slovníkové části a frekvenčního seznamu), který obsahuje 34 názvů mužského rodu a téměř stejný počet ženského rodu¹³⁸. Dále bylo nutné vhodně zkrátit seznam různých typů oblečení. I zde bylo vybráno těch 15 výrazů, které se objevují v učebnicích a zároveň v seznamu nejčastěji použitých názvů profesí v FDCL (Čermák a Křen, 2011, str. 146).

c) Pomocí vlastní analýzy v Českém národním korpusu byly zjištěny a doplněny frekventované substantivní kolokáty¹³⁹, těch sloves, která se považují podle *Prahové úrovně* za závažné pro realizaci komunikačních úloh na úrovni B1. Příslušná slovesa byla vybrána z výpisu tematických okruhů (*specifických pojmů*) v *Prahové úrovni* (např.: *vystudovat + A, jet/jezdit čím*; srov. k tomu tab. 1.3 v příloze). Takový postup umožňoval vyhledávání frekventovaných substantiv v daném spojení a přihlídnutí ke všem pozicím valenčního rámce slovesa, tedy k

¹³⁶ Nepočítaly se zvláště ženské tvary názvů národností a povolání.

¹³⁷ Podle FDCL (Čermák a Křen, 2011).

¹³⁸ Do výběru vstoupily i zde ženské tvary stejným podílem jako mužské. U slova *účetní* se přes homonymii v nominativu důsledně rozlišuje *účetní* (m) a *účetní* (f).

¹³⁹ Ve dvou případech byly zkoumány i kolokáty slovesných kolokací: kolokace *mít chuť* a *mít na sobě*. Podle Čermákovy terminologie se považují za *běžnou kolokaci uzuální* (Čermák, 2010, str. 229).

objektovým doplněním určeným vazbou řídicího slovesa a zároveň k doplnění podmětovému. Sonda byla provedena v již zmíněném subkorpusu SYNEK2010. Do této analýzy jsme nezahrnuli žádný z korpusů mluvené češtiny, poněvadž žádný kolokát náhodně vybraných 10 sloves neměl více než 10 absolutních výskytů.

Uvedme si jako příklad tematický okruh *1. osobní informace – 1.2 adresa*, který je závažný alespoň pro dvě komunikační oblasti: *cestování a domluva s místními lidmi*, popř. *orientace ve městě*. Ze seznamu navržených exponentů těchto témat bylo vybráno klíčové sloveso, v tomto případě *bydlet*. Kolokace byly vyhledávány v rámci –3 až 3 (vlevo a vpravo) od KWIC¹⁴⁰ a tříděny podle absolutní frekvence¹⁴¹. Následně byly vyřazeny jednotky, které neměly substantivní povahu (interpunkční znaménka, spojovací výrazy apod.). V jiných případech byly navíc – kde to bylo nutné – vyřazeny kolokace, které mají blízko k frazémům, nebo kolokace nehodící se k tematickému okruhu (jako např. *dát si pozor*, *dát si nohu/ruce* + předložkový výraz, *nedám si vymluvit* + substantivní výraz). Ze vzniklých kolokačních paradigmat jsme vybrali kolokáty, které mají v daném spojení absolutní frekvenci minimálně 10. Následující výčet ukazuje frekventované kolokáty slovesa *bydlet* v SYNEK2010, přičemž podtržené výrazy bylo potřeba do základního souboru substantiv doplnit, poněvadž ve výběru chyběla:

kousek, domek, čtvrť, Londýn, kolej, New York, podnájem, sousedství, okolí, Praha, centrum, vila, panelák, barák, obce, Brno, budova, palác, sídliště, střecha, předměstí, blízkost, panelový dům, okraj, Paříž, Moskva, činžák, Vinohrady, místnost, domácnost, internát, podkroví

Výhody zvolené metody spočívají v tom, že mimo časté zástupce jedné kategorie (*bydlet* + ‚budova‘ nebo ‚místo ve městě‘) odhalila frekventované výrazy, které autoři učebnic pro svou běžnost snadno opomíjí (např. *bydlet* + *v susedství...*, *bydlet* + *v blízkosti ...*). Podtržené výrazy totiž názorně ukazují, že některé typické příklady v učebnici nebo v *Prahové úrovni* vůbec nenajdeme (srov. *kolej, sídliště*).

Úplný seznam nalezených kolokátů a příslušných sloves a slovesných kolokací je uveden v tabulce 1.3 v příloze.

¹⁴⁰ Key word in context.

¹⁴¹ Zvolili jsme absolutní frekvenci, poněvadž důležitější než soudržnost obou komponentů je informace o lexikálních jednotkách, jež se často vyskytují v okolí daného slovesa, dané slovesné koncepcí.

Na závěr je třeba poznamenat, že přes všechna opatření se do souboru nedostaly (s výjimkou *Prahy* a *Brna*) názvy větších českých měst; tyto názvy však považujeme za důležité zvláště v kontextu cestování a domluvy s místními obyvateli. Proto byly doplněny alespoň jména krajských měst: *Ústí nad Labem, Liberec, Plzeň, Karlovy Vary, České Budějovice, Zlín, Olomouc, Hradec Králové, Pardubice, Jihlava, Ostrava*.

Podle popsaného postupu výběru vznikl základní soubor substantiv o rozsahu 1649 slov, mezi nimiž se nalézají vlastní jména (města, názvy národností a zemí a – pokud byly součástí kolokací – také jiná propria, např. *Harvard, Zlaté stránky*), zkratky, číslovky substantivní povahy a víceslovné pojmy (např. *bramborová kaše, jednolůžkový pokoj, letecká linka*).

Substantiva, jež označují různé pojmy stejnou formou, zůstala v souboru dvěma různými jednotkami jen tehdy, rozlišují-li se buď psaním velkým a malým písmenem, např. *země* a *Země*, nebo rodem, např. *účetní* (m) a *účetní* (f). Hlouběji se problematice víceznačnosti věnujeme v části 5.4.

5.3.4 Charakteristika základního souboru

Popis základního souboru dat za pomoci lingvistických kritérií sleduje dva cíle: jednak vydělení menšího stratifikovaného vzorku, jehož složení odpovídá složení výchozího základního souboru, a jednak hrubé roztřídění substantiv podle charakteristik – jak formálních, tak sémantických – které mohou mít vliv na frekvenci určitého morfologického tvaru v textu, např. třída jmen látkových. Substantiva základního souboru jsou značkována podle těchto charakteristik. Následně je z každé podtřídy (např. třídy abstrakt) náhodně vybrána osmina substantiv a takovým způsobem vytvářen stratifikovaný vzorek.

Hrubé třídění substantiv by mělo odrážet takové rysy substantiv, které mohou mít vliv na frekvenci morfologického tvaru, zejména hodnoty kategorie pádu a čísla¹⁴².

Nejzřetelnější je souvztažnost mezi významovou stránkou lexémů a jejich výrazem na rovině textu u gramatického čísla. Na výraz čísla má podstatný vliv sa-

¹⁴² Podle Lehmana (2004) je rod podstatných jmen kategorií neflektivní čili klasifikující. Od kategorií flektivních se liší tím, že se funkční opozice vyjadřuje pomocí kmene nikoli afixem. Gramatický rod se na rovině textu nemění, je pevně spjat s lexémem. Případá mu však závažná role v soustavě deklinačních tříd českého substantiva, poněvadž je jedním z kritérií, která přiřazují substantiva k deklinačním vzorům.

motná povaha denotátu (např. u substantiv látkových nebo abstrakt) anebo schopnost substantiv vyjádřit souborovost nebo hromadnost. Úzce spjat s výrazem jedinečnosti popř. mnohosti je též základní distinkce abstraktum x konkrétum, které se zde (nejen proto) zvlášť také věnujeme. Zvláštní jev z tohoto hlediska je ryze formální omezení pluralií tantum.

Pro orientační klasifikace substantiv dle významu zde užíváme typologii, již navrhuje Čermák (2010, str. 165n.), poněvadž dokáže odrážet obecnou kolokabilitu a vybízí k jemnějšímu třídění:

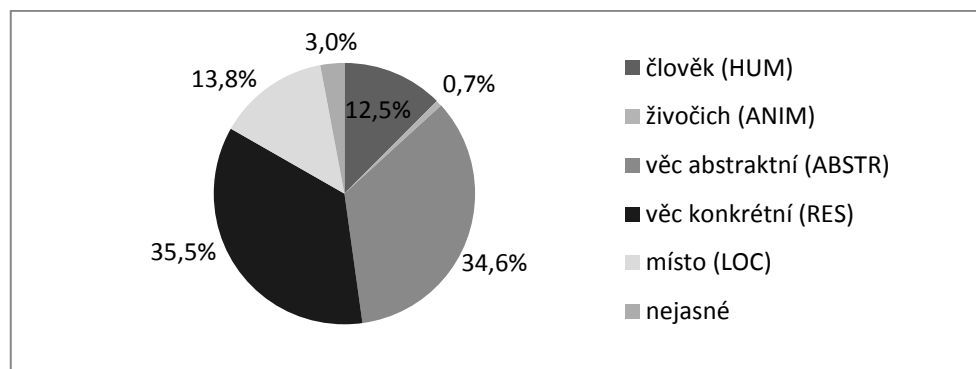
- (1) člověk (HUM(anus), *muž, učitel*)
- (2) živočich (ANIM(alis), *pes, pták*)
- (3) věc konkrétní (RES, konkrétum, *kámen, tužka*)
- (4) věc abstraktní (ABSTR(actus), abstraktum, *chytrost, ministerstvo, jízda*)
- (5) místo (LOC(us), *cesta, pohoří*)

Čermák zdůrazňuje, že distinkce abstraktum-konkrétum má škálovitý charakter a uplatňuje se v různé míře i uvnitř uvedených tříd. Toto základní rozlišení se podle něj „projevuje v zásadě odlišném chování v textu, kolokabilitě aj.“ (tamtéž).

Substantiva základního souboru byla tudíž roztržena podle typu jejich denotátu (třídy HUM, ANIM, RES, ABSTR a LOC) a dále podle povahy jejich denotátů, pokud má vliv na výraz gramatického čísla. Značkování substantiv také zohlednila klasickou distinkci propria x apelativa, a to, aby náhodný výběr odpovídal i z tohoto hlediska základnímu souboru.

5.3.4.1 Podle typu denotátu

Substantiva základního souboru se podle typu denotátu seskupují následně:



Obr. 5.1 Rozložení substantiv podle typu denotátu

Člověk (HUM)

Kromě základních (neodvozených) podstatných jmen jako *holka, kluk, muž* a názvů členů rodiny jsou v základním souboru obsaženy zejména názvy povolání a národnosti. V pojmech slovtvorby se jedná u odvozených substantiv převážně o třídy jmen konatelských (např. *chatař*), činitelských (např. *dědic, hostitel, hráč, pracovník* apod.) a obyvatelských včetně příslušných přechýlených jmen. Řadí se sem dále malý počet substantivizovaných adjektiv (např. *cestující, známý, mluvčí, příbuzný*).

V základním souboru chybějí křestní jména a příjmení. Je to dáno skutečností, že nejsou zachycena v popisu *Prahové úrovně* ani ve slovníkových dodatcích učebnic. Ponecháme je proto stranou.

Živočich (ANIM)

Podstatných jmen označujících živočichy (vyjma člověka) je celkem jen 12 (mezi nimi *domácí zvíře, krůta, losos, pstruh*).

Věci abstraktní a konkrétní

Rozlišení abstrakt a konkrét zachycuje na velice obecné úrovni vztah lexému k vnějšímu, fyzickému světu. Kritérium je závažné mj. proto, že lze předpokládat souvislost mezi konkrétností a schopností vytvářet plurálové tvary.

Za konkréta se zde považují podle definice Encyklopedického slovníku češtiny (ESČ) substantiva „mající denotát fyzické povahy (*lékař, kočka, kapesník*), tj. označující takové třídy objektů, jejichž prvky jsou smyslově vnímatelné a časoprostorově lokalizovatelné a o nichž říkáme, že existují.“ (ESČ, 2002, str. 223) Zahrnujeme do té kategorie i tzv. *kvazikonkréta* neboli *pseudokonkréta* (např. *anděl, bůh*), poněvadž lze předpokládat, že mají podobné chování (Čermák F. , 2010, str. 55).

Pokud se pohybujeme vně konkrétního textu, lze někdy jen s obtíží jednoznačně určit, jestli se jedná o konkrétum nebo abstraktum (srov. např. *hlava, země*). Čermák vyděluje případy, kde je abstrakce dána systémem, kde tedy konkréta plní v lexikonu funkci označení celé třídy denotovaných objektů a případy, kde se jedná o „odvozený význam abstraktní“ (např. *hlava* – jako kapitola knihy), jako abstrakci *sekundární* (Čermák F. , 2010, str. 55). Pro nedostatek kontextu jsme s

případnými spornými případy zacházeli jako s konkréty. Věci konkrétních je v základním souboru 585.

Abstrakta se právě uvedenými rysy jako možností lokalizace v prostoru nebo smyslovou vnímatelností nevyznačují. Vztahují se na různé aspekty našeho myšlení a vnímání světa (ESČ, 2002). Skupina jmen, která lze přiřadit k abstraktním věcem (v základním souboru je jich 570), je téměř stejně rozsáhlá jako skupina konkrétních věcí. Řadí se sem jména označující vlastnosti (*blízkost, chudoba, výška*), stavy (*počasí, povaha*), děje (*lyžování, přijímání, létání, týkáni, zákaz*), časové relace (*půlhodina, prosinec, chvíle, budoucnost*) a čísla, míry i zlomky (např. *dvojka, trojka, stovka, třetina, polovina*). I abstrakta se však mohou za určitých okolností „zkonkretizovat“, jak názorně ukazují příklady v Příruční mluvnici (PMČ): „Významy abstraktních názvů vlastností a dějů se různou měrou specifikují, posouvají a konkretizují a v přímé závislosti na tomto procesu se objevuje jejich rozčleněnost, počitatelnost: *Dopřával si těch radostí; Kupoval si sladkosti; [...]*“ (PMČ, 1995, str. 230). Sporná abstrakta jsou z tohoto hlediska některá podstatná jména s původním významem dějovosti, která se rovněž anebo dokonce převážně užívají k označení výsledku děje, jako např. *žádost, smlouva, státní zkouška, vzkaz, malba, stavba, psaní*. V takových případech bylo o přiřazení k abstraktům rozhodnuto na základě *Slovníku spisovné češtiny pro školu a veřejnost* (SSČ). Pokud heslo ve slovníku konkrétní význam zvlášť neuvádí, označili jsme substantiva jako abstrakta. Jedná se zejména o slova *zkouška, test, vzkaz, smlouva, žádost*. Pokud však slovník vykazuje též konkrétní význam, rozhodli jsme pro značkování jako *nejasné* z hlediska povahy denotátu. Týká se to slov *psaní, malba, nákup, stavba*.

Nejasné

Poněkud zvláštní postavení mají ve světle distinkce abstraktum x konkrétním popř. místo x ne-místo názvy institucí jako *banka, burza, pošta, škola*. Podobné názvy mohou být pojímány buď jako pojmenování konkrétní budovy, nebo jako pojmenování institucí, a tedy mohou nabývat abstraktního významu. Proto byla vyznačena zvlášť jako *nejasná*. Sem se řadí rovněž některá substantiva dějová (viz výš) a substantiva víceznačná jak významově, tak z hlediska distinkce abstraktnost x konkrétnost popř. životnost x neživotnost. Mohou mít konkrétní denotát, ale záro-

veň abstraktní významovou složku, např. *obsluha, postava, síť, recepce, myš, trh, sjezd, pravice*.

Místa (LOC)

Místo je základní kategorií lidského vnímání a myšlení. Ve svém základním, nepřeneseném smyslu se vztahuje na konkrétně vnímatelné prostory, kde se mohou odehrávat děje a události, ocitat se věci, lidé a zvířata. Ve své hrubé typologii F. Čermák uvádí jako příklady pro členy kategorie *místa* (LOC) *cesta, pohoří, pěšina* (Čermák, 2010, str. 166). Skupinu však blíže nevynezuje. Je ale potřeba jasně oddělit substantiva označující místa od těch, které označují věci konkrétní. Předměty jako *krabice a balíček* např. nejsou prototypickými místy, přece je možné do nich leccos umístit, něco se tam může odehrávat. Kategorie místa jsme proto ohraničili na základě dalšího kritéria. K třídě míst byla přiřazena všechna substantiva označující denotáty s prvotním významem místa, které nejsou pohyblivé a mohou být zároveň cílem nebo výchozím bodem nějakého pohybu (např. *rybník, zámek, Berlín, hala, nebe, poschodí*).

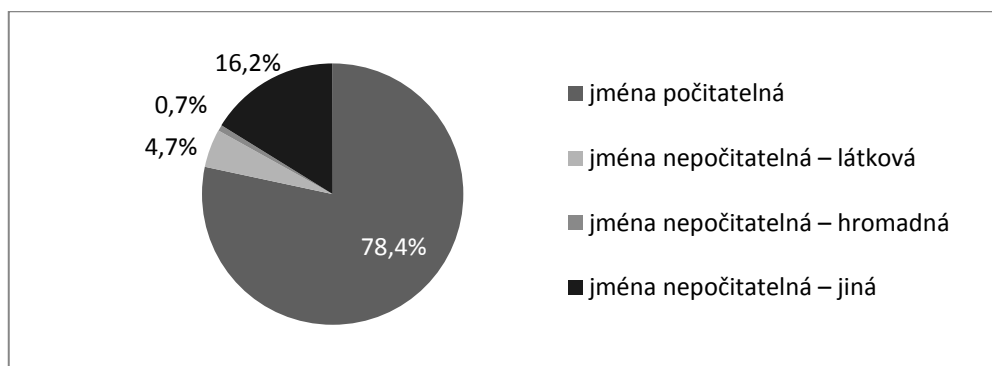
5.3.4.2 *Propria a apelativa*

I distinkce *propria* vs. *apelativa* se odráží v preferenci singulárových tvarů. *Propria* se v převážné většině případů vyskytují jen v singulárové podobě, poněvadž jejich denotáty jsou konkrétní osoby nebo entity, nikoliv třídou entit. Za velice specifických okolností však vytvoření plurálu možné je. Specifické z tohoto hlediska jsou názvy národností (etnonyma), které mohou v textu nabýt generického významu. Podle ESČ u etnonym plní tuto funkci bezpříznakově číslo množné, jako v příkladu *Angličané jsou známí klid'asi* (ESČ, 2002, str. 95). Přesto se podle ESČ tradičně přiřazují k *propiím*.

Většina *propií* v základním souboru jsou názvy zemí i měst popř. míst (celkem 63, např. *Praha, Vinohrady, Anglie*). Názvů národností je v souboru obsaženo 36 včetně přechýlených tvarů (např. *Francouz, Řekyně, Evropan*). Názvy výrobků *Zlaté stránky* a *Dobrá voda* jsme zařadili k *propiím*, ačkoliv jsme si vědomi toho, že tyto případy mohou být sporné nebo nejednoznačné, protože se mohou vztahovat i k celé třídě (ESČ, 2002, tamtéž). Dalšími *propií* jsou vlastní názvy škol nebo institucí *Harvard* a *NATO* a název planety *Venuše*.

5.3.4.3 Podle povahy denotátu

Vzhledem k sémantické povaze denotátu popř. jeho rozčleněnosti se rozdělují substantiva v základním výběru v následujícím poměru:



Obr. 5.2 Rozložení substantiv podle povahy denotátu

Substantiva počítatelná a nepočítatelná

Za počítatelná se označují substantiva, která spojuje sémantický rys rozčleněnosti (ESČ, 2002 str. 467). Oproti substantivům nepočítatelným jsou tudíž s to vyjadřovat mnohost a větší počet objektů pomocí plurálových forem.

Abstrakta, unika nebo jména látková v převážné většině plurálové tvary netvoří, poněvadž označené entity nelze rozčlenit popř. jsou jedinečné jako např. *vesmír* a *gotika* (PMČ, 1995, str. 230) a jsou tudíž nepočítatelná.

Na rovině textu však k vytvoření plurálu dojít může, např. u názvů nápojů (tedy látek), které jsou k dispozici v různých baleních nebo nádobách. Tato látková substantiva se mohou stát počítatelnými tím, že se pojmenování nádoby vynechává a počítanou jednotkou je samotná látka (např. *tři piva*). Plurálová forma u látek může ale také vyjadřovat existenci různých druhů, jak znázorňuje příklad v MSČ: „Některé minerální vody mají vyšší obsah fluoru.“ (MSČ, 2010, str. 138).

U abstrakt je zvlášť silná tendence tvoření plurálových forem, ačkoliv svými sémantickými rysy spíše patří k singulariím tantum. Akademická mluvnice češtiny (MČ, svazek 2, Petr, 1987) jmenuje celou řadu podmínek, za kterých se abstrakta mohou vyskytovat v plurálové podobě. Plurálové tvary mohou vyjadřovat různé modifikace výchozího významu, např. vlastnost se stává „počítatelnou“ tím, že se zvýrazňuje její projev (srov. <*život přece jen tu a tam nabízí drobné radosti*>). U substantiv označujících děj se mohou zdůrazňovat výsledek, místa nebo okolnost

děje (srov. <*kam až oko dohlédne, stojí další domy a stavby*>, <*máte nějaké námítky*>). Plurálové tvary též mohou dodat slovu význam souborovosti, např. <*rozvíjí budoucí učitel i své pedagogické schopnosti i dovednosti*> nebo konkrétnosti, např. <*Jessie, vy musíte napsat svoje paměti*> (MČ, svazek 2, 1987, str. 49)¹⁴³.

Abstraktní substantiva, jež pravidelně tvoří plurálové tvary přes chybějící rys rozčleněnosti, byla v základním souboru značkována jako počítatelná, např.: *podmínka, právo, povaha, základ, rozdíl, škoda*.

Substantiva schopná vyjadřovat souborovost

Dalším příznakem, který je úzce spjatý s kategorií čísla, je souborovost, tj. schopnost vyjadřovat souhrn jednotlivin jednoho druhu. S distinkcí *souborovost* x *prostý počet* současné mluvnické nepracují, uvádějí ji však Panevová a Ševčíková (2011). Schopnost vyjadřovat souborovost se projevuje především u kolektiv, u nich se váže převážně na tvary singuláru. Kolektiva jsou proto podskupinou singularií tantum. Význam souborovosti však také mohou vyjadřovat běžná substantiva, a to pomocí plurálových tvarů (viz o tom podrobněji dále).

Substantiva s prototypickým významem souborovosti

Oproti zvláštní třídě kolektiv mohou souborovost vyjadřovat i substantiva běžná, tedy schopná tvořit jak singulárové, tak i plurálové tvary, ale používají k tomu většinou plurálové tvary. Na rozdíl od běžné funkce plurálových forem nevyjadřuje souborovost větší množství počítaného předmětu, ale pár nebo soubor. Panevová a Ševčíková (2011) nepovažují souborovost dokonce jen za pouhý lexikální rys, ale za gramatikalizovanou kategorii. Nicméně připouštějí, že význam souborovosti může být pro některá substantiva prototypický, tj. „[...] tato substantiva se často užívají v plurálu, jejich plurálová forma je přitom běžně používána k označení souboru/souborů, u některých substantiv dokonce častěji než k označení pouhého většího množství jednotlivin.“ (tamtéž, str. 166). Následující třídy substantiv mají podle autorek takový prototypický souborový význam:

- části těla, např. *ruce, oči, vlasy, zuby*
- předměty určené pro tyto části těla, např. *rukavice, sluchátka*
- další předměty denní potřeby, např. *klíče, sirky, cigarety*

¹⁴³ Ukázky, které znázorňují dané jevy, jsou vyňaté z korpusu SYN2010. Výběr příkladových slov se však opírá o příklady v *Mluvnici češtiny* (MČ, svazek 2, 1987).

- rodinní příslušníci tvořící páry nebo skupiny, např. *rodiče, sourozenci*
 - a další, např. *závory, koleje, schody*
- (viz tamtéž).

V našem souboru se nachází 22 substantiv, kterým by bylo možné prototypicky souborový význam podle výše popsaných kritérií připsat, mj. *boty, nohy, rukavice, sluchátka, zápalky, záchody*. Panevová a Ševčíková (2011, str. 172) však upozorňují, že jednoznačná souvislost mezi častějším výskytem substantiv v plurálových tvarech a schopnost vyjadřovat souborový význam není vždy daná. Navíc jsou v některých příkladech hranice mezi souborovými a pomnožnými substantivy splývavé (srov. *hranice, uvozovky*). Z uvedených důvodů jsme se rozhodli při sestavování náhodného vzorku zvlášť nepřihlížet k potenci slov vyjadřovat souborovost.

Substantiva hromadná (kolektiva)

Podle ESČ tato jména „označují tvarem sg. soubor (souhrn, celek) jednotlivin téhož druhu“ (ESČ, 2002, str. 466). ESČ je proto řadí mezi *singularia tantum*. Autoři však odkazují na skutečnost, že „význam souborného množství mají i některá substantiva užívaná převážně v plurálových tvarech: *peníze, bezobratlí, Přemyslovci*“ (tamtéž). Základní soubor obsahuje celkem jenom 12 kolektiv. Mezi nimi jsou *singularia tantum* (např. *ovoce, zelenina, pití, cukroví, nábytek*), ale též slova, která vyjadřují právě souborový význam jenom plurálovými tvary, např. *peníze, potraviny, prachy, hory*. Pro jejich malý počet a proto, že se aspoň tři z nich považují navíc za *pluralia tantum*, nebylo možné třídu hromadných jmen při vydělování náhodného vzorku zvlášť zohlednit.

5.3.4.4 Podle formálních příznaků

Pluralia tantum

Zvlášť byla značkována *pluralia tantum* (celkem 31), poněvadž s sebou nesou jasné omezení vzhledem k vytvoření singulárové podoby, ačkoliv tato vlastnost není přímo daná inherentním rysem sémantiky těchto slov.

Další třídy substantiv by bylo možno vytvořit na základě formálních příznaků, které se vztahují buď k odvozenosti (deverbativní a deadjektivní substantiva) nebo

k původní přináležetosti k jiné slovní třídě (adjektiva, u kterých se připouští substantivní použití, substantiva číslovková a zájmenná). S těmito třídami zde nepracujeme, poněvadž se v této studii nepřihlíží k slovtvorným aspektům.

Substantiva neohebná sice vykazujeme, nepřihlížíme však zvlášť k jejich rovnoměrnému podílu na náhodném výběru.

5.3.5 Vytvoření náhodného vzorku

Ačkoliv projekt Českého národního korpusu nabízí značkové a lemmatizované korpusy současné psané češtiny a navíc velice výkonný prohledávač, není v rámci této studie možné podrobit celý základní soubor substantiv detailní analýze morfologických forem. Proto bylo zapotřebí vydělit náhodný, stratifikovaný vzorek. Vzorek zahrnuje osminu celku, celkem 203 lexikálních jednotek.

Aby vzorek vzhledem ke svému složení odpovídal základnímu souboru dat, vytřídili jsme za pomoci výše popsaných kritérií následující podtřídy:

		počet členů třídy	% z celku
1 HUM			
1a	člověk – propria – jména počítatelná	36	2,2
1b	člověk – apelativa – jména počítatelná	170	10,3
2 ANIM			
2a	živočich – apelativa – jména počítatelná	12	0,7
3 ABSTR			
3a	věci abstraktní – apelativa – jména počítatelná – se singulárovými i plurálovými tvary	390	23,7
3b	věci abstraktní – apelativa – jména počítatelná – pluralia tantum	3	0,2
3c	věci abstraktní – apelativa – jména nepočítatelná – se singulárovými i plurálovými tvary	181	11,0
3d	věci abstraktní – apelativa – jména nepočítatelná – pluralia tantum	3	0,2
4 RES			
4a	věci konkrétní – propria – jména nepočítatelná	4	0,2
4b	věci konkrétní – apelativa – jména počítatelná – se singulárovými i plurálovými tvary	472	28,6
4c	věci konkrétní – apelativa – jména počítatelná – pluralia tantum	13	0,8
4d	věci konkrétní – apelativa – jména nepočítatelná – látková	77	4,7
4e	věci konkrétní – apelativa – jména nepočítatelná – pluralia tantum	2	0,1
4f)	věci konkrétní – apelativa – jména nepočítatelná – jiná (včetně hromadných)	15	0,9
5 LOC			
5a	místa – propria – jména nepočítatelná – se singulárovými i plurálovými tvary	56	3,4

5b	místa – propria – jména nepočítatelná – pluralia tantum	6	0,4
5c	místa – apelativa – jména počítatelná – se singulárovými i plurálovými tvary	156	9,5
5d	místa – apelativa – jména počítatelná – pluralia tantum	3	0,2
5e	místa – apelativa – jména nepočítatelná	6	0,4
6 NEJASNÉ			
6a	nejasná – propria – jména nepočítatelná	1	0,1
6b	nejasná – apelativa – počítatelnost nelze určit	43	2,6
CELKEM		1649	100

Tab. 5.8 Podíl podtříd na základním souboru

Podstatná jména byla značkována podle příslušnosti k jedné ze zmíněných podskupin. Z každé podskupiny byla náhodně vybrána osmina lexikálních jednotek. Výběrový soubor obsahuje tedy 206 lexikálních jednotek, z nichž je 12 víceslovným pojmem. Měly-li podskupiny méně než 8 členů, byly sloučeny s nadřazenou skupinou. Bylo tomu tak zejména v případech podskupin pluralií tantum. Výsledný seznam vybraných lexikálních jednotek a rovněž celý základní soubor ukazují přehledy v příloze č. 2.1 a 2.2.

5.4 Problémy kvantifikace morfologických tvarů

5.4.1 Víceznačnost tvarů

S homonymností se v češtině setkáváme nejen na lexikální rovině, ale rovněž na rovině morfologických tvarů, a to jak uvnitř paradigmatu jednoho lemmatu (*ulice* = Nsg., Gsg., Vsg., Npl., Apl., Vpl.), tak mezi tvary různých lemmat (např. ‚tancích‘ = Lpl. lexému *tank* nebo lexému *tanec*) nebo řídce i mezi tvary patřícími k odlišnému slovnímu druhu (např. *cestující* = adjektivum nebo substantivum). Pokud je homonymie daná na rovině slovních tvarů, vypořádá se s ní morfologická disambiguace, a např. tvaru *let* za pomoci kontextu správně přiřazuje buď lemma *rok* nebo *let*. Lemmatizace a morfologická analýza probíhají v Českém národním korpusu sice s úspěšností 96 %, ale přesto jsme věnovali zvláštní pozornost problematickým případům, aby bylo možné vyřadit chybně lemmatizované nebo značkové tvary.

Ve vzorku se nacházejí jen dvě lemmata, která jsou deverbativa, jejichž tvary by automatická analýza mohla zaměnit s tvary adjektiv (vzniklých ze slovesných přičestí). Jedná se o *oslovení* a *oblečení*. Manuálně jsme proto analyzovali 6 sad

50 náhodně vybraných konkordančních řádků slova *oblečení*. Ukázalo se, že se v 13 z 300 případů jednalo o adjektivní užití. Pomocí Studentova t-rozdělení jsme odhadli pravděpodobnost výskytu této chyby v celém zkoumaném souboru na 3,66 %¹⁴⁴. Objevené adjektivní tvary byly vyřazeny.

S případem zřejmé nesprávné lemmatizace jsme se setkali několikrát. Například z výskytů slovního tvaru *mas* lemmatu *maso* s tagem <NNNP2-----A-----> lze jen 27 výskytům z 252 přiřadit lexém *maso*. Ostatních 225 výskytů jsou tvary lexému *masa*. Dále překvapil četný výskyt tvaru Gpl. u lemmatu *patro* - *pater*. Ukázalo se, že většina dokladů byly tvary druhové číslovky *patero*. Nesprávná lemmatizace byla identifikována rovněž u některých výskytů tvaru *čtvrtě* (Gsg. lemmatu *čtvrť* místo Gsg. lemmatu *čtvrť*). I v těchto případech byly příslušné konkordanční řádky z další analýzy vyloučeny.

Složitým úkolem pro morfologickou disambiguaci je rovněž rozeznání tvarů nominativu a akuzativu u maskulin neživotných a neuter. Poněvadž by se podrobnější rozbor týkal více než poloviny slov náhodného vzorku, neprovedli jsme zvláštní analýzu, a počítali tedy s již zmíněnou obecnou chybností značkování cca. 4 %.

5.4.2 Víceznačnost lexémů a víceslovné lexémy

Když sečteme výskyty pádových tvarů lemmatu, bereme v potaz, že frekvenční distribuce souhrnně odráží užití slova ve všech významových odstínech, případně i odlišné významy a zároveň užití terminologických kolokací. Jak již bylo diskutováno v kapitole 3, měla by se tato víceznačnost ukázat na rozličném kontextu. Vymezení lexikální jednotky za pomoci rozličných kontextů není jednoduchým úkolem. Podobně není možné jednoznačně identifikovat u kolokací jejich ustálenost či systémovou platnost. Jak ukazuje Čermák (2010, str. 315nn.), hranice mezi kolokacemi ustálenými a neustálenými není ostrá. Takový přechodný typ představují *běžné kolokace* (*gramaticko-sémantické kombinace*), jako např. *letní dovolená*, *místní víno* (tamtéž, str. 316). Co je ustálené, tedy součást systému, a co je „jen“ běžné a typické, v nemalé míře určuje frekvence kombinací.

¹⁴⁴ Konkrétně je odhadovaný počet výskytů této chyby ve zkoumaném souboru při 5% pravděpodobnosti omylu 32.56 (3.667 %) s tolerancí ± 27.439 výskytů (tj. ± 3.09 procentního bodu počítáno z celkové populace). Počet byl spočítán pomocí kalkulačky spolehlivosti náhodných vzorků přístupné na <http://ucnk.ff.cuni.cz/bonito/kalkulacka.php>.

Protože by přesahovalo rámeček této práce zkoumat morfologické profily zvlášť pro všechny ustálené a běžné kolokace substantiv našeho vzorku, soustředíme se na úhrnné užití lemmatu. Nehledě na to jsou v náhodném vzorku však obsažené i víceslovné lexémy (např. *jednolůžkový pokoj*). Nechceme se ale otázce úplně vyhnout a budeme ve vyhodnocení dat na vybraných příkladech diskutovat, do jaké míry se morfologické profily kolokací rozlišují¹⁴⁵.

Důsledně jsme nevyřadili ani žádné z konkordancí slov, které mohou označit dva povahově odlišné denotáty (např. u slov *sjezd*, *krám*, *pravice*). Při posouzení jejich profilu ale k této problematice přihlídneme.

5.4.3 Vytvoření morfologických profilů

Pokud chceme sestavit profily užití lemmat na základě slovních tvarů seřazených podle frekvence, čelíme otázce, kolik tvarů do profilu zahrnujeme. Poněvadž nás zajímají nejen preference jednotlivých slov realizovat se v určitém pádu, ale z didaktického hlediska rovněž frekvence výskytu celého slovního tvaru vyjadřujícího zároveň číslo, zohledňujeme také číslo substantiv. Jak již bylo uvedeno v kapitole 4.2, Těšitelová (1990) konstatuje, že, pokud se nějaký jazykový jev uspořádá podle sestupné frekvence, obvykle tři nanejvýš čtyři jednotky jsou relativně nejfrekventovanějšími, kdežto frekvence zbylých rychle klesá. Pro ověření této hypotézy na zde zkoumaném jevu jsme vypočetali pro substantiva pilotního vzorku kumulované frekvence jejich pádových tvarů a porovnávali, kolik procent všech slovních tvarů lemmatu je třemi tvary pokryto. Výsledkem analýzy bylo zjištění, že se substantiva z tohoto pohledu opět velmi liší. Některá lemmata snadno dosahují třemi tvary pokrytí nad 85 %, např. *Praha* (94,6 %), *Brno* (92 %), *venkov* (88,3 %), *ekonomie* (88,2 %) a *mluvčí* dokonce pouhými dvěma tvary 90,2 %. Oproti tomu mají lemmata s rovnoměrnějším rozložením pádů se třemi pády značně menší pokrytí svého celkového užití, např. *kniha* (51,4 %), *novinář* (57,2 %), *pítí* (60,5 %), *dům* (60,2 %), *byt* (60,7 %), *sklenička* (65 %).

Zdá se proto, že typické užití zachytíme lépe, když stanovíme určité procento pokrytí užití lemmatu jako hraniční hodnotu. Zároveň je potřeba najít v rámci praktických možností postup, jehož pomocí můžeme substantiva vzhledem k frek-

¹⁴⁵ V této souvislosti se probírají zvlášť i propriální kolokace substantiv, která jsou jinak apelativa (Bílý dům, Dům knihy, Hotel XXX).

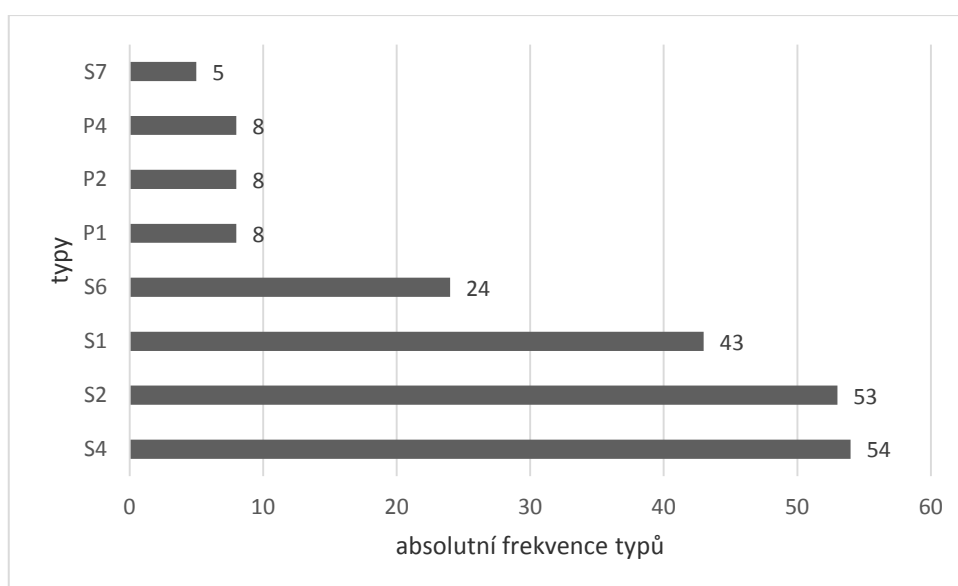
venčnímu pořadí porovnávat. K tomuto účelu je vhodné a nutné vycházet z omezeného počtu tvarů. Substantiva proto porovnááme nejprve na základě dvou a v dalším kroku tří nejfrekventovanějších tvarů lemmat. Tento počet může být vodítkem, poněvadž 98 % všech substantiv pilotního vzorku pokrývají třemi¹⁴⁶ nejčastějšími tvary 50 % všech výskytů. Pokud jde o porovnávání substantiv vzhledem k rovnoměrnosti frekvenční distribuce tvarů, přikláníme se k použití hranice 85 % pokrytí, protože – jak bylo výše ukázáno - dokáže dostatečně odrážet odlišnost frekvenční distribuce.

¹⁴⁶ Z celkem čtrnácti tvarů.

6 Vyhodnocení

6.1 Typy profilů - obecně

Jak už jsme zjistili na pilotním vzorku, lze substantiva tohoto stratifikovaného vzorku rozřadit podle morfologického profilu, tj. podle tvarů, v nichž se očitají nejčastěji. Přihlížíme-li jen k první pozici frekvenčního rozložení, rozdělují se substantiva na osm obecných typů, jejichž frekvence znázorňuje obrázek 6.1. Typy označujeme v následujícím vyhodnocení dat pro lepší orientaci zkratkou <S+číslo pádu> nebo <P+číslo pádu>, přičemž S = singulár a P = plurál¹⁴⁷.



Obr. 6.1 Četnost typů morfologických profilů podle první pozice ve frekvenčním pořadí tvarů

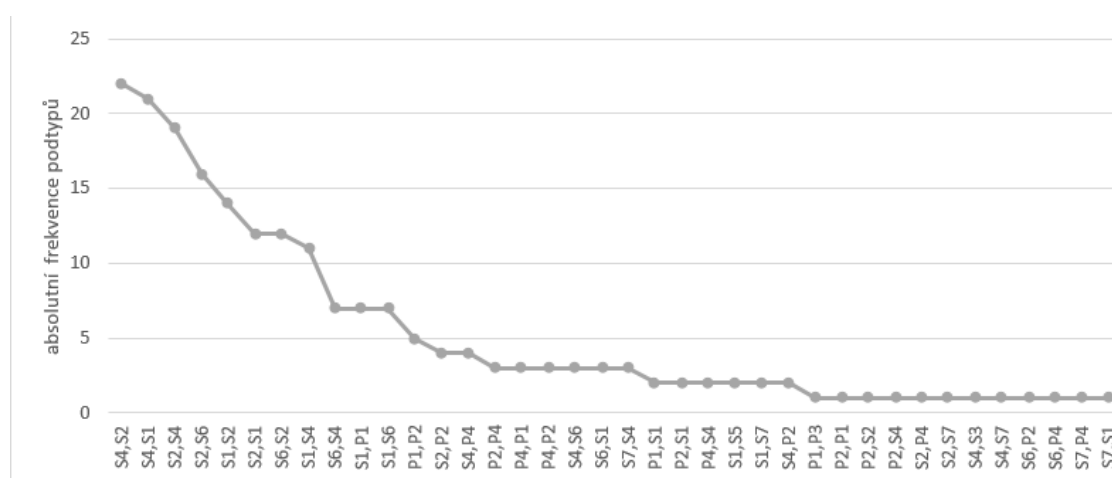
Na rozdíl od pilotního vzorku patří nejvíc substantiv (26,6 %) k obecnému typu S4, tedy k lexémům nejčastěji se vyskytujícím ve tvaru *Asg*. Téměř stejně tolik substantiv nacházíme v korpusu nejčastěji ve tvaru *Gsg*. (26,1 %). *NSg*. jako nejfrekventovanější tvar pozorujeme jen u 21,8 % zkoumaných lexémů. Data náhodného vzorku tedy lehce upravují výsledky pilotní studie vzhledem k počtu substantiv, jež se začleňují do jednoho nebo druhého typu. Ukazují, že objem typů je sice proměnlivý, ale jejich počet je stále omezený. Tento vzorek o rozsahu 203 substantiv si vystačuje s osmi obecnými typy¹⁴⁸. I když z něj vytvoříme náhodné skupiny o rozsahu 50 slov, ustaluje se počet typů pro všechny menší vzorky na

¹⁴⁷ Typ S1 je tedy skupina lexémů, jež se vyskytují nejčastěji v nominativu singuláru.

¹⁴⁸ Z pilotního vzorku o rozsahu 50 slov jsme vytvořili pět typů podle nejfrekventovanějšího morfologického tvaru.

osm. Po porovnání výsledků pilotní a vlastní studie lze pro základní výběr substantiv předpovědět, že počet typů s rostoucím objemem vzorků může sice narůstat, že však s vysokou pravděpodobností zůstane omezený. Můžeme říci, že alespoň 85 % všech substantiv se vyskytuje nejčastěji buď v A_{sg.}, G_{sg.}, N_{sg.} nebo L_{sg.}¹⁴⁹. Je ovšem třeba mít na zřeteli, že se objem obecných typů bude lišit v závislosti na povaze substantiv zvoleného vzorku.

Následně byly vytvořeny typy podle druhé pozice ve frekvenčním žebříčku forem. Vzniklo 38 podskupin¹⁵⁰, pro něž ukazujeme poměr počtu typů k jejich objemu na obrázku 6.2¹⁵¹.



Obr. 6.2 Četnost typů morfologických profilů podle prvních dvou pozic ve frekvenčním pořadí tvarů

Z našich dat vyplývá, že 50 % substantiv náleží k malému počtu nejfrekventovanějších typů, jichž je celkem 6: <S4,S2>, <S4,S1>, <S2,S4>, <S2,S6>, <S1,S2> a <S2,S1>.

6.2 Roztřídění substantiv

Na začátku se klade otázka, na čem zakládáme roztržidění substantiv. Mohli bychom se shodnout na třech nejfrekventovanějších tvarech, a to s již uvedeným argumentem, že by tři tvary jednoho lexému pokrývaly v naprosté většině alespoň

¹⁴⁹ Tvrzení platí v našem vzorku pro 85,7 %, v pilotním vzorku dokonce pro 94 % lexémů.

¹⁵⁰ Poněvadž založíme další vyhodnocení převážně na dvou nejfrekventovanějších tvarech, užíváme pro podskupiny s kombinací dvou nejčastějších tvarů označení *typ*, zatímco skupiny, které se vytvářejí na základě jen prvního nejfrekventovanějšího tvarů, označujeme jako *obecný typ*.

¹⁵¹ U typů rozlišujeme psaní <S1,S2>, které k pořadí tvarů na frekvenčním žebříčku přihlíží, a psaní <S1-S2>, které pořadí dotyčných tvarů opomíjí.

50 % jeho užití. Počet typů založených na třech tvarech bez přihlídnutí k pořadím je přirozeně větší než počet typů založených na dvou (35 vs. 24), a tím je vyšší i číslo okrajových typů (22 typů se vyskytuje jen jednou nebo dvakrát). Rozhodli jsme se i z praktických důvodů pro popis a roztřídění substantiv po prvních dvou pozicích jejich frekvenčního rozložení, poněvadž s tímto postupem zachycujeme 192 z 203 substantiv ve třídách s max. 3 členy.

Pokud bychom přihlíželi k pořadí na prvních dvou nebo třech pozicích, zvýšil by se počet podskupin značně (na 38), což ztěžuje jejich vyhodnocení a popis. Podrobný rozbor tříd ukáže, zda je jemnější roztřídění náležité. Třídy nebo typy substantiv založených na jejich dvou nebo max. třech nejfrekventovanějších slovních tvarech budeme označovat za morfologické profily substantiv.

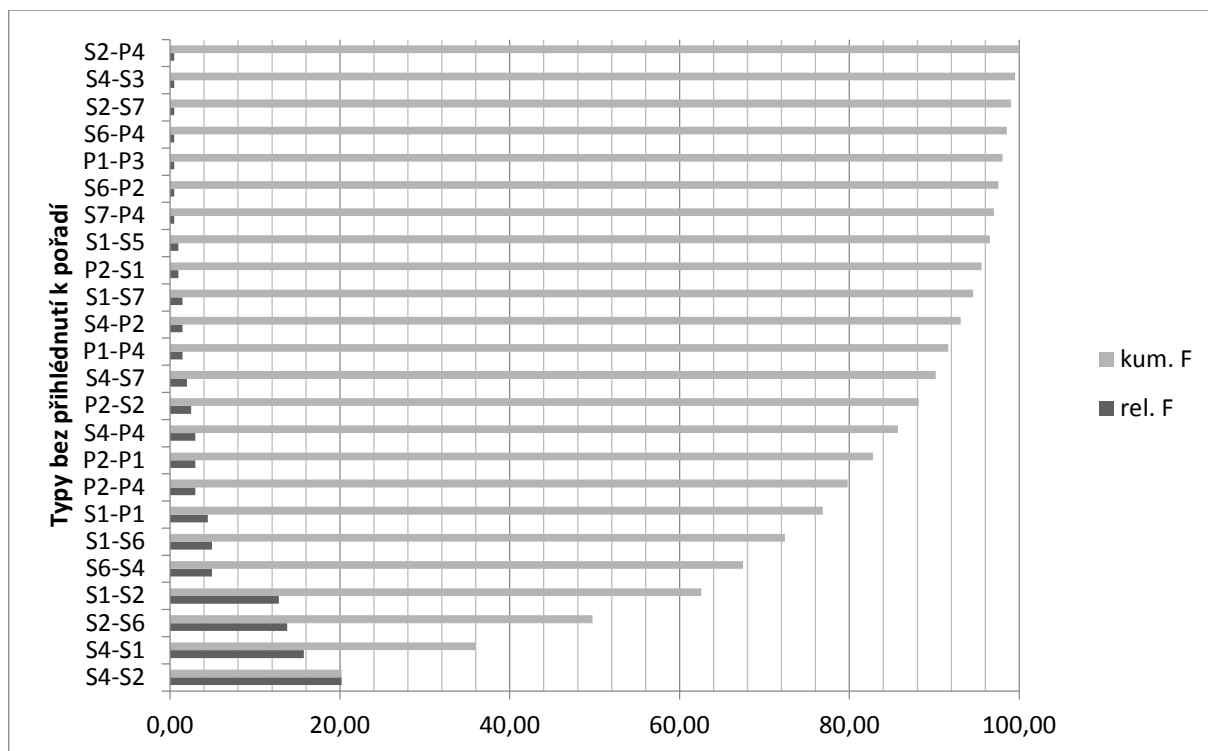
Za daných parametrů¹⁵² se vyděluje následujících 24 typů:

č.	Typy podle dvou nejfrekv. tvarů	abs. F	rel.F	č.	Typy podle dvou nejfrekv. tvarů	abs. F	rel. F
1	S4-S2	41	20,20	13	P1-P4	3	1,48
2	S4-S1	32	15,76	14	S4-P2	3	1,48
3	S2-S6	28	13,79	15	S1-S7	3	1,48
4	S1-S2	26	12,81	16	P2-S1	2	0,99
5	S6-S4	10	4,93	17	S1-S5	2	0,99
6	S1-S6	10	4,93	18	S7-P4	1	0,49
7	S1-P1	9	4,43	19	S6-P2	1	0,49
8	P2-P4	6	2,96	20	P1-P3	1	0,49
9	P2-P1	6	2,96	21	S6-P4	1	0,49
10	S4-P4	6	2,96	22	S2-S7	1	0,49
11	P2-S2	5	2,46	23	S4-S3	1	0,49
12	S4-S7	4	1,97	24	S2-P4	1	0,49
	celkem	183			celkem	20	

Tab. 6.1 Typy morfologických profilů podle prvních dvou pozic frekvenčního pořadí

Obrázek 6.3 ukazuje na svislé ose frekvenční rozložení typů. Na vrchním konci distribuce (absolutní frekvence 41 – 9) se nachází 7 typů, které můžeme s kumulovanou frekvencí 76,85 % považovat za typické. Ve středním pásmu (abs. frekvence 6 až 3) se ocitá 8 typů a na spodním kraji 9 typů (frekvence 2 – 1).

¹⁵² Bez přihlídnutí k pořadí na prvních dvou pozicích frekvenčního pořadí tvarů jednotlivých lexémů.



Obr. 6.3 Relativní a kumulovaná frekvence typů morfologických profilů

V následující části se budeme důkladněji a individuálně věnovat nejfrekventovanějším typům a souhrnně popisovat nízké pásmo frekvenční distribuce. Typy ze středního frekvenčního spektra a jejich zástupci jsou znázorněny v příloze 3.3.

6.3 Frekventované typy morfologických profilů

6.3.1 Typ <S4-S2>

V tomto nejobsáhlejšímu typu se nacházejí především jména abstraktní a konkrétní. Nenalezneme zde žádná označení míst (s výjimkou slova *chata*), institucí, osob ani zvířat. Poměr abstrakt a konkrétní je poměrně vyrovnaný, obdobně i počty jmen počítatelných (21) a nepočítatelných (20). Výrazněji se mezi konkréty vyjmají názvy oborů lidské činnosti (4) a mezi abstrakty jména látková (v tomto typu se nachází osm z deseti). Náplň typu znázorňuje tabulka 6.2.

KATEGORIE	SUBSTANTIVA TYPU <S2-S4>
ABSTRAKTA (15)	<i>barva, ekonomie, fyzika, konkurz/konkurs, mír, odpočinek, osmička, pohyb, práce, tělocvik, výtvarná výchova, vzdělání, zaměstnání, život</i>
KONKRÉTA (25)	<i>alkohol, auto, benzín, bunda, čaj, flaška, jídlo, mapa, maso, muzikál, nádobí, oblečení, zubní pasta, pivo, plášť, průkaz, pult, řidičák, šálek, sešit, sklenka, sprcha, smetana, víno, zmrzlina</i>
NEJASNÁ (1)	<i>smlouva</i>
MÍSTA (1)	<i>chata</i>

Tab. 6.2 Substantiva typu <S2-S4> a absolutní počet podle kategorie

Pokusili jsme se o jemnější roztřídění pomocí třetího nejfrekventovanějšího tvaru. Vznikají morfologické profily <S2-S4,S1>, do nichž se zařazuje nejvíc substantiv (25), <S2-S4,S6> se sedmi členy (*konkurz, zaměstnání, život, bunda, mapa, plášť, chata*), <S2-S4,S7> se čtyřmi členy (*flaška, pult, sklenka, smetana*) a <S2-S4,P2> se třemi členy (*auto, barva, průkaz*). Abstrakta se přitom zkonzcentrují do prvních dvou skupin: *ekonomie, fyzika, mír, muzikál, osmička, pohyb, tělocvik, výtvarná výchova, vzdělání* v <S2-S4,S1> a *konkurz, zaměstnání, život* v <S2-S4,S6>. Konkréta se oproti tomu rozkládají po více skupinách. Celkově však musíme konstatovat, že jemnější rozdělení neposkytuje velkou pomoc při další specifikaci. Podívali jsme se proto příkladově na kolokační profily některých substantivních lexémů s morfologickým profilem <S2-S4,S1>.

K tomuto účelu byly vyhledávány kolokáty dvou nejčastějších morfologických tvarů, tedy Gsg. a Asg., v levém a pravém kontextu od KWIC v rozsahu od -3 do 3. Nalezené kolokáty byly potom seřazené podle jejich absolutní frekvence v kontextu zkoumaného tvaru v sestupném pořadí. Záměrem bylo zjistit, zda substantiva sdílejí obdobné kolokační vzorce, které by odůvodňovaly častý výskyt genitivních a akuzativních tvarů. Rozsah, ve kterém jsme kolokace vyhledávali, byl zvolen poměrně široce, aby bylo možné zahrnout i kolokace se slovesy, které se nutně nemusí ocitat v bezprostředním kontextu. Pracujeme zde s absolutními frekvencemi, nikoliv kolokačními měrami, poněvadž v kontextu této práce je soudržnost jednotlivých členů kolokace méně důležitá než konkrétní výskyt předložek, substantiv a sloves, jež mají vliv na funkci vyjádřenou zkoumaným substantivem. Seznam kolokátů, byl manuálně vyhodnocen. Tabulka 3.2 v příloze podává přehled o frekventovaných kolokacích zkoumaných lexémů. Zde představujeme výsledky souhrnně.

Srovnávání kolokací je poměrně komplexním úkolem. Snažili jsme se proto tříditi nalezené kombinace substantiv na obecnější úrovni. Řídili jsme se přitom Čermákovou typologií lexikálních kombinací (Čermák, 2010, str. 314), jež kombinace slov zachycuje na ose ‚systémové – textové‘ a zároveň na ose ‚pravidelné – nepravidelné‘. Takovým způsobem je možné soustředit srovnávání na frekventované a typické kombinace, jež mohou vysvětlovat častý výskyt genitivních nebo akuzativních tvarů.

Rozlišujeme tudíž:

- systémové, pravidelné kombinace: termínové a propriální kolokace (víceslovné termíny a propria)
- textové, pravidelné: běžné kolokace
- a přechodný typ textově-systémových kombinací: běžné kolokace uzuální

Není vždy možné jednoznačně určit, zda se u některých kombinací jedná „jenom“ o typickou kolokaci nebo „již“ o víceslovný termín (např. *límeč bundy* v textilovním průmyslu jistě termínem je). U běžných kolokací typu <S+Prep+S¹⁵³> nebo <S+S> jsme se pokusili nalézt jemnější kategorizaci na základě sémantické funkce prostého nebo předložkového genitivu resp. akuzativu, např. pracujeme s vlastností genitivu vyjadřovat část z celku nebo cíl či účel.

Kolokační profil genitivních tvarů

Tvar Gsg. je součástí víceslovného pojmenování

S-S	<p><i>ekonomie (Katedra ekonomie, Ústav ekonomie, profesor ekonomie)</i> <i>fyzika (Ústav fyziky/fysiky, profesor fyziky)</i> <i>tělocvik (učitel / učitelka tělocviků, hodina tělocviků)</i> <i>výtvarná výchova (učitel / učitelka tělocviků, hodina výtvarné výchovy)</i> <i>vzdělání (úroveň vzdělání)</i> <i>mír (náměstí Míru, třída Míru, Nobelova cena míru, Závod míru)</i> <i>bunda (kapsa, límeč, rukáv bundy)</i></p>
-----	---

Tvar Gsg. je součástí běžné kolokace uzuální

S-Prep-S V-Prep-S	<p>do <i>pohyb (dát se do, uvést do pohybu)</i></p> <p>od <i>čaj (pytlík od, bedna od), zubní pasta (tuba od), zmrzlina (kornout od)</i> <i>sprcha (hadice od, závěs od)</i></p> <p>z <i>benzín (daň z), muzikál (píseň z), tělocvik (známka z)</i></p>
----------------------	--

Tvar Gsg. je součástí běžné kolokace: S+Prep+S a S+S

část celku	<p><i>alkohol (promile, množství, láhev alkoholu), benzín (litr benzínu), čaj (šálek, sklenka, trochu, hrnek, konvice čaje), maso (g, kilo, kus, kousek, plátek, porce masa),</i> <i>zubní pasta (tuba zubní pasty), zmrzlina (kopeček, porce zmrzliny)</i> <i>mapa (část mapy)</i></p>
zdroj / původ	<p><i>pohyb (radost z, zážitek z pohybu), vzdělání (radost ze vzdělání)</i></p>
objekt	<p><i>alkohol (pití, požití, konzumace alkoholu), benzín (spotřeba benzínu), čaj (pití, příprava, účinek čaje), maso (prodej masa)</i> <i>mír (uzavření míru), vzdělání (investice do, investovat do vzdělání, dosáhnout vzdělání), fyzika (úvod do fyziky), muzikál (inscenace muzikálu)</i></p>
cíl	<p><i>osmička (postup do osmičky)</i></p>

¹⁵³ S = substantivum, Ad = adjektivum, P = zájmeno (pronomen), Prep = předložka (prepozice), V = sloveso (verbum)

Tvar Gsg. je součástí běžné kolokace: V+S nebo V+Prep+S

V+Prep+S	<p><i>alkohol, benzín, čaj, maso (V+ z čaje, z masa, z alkoholu, z benzínu)</i> <i>benzín (přimichávat do, přidávat do benzínu)</i> <i>bunda (dát do bundy, vytáhnout z bundy)</i> <i>čaj (přidávat do čaje, dát si do čaje),</i> <i>sprcha (jít, vlézt do sprchy)</i> <i>osmička (dostat se do osmičky)</i> <i>mapa (zmizet z mapy, nakreslit do mapy, podívat se do mapy), muzikál</i> <i>(obsadit do muzikálu)</i> <i>sprcha (vylézt, vyjít ze sprchy)</i> <i>vzdělání (investovat do vzdělání)</i></p>
V+ přímý objekt	<p><i>vzdělání (dosáhnout vzdělání)</i></p>

Tvar Gsg. je součástí běžné kolokace (uzuální): adverbialní výraz

<p>bez <i>alkohol (bez alkoholu), maso (bez masa)</i> <i>pohyb (sedět bez pohybu), mapa (bez mapy)</i> v době <i>mír (v době míru)</i> podle <i>mapa (podle mapy)</i></p>
--

Nejvýraznější pro genitivní tvary analyzovaných substantiv jsou běžné kolokace typu <S+Prep+S> a <S+S>. Velkou skupinu tvoří kombinace s kvantifikátory, kde tvar Gsg. vyjadřuje spočítanou látku. Nepřekvapuje, že tu nalézáme převážně jména látková (včetně kapalin), např. <číselný výraz> *promile alkoholu*¹⁵⁴, *množství alkoholu*, <číselný výraz> *litr benzínu, trochu čaje, sklenka čaje, kus masa, kopeček zmrzliny* apod. U osmi ze sedmnácti slov pozorujeme frekventované kombinace typu <S+S>, kde genitivní tvar vyjadřuje objekt děje, specifikovaná deverbativním substantivem, např. *požití alkoholu, pití čaje, prodej masa, uzavření míru*. V těchto kombinacích najdeme jak abstrakta, tak konkréta.

Mnoho substantiv (12) se běžně kombinuje s předložkami *do* a *z*, a to zpravidla se slovesy, u nichž má předložkový výraz často význam směru či původu, např. *přimichávat do benzínu, vlézt do sprchy, vylézt ze sprchy, dát si do čaje, zmizet z mapy*.

V neposlední řadě jsou genitivní tvary analyzovaných substantiv součástí ustálených, pravidelných kombinací, tedy víceslovných termínů a proprií. Týká se to především názvů oborů (*ekonomie, fyzika, tělocvik, výtvarná výchova*), ale také abstrakt *mír* a *vzdělání*. Kolokace mají často pro dotyčná slova vysokou frekvenci,

¹⁵⁴ Pro první člen kolokace uvádíme jako lemma.

např. *profesor (+Ad) ekonomie* (F= 59; 18 % všech výskytů tvarů Gsg¹⁵⁵), *profesor fyziky* (62; 9 %), *hodina tělocviku* (39; 28 %), *učitel/učitelka tělocviku* (43; 30,9 %), *učitel/učitelka výtvarné výchovy* (15; 26,3 %), *náměstí Míru* (217; 11,4 %), *Nobelova cena míru* (103; 9,5 %), *kapsa bundy* (52; 14 %).

Ojedinele jsme pozorovali, že genitivní tvar je součástí adverbialního výrazu, a to zejména u slov *mapa (podle mapy)*, *mír (v době míru)*, *pohyb (bez pohybu, neschopen pohybu)*, *alkohol (bez alkoholu)*, *maso (bez masa)*.

Kolokační profil akuzativních tvarů

Tvar A sg. je součástí víceslovného pojmenování

S-Prep-S V-Prep-S	<p>za <i>fyzika (Nobelova cena za fyziku), ekonomie (Nobelova cena za ekonomii)</i> <i>mír (cena za mír, bojovník za mír)</i></p> <p>pro <i>vzdělání (ministerstvo/ministr pro vzdělání, Ústav pro vzdělání, výbor pro vzdělání), fyzika (Ústav pro teoretickou fyziku, časopis pro fyziku)</i></p> <p>na <i>alkohol (test, zkouška na)</i></p>
----------------------	--

Tvar A sg. je součástí běžné kolokace: V+S nebo V+Prep+S

přímý objekt	<p><i>alkohol (pít, mít, popíjet alkohol)</i> <i>benzín (přijet, zastavit pro benzín), mír (uzavřít, chtít, přinést), muzikál (uvádět),</i> <i>bunda (mít, mít na sobě, sundat, svléknout, vzít bundu)</i> <i>čaj (vařit/uvařit, dát si, dát někomu čaj, udělat, připravit /připravovat, popíjet, vypít, přinést čaj, chtít čaj, nalít čaj, nabídnout čaj, dopít),</i> <i>ekonomie (studovat, vystudovat),</i> <i>fyzika (vystudovat, studovat)</i> <i>mapa (vytvořit, ukázat, najít, studovat, rozložit mapu),</i> <i>maso (jíst, nakrájet, přidat, osolit, vložit, dát, dát si maso, vyjmout, opéci, promíchat, prodávat),</i> <i>mír (uzavřít, přinést, chtít)</i> <i>pohyb (sledovat, udělat, vidět, zhlédnout),</i> <i>sprcha (dát si, pustit, mít sprchu),</i> <i>tělocvik (mít, mít rád),</i> <i>výtvarná výchova (učit, vyučovat)</i> <i>vzdělání (doplnit, poskytovat, dostat, dokončit, zajistit)</i> <i>zmrzlina (dát si, koupit, jíst)</i></p>
předložkový akuzativ	<p>za <i>alkohol (zaplatit za), benzín (zaplatit za), bunda (chytit za, tahat za), maso (zaplatit za)</i></p> <p>pro <i>benzín (přijet/jet pro, zastavit pro benzín)</i></p> <p>na</p>

¹⁵⁵ První číslo udává absolutní frekvenci kolokace, druhé poměr výskytu kolokace ke všem výskytům genitivního tvarů dotyčného substantiva.

čaj (postavit na čaj, jít/přijít na čaj), ekonomie (přihlásit se na, dostat se na ekonomii), muzikál (lákat, zvát/pozvat na muzikál), osmička (sázet na osmičku), pohyb (reagovat na, čekat na pohyb), zmrzlina (jít na zmrzlinu)

o

ekonomie (mít zájem o, zajímat se o ekonomii)

pod

bunda (schovat pod, dát pod bundu), sprcha (vlézt pod sprchu)

Tvar *Asg.* je součástí běžné kolokace (uzuální): *S+Prep+S*

cíl / účel

za

alkohol (cena za alkohol), benzín (poplatky za benzín), maso (peníze za maso)

pro

mír (partnerství pro mír), muzikál (zpěvák, kostým pro muzikál), vzdělání (svoboda, motivace pro vzdělání), ekonomie (význam, nástroje pro ekonomii), fyzika (význam, nadání pro fyziku), tělocvik (plavky, věci na tělocvik)

na

alkohol (reklama na alkohol), benzín (daň na, příspěvek na, poukaz na benzín), čaj (voda na, konvice na, hrnek na čaj), ekonomie (odborník na ekonomii), zmrzlina (reklama na zmrzlinu), zubní pasta (reklama na zubní pastu)

U kolokačních profilů akuzativních tvarů je nápadné, že všechna substantiva tvoří běžné kolokace se slovesy, které mohou být pro dotyčný tvar nebo i celé lemmata velmi frekventované, např. *pít alkohol* (247; 19 %), *vařit/uvařit čaj* (224; 8,6 %), *uzavřít mír* (91; 7,5 %), *jíst maso* (174; 6,2 %), *tankovat/natankovat benzín* (47; 6,3 %), *dát si sprchu* (74; 21,1 %), *koupit zmrzlinu* (34; 7,6 %). Výrazný profil mají opět názvy oborů, např. *studovat/vystudovat ekonomii* (65; 31,1 %), *studovat/vystudovat fyziku* (44; 15,5 %), *učit/vyučovat výtvarnou výchovu* (7; 33,3 %). Pro tři látková jména (*alkohol, benzín, maso*) jsou frekventované kolokace se slovesem *platit/zaplatit za*, popř. se substantivním výrazem obdobného významu (např. *cena za, poplatek za, peníze za*). Více než polovina zkoumaných slov (11) je součástí kolokace s adnominální předložkou typu <S+na+S> nebo <S+pro+S >, v nichž akuzativní tvary vyjadřují především cíl nebo účel, např. *partnerství pro mír, význam pro ekonomii, význam pro fyziku, zpěvák pro muzikál, věci na tělocvik, voda na čaj, odborník na ekonomii*. Nejednou jsme se setkali s kolokací se substantivem *reklama* (např. *reklama na zubní pastu, na alkohol, na zmrzlinu*). U těchto kolokací se jedná o kolokace běžné, u některých by se však dalo uvažovat o jejich ustálenosti. Nakonec jsme narazili také na užití tvarů abstrakt *fyzika, ekonomie, mír* a *vzdělání* ve víceslovných pojmenováních, resp. termínových a propriálních kolokacích (srov.: *Nobelova cena za fyziku, Nobelova cena za ekonomii, cena za mír, Ministerstvo/ministr pro vzdělání, Ústav pro vzdělání, výbor pro vzdělání*).

Shrnutí

Předběžně můžeme po prozkoumání prvního typu říci, že kolokace morfologického tvaru dávají dostatečně specifickou informaci, aby bylo možné srovnávat různé lexémy vzhledem k jejich chování v textu. Ve zkoumané třídě substantiv s profilem <S2-S4> bylo možné identifikovat a popsat několik obecnějších kolokačních vzorců. Ty ovšem platí vždy jen pro několik zástupců třídy. Nejvíce výrazný kolokační profil mají jména látková a názvy oborů zahrnuté v této třídě. Jména abstraktní a konkrétní jinak nevynikají zvláštním profilem.

V rámci praktických možností této práce bohužel není možné stejným způsobem vyhodnotit kolokace pro všechny třídy, ačkoliv jenom důkladné porovnávání kolokačních profilů všech typů dokáže zviditelnit, v jakých kolokačních vzorech se rozlišují. Soustředujíc se na didaktický záměr této práce se ale vymežíme na analýzu a znázornění několik málo tříd, které vyplývají z frekventovaných typů morfologických profilů.

Skupina názvů oborů

Poněvadž se kolokační profily tvarů *Asg.* a *Gsg.* názvů oborů z mnoha hledisek podobají, bylo na dalších příkladech ze základního souboru ověřováno, zda mají podobné morfologické profily. Z analýzy vyplývá, že názvy oborů lze přiřadit dvěma skupinám, první s profilem <S2, S4, S1> a druhé s profilem <S2, S1, S4>.

<S2, S4, S1>	<S2, S1, S4>
<ul style="list-style-type: none">▪ <i>ekonomie</i>▪ <i>fyzika</i>▪ <i>hudební výchova</i>▪ <i>medicína</i>▪ <i>tělocvik</i>▪ <i>výtvarná výchova</i>	<ul style="list-style-type: none">▪ <i>biologie</i>▪ <i>chemie</i>▪ <i>filozofie</i>▪ <i>literatura</i>▪ <i>matematika</i>▪ <i>sociologie</i>

U všech názvů oborů je jednoznačně preferovaným tvarem genitiv singuláru. Lexémy se liší jen v tom, zda je druhým nejfrekventovanějším tvarem *Asg.* nebo *Nsg.*

Skupina substantiv látkových

V typu <S2-S4> se nachází většina analyzovaných podstatných jmen látkových (80 %). Viděli jsme, že zejména genitivní tvary tvoří podobné kolokace. Proto zde znázorňujeme morfologické profily dalších podstatných jmen látkových ze základního souboru. Rýsují se následující morfologické profily:

<S2-S4>		<S2-S1>	<S2-S7>
<S2, S4, S1>	<S4, S2, S1>	<S2, S1, S4>	<S2, S7, S1>
▪ <i>alkohol</i>	▪ <i>čaj</i>	▪ <i>zelí</i>	▪ <i>hořčice</i>
▪ <i>bavlna</i>	▪ <i>jídlo</i>	▪ <i>sůl</i>	<S2, S7, S4>
▪ <i>benzín</i>	▪ <i>káva</i>	▪ <i>olej</i>	▪ <i>máslo</i>
▪ <i>maso</i>	▪ <i>pivo</i>		▪ <i>majonéza</i>
▪ <i>víno</i>	▪ <i>zubní pasta</i>		
▪ <i>zlato</i>	▪ <i>zmrzlina</i>		

Podobně jako názvy oborů dávají podstatná jména látková jednoznačně přednost tvaru Gsg. Viděli jsme, že se často kombinují s různými kvantifikátory, genitiv tedy zde má často funkci vyjadřovat část z celku (genitiv partitivní). Jen skupina nápojů se vykytuje častěji v Asg., ocitají se tedy častěji v roli přímého objektu. U slov *hořčice*, *máslo* a *majonéza* je nápadná důležitá role instrumentálu. Pohled na typické kolokace tuto preferenci vysvětluje: *s hořčicí* (52; 42,2 %), *chléb s máslem* (90; 10,4 %), *potřít máslem* (59; 6,8 %), *s majonézou* (52; 54,1 %).

6.3.2 Typ <S4-S1>

Do tohoto typu se řadí především věci abstraktní a konkrétní. S výjimkou lemmat *Němka*, *Čiňanka* a *jednolůžkový pokoj* zde téměř nenacházíme označení osob ani míst.

KATEGORIE	SUBSTANTIVA TYPU <S4-S1>
ABSTRAKTA (19)	<i>fakt, chyba, letecká linka, návrh, nebezpečí, oslovení, předpověď počasí, příběh, rande, školné, sobota, spojení, stolní tenis, vysvětlení, vzkaz, zájem, zážitek, zpráva, ztráta</i>
KONKRÉTA (10)	<i>čárka, polévka, klobása, lampička, lžička, náplast, obraz, párek, šála, sanitka</i>
OSOBY (2)	<i>Čiňanka, Němka</i>
MÍSTO (1)	<i>jednolůžkový pokoj</i>

Tab. 6.3 Substantiva typu <S4-S1> a absolutní počet podle kategorie

Zajímavé je, že se u lexémů této skupiny jedná téměř výlučně o jména počítatelná (s dvěma výjimkami), tj. i abstrakta v této skupině nejsou typickými zástupci své kategorie v tom smyslu, že počítatelnost nepovažujeme za typický rys abstrakt.

Přihlížíme-li k třetí pozici ve frekvenčním pořadí tvarů, vyděluje se skupina slov, jež se frekventovaně vyskytují i v nějakém plurálovém tvaru. Jedná se o lexémy: *čárka*, *chyba*, *vzkaz*, *zážitek*, *zpráva* (<S4-S1,P4>), *lžička* (<S4-S1,P1>) a *klobása* (<S4-S1,P2>). Typickým profilem je však <S4-S1, S2>, jenž má celkem 17 substantiv.

Kombinace s instrumentálem sg. na třetí pozici pozorujeme pro lexémy *fakt*, *náplast*, *šála*, *Němka* a *Čínanka*. Kombinace s tvarem Lsg. signalizuje lokální významovou složku u slov: *jednolůžkový pokoj*, *rande* a *spojení*.

6.3.3 Typ <S2-S6>

60,7 % zástupců tohoto typu tvoří substantiva s primárně lokálním významem (viz tab. 6.4). Pochopitelně se nejčastěji vyskytují s předložkami vyjadřujícími prostorové relace (*v*, *na*, *do*). Většina jmen přitom volí předložky *v* a *do* pro výraz místa popř. směru, což vysvětluje častý výskyt právě lokálových a genitivních tvarů (např. výraz *do přízemí* pokrývá 60,5 % všech výskytů tvaru Gsg. lemmatu *přízemí*). Výjimkou je lemma *venkov*, jež zavděčuje hojný výskyt genitivních tvarů zejména kolokacím jako např. *rozvoj venkova* (167), *obnova venkova* (102).

KATEGORIE	SUBSTANTIVA TYPU <S2-S6>
ABSTRAKTA (5)	<i>čeština, sousedství, třicítka</i> <i>pojmy časové: prosinec, říjen</i>
KONKRÉTA (1)	<i>křeslo</i>
NEJASNÉ (5)	<i>fakulta, gymnázium, krám, ovzduší, recepce</i>
MÍSTA (17)	<i>centrum, čtvrť, dům, hotel, Londýn, obchodní dům, opravna, panelák, patro, pokoj, přízemí, suterén, ubytovna, ulice, venkov, vesnice, vila</i>

Tab. 6.4 Substantiva typu <S2-S6> a absolutní počet podle kategorie

Lokální významovou složku poznáváme ovšem i u ostatních substantiv této skupiny. Lokálový tvar lemmatu *sousedství* se kombinuje v 98 % všech výskytů s předložkou *v*, např. *v těsném sousedství* (*domu, kostela*), *dům v sousedství*, *bydlet v sousedství*. Obdobně výrazné je spojení s předložkou *na* u slova *fakulta* (93 % všech výskytů lokálového tvaru). Slovo *recepce* je zastoupené v konkordancích lokálového tvaru v obou jeho významech. U obou významů zdůrazňují kolokace

zejména lokální význam, srov.: *být/zastavit se v recepci, muž v recepci, na recepci u strýce / v Berlíně*. Pro úplnost je třeba se zmínit o tom, že u lemmat *fakulta, recepce* a *gymnázium* užití předložky *do* na výraz směru kolísá s užitím předložky *na* (+A).

Přibrání třetí pozice ve frekvenčním pořadí rozděluje substantiva především do dvou profilů:

- <S2-S6, S1> | např.: *centrum, čeština, čtvrť, hotel, Londýn, opravna, prosinec* (celkem 14) a
- <S2-S6, S4> | např.: *ovzduší, třicítka, křeslo, krám, gymnázium, recepce, dům, venkov* (celkem 12)

V tomto typu se nachází dva názvy měsíců. Na dalších příkladech ze základního souboru prozkoumáme, zda mají shodný morfologický profil. Ze substantiv se poněkud vyjímá slovo *čeština*. Prozkoumali jsme proto jeho kolokační profil a podívali jsme se, zda obdobný morfologický profil mají i ostatní názvy jazyků. Označením míst se budeme souhrnně věnovat v části 6.5.

Skupina názvů měsíců

Nejfrekventovanějším pádem pro měsíce je genitiv, jenom u měsíců *březen, červen* a *září* jsou tvary lokálu četnější. Lokál je však vždy zastoupený mezi třemi nejfrekventovanějšími tvary. Měsíce *březen, září* a *listopad* se mimo to vyskytují častěji v Asg. než v Nsg. Celkově ale můžeme říct, že typickým profilem názvů měsíců je <S2-S6, S1>, srov. následující přehled o morfologických profilech:

<S2-S6>		<S2-S4>
<S2-S6, S1>	<S6, S2, S4>	<S2, S4, S6>
▪ <i>leden</i>	▪ <i>březen</i>	▪ <i>listopad</i>
▪ <i>únor</i>	▪ <i>září</i>	
▪ <i>duben</i>		
▪ <i>květen</i>		
▪ <i>červen</i>		
▪ <i>červenec</i>		
▪ <i>srpen</i>		
▪ <i>říjen</i>		
▪ <i>prosinec</i>		

Příkladově zde uvádíme frekventované kolokáty tvarů Gsg. a Lsg. lexému *únor*¹⁵⁶.

¹⁵⁶ Výraz v závorkách udává absolutní frekvenci kolokace (s lemmatem). Soustředíme kolokáty, které ovlivňují tvar lemmatu.

<i>kolokáty</i>	<i>příklady</i>
,číslovky' (POS C) (3277)	<i>Vlastní soutěž se uskuteční ve dnech 29. ledna až 1. února 2007 v areálu Fakulty chemické Vysokého učení technického v Brně.</i>
	<i>z (78), např.: Zápis z 27. února 1937.</i>
od (670)	<i>den (133), např. Kompletní článek vyšel v Computeru 2/95 dne 4. února.</i>
do (516)	<i>Pracovali jsme na tom od února roku 2003.</i>
konec (447)	<i>Trvá do 28. února.</i>
polovina (178)	<i>na konci února (115)</i>
začátek (172)	<i>v polovině února (125)</i>
rok (171)	<i>začátkem února (78), na začátku února (56)</i>
	<i>Od února letošního roku máte novou funkci.</i>

Tab. 6.5 Kolokace tvaru Gsg. lemmatu *únor*

<i>kolokáty</i>	<i>příklady</i>
v (2365)	<i>V únoru to budou tři roky.; letos v únoru (221); vloni v únoru (78); začít v únoru (68), např.: Náš projekt začal v únoru letošního roku.</i>
po (203)	<i>Po únoru 1948 fotbal pokračoval.</i>

Tab. 6.6 Kolokace tvaru Lsg. lemmatu *únor*

Skupina názvů jazyků

Z frekventovaných kolokátů slova *čeština* je zřejmé, že častý výskyt slova v genitivu je dán především běžnou kolokací se slovesem *přeložit/překládat do* a příslušným deverbativem *překlad do/z*. Mezi kolokacemi se nacházejí ale i více méně ustálená spojení, srov. *slovník češtiny, výuka češtiny a znalost češtiny*. U kolokací lokálového tvaru jasně dominuje kombinace s předložkou *v* (*být v češtině, vyjít v češtině, apod.*), srov. k tomu tab. 6.7 a 6.8.

<i>kolokáty</i>	<i>příklady</i>
do (344)	<i>přeložit / překládat do (143): Ale já bych to musel do češtiny přeložit.</i>
z (134)	<i>překlad do češtiny (41), např.: K překlada do češtiny zbývá ještě hodně věcí.</i>
včetně (18)	<i>Otázek z češtiny a angličtiny se moc nebojí. Maturitní písemka z češtiny bude o něco náročnější než dnes.</i>
podpora (26)	<i>Je k dispozici v devíti jazycích včetně češtiny.</i>
výuka (26)	<i>Ohledně podpory češtiny by vás měl informovat prodejce.</i>
znalost (23)	<i>Výuka češtiny tu byla na několik let přerušena.</i>
slovník (17)	<i>Cizinci ve škole musí prokázat znalost češtiny.</i>
	<i>Citovaný Slovník spisovné češtiny dudlík charakterizuje jako předmět k dumlání pro konejšení nemluvnat.</i>

Tab. 6.7 Kolokace tvaru Gsg. lemmatu *čeština*

¹⁵⁷ Číslo v závorce udává absolutní počet výskytů tvaru.

kolokáty příklady

v (761)	v češtině, ve starší češtině být v češtině (69) vyjít v češtině (53): např. Ačkoliv nevyšla v češtině , může pro nás být zajímavá. v angličtině a češtině (46), např.: Stránky jsou dvojjazyčné, v češtině a angličtině.
o (38)	FAQ (tedy často kladených otázek) jsou zajímavosti o češtině .

Tab. 6.8 Kolokace tvaru Lsg. lemmatu *čeština*

Náhled do frekvenční distribuce dalších názvů jazyků ze základního souboru ukazuje, že se morfologické profily velice podobají. Preferovanými profily jsou <S2,S6,S4> a <S6,S2,S4>. Mezi nejfrekventovanějšími tvary se ocitají vždy Gsg. a Lsg.

<S2-S6>		<S2-S4>
<S2-S6, S4>	< S6, S2, S1>	<S2, S4, S6>
▪ angličtina	▪ arabština	▪ ruština
▪ němčina		
▪ italština		
▪ polština		

Přehled zároveň znázorňuje, že frekvenční distribuce názvů jazyků a měsíců je velmi podobná.

6.3.4 Typ <S1-S2>

Tento typ je poměrně rozčleněný. Nacházíme v něm substantiva všech kategorií. Mezi konkréty se vyjímají dopravní prostředky (chybí zde jenom *auto* a *sanitka*). Abstrakta je oproti tomu těžko zachycovat obecnějším pojmem, nalézají se mezi nimi názvy oborů, stativa a události. Jedná se však především o jména nepočítatelná. Velmi malá skupina živočichů je zde zastoupená v úplnosti.

KATEGORIE	SUBSTANTIVA TYPU <S1-S2>
ABSTRAKTA (9)	<i>biologie, chaos, chemie, náboženství, nuda, průmysl, spor, velikost</i> názvy oborů: <i>biologie, chemie</i>
KONKRÉTA (7)	<i>autobus, hudba, chléb, karavan, kniha, nádrž, tramvaj, vrtulník</i>
NEJASNÉ (2)	<i>podnik, policie</i>
ŽIVOČICHOVÉ (2)	<i>losos, treska</i>
OSOBY (5)	<i>advokát, ministr, mluvčí, ředitel, teta</i>
MÍSTA (1)	<i>město</i>

Tab. 6.3 Substantiva typu <S2-S2> a absolutní počet podle kategorie

Podle třetí pozice frekvenčního rozložení tvarů se dělí substantiva na profily <S1-S2,S4>, <S1-S2,S7> a <S1-S2,P2>, přičemž první je typický. Můžeme k němu přiřadit 14 lexémů. Tento profil je výrazný zejména pro abstrakta: *biologie, chaos, chemie, hudba, náboženství, nuda, průmysl, spor, velikost*.

Profily označení živočichů a osob se budeme blíže zabývat v části 6.5.

6.3.5 Typ <S6-S4>

V tomto typu se nacházejí pojmy časové, i když v malém počtu, a substantiva s významovou složkou místa (*louka, podlaha, přejezd, trh*).

KATEGORIE	SUBSTANTIVA TYPU <S6-S4>
ABSTRAKTA (4)	<i>chvíle, půlhodina, smysl, souvislost</i>
KONKRÉTA (5)	<i>louka, podlaha, přejezd, telefonní seznam, večere</i>
NEJASNÁ (1)	<i>trh</i>

Tab. 6.10 Substantiva typu <S6-S4> a absolutní počet podle kategorie

Užití v předložkové skupině s lokálním významem je pro některá ze substantiv typické a dalo by se říct primární. Např. spojení *na trhu* pokrývá 96 % všech výskytů lokálového tvaru a 33,3 % všech výskytů celého lemmatu, *na přejezdu* se vyskytuje 152krát, tedy v 95 % všech konkordancích lokálového tvaru a v 18 % všech morfologických tvarů lemmatu. Na rozdíl od lexémů s lokálním významem v typu <S2-S6> se vyznačují lemmata *louka, podlaha, přejezd* a také *večere* a *trh* tím, že k výrazu směru používají předložkový akuzativ, tedy *na+A*, např. *na (+Ad) louku* (213; 44,7 %).

U substantiv *mysl* a *souvislost* je nápadná vysoká míra pokrytí, jíž dosahují dvěma nejfrekventovanějšími tvary. Tvary Lsg. a Asg. pokrývají 71,4 % všech užití slova *mysl* a 70,7 % všech užití slova *souvislost*, u něhož dokonce 62,9 % výskytů připadá jen na Lsg. Důvodem jsou poměrně silné kolokace, které tvoří lokálové tvary obou lexémů s předložkou *v*, a tvar akuzativu slova *mysl* se slovesem *mít*, srov.:

- *v (+Ad/P) souvislosti* (5440; 98,8 %), z toho: *v souvislosti s* (4067; 73,9 %)
- *v (+Ad/P) smyslu* (3878; 70,9 %), *v (+Ad) slova smyslu* (1009; 18,5 %)
- *mít mysl* (2556; 33,9 %) a *mít mysl pro* (1310; 17,4 %).

Zejména u spojení v *souvislosti* s můžeme pro jeho vysokou frekvenci mluvit o ustálené kolokaci, kterou lze považovat za víceslovnou předložku (srov. Čermák, 2010, str. 175).

Pokud přibíráme třetí pozici, vznikají profily <S6-S4, S1> (*louka, půlhodina, smysl*) a <S6-S4, S2> (*chvíle, podlaha, přejezd, telefonní seznam, trh*). Bereme-li v potaz tři nejfrekventovanější tvary bez přihlídnutí k pořadí, bude zřejmé, že substantiva těchto skupin mají blízko k typu <S2-S6>.

6.3.6 Typ <S1-S6>

Kombinace tvarů lokálu a nominativu singuláru je typická pro názvy měst. Do tohoto typu se řadí všechna města s výjimkou *Londýna* (s profilem <S6,S2,S1>). Hyperonymum *město* však tento profil nesdílí, vyskytuje se nejčastěji v N a G singuláru.

KATEGORIE	SUBSTANTIVA TYPU <S1-S6>
NEJASNÉ (3)	<i>pravice, sjezd, vedro</i>
MÍSTA (7)	<i>Bratislava, Brno, Liberec, Praha, Ústí, Washington, Ženeva</i>

Tab. 6.11 Substantiva typu <S1-S6> a absolutní počet podle kategorie

Profil názvů měst můžeme označit za poměrně výrazný. Po přibrání třetí pozice je můžeme přiřadit buď profilu <S1,S6,S2> nebo <S6,S1,S2>. Typické kolokace třídy názvů měst ukážeme v části 6.5.

K typu dále patří slova, jež jsou vzhledem k distinkci abstrakta-konkréta nejasná, zejména proto, že mohou označovat povahově odlišné denotáty, srov. definici ve *Slovníku spisovného jazyka českého*¹⁵⁸:

pravice: a) ‚pravá ruka‘ b) ‚pravá strana‘ c) ‚konzervativní, reakční orientace v politickém i kulturním životě‘

sjezd: a) ‚shromáždění lidí, kt. se sjeli za společnými zájmy‘, b) ‚jízda dolů‘ a c) ‚cesta, po kt. se sjíždí‘.

Zajímá nás, do jaké míry se odrážejí odlišné významy v kolokačním profilu slov.

V převážné většině svých výskytů se nachází tvar Lsg. lemmatu *pravice* v kombinaci *po pravici* (318; 83,2 % např. *sedí na čestném místě po pravici krále*), a odráží tak zejména užití ve smyslu b), tedy primárně lokální význam. Mezi

¹⁵⁸ Srov. *pravice*, Ústav pro jazyk český AV ČR, 2015; *sjezd*, Ústav pro jazyk český AV ČR, 2015.

konkordancemi Nsg. narážíme na kolokace s adjektivy *česká* (28; 7,3 %), *krajní* (23; 6 %) a *politická* (11; 2,8 %), jež svědčí o užití ve smyslu c).

Nejvýznamnější kolokací lemmatu *sjezd* v lokálu singuláru je kombinace s předložkou *na* (593; 65,1 %). Frekventované kolokáty jak např. *strana, ČSSD, spisovatel, svaz* naznačují, že se slovo užívá zde především v nahoře uvedeném smyslu a). Kombinace typů *v(e)* (+Ad) *sjezdu* jsou třetí nejčetnější kolokací (138; 15,1 %). Lemma se zde užívá jen ve smyslu b), např. *Ve sjezdu sbírala světové medaile*. Kolokační profil nominativního tvaru se projevuje jako charakteristický pro užití slova ve smyslu a), z 40 nejčastějších kolokátů ukazuje 21 na užití ve významu ‚shromáždění lidí‘ či ‚kongres‘ (např. *konat se, strana, spisovatel, komunistický, svolat*), a jen čtyři vyloženě na užití ve smyslu b) a c).

S opatrností bychom tedy mohli shrnout, že se víceznačnost lemmatu *pravice* zrcadlí v jeho morfologickém profilu. Pro lemma *sjezd* však není jasné, zda by se profily založené na odlišných významech lemmatu lišily alespoň s přihlédnutím k dvěma prvním pozicím frekvenční distribuce.

6.3.7 Typ <S1-P1>

Tento typ zachycuje jenom substantiva sémantické třídy „člověk“, shromažďuje dokonce nejvíce podstatných jmen této kategorie. Jedná se o následující lexémy:

Angličan, ekonom, herec, lékař, manažer, podnikatel, Rakušan, řidič, učitel.

Je to typický profil pro názvy povolání. Tyto lexémy se navíc úplně shodují v pořadí tří nejfrekventovanějších tvarů: <S1,P1,P2>. Výjimkou je slovo *Angličan*, jež se sice také nejčastěji vyskytuje v dotyčných tvarech, ale v odlišném pořadí: <P1,S1,P2>. Mimo substantiva *Angličan* se v tomto typu nachází další název národností. Příslušníků národností je v náhodném vzorku málo, přitom je tato skupina ze sémantického hlediska dobře vymežitelná. Ověřili jsme proto na dalších příkladech základního souboru, jaký má tato skupina substantiv typický profil. Podíváme se nejdříve jen na maskulina. Femininům se budeme blíže věnovat v části 6.5.1.

Skupina příslušníků národností

Z následujícího přehledu morfologických profilů vyplývá, že třemi nejfrekventovanějšími tvary lexémů označujícími příslušníky národností jsou Npl., Nsg. a Gpl.

Neplatí to jen pro slova *Švýcar* a *Švéd*, jež místo Gpl. upřednostňují Asg. nebo Ipl. Pro hrubou polovinu zkoumaných lemmat (55 %) je typickým profilem <P1,S1,P2>.

<S1-P1>		<P1-P2>	
S1,P1,P2	P1,S1,P2	S1,P1,P7	P1,P2,S1
<ul style="list-style-type: none"> ▪ <i>Kanaďan</i> ▪ <i>Polák</i> 	<ul style="list-style-type: none"> ▪ <i>Angličan</i> ▪ <i>Američan</i> ▪ <i>Čech</i> ▪ <i>Číňan</i> ▪ <i>Francouz</i> ▪ <i>Japonec</i> ▪ <i>Němec</i> ▪ <i>Rakušan</i> ▪ <i>Slovák</i> ▪ <i>Španěl</i> 	<ul style="list-style-type: none"> ▪ <i>Švýcar</i> 	<ul style="list-style-type: none"> ▪ <i>Evropan</i> ▪ <i>Maďar</i> ▪ <i>Rus</i> ▪ <i>Řek</i>
S1,P1,S4			
<ul style="list-style-type: none"> ▪ <i>Švéd</i> 			

Častý výskyt substantiv v množném čísle je dán tím, že se jimi označuje celý národ. Nepřekvapuje, že zmíněné tvary nominativu a genitivu jsou nejčastěji zastoupené v publicistice. Jednotné číslo nominativu je jen u slov *Kanaďan*, *Švéd*, *Švýcar* a *Polák* četnější. Pohled na kolokace prozrazuje, že se tvar *Polák* často vyskytuje s křestními jmény, což svědčí o tom, že se k lemmatu přiřadilo i příjmení *Polák*¹⁵⁹. To může vysvětlit častý výskyt Nsg. Ve srovnání s jinými substantivy méně frekventovaná lemmata *Kanaďan* (s celkem jen 500 výskytů), *Švéd* (486 výskytů) a *Švýcar* (590 výskytů) se oproti tomu skutečně často užívají i k označení jednotlivého zástupce národů.

6.3.8 Okrajové jevy

Tabulka 6.12 znázorňuje morfologické profily, které se vyskytují v náhodném vzorku jen jednou nebo maximálně dvakrát.

TYP	SUBSTANTIVA Z NÁHODNÉHO VZORKU
P2-S1	<i>pacient, umělec</i>
S1-S5	<i>slečna, táta</i>
S7-P4	<i>koniček</i>
S6-P2	<i>lokalita</i>
P1-P3	<i>novinář</i>
S6-P4	<i>léto</i>
S2-S7	<i>okno</i>
S4-S3	<i>pití</i>
S2-P4	<i>stereotyp</i>

¹⁵⁹ S tím se musí počítat také u lemmat *Čech* a *Němec*, jejichž celkový počet je však mnohem vyšší, srov. 12 610 (*Čech*) a 9 529 (*Němec*) vs. 2 741 (*Polák*).

Tab. 6.12 Přehled typů s nízkou frekvencí

Příkladově pozorujeme kolokace vybraných lexémů, zda nám neobjasňují nezvyklý častý výskyt morfologických tvarů, jež se ocitají v celkovém rozložení pádových tvarů spíš na zadních místech.

Pacient a umělec

Profily lemmat *pacient* a *umělec* mají blízko k substantivům typu <S1-P1>, přihlížíme-li k třetímu nejfrekventovanějšímu tvaru, jímž je Npl. Liší se od nich tím, že preferovaným tvarem je Gpl. Zvláště frekventovaný je tvar v kolokaci s předložkou *u* (317), dále je nápadný častý výskyt ve spojeních *počet pacientů* nebo *procento pacientů* (174). Podobně vyčnívají právě kolokace s kvantifikátory (*řada, mnoho, většina* – celková frekvence 137) v kolokačním profilu lemmatu *umělec*.

Lokalita

Lexém *lokalita* patří k sémantické třídě míst. Obdobně jako převážná většina substantiv této třídy je nejčetněji zastoupený v Lsg. (srov. typ <S2-S6>). Typicky se tedy vyskytuje ve spojení s předložkami *v* a *na* (viz tab. 6.13). Druhým nejfrekventovanějším pádem je genitiv plurálu. Frekventované kolokáty znázorněné v tabulce 6.14 objasňují, že genitivní tvar slouží nejen k vyjádření původu nebo cíle, ale také poměrně často k výrazu části z celku.

Tvar Lsg. (1173)

<i>kolokáty</i>	<i>příklady</i>
<i>v</i> (988)	<i>Celkem devět stromů zmizelo ze sadů v lokalitě Polanka v Třebíči. Město investorům v této lokalitě vychází vstríc.</i>
	<i>nacházet se v (+Ad/P) lokalitě (16): v jaké lokalitě se byt nachází</i>
	<i>v lokalitě u (56): Řidiči tak zaplatí za parkování v lokalitě u nemocnice a u hlavní pošty.</i>
<i>na</i> (87)	<i>Chce mít na této lokalitě vlastní elektrárnu anebo alespoň společný podnik v každém případě. Hodně záleží na lokalitě.</i>

Tab. 6.13 Kolokace tvaru Lsg. lemmatu *lokalita*

Tvar Gpl. (842)

<i>kolokáty</i>	<i>příklady</i>
<i>z</i> (262)	<i>Dřevo pochází z různých lokalit Moravy.</i>
	<i>žádná/jedna z (+Ad/P) lokalit (39): Jedné z nejhezčích lokalit v blízkosti metropole. Žádná z těchto lokalit není komerčně atraktivní.</i>
<i>do</i> (91)	<i>Je určitě dobře, že se do těchto zničených lokalit vrací les.</i>
<i>několik</i> (35)	<i>Ve Slezské Ostravě je několik problémových lokalit.</i>
<i>většina</i> (29)	<i>Většina lokalit zanikla.</i>

počet (28)	<i>Lze očekávat mírný nárůst počtu lokalit.</i>
řada (21)	<i>Na místě stále pokračují záchranné práce, v řadě lokalit je ale komplikuje špatný přístup do zasažených oblastí.</i>

Tab. 6.14 Kolokace tvaru Gpl. lemmatu *lokalita*

Léto

Kolokace dvou nejčtenějších tvarů lemmatu *léto* dobře zviditelňují odlišné významové odstíny substantiva. Zatímco u lokálových tvarů dominuje kolokace *v létě* (5909; 94,9 %) s významem ‚časového úseku o rozsahu konkrétního ročního období‘, převažuje mezi konkordancemi tvarů Apl. užití slova ve smyslu ‚dlouhý časový úsek‘. Srov. následující kolokace, které obnáší celkem 43,3 % všech užití tohoto tvaru, srov.:

- *za (+Ad) léta* (861): **Za celá dlouhá léta** praxe nepamatoval takovou hrůzu. Páteční večer byl jedním z nejdramatičtějších v Polsku **za poslední léta**.
- *po (+Ad) léta* (851): Působí nerušená **po léta**, její členové jen kvetou a kynou.
- *na (+Ad) léta* (593): **Na dlouhá léta** to byla marná naděje.

6.4 Věci abstraktní a konkrétní

Je známo, že se abstrakta a konkréta liší v tom, které kolokační vzorce a typy valence preferují (srov. Čermák, 2005 a Čermáková, 2009, str. 99nn.). Čermák (2005, str. 145) konstatuje, že konkréta nemají stejné funkční, syntaktické a kolokační vzorce, zatímco abstrakta mají několik obecných vzorců, které jsou více strukturované. Položili jsme si proto otázku, do jaké míry se rozlišují označení věcí konkrétních a abstraktních i ve frekvenci výskytu nějakého pádového tvaru.

Na první pohled se jména abstraktní a konkrétní vzhledem k morfologickému profilu liší jen málo. Rozdělují se na téměř obdobný počet typů (15 u konkrét, 13 u abstrakt) a výrazně častěji se vyskytují v profilech <S1-S4>, <S2-S4> a <S1-S2>. Až na druhý pohled spatříme, že jsou v těchto skupinách zastoupené v jiném poměru. U abstrakt je nejsilnější typ <S1-S4> (31,1 %), u konkrét <S2-S4> (34,2 %), jenž je u abstrakt druhou nejsilnější skupinou (23,0 %). Skupina <S1-S2> je u obou skupin třetím nejčastějším typem (13,1 % u konkrét a 13,1 % u abstrakt). S výskytem celkem méně čtenějších typů s plurálovými tvary na první nebo druhé pozici frekvenčního pořadí jsme počítali spíše u jmen konkrétních. Náleží k nim 19,1 % všech konkrét, zejména k typům <P1-P2>, <P1-P4>, <S4-P4>. Naproti tomu je překvapující, že se i 14,6 % všech zkoumaných abstrakt přednostně vy-

skytuje v plurálových tvarech, např. *požadavek (P2,P4)*, *léto (P4)*, *koniček (P4)*, *zločin (P2)*. Substantiva s řídkým a individuálním profilem nalézáme u obou skupin (např.: *stereotyp*, *koniček*, *okno*, *pítí*).

typ	konkrétní		abstraktní		nejasné	
	příklady ¹⁶⁰	rel. F	příklady	rel. F	příklady	rel. F
P1-P2	<i>banán, hruška</i>	4,1 %	-	-	-	-
P1-P4	<i>nůžky, řízek, rukavice</i>	4,1 %	-	-	-	-
P2-P4	<i>kalhoty, květina</i>	4,1 %	<i>požadavek</i>	4,9 %	-	-
S1-S2	<i>kniha, nádrž, tramvaj</i>	11,0 %	<i>chaos, chemie</i>	13,1 %	<i>policie, podnik</i>	18,2 %
S1-S4	<i>klobása, lampička</i>	15,1 %	<i>návrh, nebezpečí</i>	31,1 %	-	-
S1-S6		-	-	-	<i>vedro, sjezd</i>	18,2 %
S1-S7	<i>tatarská omáčka</i>	1,4 %	<i>heslo</i>	1,6 %	-	-
S2-P2	<i>vůz</i>	1,4 %	<i>milimetr, zločin</i>	3,3 %	-	-
S2-P4		-	<i>stereotyp</i>	1,6 %	-	-
S2-S4	<i>benzín, bunda, čaj</i>	34,2 %	<i>barva, ekonomie</i>	23,0 %	<i>smlouva</i>	9,1 %
S2-S6	<i>křeslo</i>	1,4 %	<i>čeština, prosinec</i>	8,2 %	<i>fakulta, recepte¹⁶¹</i>	36,4 %
S2-S7	<i>okno</i>	1,4 %	-	-	-	-
S3-S4	<i>pítí</i>	1,4 %	-	-	-	-
S4-P2	<i>sklenička</i>	2,7 %	<i>půllitr</i>	1,6 %	-	-
S4-S6	<i>louka, podlaha</i>	6,8 %	<i>chvilé, smysl</i>	6,6 %	<i>pravice, trh</i>	18,2 %
S4-S7	<i>bradka, čepice, hlava</i>	4,1 %	<i>součást</i>	1,6 %	-	-
S4-P4	<i>houska, písnička</i>	6,8 %	-	-	-	-
S6-P4		-	<i>léto</i>	1,6 %	-	-
S7-P4		-	<i>koniček</i>	1,6 %	-	-
celkem	62	100 %	73	100 %	11	100 %

Tab. 6.15 Morfologické profily (typy) substantiv konkrétních a abstraktních a vzhledem k distinkci nejasných substantiv

Na základě dat náhodného vzorku nemůžeme zjistit, zda se abstrakta jako třída výrazněji liší od konkrétních ve svých morfologických profilech, pokud přihlídneme k prvnímu a druhému nejfrekventovanějšímu slovnímu tvaru. Nejvíce abstrakt se nejčastěji vyskytuje buď v Nsg. nebo Asg., nejvíce konkrétní buď v Gsg. nebo Asg. Počet těch lexémů, které upřednostňují nějaký akuzativní tvar (ať plurálový nebo singulárový), je v poměru vyšší u věcí abstraktních než u věcí konkrétních (celkem 45 % konkrétní vs. 79,5 % abstrakt). Nsg. nebo Npl. je upřednostňovaným tvarem u 45,9 % konkrétní a 35,6 % abstrakt.

U poměrně malé skupiny substantiv, jež jsou vzhledem k distinkci abstrakta-konkrétní nejasná, s obtížností pozorujeme výraznou preferenci pro určitý typ. Je-

¹⁶⁰ Pro úplný přehled substantiv viz příloha 3.1.

¹⁶¹ Dále také: *gymnázium, ovzduší*. Těmito příklady je výčet skupiny „nejasných“ úplný.

jich profil zrcadlí spíš právě odlišné významové složky, např. instituce *fakulta* a *gymnázium* se obdobně jako nejvíce označení míst vyskytují přednostně v lokálu a genitivu singuláru, *policie* naproti tomu preferuje oblíbený profil označení osob. Lokální význam odráží i profil substantiv *ovzduší*, *pravice*, *recepce*, *sjezd*, *trh* a *vedro*.

6.5 Místa, osoby a živočichové

6.5.1 Kategorie „člověk“ (Hum)

Kategorie názvů osob vyniká hned několika způsoby. Jako jediná ze zde analyzovaných kategorií vůbec tvoří typ <S1-P1>, který je navíc nejfrekventovanějším této skupiny substantiv. Názvy osob se výrazně často vyskytují v nominativních tvarech, což vzhledem k tomu, že se jedná o jména činitelská a konatelská často plní roli agenta a tudíž subjektu ve větě, nepřekvapuje. Pro téměř všechna zde zkoumaná substantiva jsou nejčastějším tvarem buď tvary Nsg. nebo Npl. (výjimkami jsou *umělec* a *pacient*, u nichž je jen Gpl. častější). Typickými profily jsou <S1-P1> a <S1-S2>, jež má celkem 56 % všech substantiv této kategorie. Pro úplný výčet profilů a jejich zástupců viz tabulku 6.16.

Jako jediná kategorie tvoří skupina osob profil s vokativem, typickými zástupci jsou jména *slečna* a *táta*. Dalším velice individuálním profilem je typ P1-P3 slova *novinář*.

Typ	Substantiva z náhodného vzorku	abs. F	rel. F
P1-P2	<i>politik, Rus, voják</i>	3	12 %
P1-P3	<i>novinář</i>	1	4 %
S1-P1	<i>Angličan, ekonom, herec, lékař, manažer, podnikatel, Rakušan, řidič, učitel</i>	9	36 %
S1-P2	<i>pacient, umělec</i>	2	8 %
S1-S2	<i>advokát, ministr, mluvčí, ředitel, teta</i>	5	20 %
S1-S4	<i>Čiňanka, Němka</i>	2	8 %
S1-S5	<i>slečna, táta</i>	2	8 %
S1-S7	<i>vdova</i>	1	4 %
celkem		25	100 %

Tab. 6.16 Morfologické profily (typy) substantiv kategorie „člověk“ (HUM)

Přehled zviditelňuje skutečnost, že u etnonym mají plurálové tvary maskulin životných význam generický. Hrají tudíž pro něj významnou roli. Oproti tomu se přechýlené ženské tvary zřejmě častěji vyskytují v čísle jednotném, a to v N nebo A. Tvary množného čísla Npl. a Apl. se u nich ocitají na frekvenčním žebříčku

teprve na páté a šesté pozici. Obdobně v našem vzorku přináležejí k typům se singulárovými tvary i jiná substantiva označující výlučně ženské osoby.

6.5.2 Kategorie „živočich“ (ANIM)

Kategorie živočichů je příliš malá, abychom z příkladů mohli odvodit obecné závěry. Jediní zástupci, názvy ryb *losos* a *treska*, mají téměř totožné frekvenční pořadí morfologických tvarů, a to <S2, S1, P2, S4>. Kolokační profily jejich genitivních tvarů prokazují rovněž velkou míru shodnosti a nastiňují užití obou označení zejména v tematickém okruhu jídlo a stravování.

*Tvary genitivu singuláru*¹⁶²

<i>losos</i> (164)	<i>treska</i> (55)
<ul style="list-style-type: none"> ▪ z (+Ad) <i>lososa</i> (37), <i>filet</i> z (+Ad) <i>lososa</i> (14), <i>g filetu</i> z <i>lososa</i> (7) ▪ <i>g</i>¹⁶³ (+Ad) <i>lososa</i> (14) ▪ <i>plátek lososa</i> (20) ▪ <i>porce lososa</i> (15) ▪ <i>kousek lososa</i> (12) 	<ul style="list-style-type: none"> ▪ z (+Ad) <i>tresky</i> (20), <i>filet</i> z <i>tresky</i> (9), <i>filé</i> z <i>tresky</i> (7) ▪ <i>g</i> (+Ad) <i>tresky</i> (4) ▪ <i>plátek tresky</i> (4)

Tab. 6.17 Nejčastější kolokace genitivních tvarů lexémů *losos* a *treska*

6.5.3 Kategorie „místo“ (LOC)

Skupina jmen označující místa má velice výrazný profil. Nejtypičtějším typem je <S2-S6>. Zařazuje se sem 58,1 % všech substantiv s primárně lokálním významem. Dalších 22,6 % kombinuje frekventovaný lokál s nominativem singuláru (S1-S6). Pro celkem 83,9 % všech substantiv této kategorie je lokálový tvar v množném nebo jednotném čísle mezi prvními dvěma nejfrekventovanějšími. V této kategorii se jednoznačně rýsují zvláštní profily jednak pro třídu názvů měst (S1-S6), tedy *propria*, a jednak pro *apelativa* s primárně lokálním významem (S2-S6).

Typ	Substantiva z náhodného vzorku	abs. F	rel. F
S1-S2	<i>město</i>	1	3,2 %
S1-S6	<i>Bratislava, Brno, Liberec, Praha, Ústí, Washington, Ženeva</i>	7	22,6 %
S2-P2	<i>byt, domek</i>	2	6, %
S2-S4	<i>chata</i>	1	3,2 %
S2-S6	<i>centrum, čtvrť, dům, hotel, krám, Londýn, obchodní dům, opravena, panelák, patro, pokoj, přízemí, suterén, ubytovna, ulice, venkov, vesnice, vila</i>	18	58,1 %
S4-P4	<i>toaleta</i>	1	3,2 %
S6-P2	<i>lokality</i>	1	3,2 %
celkem		31	100 %

Tab. 6.18 Morfologické profily (typy) substantiv kategorie „Místo“ (LOC)

¹⁶² Čísla v závorce udávají absolutní počet výskytů.

¹⁶³ Zkratka jednotky *gram*.

Podrobnější náhled do distribuce lokálových a nominativních tvarů u proprií získáme, když přibíráme další názvy zemí a měst ze základního souboru a srovnáváme první tři pozice frekvenčního pořadí.

Profil	Substantiva označující města a země
S1, S6, S2	<i>Bratislava, Brno, Liberec, Olomouc, Ostrava, Praha, Ústí, Washington, Ženeva</i>
S6, S2, S1	<i>Anglie, Budapešť, Kanada, Londýn, Německo, Paříž, Polsko, Řecko, Řím, Španělsko</i>
S6, S1, S2	<i>Česko</i>

Tab. 6.19 Morfologické profily substantiv označujících města a země

Zdá se, že profil <S1,S6,S2> je typický pro názvy měst, názvy zemí oproti tomu upřednostňují profil <S6,S2,S1> (s výjimkou *Česka*). Některá města se však chovají obdobně jako názvy zemí. Kolokační profily měst *Liberec* a *Olomouc* ukazují příkladně, proč je nominativ sg. tak frekventovaný zejména pro města v České republice.

<i>tvary nominativu singuláru¹⁶⁴</i>	
<p>Olomouc (2966)</p> <ul style="list-style-type: none"> ▪ » - « (972), » . « (790), » , « (772), » / « (223) ▪ » (« (172) a ») « (149) ▪ <i>být</i> (202) ▪ <i>Sigma Olomouc</i> (189) ▪ <i>HFK Olomouc</i> (133) ▪ <i>město Olomouc</i> (70) 	<p>Liberec (3125)</p> <ul style="list-style-type: none"> ▪ » - « (1086), » . « (989), » , « (731), » (« (183) a ») « (183), » / « (126) ▪ <i>být</i> (255) ▪ <i>Slovan</i> (169) ▪ <i>město</i> (87) ▪ <i>Slavia</i> (73)

Tab. 6.20 Nejčastější kolokace nominativních tvarů proprií Olomouc a Liberec

Náhled do konkordančních řádků neobvykle často vyskytující se kombinace s interpunkčními znaménky prozrazuje, že nominativní tvar je frekventovaný ve sportovních zprávách zejména ze světa fotbalu. To potvrzují i kolokace s proprií jako *Sigma* nebo *Slovan*, u nichž Nsg. vyjadřuje funkci jmenovací. Vyhledávali názvy měst jen v subkorpusech BEL a ODB, vypadá pořadí na frekvenčním žebříčku již jinak, a to následující: <S6,S1,S2>.

Bezesporu je pro názvy měst důležitým pádem nominativ i pro svou funkci jmenovací (např.: *K turisticky zajímavým místům patří město Olomouc s městskou památkovou rezervací.*) Vysoká frekvence Nsg. např. u měst *Praha* a *Washington* způsobuje mimo častý výskyt v čele novinových zpráv (srov.: *WASHINGTON Bílý dům včera kvůli obavám, že do ...*) také metonymické užití názvů města jako

¹⁶⁴ Čísla v závorce udávají absolutní počet výskytu.

„místo těch, kteří v něm působí a rozhodují“ (např.: *Washington tvrdí, že rozšířený protiraketový štít bude chránit USA.*) Můžeme shrnout, že výskyt nominativu v této kategorii v poměrně velké míře závisí na funkčním stylu. Stabilním jádrem profilů kategorie míst je lokál a genitiv¹⁶⁵.

6.6 Morfologický profil substantiv v mluveném jazyce

V kapitole 4 jsme se již věnovali závislosti frekvence substantivních forem na sféře užívání. Ze srovnání rozložení pádů v korpusu SYN2005 a jejich rozložením v korpusu mluvené češtiny PMK vyplynulo, že N, G a A jsou sice nejpočetnějšími pády jak ve psané, tak v mluvené podobě, že však právě u nich zaznamenané velké rozdíly v číselném zastoupení v obou sférách. Zatímco poměr G a N je v korpusu psané češtiny téměř vyrovnaný, převládá v korpusu PMK jednoznačně N s relativní frekvencí 34,8 %¹⁶⁶ oproti relativní frekvenci G s 18,3 %. Četněji zastoupený v PMK je i akuzativ s frekvencí 25,8 % oproti 19,2 % v korpusu SYN2005. Rovněž jsme viděli, že se pořadí ve středním a nízkém pásmu frekvenčního rozložení pádů (tedy u instrumentálu, lokálu, dativu a vokativu) jen málo liší v závislosti na mluvené versus psané formě textů.

S přihlédnutím k závěrům z obecného rozložení substantivních tvarů v mluvené češtině nyní vyvstává otázka, zda obecnou tendenci k četnějšímu výskytu tvarů *Asg.* a řídkému výskytu *Gsg.* pozorujeme u všech substantiv podobně a do jaké míry se mění složení nejfrekventovanějších tvarů u jednotlivých substantiv v korpusu mluvené češtiny.

Analýzu opíráme o menší vzorek o rozsahu 28 lexémů vytvořený na základě pilotního vzorku. Základní závěry o rozložení pádových tvarů u jednotlivých lexémů byly prezentované již v Schöne (2011, viz tam i pro úplný výčet zahrnutých lemmat). Byl zjištěn patrný rozdíl mezi frekvenčním pořadím v psaném a v mluveném projevu i s přihlédnutím k jednotlivým lexémům, zejména je však nápadný celkem častější výskyt akuzativu. Tvary akuzativu stoupaly na frekvenčním žebříčku 15 slov. U hrubé poloviny slov (57,1 %) došlo ke změně na první pozici. Totožné složení pádových tvarů bylo zjištěno v pásmu, které pokrývá 85

¹⁶⁵ U názvů měst a zemí není smysluplné rozlišovat profily podle čísla, poněvadž několik názvů měst tvary singulárové nemá.

¹⁶⁶ Frekvence nominativu v PMK podle Šonkové (2008, str. 55).

% užití lemmatu, jen u 18 % substantiv. 46,4 % se lišilo jen jedním tvarem a 25 % dvěma tvary (Schöne, 2011, str. 235).

Pokud se podíváme na první tři pozice frekvenční distribuce - při omezení na tři nejfrekventovanější jevy - zjišťujeme, že přesná polovina lexémů upřednostňuje i v mluvených textech stejné pádové tvary jako v psaných textech, i když se číselné zastoupení jednotlivých pádových tvarů a tím i jejich pořadí v čele frekvenčního rozložení mění. Pravda je, že A sg. stoupá v žebříčku u 38 % lemmat, a to především na úkor méně četnějších tvarů G sg. (např. u lemmat *maso, víno, chléb/chleba, chemie*) a málokdy i N sg. (*akademie, fakulta*). U 84 % analyzovaných lemmat je jeho relativní frekvence vyšší než v korpusu SYNEK2010. Celkově mají vyšší relativní frekvenci i tvary N sg. (u 66 % lemmat, např. *čaj, jídlo, pití, podnikatel, učitel, víno*). Do trojky preferovaných tvarů mohou nastoupit ale i tvary G pl. (*klavír, pivo*), N pl. (*sklenička*) D sg. (*doktor, jídlo*) nebo D pl. (*kniha*). V rámci možností této práce se nám nepodařilo nalézt obecnou tendenci v tom, jak se liší užití jejich tvarů v mluveném a v psaném korpusu. I zde jsou rozdíly velmi individuální.

7 Závěr

7.1 Souhrn výsledků korpusové analýzy a metodické aspekty

Analýza substantiv náhodného vzorku o rozsahu 203 slov potvrdila hypotézu pilotní studie, že je možné roztřídit substantiva na základě jejich morfologických profilů, tj. jejich nejfrekventovanějších dvou až tří slovních tvarů. Porovnání výsledků pilotní a vlastní studie ukázalo, že objem typů, tedy počet substantiv, která se nejčastěji vyskytují v tom či onom pádu, je proměnlivý. Liší se v závislosti na povaze substantiv zvoleného vzorku. Nejfrekventovanějším pádem pro většinu slov byl genitiv a ve větším náhodném vzorku akuzativ. Celkově lze shrnout, že nominativ není a priori typickým pádem z hlediska užití konkrétních jednotlivých substantiv.

Roztřídění substantiv bylo založeno na dvou nejfrekventovanějších morfologických tvarech lemmat. Pomocí těchto profilů se nám podařilo zachytit substantiva v sedmi vysoko frekventovaných, osmi středně a devíti nízko frekventovaných typech. K sedmi nejčastějším typům lze přiřadit celkem 76,85 % zkoumaných substantiv. U jednotlivých typů bylo dále ověřováno, zda se substantiva seskupují po přibrání další, třetí pozice frekvenčního pořadí slovních tvarů.

Největší skupina substantiv, typ <S2-S4>, obsahuje přirozeně nejvíce takových podtypů a zároveň i povahově celkem odlišné lexémy. Tento typ je však typický pro jména látková a názvy oborů. Typ <S4-S1> lze sémanticky těžko podchytit. Je to skupina slov převážně počitatelných a těch abstrakt, jejichž denotát může nabýt i konkrétnější povahy (např. u lemmat *návrh*, *oslovení*, *vzkaz*). Zřejmou přednost dávají názvy měsíců a s nimi i názvy jazyků typu <S2-S6>. Jinak je tento profil specifický pro substantiva s lokálním významem. Přibrání třetí pozice jak u typu <S2-S6>, tak u typu <S6-S4> ukazuje, že substantiva těchto skupin upřednostňují singulárové tvary lokálu, nominativu a akuzativu, přičemž lokálový tvar je vždy buď prvním, nebo druhým nejčastějším tvarem. Kombinace tvarů lokálu a nominativu singuláru typu <S1-S6> je typická pro názvy měst, které mají i po přibrání třetí pozice celkem výrazný profil <S1,S6,S2> nebo <S6,S1,S2>. Typ <S1-P1> nápadně hromadí nejvíce názvů osob, a to jak označení povolání, tak příslušníků národností.

V rámci omezeného náhodného vzorku se tedy podařilo identifikovat morfologické profily, které můžeme označit jako typické pro určitou sémantickou třídu substantiv, např. názvy míst, měst, jazyků nebo příslušníků národností. Na začátku jsme si také položili otázku, zda se obecnější sémantické kategorie, tj. věci konkrétní a abstraktní, lidé, živočichové a místa¹⁶⁷, projevují i v odlišných preferencích morfologických profilů.

Vzhledem k distinkci abstraktum-konkrétum se ukázalo, že typy, k nimž se přiklání nejvíce abstrakt, jsou identické s typy, které upřednostňují i konkréta, i když jsou v nich zastoupené v odlišných poměrech. Celkem je nominativ o něco častěji preferován věcmi konkrétními, mezi abstrakty je více lemmat nejčastěji zastoupených v akuzativu. U obou tříd najdeme v čele frekvenčního žebříčku i profily s plurálovými tvary. Zřejmě se třídění dle předložených morfologických profilů nehodí k tomu, aby odráželo odlišné chování konkrét a abstrakt. Spíš poukazuje na podobnosti v užití, které mohou být dány např. zvýrazněním konkrétního významu jinak abstraktních substantiv. Nápadné bylo například, že je několik málo abstrakt, jež se preferovaně vyskytují v textu v plurálových tvarech (např. *požadavek, přesčas, léto, koníček*), a že se ve skupině s profilem <S1-S4> nacházejí především abstrakta, jejichž význam se může v textu zkonkretizovat (např. *letecká linka, předpověď počasí, zpráva*).

Substantiva sémantické kategorie „člověk“ (HUM) se oproti tomu odlišují od ostatních kategorií hned třemi způsoby. Vytváří profil, totiž <S1-P1>, jenž slova jiných kategorií nemají. Je v něm zahrnuto nejvíce zástupců této kategorie. Jinak platí pro všechny názvy osob, že jeden z prvních dvou preferovaných tvarů je tvarem nominativním. Není divu, že jako jediná kategorie tvoří profil s vokativem. Z kategorií živočichů bylo ve vzorku příliš málo zástupců, aby bylo z jejich preferencí pádových tvarů možné odvodit závěry pro celou kategorii. Neudivuje výsledek, že se substantiva označující místo vyznačují lokálem singuláru jako nejčetnějším tvarem. Pro tuto kategorii byly identifikovány typy <S2-S6> a <S1-S6> jako typické profily, přičemž poslední je preferován zejména názvy měst. Kromě lokálu, který však v některých profilech těchto substantiv mezi dvěma nejčastějšími tvary není (např. u lemmat *město, byt, domek, lokalita*), je důležitým pádem zejména genitiv.

¹⁶⁷ Srov. Čermák, 2010, str. 165n.

Dalo by se říci, že právě kategorie ‚místo‘, a ‚člověk‘ ukazují preference pro určité morfologické profily. Celkem lze shrnout, že vzniklé typy naznačují preferenci pro určité sémantické složky, nejsou však vždy jasně a ostře vymezené vzhledem k sémantické povaze jejich členů.

Závislost frekvence pádů na funkčním stylu a sféře užívání

V kapitole 4 jsme ukázali, že obecná frekvenční distribuce pádů je závislá na funkčním stylu a sféře užívání, zvláště frekvence genitivu, akuzativu, ale i vokativu. Kolísání frekvence pádů ovlivňuje mnoho faktorů, zejména podíl určitých slovtvorných typů, užití deverbálních substantiv a pasivních konstrukcí.

Na pilotním i na větším náhodném vzorku jsme viděli, že četnost, s níž se určitý lexém vyjadřuje v konkrétním pádu, je do značné míry individuální jev. Obecnou tendenci, že by výskyt konkrétního pádu závisel významně i na funkčním stylu, jsme pozorovali v rámci pilotního průzkumu jen pro vokativ a dativ. Ovšem i tu záleží na sémantice slova, zda a v jakém počtu tyto pády hrají vůbec roli ve frekvenční distribuci tvarů. V celkovém přehledu frekvenční distribuce pádů se tyto pády ocitají ve všech stylech a také v mluvené češtině na spodním konci frekvenčního žebříčku. Jinak jsou změny ve frekvencích jednotlivých pádů v tom či onom subkorpusu významné vždy jen pro konkrétní slova. Ani v užití slov v mluvené češtině se nám nepodařilo identifikovat výraznější tendenci, která by byla platná pro všechna zkoumaná substantiva, než tu, již pozorujeme při souhrnném pohledu na frekvence pádů: tvary nominativu a akuzativu jsou v mluvené podobě češtiny celkem četněji zastoupené než v psané podobě. Vliv mluvenosti na frekvenční pořadí morfologických tvarů je však silný. Jen u velmi málo lemmat zůstane frekvenční pořadí totožné. Jen 18 % slov pokrývá 85 % všech jejich užití stejnými morfologickými tvary. Ovšem alespoň u poloviny zkoumaných slov se nemění celkové složení tří nejfrekventovanějších tvarů. Jak bylo již naznačeno, stoupá u mnoha slov v pořadí akuzativ. Které pády se „v jeho prospěch“ posouvají na další pozice frekvenčního pořadí, je opět pro jednotlivé lexémy velmi různé. Ačkoliv tedy pády celkově (bez přihlídnutí k jednotlivým lexémům), obzvláště genitiv, akuzativ, vokativ a v omezené míře i dativ, reagují na změnu funkčního stylu a stylu projevu, mají změny ve frekvencích pádových tvarů konkrétních lexémů spíše individuální ráz. Nápadný pro více substantiv je častější výskyt akuzativních forem v mluvené češtině.

Morfologické profily víceslovných pojmenování a víceznačných lexémů

Vycházeli jsme z toho, že kontext jazykového jevu významně přispívá k utváření jeho vlastností, včetně jeho gramatických kategorií nebo i pragmatických vlastností (srov. Cvrček, 2013, str. 23) a že tudíž slovní tvar, jenž prokazuje konkrétní gramatické vlastnosti a zahrnuje v sobě možné funkce v textu, leccos napovídá o svém kontextu.

Za těchto předpokladů by se měly morfologické profily víceslovných pojmů rozlišit od profilů příslušných základních lemmat. Bylo by potřeba více prostoru, než je nám k dispozici v této práci, věnovat se důkladně této problematice. Předběžně a příkladně jsme zjistili morfologické profily několika příkladů víceslovných pojmenování a jejich běžné kolokace. Příklady ukazují, že se terminologické kolokace mohou svým profilem výrazně lišit od pouhého jednoslovného výrazu (např. *jídelní lístek*, *frézovací hlava*, *jednolůžkový pokoj*). U běžných kolokací nesystémových tomu tak být nemusí, srov. např. *slazený čaj*, *horký čaj* vs. *zelený čaj* (viz tabulku 7.1).

podobný morfologický profil	odlišný morfologický profil
<i>příběh</i> <S1,S4,S2> — <i>životní příběh</i> <S4,S1,S2>	<i>pokoj</i> <S2,S6,S4> — <i>jednolůžkový pokoj</i> <S4,S1,S6>
<i>čaj</i> <S4,S2,S1> — <i>horký čaj</i> <S4,S2,S7>, <i>slazený čaj</i> <S4,S2,S7>, <i>bylinkový čaj</i> <S4,S2,S1>	<i>čaj</i> <S4,S2,S1> — <i>zelený čaj</i> <S1,S2,S4>
<i>fakulta</i> <S2,S6,S1> — <i>Filozofická fakulta</i> <S6,S2,S4>	<i>Filozofická fakulta</i> <S6,S2,S4> — <i>ekonomická fakulta</i> <S1,S2,S4>
<i>polévka</i> <S4,S1,S2> — <i>gulášová polévka</i> <S1,S4,S7>	<i>lístek</i> <P4,S4,P1> — <i>jídelní lístek</i> <S4,S6,S2>
	<i>čeština</i> <S2,S6,S1> — <i>spisovná čeština</i> <S2,S7,S4>, <i>lámaná čeština</i> <S7,S6,S1>
	<i>hlava</i> <S7,S4,S2> — <i>frézovací hlava</i> <S1,S7,S2>

Tab. 7.1 Morfologické profily několik běžných a terminologických kolokací

Bylo by náležité zkoumat konkrétní kontext těchto kolokací, aby bylo možné učinit závěry o souvislosti mezi ustáleností nějaké kolokace, jejím kontextem a morfologickým profilem.

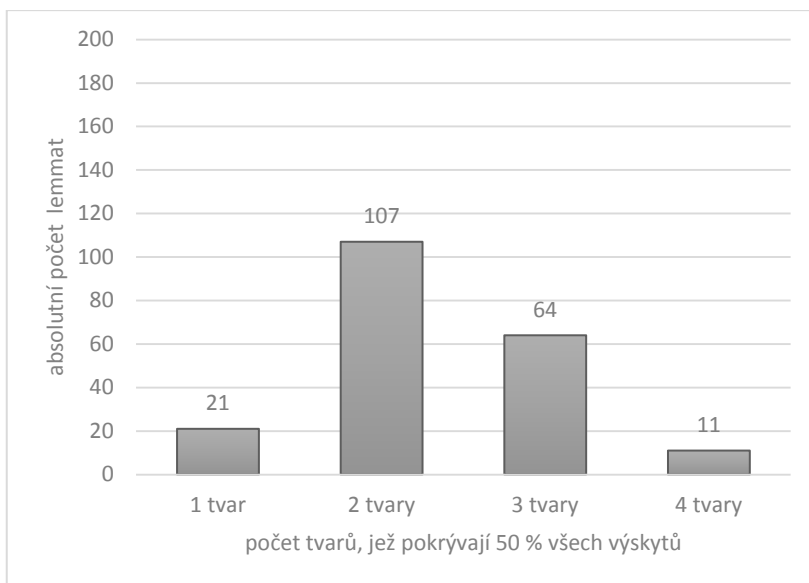
Na druhé straně morfologické profily a kolokace lemmat s odlišnými významy či různými významovými odstíny doložily, že morfologický profil může být i signálem právě této sémantické rozrůzněnosti. Tak se např. slova nejasná, popř. více-

značná vzhledem k sémantickým kategoriím věci abstraktní, konkrétní a místa (např. *fakulta, gymnázium, recepce, krám, pravice, sjezd, smlouva, práce*) ocitají podle toho, které z významových složek se u nich v textu zvláště zvyrazňují, buď v profilech s častými lokálovými tvary typickými pro označení míst obecných <S2-S6> (*fakulta, gymnázium, recepce, krám*) a označení názvy měst <S1-S6> (*pravice, sjezd*) nebo v profilech s častými výskyty akuzativu <S4-S2> (*smlouva, práce*).

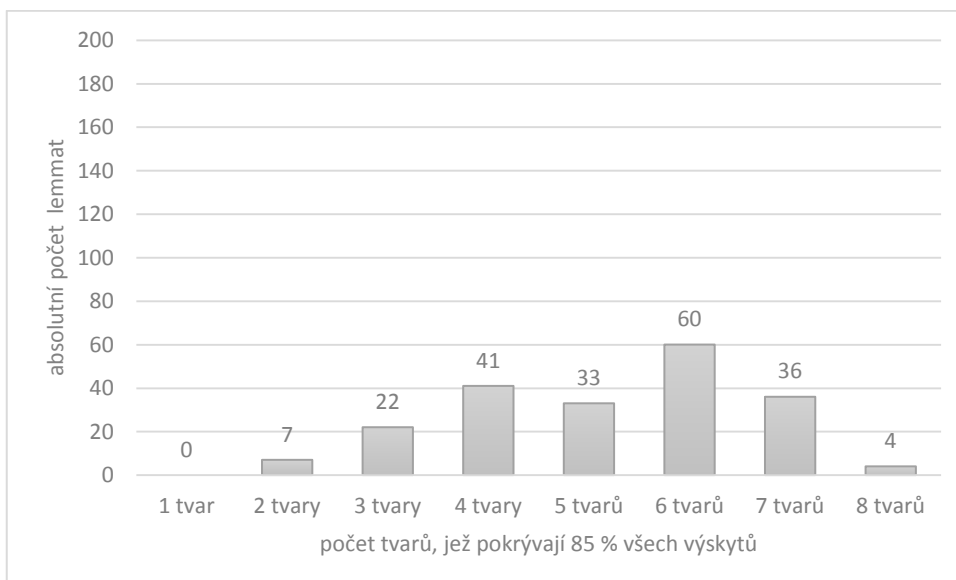
Informační hodnota zvoleného morfologického profilu

Vzhledem k proveditelnosti porovnávání frekvenční distribuce tvarů u jednotlivých lexémů jsme zvolili profil, který zohledňuje první dva až tři nejfrekventovanější tvary substantiv. Závěrem můžeme říci, že takto vytvořené profily dostatečně odráží podobné chování specifických sémantických skupin (jako názvy měst, povolání, jazyků, míst). Vzhledem k dalším konkrétům a abstraktům lze předběžně konstatovat, že profily mohou alespoň pobízet jemnější seskupení abstrakt vzhledem k tomu, jak silně je v nich zastoupená konkrétní významová složka.

V kapitole 5 bylo diskutováno, zda hypotéza, že jsou první tři nebo čtyři jevy na sestupně řazeném frekvenčním žebříčku obvykle relativně nejfrekventovanějšími, platí i pro zde zkoumaný výskyt morfologických tvarů. Proto byla pro každé lemma náhodného vzorku spočítána kumulovaná frekvence jeho tvarů (seřazených podle frekvence). Výsledky dávají najevo, že 192 z 203 lemmat (95 %) pokrývá prvními třemi nejfrekventovanějšími tvary 50 % všech svých výskytů v textu, zbylých jedenáct lemmat k tomu potřebuje čtyři tvary. Bereme-li v potaz, že 90,6 % substantiv má celkem devět tvarů, 82,2 % z nich má deset a 72,4 % dvanáct tvarů, mohli bychom s opatrností říci, že se - v souladu se Zipfovým zákonem a hypotézami Těšitelové (srov. Těšitelová 1990 a kapitolu 4) - nejvíce tvarů nachází na nejvyšších třech (max. čtyřech) pozicích ve frekvenčním pořadí. Nevyřešíme zde ovšem otázku, zda můžeme zmíněných 50 % všech tvarů lemmatu považovat za jeho jádro – za to, co je pro něj typické. Obrázky 7.1 a 7.2 poskytují přehled o různých počtech tvarů, které jsou pro lemmata potřebné k pokrytí 50 % a 85 % všech jejich výskytů.



Obr. 7.2 Pokrytí 50 % všech výskytů lemmat



Obr. 7.3 Pokrytí 85 % všech výskytů lemmat

Z obrázku 7.2 je patrné, že 22 lemmatům stačí dokonce 3 nejčastější tvary na to, aby pokryly 85 % všech svých výskytů v textech psané češtiny. Jedná se o:

*Brno, Bratislava, Ústí, Washington, Ženeva, Praha, Londýn, ředitel, tělo-
cvik, výtvarná výchova, ekonomie, centimetr, maso, školné, suterén, nádo-
bí, nůžky, chléb (chleba), bradka, benzín, chvíle.*

Jsou to substantiva různých morfologických profilů. Poznáváme však mnoho zá-
stupců typu <S1-S6>, v němž jsou zahrnuté zejména názvy měst.

Na 7 lexémů náhodného vzorku narazíme v textech obvykle jen ve dvou pro ně velice typických tvarech, jež je reprezentují v 85 % všech jejich výskytů. Zároveň pokrývají jedním z tvarů již 50 % všech užití. Tímto rysem se vyznačují lexémy

mluvčí (m) <S1,S2>, *přízemí* <S6,S2>, *prosinec* <S2,S6>, *říjen* <S2,S6>, *sobota* <S4,S1>, *součást* <S7,S4> a *sousedství* <S6,S2>.

Tuto zvlášť vysokou frekvenci jednoho nebo dvou tvarů bychom mohli interpretovat tak, že tyto tvary jsou nejlepšími reprezentanty lexému. Zároveň by však také bylo myslitelné - ve smyslu Sinclaire - považovat dotyčný tvar za lexikální jednotku samu pro sebe, pokud vycházíme z toho, že tvary nereflektují automaticky chování celého lemmatu (srov. Sinclair 1991, str. 8). Právě u těchto slov by bylo proto zajímavé a náležité podrobit kontext zvlášť frekventovaného tvaru kolokační analýze.

7.2 Didaktické aspekty prezentace substantivních tvarů

Z výzkumů osvojovacích procesů vyplývá, že osvojování jazyka je pevně spjata s vyhodnocováním informací o frekvenci a distribuci forem v textu. Zkušenost s frekvencí a distribucí jazykových jednotek a jejich struktur řídí jejich kognitivní reprezentaci a určuje způsob a rychlost přístupu ke konstrukcím a tvarům. Vymezení toho, co je základní jednotkou, kterou ukládáme v paměti, a otázka, které typické kombinace tyto jednotky tvoří a jak jsou nám přístupné, zaměstnává jak jazykovědce, tak i odborníky aplikované lingvistiky. Pojmy *konstrukce* (je užíván zejména zástupci tzv. *construction grammar* a *usage-based grammar*), *lexical item* (je diskutován v kapitole 3 v rámci vymezení významu pomocí kontextu) a *kolokace* (je jedním ze základních pojmů korpusové lingvistiky, jejichž pomocí jsme vyhodnocovali a popisovali některé frekventované tvary) osvětlují různé aspekty problematiky vzájemné souvislosti významu s formou. Všem je společné, že zahrnují prvky všech jazykových rovin, rozdíl mezi slovy gramatickými a lexikálními je v různé míře upozaděný. Na pozadí diskuzí vedených ve zmíněných oblastech lze pro výuku morfologie učinit závěr, že by měla přihlédnout ke konkrétním tvarům lexémů, jejich běžným kolokacím a k jejich distribuci v jazykovém úzu. Konkrétní tvar v pro něj typickém kontextu by se proto měl dostat více do centra pozornosti výukového procesu a učebních materiálů.

Jedním z výchozích bodů dané práce byla otázka, které slovní tvary jsou v didakticky orientovaném popisu substantiva závažná pro přiměřenou prezentaci substantivního lexému. Viděli jsme, že nominativ je sice v celkovém rozložení pádových tvarů všech substantiv, a to zejména v beletristických a publicistických textech, nejčastějším tvarem, ale jen pro hrubou třetinu substantiv nejběžnějším pádovým tvarem. Pokusili jsme se proto v této práci identifikovat typické užití lexémů, tedy jejich frekventované formy a zároveň kontext, v němž se často ocitají. Frekvence tvarů a jejich kombinace ve frekvenčním pořadí nám sloužily jako prostředek k rozřazení substantivních lexémů. Pokusili jsme se identifikovat typické zástupce skupin a zároveň typické morfologické profily sémantických tříd. Na kolokacích substantiv jednoho typu bylo ověřováno, do jaké míry mohou být morfologické profily signálem společně preferovaného kontextu těchto substantiv. V náhodném vzorku se vytvořilo celkem sedm vysoko frekventovaných typů kombinací nečastějších morfologických tvarů. Ukázalo se, že takovýto morfologický profil může být typický pro určitou sémantickou třídu substantiv. Zároveň by se dalo říci, že pro některé sémantické třídy substantiv jsou určité profily prototypické. Pro následující skupiny substantiv obsažené v základním souboru této studie je tomu takto:

- názvy oborů (lidské činnosti): <S2,S4,S1> a <S2,S1,S4>
- názvy měst: <S1-S6>, konkrétně <S1,S6,S2> a <S6,S1,S2>
- názvy zemí: <S6-S2>, konkrétně <S6,S2,S1>
- názvy měsíců: <S2-S6>, konkrétně <S2,S6,S1> a <S6,S2,S1>
- názvy jazyků: <S2-S6>, konkrétně <S2,S6,S4> a <S6,S2,S4>
- názvy příslušníků národností: <S1-P1>, konkrétně: <P1,S1,P2>
- obecné názvy míst: <S2-S6>
- obecné názvy osob: <S1-P1> a <S1-S2>
- substantiva látková: <S2-S4>, konkrétně: <S2,S4,S1> a <S4,S2,S1>

Při sestavování látky pro učebnice nemůže být frekvence lemmat ani tvarů vždy předním kritériem výběru. Poněvadž texty, které se využívají v učebních materiálech, nabízejí vlastní konstrukce k probranému tématu, není vždy možné ani smysluplné texty frekvenčním profilům slov přizpůsobit. Avšak máme za to, že informace o morfologických profilech lze zohledňovat při sestavování slovní zásoby pro cvičení popř. pro tematické přehledy, při sestavování slovníkové části

resp. doplněk učebnic, ale také při explikaci funkcí pádů a konstrukcí typických pro daný pád, resp. typických zástupců těchto konstrukcí.

Podle našeho názoru mohou být identifikace morfologického profilu lemmatu, tj. jeho nejtypičtějších tvarů, a zjišťování častých kolokací vhodným způsobem vymezování náležitých případů jeho užití. Máme za to, že by se ve slovníkových částech učebnic měly uvádět mimo tvaru Nsg. také alespoň dva nejfrekventovanější tvary pokud možno s minimálním, zobecnovaným kontextem¹⁶⁸. U příslušníků sémantických tříd, jež spojuje podobný morfologický a sémantický profil, by se to mohlo uskutečnit pomocí odkazu na příkladový kolokační profil jednoho zástupce třídy. Předpokladem takového přístupu je samozřejmě analýza kolokačních profilů více slov než těch, která mohla být prozkoumaná v této práci. Příkladně zde znázorníme navržený přístup na několika příkladech z náhodného vzorku¹⁶⁹. Je potřeba doplnit, že by se výběr příkladových kolokací měl řídit v ideálním případě frekvencí výskytu nejenom v korpusu psané, ale také mluvené češtiny. Zde vybrané kolokace jsou však založené jen na úzu v psané češtině.

lemma	profil
<i>alkohol</i>	[Gsg. <i>bez alkoholu</i> , Asg. <i>pít alkohol</i>]
<i>Angličan</i>	[Npl. <i>Angličané jsou</i> , Nsg. <i>Angličan byl</i>] → NÁRODNOSTI
<i>benzín</i>	[Gsg. <i>litr benzínu</i> , Asg. <i>za benzín</i>]
<i>bunda</i>	[Asg. <i>mít na sobě bundu</i> , Gsg. <i>dát do bundy</i>]
<i>čeština</i>	[Gsg. <i>přeložit do češtiny</i> , Lsg. <i>v češtině</i>] → JAZYKY
<i>ekonomie</i>	[Gsg. <i>profesor ekonomie</i> , Asg. <i>vystudoval ekonomii</i>] → OBORY
<i>losos</i>	[Gsg. <i>plátek lososa</i> , Nsg. <i>losos je</i>]
<i>maso</i>	[Gsg. <i>kus masa</i> , Asg. <i>nejí maso</i>]
<i>pacient</i>	[Gpl. <i>mnoho pacientů</i> , Nsg. <i>pacient má</i>]
<i>recepce 1</i>	[Lsg. <i>žena v recepci</i> , Gsg. <i>do recepce</i>]
<i>recepce 2</i>	[Lsg. <i>být na recepci</i>]
<i>smysl</i>	[Asg. <i>nemá smysl</i> , Lsg. <i>v tom smyslu</i>]
<i>sousedství</i>	[Lsg. <i>v sousedství</i> , Gsg. <i>lidé ze sousedství</i>]
<i>zmrzlina</i>	[Asg. <i>koupit zmrzlinu</i> , Gsg. <i>kopeček zmrzliny</i>]

¹⁶⁸ V přehledu uvádíme příklady z absolutně nejčastějších kolokátů v kontextu -3 3 od KWIC v podobě *word*. Dalo by se uvažovat o tom, zda v konkrétních případech není lepší uvést lemma kolokátu.

¹⁶⁹ Následující přehled slouží ke znázornění přístupu obohacení informace u substantivních hesel v slovníkové části učebnic, nepředstavuje však úplný návrh na podobu takového seznamu. Chybí zde kromě překladu do výchozího či mateřského jazyka např. informace o gramatickém rodu, již lze znázornit i barevným označením. Také by se dalo uvažovat o doplnění informace o příslušnosti k deklinačnímu vzoru nebo informace o sémantických vztazích mezi slovy.

Tvaru nominativu singuláru ovšem nadále připadá důležitá role v učebnicích. Jednak pomáhá při přiřazování substantiva k deklinačním vzorům, jednak slouží jako lemma v slovníkové části učebnic. Nicméně by si měl student být vědom toho, že nominativ singuláru není automaticky typickým tvarem jedné lexikální jednotky a že tuto roli častěji zaujímají tvary genitivu a akuzativu singuláru a v nemalých případech i tvar Lsg. nebo Npl.

Dále shledáváme smysluplným vytvořit i uvnitř lekcí více prostoru syntagmatické stránce substantiv. Tak by se například kolokační profily výše zmíněných sémantických tříd mohly stručně shrnout u příslušných tematických okruhů.

V kapitole 2 jsme diskutovali různé přístupy autorů učebnic k prezentaci deklinačních tříd a k odstupňování komplexního pádového systému češtiny, popř. k lepší parcelaci látky¹⁷⁰. V mnoha učebnicích se nám jevílo jako problematické, že explikace funkce nějakého pádu často postrádá dostatek příkladových konstrukcí. Na pozadí závažnosti prototypu, tj. typického zástupce nějaké pádové funkce nebo konkrétní konstrukce, pro osvojovací proces, je podle našeho mínění záhodno identifikovat typická substantiva, jež určitou pádovou funkci často vyjadřují, a jejich typické kolokace, jež mohou sloužit jako příklad této funkce. Považujeme za důležité poskytovat studentům k analýze funkcí pádů dostatek příkladů, a to substantivní tvary různých lexémů. Takovým způsobem studenti dostávají již „hotové“ a správné konstrukce, jejichž pomocí mohou navíc sami objevovat pravidelnosti deklinačního systému. Při řešení takových úkolů a při práci s textem je potom záhodno obracet pozornost studentů na to, že v určitých konstrukcích vázaných na konkrétní syntaktickou funkci (jako např. vyjádření data, cíle nebo účelu cesty apod.), v nichž tak často chybují¹⁷¹, mohou být prototypické i jiné pádové tvary než nominativ.

¹⁷⁰ Např. výběr pouze frekventovaných zástupců méně častých vzorů a jejich kolokací pro nácvik substantiv v prvních lekcích (Holá 2006; Holá a Bořilová, 2011).

¹⁷¹ Srov. výsledky analýzy chyb v kapitole 2.

Literatura

ALEGRE, M. a GORDON, P. 1999. Frequency Effects and the Representational Status of Regular Inflections. *Journal of Memory and Language* 40. 1999, roč. 40, č.1, s. 41-61. ISSN 0749-596X.

BARTOŇ, T.; CVRČEK, V. et al. 2009. *Statistiky češtiny*. Praha : Nakladatelství Lidové noviny, 2009. ISBN 978-80-7106-594-4.

BAUSCH, K., et al. (eds.). 2006. *Aufgabenorientierung als Aufgabe : Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*, Tübingen : Narr, 2006. ISSN 0175-7776.

BEDŘICHOVÁ, Z.; ŠEBESTA K.; ŠORMOVÁ, K. 2011. Podoba a využití korpusů jinojazyčných a romských mluvčích češtiny: CZESL a ROMi. In ČERMÁK, F. (ed.) *Korpusová lingvistika Praha 2011 : 2 Výzkum a výstavba korpusů*, Praha : Nakladatelství Lidové noviny, 2011 s. 93-104. ISBN 978-80-7422-115-6.

BENEŠOVÁ-GRUNDOVÁ, D. 2006. Jak začlenit sociokulturní aspekty do výuky anglicky mluvících studentů : příklad české verze programu Euromobil. In ŠTINDLOVÁ, B.; ČEMUSOVÁ, J. (ed.) *Sborník Asociace učitelů češtiny jako cizího jazyka (AUČCJ)*, 2005-2006. Akropolis, 2006, s. 113-126. ISBN 80-86903-37-0.

BIBER, D. 1993. Representativeness in Corpus Design. *Lit Linguist Computing*. Vol. 8, No. 4, 1993, s. 243-257. ISSN 0268-1145

BISCHOFOVÁ, J.; H. CONFORTIOVÁ; J. CVEJNOVÁ; et al. 2001. *Prahová úroveň : čeština jako cizí jazyk*. Strasbourg : Council of Europe, 2001. ISBN 92-871-4761-2.

BOERS, F.; DECONINCK, J.; LINDSTROMBERG, S. 2010. Choosing motivated chunks for teaching. In BOERS, F. et al. (ed.) *Fostering language teaching efficiency through cognitive linguistics*. Berlin : Mouton de Gruyter, 2010, s. 239–256. ISBN 978-3-11-024583-7

BORDAG, Denisa. 2006. *Psycholinguistische Aspekte der Interferenzerscheinungen in der Flexionsmorphologie des Tschechischen als Fremdsprache*. Hildesheim : Georg Olms Verlag AG, 2006. ISBN 3487130254.

- BOYD, A. a HANA, J. et al. 2014. The MERLIN corpus : Learner Language and the CEFR. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, May 26-31, 2014*. European Language Resources Association (ELRA). 2014, s. 1281-1288. ISBN 978-2-9517408-8-4
- BOZDĚCHOVÁ, I.; ŠAŠKOVÁ-PIERCE, M. 1995. Vyučování jazyků komunikativní metodou. *Časopis pro moderní filologii, roč. 77*. 1995, s. 30-38. ISSN 0862-8459
- BURANI, C., SALMASO, D., & CARAMAZZA, A. 1984. Morphological structure and lexical access. *Visible Language 4*. 1984, s. 348–358. ISSN 0022-2224
- BYBEE, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change 14*. 2002, s. 261-290. ISSN 1469-8021.
- BYBEE, J. 2006. From usage to grammar: the mind's response to repetition. *Language*. Vol. 82, No. 4. 2006, s. 711-733. ISSN 1574-020X.
- BYBEE, J. 2008. Usage-based grammar and second language acquisition. In ROBINSON, P.; ELLIS N. C. (ed.) *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York : Routledge, 2008, s. 216-236. ISBN 0805853510.
- ČADSKÁ, M. et al. 2005. *Čeština jako cizí jazyk. Úroveň A2*. Praha : Ministerstvo školství, mládeže a tělovýchovy České republiky, 2005. ISBN 802601460X.
- ČERMÁK, F. 2006. Kolokace v lingvistice. In ČERMÁK, F.; ŠULC, M. (ed.) *Kolokace*. Praha : Nakladatelství Lidové noviny, Ústav Českého národního korpusu, 2006.
- ČERMÁK, F. 2010. *Lexikon a sémantika*. Praha : Nakladatelství lidové noviny, 2010. ISBN 978-80-7422-020-3.
- ČERMÁK, F., J. HOLUB. 2005. *Syntagmatika a paradigmatika českého slova. I Valence a kolokabilita*. Praha : Nakladatelství Karolinum, 2005.

- ČERMÁK, F.; KŘEN, M. 2005. Large Corpora, Lexical Frequencies and Coverage of Texts. In *Proceedings from the Corpus Linguistics Conference Series*, Vol. 1, no. 1. 2005, s. 1-8. ISSN 1747-939.
- ČERMÁK, F.; KŘEN, M. 2011. *A Frequency Dictionary of Czech : Core Vocabulary for Learners*. New York : Routledge, 2011. ISBN 13: 978-0-415-57661-1.
- ČERMÁKOVÁ, A. 2009. *Valence českých substantiv*. Praha : Nakladatelství Lidové noviny, 2009. ISBN 978-80-7106-426-8.
- CVEJNOVÁ, J. 2006. Úroveň B1 - Prahová úroveň. In ŠTINDLOVÁ, B.; ČEMUSOVÁ, J. (ed.) *Sborník Asociace učitelů češtiny jako cizího jazyka (AUČCJ)*, 2005-2006. Akropolis, 2006, s. 27-38. ISBN 80-86903-37-0.
- CVRČEK, V. et al. 2010. *Mluvnice současné češtiny (MSC) : 1 / Jak se píše a jak se mluví*. Praha : Karolinum, 2010. ISBN 978-80-246-1743-5.
- CVRČEK, V.; RICHTEROVÁ, O. 2015. 'cnk:syn2010'. *Průručka ČNK*. [Online], 17.02.2015 [cit. 12.04.2015]. Dostupný z WWW <<http://wiki.korpus.cz/doku.php?id=cnk:syn2010&rev=1424167381>>.
- CVRČEK, V.; RICHTEROVÁ, O. 2013. 'pojmy:reprezentativnost'. *Průručka ČNK* [Online]. 13.09.2013 [cit. 12.04.2015]. Dostupný z WWW <<http://wiki.korpus.cz/doku.php?id=pojmy:reprezentativnost&rev=1379079160>>.
- CVRČEK, V.; RICHTEROVÁ, O. 2015 (ed.). *Průručka ČNK* [Online]. 07.04.2015 [cit. 11.04.2015] Dostupný z WWW <<http://wiki.korpus.cz/doku.php?id>manualy:kontext:index&rev=1428418894>>.
- CVRČEK, V. 2013. *Kvantitativní analýza kontextu*. Praha : Nakladatelství Lidové noviny, 2013. ISBN 978-80-7422-264-1.
- DANEŠ, F.; HLAVSA, Z.; GREPL, M. 1987. *Mluvnice češtiny : 3 Skladba*. 1. vyd. Praha : Academia, 1987.
- DÍAZ-NEGRILLO, A.; FERNÁNDEZ-DOMÍNGUEZ, J. 2006. Error-coding systems for learner corpora. *Revista Española de Lingüística Aplicada* 19. 2006, s. 83-102. ISSN 0213-2028
- EK, J.A.; Trim, J.L.M. 1980. *Threshold level*. Strasbourg : Council of Europe. Council for Cultural Co-operation, 1980. ISBN 978-0080245881.

- EK, J.A. 1986. *Objectives for Foreign Language Teaching*. Strasbourg : Council of Europe, 1986. ISBN 9287109206.
- ELLIS, N.C. 2002. Frequency Effects in Language Processing : A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition*, Vol. 24, No. 2. 2002, s. 143-188.
- ELLIS, N.C. 2014. Frequency-based accounts of Second Language Acquisition. In GASS, S.M.; MACKEY, A. *The Routledge Handbook of Second Language Acquisition*. New York : Routledge, 2014, s. 193-210. ISBN 9780415709811.
- FIRTH, J.R. 1969. Personality and Language in Society. In FIRTH, J.R. *Papers in Linguistics 1934-51*. London : Oxford University Press, 1969.
- FRANCIS, G. 1993. A Corpus-Driven Approach to Grammar - Principles, Methods and Examples. In BAKER, M. et al. *Text and Technology : In Honour of John Sinclair*. Philadelphia, Amsterdam : John Benjamins, 1993. ISBN 978-1556194948.
- GOLDBERG, A. E. 2006. *Constructions at Work*. Oxford, New York : Oxford University Press, 2006. ISBN 0199268517.
- GOLDBERG, A.E.; CASENHISER, D. 2008. Construction learning and second language acquisition. In ROBINSON, P.; ELLIS, N.C. (ed.) *Handbook of Cognitive Linguistics and Second Language Acquisition*. London : Routledge, 2008, s. 197-215. ISBN 0805853510.
- HALLIDAY, M.A.K.; MATTHIESSEN, C. 2014. *Halliday's introduction to functional grammar*. 4. vydání. London : Routledge, 2014. ISBN 97810415826280.
- HALLIDAY, M.A.K. 1991. Corpus Studies and Probabilistic Grammar. In AIJMER, K.; ALTENBERG, B. (ed.) *English Corpus Linguistics*. London : Longman, 1991. ISSN 1610-868X.
- HANKS, P. 2000. Contributions of Lexicography an Corpus Analysis to a Theory of Language Performance. *Proceedings of Euralex 2000*. 2000. s. 3-13. Dostupný také z WWW <http://www.euralex.org/proceedings-toc/euralex_2000/>
- HOLÁ, L. a BOŘILOVÁ, P. 2009b. Prezentace učebnice Česky krok za krokem 2. In HÁDKOVÁ, M. (ed.) *Usta ad Albim BOHEMICA*. roč. IX, č.1, 2009, s. 76-

82. Dostupný také z WWW <http://www.pf.ujep.cz/files/data/KBO_casopis2009-1.pdf>. ISSN 1802-825X.

HOLÁ, L.; HANKOVÁ, D. 2014. K výuce gramatiky v češtině pro cizince. *Czech step by step* [Online]. [cit. 12.11.2014]. Dostupný z WWW <http://www.czechstepbystep.cz/clanky/k_vyuce_gramatiky.html>.

HRDLIČKA, M. 2009. *Gramatika a výuka češtiny jako cizího jazyka : K prezentaci gramatiky českého jazyka v učebnicích češtiny pro cizince*. Praha : Karolinum, 2009. ISBN 978-80-246-1527-1.

HUNSTON, S. FRANCIS, G. 2000. *Pattern Grammar : A corpus-driven approach to the lexical grammar of English*. Amsterdam : John Benjamins, 2000. ISBN 90-272-2273-8.

PANEVOVÁ, J.; ŠEVČÍKOVÁ, M. 2011. Jak se počítají substantiva v češtině : poznámky ke kategorii čísla. *Slovo a slovesnost*, roč. 72, č.3, 2011, s. 163-176. ISSN 0037-7031.

JELÍNEK, J.; BEČKA, J. V.; TĚŠITELOVÁ, M. 1961. *Frekvence slov, slovních druhů a tvarů v českém jazyce*. Praha : Státní pedagogické nakladatelství, 1961.

JELÍNEK, T.; PETKEVIČ, V. 2011. Systém jazykového značkování korpusů současné psané češtiny. In PETKEVIČ, V.; ROSEN, A. (ed.). *Korpusová lingvistika Praha 2011 : 3 Gramatika a značkování korpusů*. Praha : Nakladatelství Lidové noviny, 2011, s. 154-170. ISBN 978-80-7422-116-3.

JOHNSON, K. 2009. Foreign language syllabus design. In KNAPP, K.; SEIDLHOFER, B. (ed.). *Handbook of Foreign Language Communication and Learning*. De Gruyter Mouton, 2009, ISBN 978-3-11-021424-6.

KARLÍK, P. et al. (ed.) 2002. *Encyklopedický slovník češtiny (ESČ)*. Praha : Nakladatelství lidové noviny, 2002. ISBN 80-7106-484-X.

KARLÍK, P. 2000. Hypotéza modifikované valenční teorie. *Slovo a slovesnost*. 2000, Roč. 61, č. 3, str. 170-189. ISSN 0037-7031.

KARLÍK, P.; NEKULA, M.; RUSÍNOVÁ, Z. 1995. *Příruční mluvnice češtiny (PMČ)*. Praha: Nakladatelství Lidové noviny, 1995. ISBN 80-7106-134-4.

- KEMPE, V.; MACWHINNEY, B. 1998. The Acquisition of Case-Marking by Adult Learners of Russian and German. *Studies in Second Language Acquisition*, Vol. 20, Cambridge University Press, 1998, s. 543-587. ISSN 1470-1545.
- KENNEDY, Graeme. 2007. Phraseology and language pedagogy : Semantic preference associated with English verbs in the British National Corpus. In MENUIER, F.; GRANGER, S. (ed.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam : John Benjamins, 2007, s. 21-41. ISBN 978-9027-232441.
- KLEIN, W. 1990. *Second Language Acquisition*. Cambridge : Cambridge University Press, 1990. ISBN 9780521268790.
- LEWIS, M. 2002. *The Lexical Approach. The State of ELT a Way Forward*. Boston : Thomson Publishing, 2002. ISBN 0-90671-99-X.
- MAHLBERG, M. 2005. *English General Nouns. A corpus theoretical approach*. Amsterdam : John Benjamins, 2005. ISBN 9789027222916.
- MCENERY, T., XIAO, R.; YUKIO, T. 2006. *Corpus-Based Language Studies : An Advanced Resource Book*. London : Routledge, 2006. ISBN 0415286239.
- MERLIN PROJECT. 2014. *Merlin* [Online]. 2014 [cit. 27.05.2015], Annotation scheme. Dostupný z WWW <<http://commul.eurac.edu/dev/merlin/php/docs/MERLIN-annotation-scheme.pdf>>.
- MEYER, Ch.F. 2002. *English Corpus Linguistics. An introduction*. Cambridge : Cambridge University Press, 2002. ISBN 0521808790.
- PETR, J. et al. 1987. *Mluvnice češtiny 2 : Tvarosloví (MČ2)*. Praha : Academia, 1987.
- PROJEKT EUROMOBIL. 2014. *Projekt Euromobil* [Online]. 2014 [cit. 10.04.2014], Projekt. Dostupný z WWW <<http://www.euromobil.org/cestina/project.php>>.
- POLDAUF, I.; ŠPRUNK, K. 1968. *Čeština jazyk cizí. Mluvnice češtiny pro cizince*. Praha : Státní pedagogické nakladatelství, 1968.
- RAMA, J.L.; AGULLÓ, G.L. 2012. The role of grammar teaching : from Communicative approaches to the Common European Framework of Reference

for Languages. *Revista de Lingüística y Lenguas Aplicadas*, Vol. 7, 2012, s. 179-191. Dostupný také z WWW <<http://ojs.upv.es/index.php/rdlyla/issue/view/137>>. ISSN 1886-6298.

REZNICEK, M.; LÜDELING, A.; HIRSCHMANN, H. 2013. Competing target hypotheses in the Falko corpus : A flexible multi-layer corpus architecture. In DÍAS-NEGRILLO, A. et al. (ed.) *Automatic treatment and Analysis of Learner Corpus Data*. 2013, s. 101-123. ISBN 9789027203663.

ROBINSON, P.; ELLIS, N.C. 2008. Conclusion - Cognitive Linguistics, Second language Acquisition and L2 instruction : issues for research. In ROBINSON, P.; ELLIS, N.C. (ed.) *Handbook of Cognitive Linguistics and Second Language Acquisition*. London : Routledge, 2008, s. 489-545. ISBN 0805853510.

SCHMID, S. 1997. The Naturalness Differential Hypothesis : Cross-linguistic Influence and Universal Preferences in Interlanguage Phonology and Morphology. *Folia Linguistica*. XXX13-4, 1997, s. 331–348. ISSN 0165-4004.

SCHÖNE, K. 2011. Úvahy o prezentaci substantivního skloňování v učebnicích češtiny z frekvenčního hlediska. In ČERMÁK, F. (ed.). *Korpusová lingvistika Praha 2011*. Praha : Nakladatelství Lidové noviny, 2011. ISBN: 97880-7422-115-6.

ŠEBESTA, K. 2012. Parametry žakovských korpusů a CzeSL. In ŠEBESTA, K.; ŠKODOVÁ, S. et al. *Čeština – cílový jazyk a korpusy*. Liberec, 2012, s. 13-34. ISBN 978-80-7372-848-9.

SELINKER, L. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching*, Vol. 10, no.3. De Gruyter Mouton, 1972, s. 31-54. ISSN 1613-4141.

RADA EVROPY. COUNCIL FOR CULTURAL CO-OPERATION. MODERN LANGUAGES DIVISION. 2006. *Společný evropský referenční rámec pro jazyky : Jak se učíme jazykům, jak je vyučujeme a jak v jazycích hodnotíme (SERR)*. Překlad Ivanová, J. et al. 2. české vydání. Olomouc : Univerzita Palackého, 2006. ISBN 80-244-1425-2.

SINCLAIR, J. 1991. *Corpus, concordance, collocation*. Oxford : Oxford University Press, 1991. ISBN 9780194371445.

- SINCLAIR, J. 2004. *Trust the text. Language, corpus and discourse*. London : Routledge, 2004. ISBN: 0-415-31768-1.
- SINCLAIR, J.; RENOUF, A. 1988. A lexical syllabus for language learning. In CARTER, R.; McCARTHY, M. *Vocabulary and Language Teaching*. Longman, 1988, s. 140-158. ISBN 0582553822.
- ŠINDELÁŘOVÁ, J. 2010. Čeština jako cizí jazyk dle SERR. *Metodický portál : Články* [Online]. 15.03.2010 [cit. 2015-04-14]. Dostupný z WWW <<http://clanky.rvp.cz/clanek/c/Z/8143/CESTINA-JAKO-CIZI-JAZYK-DLE-SERR.html>>. ISSN 1802-4785.
- SKOUMALOVÁ, H. 2011. Porovnání úspěšnosti tagování korpusu. In PETKEVIČ, V.; ROSEN, A. (ed.) *Korpusová lingvistika Praha 2011 : 3 Gramatika a značkování korpusů*. Praha : Nakladatelství Lidové noviny, 2011, s. 199-207. ISBN 978-80-7422-116-3.
- ŠONKOVÁ, J. 2008. *Morfologie mluvené češtiny: Frekvenční analýza*. místo neznámé : Nakladatelství Lidové noviny, 2008. ISBN 978-80-7106-956-0.
- ŠTINDLOVÁ, B. et al. 2014. CzeSL – an error tagged corpus of Czech as a second language. *Language Resources and Evaluation*. Vol. 48, Issue 1. 2014, s. 65-92. ISSN 1574-020X.
- ŠTINDLOVÁ, B. 2012. Chybové taxonomie a možnosti chybové anotace v žákovských korpusech. In ŠEBESTA, K.; ŠKODOVÁ, S. *Čeština – cílový jazyk*. Liberec, 2012, s. 35-60. ISBN 978-80-7372-848-9.
- TĚŠÍTELOVÁ, M. 1985. *Kvantitativní charakteristiky češtiny*. Praha : Academia, 1985.
- TĚŠÍTELOVÁ, M. 1990. On Semantics of Nouns from the Quantitative Point of View. *Prague Studies in Mathematical Linguistics*. Vol. 10, 1990, str. 9-24. ISSN 0165-7712.
- TĚŠÍTELOVÁ, M. 1980. *Využití statistických metod v gramatice*. Praha : Academia, 1980.
- THORNBURY, S. 1999. *How to teach grammar*. Harlow : Longman, 1999. ISBN 978-0582339323.

TOGNINI-BONELLI, E. 2001. *Corpus Linguistics at work*. Amsterdam : John Benjamins, 2001, ISBN 1588110613.

VALIŠOVÁ, P. 2013. Učebnicový korpus a jeho využití pro výuku češtiny jako cizího jazyka. In *Gramatika a korpus 2012 4. mezinárodní konference, Praha, 28.-30. listopadu 2012*. Hradec Králové: Gaudeamus, 2013. s. 1-8, ISBN 978-80-7435-243-0.

WISNIEWSKI, K.; et al. 2013. MERLIN: An online trilingual learner corpus empirically grounding the European Reference Levels in authentic learner data. In *ICT for Language Learning 2013, Conference Proceedings*. Florence : Libreriauniversitaria.it. Edizioni, 2013. ISBN: 978-88-6292-423-8.

WRAY, Alison. 2002. *Formulaic language and the lexicon*. Cambridge : Cambridge University Press, 2002. ISBN 9780521022125.

Použité a citované korpusy a slovníky

ČERMÁK, F., ADAMOVIČOVÁ, A.; PEŠIČKA, J. 2001. *Pražský mluvený korpus (PMK) : přepisy nahrávek pražské mluvy z 90. let 20. století* [online]. Ústav Českého národního korpusu FF UK. [cit. 2011-04-01]. Dostupný z WWW <<http://www.korpus.cz>>.

ČERMÁK, F. et al. 2005. *SYN2005 : žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2005. Dostupný z WWW <<http://www.korpus.cz>>.

KŘEN, M. et al. 2010. *SYN2010 : žánrově vyvážený korpus psané češtiny* [Online]. Ústav Českého národního korpusu FF UK. Praha 2010. Dostupný z WWW <<http://www.korpus.cz>>.

PROJECT MERLIN. *MERLIN - Multilingual Platform for European Reference Levels : Interlanguage Exploration in Context*. 2014 [cit. 28.04.2015]. Dostupný z WWW <<http://merlin.platform.eu>>.

ŠEBESTA, K. et al. 2014. *CZESL-PLAIN : akviziční korpus psané češtiny, zvl. přepisů písemných projevů nerodilých mluvčích, verze 2 z 22. 1. 2014*. Ústav Českého národního korpusu FF UK, Praha 2012. Dostupný z WWW <<http://www.korpus.cz>> a <<http://akces.ff.cuni.cz>>.

WACLAWIČOVÁ, M., et al. 2008. *ORAL2008 : sociolingvisticky vyvážený korpus neformální mluvené češtiny* [Online]. Ústav Českého národního korpusu FF UK. Praha 2008. Dostupný z WWW <<http://www.korpus.cz>>.

Pravice. In ÚSTAV PRO JAZYK ČESKY *Internetová jazyková příručka : Slovník spisovného jazyka českého (SSJČ)* [Online]. Akademie věd ČR, Ústav pro jazyk český. 2015 [cit. 28.4.2015]. Dostupný z WWW <<http://prirucka.ujc.cas.cz/?slovo=pravice&Hledej=Hledej>>

Szjed. In ÚSTAV PRO JAZYK ČESKY *Internetová jazyková příručka : Slovník spisovného jazyka českého (SSJČ)* [Online]. Akademie věd ČR, Ústav pro jazyk český. 2015 [cit. 28.4.2015]. Dostupný z WWW <<http://prirucka.ujc.cas.cz/?slovo=szjed&Hledej=Hledej>>

Analyzované učebnice češtiny pro cizince

AMOROVÁ, V. *Vítáme vás! Ein Tschechischlehrwerk für Anfänger*. Hueber, 1998. ISBN 3190051577.

CONFORTIOVÁ, H.; CVEJNOVÁ, J.; ČADSKÁ, M. 2003. *Učebnice češtiny pro výuku v zahraničí. I. a II. díl*. Praha : Karolinum, 2003. ISBN 978-80-246-0530-2.

HASIL, J.; HÁJKOVÁ, E.; HASILOVÁ, H. 2007. *Brána jazyka českého otevřená*. Praha : Karolinum, 2007. ISBN 9788024613376.

HIRSCHMANN, Irena. 2011. *Lehrbuch der tschechischen Sprache I*. Hamburg : Buske, 2011. ISBN: 3875482352.

HOLÁ, L. a BOŘILOVÁ, P. 2009a. *Česky krok za krokem 2*. Praha : Akropolis, 2009. ISBN 978-80-86903-92-7.

HOLÁ, L. a BOŘILOVÁ, P. 2010. *Čeština expres 1 (A1/1)*. Praha : Akropolis, 2010. ISBN 978-80-87310-13-7.

HOLÁ, L. a BOŘILOVÁ, P. 2011. *Čeština expres 2. (A1/2)*. Praha : Akropolis, 2011. ISBN 9788087481271.

HOLÁ, L. a BOŘILOVÁ, P. 2013. *Čeština expres 3. (A2/1)*. Praha : Akropolis, 2013. ISBN 978-80-7470-032-3.

HOLÁ, Lída. 2006. *Tschechisch Schritt für Schritt*. Praha : Akropolis, 2006. ISBN: 978-80-86903-35-4.

MAIDLOVÁ, J.; NEKULA, M. 2007. *Tschechisch kommunikativ 1*. Stuttgart : Schmetterling, 2007. ISBN 3-89657-860-X.

MAIDLOVÁ, J.; NEKULA, M. 2013. *Tschechisch kommunikativ 2*. 2. vydání. Stuttgart : Schmetterling, 2013. ISBN 3-89657-864-2.

NEKOVÁŘOVÁ, A. 2010. *Čeština pro život*. Praha : Akropolis, 2010. ISBN 978-80-86903-80-4.

REMEDIOSOVÁ, H.; ČECHOVÁ, E.; PUTZ, H. 2002. *Wollen Sie Tschechisch sprechen? 1. Teil : Chcete mluvit česky? 1. díl*. 6. vydání. Liberec : Harry Putz, 2002. ISBN 80-902165-9-5.

SODEYFI, H.; NEWERKLA, S. 2002. *Tschechisch. Faszination der Vielfalt*. Wiesbaden : Harrasowitz, 2002. ISBN 978-3-447-05084-5.

SÝKOROVÁ, B.; DAVIDOVÁ, H. 2013. *Cvičebnice češtiny, čeština jako cizí jazyk*. Polyglot, 2013. 978-80-86195-60-5.

Příloha

1. Výběr substantiv

1.1 Komunikační úlohy a tematické okruhy

Přehled oblastí užívání jazyka, přiřazených komunikačních úloh a tematických okruhů podle *Prahové úrovně (PrÚ)*

Oblasti užívání jazyka identifikované v dotazníkovém šetření	Komunikační úlohy a cíle podle tzv. <i>Rozšířené charakteristiky</i>¹⁷² v <i>Prahové úrovni</i>	Přiřazené tematické okruhy z PrÚ (tzv. <i>specifické pojmy</i>)	Názvy <i>otevřených položek</i>¹⁷³
obstarání jídla (nakupování, výběr v menze, objednávání v restauraci)	1.3 stravování 2. vyjadřovací schopnosti v oblasti společenského styku se zvláštním ohledem na: záliby, ceny, jídlo a pití	3.1 každodenní život - doma 9.1 druhy obchodů 9.2 nakupování 9.3 potraviny 9.5 kuchyňské předměty 10.1 jídlo a pití 10.2 stravování mimo domov	druhy drůbeže druhy pečiva druhy masa druhy ovoce druhy ryb druhy zeleniny názvy obchodů přibližné jednotky objemu výrobky z masa
domluva s místními lidmi (hlavně mimo hl. města se zaměstnanci dopravních a jiných veřejných služeb, orientace ve veřejném prostoru)	1.5 Veřejná doprava 1.6 soukromá doprava 1.7 využití informačních služeb 1.8 návštěva veřejných zařízení 4. student je obeznámen s důležitými společenskými konvencemi (jazykovými), např. způsob oslovování, ukončení kontaktu, střídání	2.5 (dům a domácnost) služby 9.2 nakupování 11.1 Pošta 11.2 Telefon 11.4 banka 11.5 policie 11.8 autoservis	druhy představení kulturní zařízení názvy obchodů přibližné jednotky objemu

¹⁷² Pro podrobný popis komunikačních cílů v *Prahové úrovni* viz kapitola 3 „Rozšířená charakteristika“ (Bischofová, 2001, str. 3).

¹⁷³ Je třeba poznamenat, že seznam *názvů otevřených položek* postrádá podle našeho mínění některé pojmové kategorie. Nezachycuje zvláště např. tvary a součásti krajiny, jídla a další potraviny (jsou zde uvedeny jenom jednotlivé součásti nikoli však výrobky potravinářské) nebo technické vybavení a komunikační prostředky.

Oblasti užívání jazyka identifikované v dotazníkovém šetření	Komunikativní úlohy a cíle podle tzv. <i>Rozšířené charakteristiky</i> ¹⁷² v <i>Prahové úrovni</i>	Přiřazené tematické okruhy z PrÚ (tzv. <i>specifické pojmy</i>)	<i>Názvy otevřených položek</i> ¹⁷³
	v dialogu atd.	11.9 benzinová stanice 11.10 další druhy služeb 13.1 porozumění a vyjadřování	
orientace ve městě	1.5 veřejná doprava 1.6 soukromá doprava 1.7 využití informačních služeb 1.11 hledání cesty	1.2 adresa 5.1 veřejná doprava obecně 5.3 vlaková doprava 5.5 taxislužba 5.6 soukromá doprava 5.7 silniční provoz 9.1 druhy obchodů 10.2 stravování mimo domov 12.1 hledání cesty	dopravní prostředky názvy obchodů
nakupování všeobecně	1.4 nákupy, spotřební zboží 1.9 použití veřejných služeb 1.9.1 pošta 1.9.3 banky 2. vyjadřovací schopnosti v oblasti společenského styku se zvláštním ohledem na: záliby, zboží, nákupy, ceny, jídlo a pití	9.1 druhy obchodů 9.2 nakupování 9.3 potraviny 9.4 oblečení, móda 9.6 kouření 9.7 placení 11.1 pošta 11.4 banka	názvy obchodů oblečení a obuv doplňky názvy mincí a bankovek přibližné jednotky objemu dílové číslovky
cestování	1.1.1 vstup do země 1.1.2 celní záležitosti 1.1.3 bezpečnostní složky 1.1.4 policie, dopravní policejní služba 1.2.1 krátkodobé ubytování 1.3 stravování 1.5 veřejná doprava 1.6 soukromá doprava	1 osobní údaje 2.1 typy ubytování 2.7 okolí 2.8 flóra a fauna 4.1 volný čas a zábava 4.7 sport 5.1 veřejná doprava obecně 5.2 letecká doprava	dopravní prostředky jména hlavních měst jména kontinentů jména národností jména zemí názvy jazyků typy ubytování historické lokality a místa turistického

Oblasti užívání jazyka identifikované v dotazníkovém šetření	Komunikativní úlohy a cíle podle tzv. <i>Rozšířené charakteristiky</i> ¹⁷² v <i>Prahové úrovni</i>	Přiřazené tematické okruhy z PrÚ (tzv. <i>specifické pojmy</i>)	Názvy <i>otevřených položek</i> ¹⁷³
	1.7 využití informačních služeb 2. vyjadřovací schopnosti v oblasti společenského styku se zvláštním ohledem na: cestování, navštívená místa, oblasti, pamětihodnosti, počasí	5.3 vlaková doprava 5.4 lodní doprava 5.5 taxislužba 5.6 soukromá doprava 5.7 silniční provoz 5.8 prázdniny, dovolená 5.9 ubytování 5.10 zavadla 5.11 vstup do země a její opuštění 5.12 cestovní doklady 14.1 počasí a podnebí	zájmu jména měsíců vybavení bytu místnosti
pochopení místního způsobu života	4. student je obeznámen s důležitými společenskými konvencemi (jazykovými i nejjazykovými) 6. Student je do určité míry obeznámen s charakteristickými rysy kultury v zemi, kde je daný jazyk užíván, zvláště s těmi, které se týkají: každodenního života, životních podmínek, mezilidských vztahů, hlavních hodnot a postojů		kulturní zařízení
studium	1.10 školská a vzdělávací zařízení 1.10.1 v roli studenta 2. vyjadřovací schopnosti v oblasti společenského styku se zvláštním ohledem na: vzdělání, zájmy, jazyky, učení jazykům a obtíže s nimi spojené	1.10 zaměstnání 3.3 3.1 každodenní život - příjem 3.4 3.1 každodenní život -vzdělání 3.5 3.1 každodenní život - plány 4.2 koníčky a zájmy 8.1 škola, vzdělání, výchova 8.2 vyučovací předměty 8.3 kvalifikace, zkoušky 13.1 jazyky, porozumění a vyjadřování	společenské vědy názvy předmětů názvy pracovišť názvy povolání

Oblasti užívání jazyka identifikované v dotazníkovém šetření	Komunikativní úlohy a cíle podle tzv. <i>Rozšířené charakteristiky</i> ¹⁷² v <i>Prahové úrovni</i>	Přiřazené tematické okruhy z PrŮ (tzv. <i>specifické pojmy</i>)	Názvy <i>otevřených položek</i> ¹⁷³
hledání ubytování	1.2.2 ubytování při dočasném pobytu	2.1 typy ubytování 2.2 prostory 2.3 zařízení bytu 2.4 náklady na byt 2.5 služby (dům a domácnost) 2.6 vybavení a úklid	názvy místností vybavení bytu

1.2 Náplň názvů otevřených položek

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i> ¹⁷⁴	názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i>	názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1a 2</i>
<i>dílové číslovky</i>	<i>čtvrt</i> <i>půlka</i> <i>třetina</i> <i>čtvrtka</i>	<i>druhy zeleniny</i>	<i>brambory</i> <i>čočka</i> <i>fazole</i> <i>hlávkový salát</i> <i>hrášek</i> <i>květák</i> <i>mrkev</i> <i>okurka</i> <i>paprika</i> <i>rajčata</i> <i>špenát</i> <i>zelí</i> <i>brokolice</i> <i>česnek</i>	<i>názvy mincí a bankovek</i>	<i>bankovka</i> <i>tisícovka</i>

¹⁷⁴ Holá, 2006; Holá a Bořilová, 2009

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i> ¹⁷⁴
doplňky	<i>hodinky kabelka peněženka taška šála šátek kravata šperky batoh náušnice pásek</i>
dopravní prostředky	<i>auto kolo motorka nákladní auto skútr áčko céčko letadlo loď taxík tramvaj vlak</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i>
	<i>pórek čočka</i>
historické lokality a místa turistického zájmu	<i>hrad zámek katedrála chrám skanzen zoo věž</i>
jídla a další potraviny	<i>bramborový salát bramborová polévka gulášová polévka hovězí polévka knedlík kroketa / krokety kuřecí polévka majonéza omáčka palačinka roláda svíčková omáčka tatarská omáčka tuk vajíčko těstoviny</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1a 2</i>
názvy místností	<i>koupelna ložnice kuchyně obývací pokoj záchod toaleta komora lodžie loft předsíň sklep</i>
názvy obchodů	<i>potraviny pekařství cukrárna řeznictví uzenářství ovoce a zelenina papírnictví drogerie oděvy tabák obuv supermarket trafika bižuterie čistírna hodinářství</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i> ¹⁷⁴
druhy drůbeže	kuře kachna husa krůta

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i>
	<i>tvaroh</i>
jména hlavních (a velkých měst) měst	Benátky Berlín Bratislava Budapešť Drážďany Kodaň Lipsko Londýn Mnichov Moskva Paříž Praha Řím Varšava Víděň Ženeva

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1a 2</i>
	<i>krejčovství</i> <i>opravna</i> <i>optika</i> <i>knihkupectví</i>
názvy povolání	dělník, dělnice doktor, doktorka ekonom, ekonomka fotograf, fotografka hasič, hasička herec, herečka historik, historička hudebník, hudebnice kadeřník, kadeřnice lékař, lékařka malíř, malířka manažer, manažerka novinář, novinářka obchodník, obchodnice podnikatel, podnikatelka policista, policistka politik, politička právník, právnička ředitel, ředitelka režisér, režisérka skladatel, skladatelka soudce, soudkyně učitel, učitelka účetní (m), účetní (f) umělec, umělkyně

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i> ¹⁷⁴
druhy masa	<i>telecí maso</i> <i>hovězí maso</i> <i>vepřové maso</i> <i>skopové maso</i>
druhy nápojů	<i>džus</i> <i>minerálka</i> <i>víno</i> <i>pivo</i> <i>voda</i> <i>čaj</i> <i>Dobrá voda</i> <i>kakao</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i>
jména kontinentů	<i>Afrika</i> <i>Amerika</i> <i>Asie</i> <i>Austrálie</i> <i>Evropa</i>
jména měsíců	<i>leden</i> <i>únor</i> <i>březen</i> <i>duben</i> <i>květen</i> <i>červen</i> <i>červenec</i> <i>srpen</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1a 2</i>
	<i>úředník, úřednice</i> <i>voják, vojačka</i> <i>zemědělec, zemědělkyně</i> <i>zpěvačka, zpěvák</i>
názvy předmětů	<i>český jazyk</i> <i>literatura</i> <i>historie</i> <i>dějepis</i> <i>zeměpis</i> <i>matematika</i> <i>fyzika</i> <i>chemie</i> <i>biologie</i> <i>tělesná výchova</i> <i>hudební výchova</i> <i>výtvarná výchova</i> <i>dějiny</i> <i>přírodopis</i> <i>tělocvik</i> <i>účetnictví</i>
názvy sportu a her	<i>fotbal</i> <i>hokej</i> <i>stolní tenis</i> <i>tenis</i> <i>lyžování</i> <i>plavání</i> <i>golf</i> <i>dostihy / dostih</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i> ¹⁷⁴
	<i>kola</i> <i>koňak</i> <i>mléko</i> <i>šampaňské</i>
druhy ovoce	<i>banán</i> <i>broskev</i> <i>citrón / citron</i> <i>grapefruit</i> <i>hroznové víno</i> <i>hruška</i> <i>jablko</i> <i>jahody</i> <i>meruňka</i> <i>pomeranč</i> <i>ořechy</i>
druhy pečiva	<i>toastový chléb</i> <i>celozrnný chléb</i> <i>rohlík</i> <i>houska</i> <i>veka</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i>
	<i>září</i> <i>říjen</i> <i>listopad</i> <i>prosinec</i>
jména národností	<i>Čech, Češka</i> <i>Evropan, Evropanka</i> <i>Američan, Američanka</i> <i>Němec, Němka</i> <i>Francouz, Francouzka</i> <i>Rus, Ruska</i> <i>Angličan, Angličanka</i> <i>Ital, Italka</i> <i>Slovák, Slovenka</i> <i>Rakušan, Rakušanka</i> <i>Japonec, Japonka</i> <i>Číňan, Číňanka</i> <i>Polák, Polka</i> <i>Španěl, Španělka</i> <i>Řek, Řekyně</i> <i>Švéd, Švédka</i> <i>Švýcar, Švýcarka</i> <i>Kanaďan, Kanaďanka</i> <i>Maďar, Maďarka</i>
jména zemí	<i>Česká republika</i> <i>Evropa</i> <i>Spojené státy americké / USA</i> <i>Spolková republika Německo</i> <i>Francie</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1a 2</i>
	<i>kulečník</i> <i>šipky</i> <i>kolečkové brusle</i> <i>rodeo</i> <i>balet</i>
oblečení a obuv	<i>bota</i> <i>šaty</i> <i>kalhoty</i> <i>košile</i> <i>kabát</i> <i>plášť</i> <i>sukně</i> <i>oblek</i> <i>bunda</i> <i>oděv</i> <i>rukavice</i> <i>čepice</i> <i>sako</i> <i>ponožka</i> <i>tričko</i>
přibližné jednotky objemu	<i>balík</i> <i>baliček</i> <i>hrnek</i> <i>hrneček</i> <i>sklenice</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i> ¹⁷⁴	názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i>	názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1a 2</i>
	<i>bageta</i> <i>cukroví</i> <i>dort</i> <i>sušenka</i> <i>buchta</i> <i>vánočka</i>		<i>Rusko</i> <i>Velká Británie / Anglie</i> <i>Itálie</i> <i>Slovensko</i> <i>Rakousko</i> <i>Japonsko</i> <i>Čína</i> <i>Polsko</i> <i>Španělsko</i> <i>Řecko</i> <i>Švédsko</i> <i>Švýcarsko</i> <i>Kanada</i> <i>Maďarsko</i>		<i>sklenička</i> <i>šálek</i> <i>talíř</i> <i>láhev</i> <i>kelímek</i>
druhy představení	<i>balet</i> <i>koncert</i> <i>film</i> <i>muzikál</i> <i>opera</i> <i>revue</i> <i>varieté</i>	kulturní zařízení	<i>kino</i> <i>divadlo</i> <i>kabaret</i> <i>noční klub</i> <i>cirkus</i> <i>diskotéka</i> <i>výstava</i>	společenské vědy	<i>filozofie</i> <i>ekonomie</i> <i>sociologie</i> <i>právo</i> <i>psychologie</i>
druhy ryb	<i>kapr</i> <i>pstruh</i> <i>losos</i> <i>makrela</i> <i>treska</i>	názvy jazyků	<i>čeština</i> <i>angličtina</i> <i>francouzština</i> <i>němčina</i> <i>španělština</i> <i>italština</i> <i>ruština</i> <i>polština</i>	typy ubytování	<i>hotel</i> <i>ubytovna</i> <i>penzion</i> <i>kemp</i> <i>stan</i> <i>chatka</i> <i>chata</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i> ¹⁷⁴
výrobky z masa	<i>salám</i> <i>šunka</i> <i>párek</i> <i>klobása</i> <i>biftek</i> <i>guláš</i> <i>řízek</i> <i>slanina</i> <i>buřt</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1 a 2</i>
	<i>hindština</i> <i>japonština</i> <i>portugalština</i> <i>spisovná čeština</i> <i>obecná čeština</i>
základní číslovky	<i>nula</i> <i>jednička</i> <i>dvojka</i> <i>trojka</i> <i>čtyřka</i> <i>pětka</i> <i>šestka</i> <i>sedmička</i> <i>osmička</i> <i>devítka</i> <i>desítka</i> <i>dvacítká</i> <i>třicítka</i> <i>čtyřicítka</i>

názvy otevřených položek	substantiva ze slovníkové části učebnic <i>Česky krok za krokem 1a 2</i>
názvy příbuzenských vztahů	<i>dítě</i> <i>otec</i> <i>matka</i> <i>syn</i> <i>táta</i> <i>rodič</i> <i>máma</i> <i>bratr, brácha</i> <i>babička</i> <i>manžel</i> <i>dcera</i> <i>manželka</i> <i>maminka</i> <i>sestra</i> <i>mamka</i> <i>děda</i> <i>tatínek</i> <i>teta</i> <i>taťka</i> <i>potomek</i> <i>dědeček</i> <i>strejda</i> <i>ségra</i> <i>strýc</i>

1.3 Časté kolokáty sloves a slovesných vazeb vystihující specifické pojmy *Prahové úrovně*

SPECIFICKÉ POJMY	SLOVESA A SLOVESNÉ VAZBY	NEJFREKVENTOVANĚJŠÍ KOLOKÁTY, JEŽ NEJSOU OBSAŽENÉ V ZÁKLADNÍM SLOVNÍM REJSTŘÍKU <i>PRAHOVÉ ÚROVNĚ</i>
1 osobní údaje	<i>podepsat/podepisovat, bydlet, (za)volat D, jmenovat se, narodit se, být/pocházet z G</i> <i>být z – se nezkoumá pro velké množství neodpovídajících dat</i>	<i>rod, svět, planeta, táta, právník, máma, šéf, advokát</i>
1.10 zaměstnání	<i>být N/I (chci být.../se stát)</i> <i>být N/I – se nezkoumá pro velké množství neodpovídajících dat</i>	viz 3.5
1.2 adresa	<i>bydlet kde</i>	<i>kousek, domek, čtvrť, Londýn, kolej, New York, podnájem, sousedství, okolí, Praha, centrum, vila, panelák, barák, obce, Brno, budova, palác, sídliště, střecha, předměstí, blízkost, panelový dům, okraj, Paříž, Moskva, činžák, Vinohrady, místnost, domácnost, internát, podkrovní</i>
10.1 jídlo a pití	<i>jíst/najíst se, pít/napít se, snídat/nasnídat se, obědvat/naobědvat se, večeřet/navečeřet se, dusit, grilovat, loupat, krájet, míchat, mixovat, mlít, péct</i>	<i>sendvič, potraviny, strava, sušenky, pizza, trouba</i>
10.2 stravování mimo domov	<i>obsloužit/obsluhovat, vybrat (si)/vybírat (si), dát(si)/dávat(si), objednat(si)/objednávat(si)</i>	<i>panák, sklenka, whisky, sendvič, doušek, bar, brandy, pizza, šluk, hlt, možnost</i>
11.1 Pošta	<i>vyzvednout/vyzvedávat</i>	-
11.10 další druhy služeb	<i>dám si - opravit, vyčistit, ostříhat, vyprat</i>	-
11.2 Telefon	<i>vytočit/vytáčet</i>	-
11.4 banka	<i>(vy)měnit, proplatit/proplácet, půjčit/půjčovat (si)</i>	<i>zámek dolary</i>
11.5 policie	<i>oznámit/oznamovat, ztratit/ztrácet</i>	<i>novina, příchod, zvěst, ostatní, dav, stop, orientace, polovina, zbytek</i>
11.8 autoservis	<i>opravit/opravovat</i>	-
11.9 benzinová stanice	<i>(na)tankovat, dolít olej, nahustit pneumatiku, (z)kontrolovat</i>	-
12.1 hledání cesty	<i>dostat se kam, jít (chodit) / jet (jezdit) + na, do, k, kolem, před, po</i>	<i>střed, sever, jih, západ, vězení, park, nebe, krám, klášter, hala, akce, salon, komora, kabina, biograf, synagoga, bar, bazén, bod, brána, branka, budova, chodba, chodník, cíl, dvůr, gynekolog, hala, krajnice, mše, nákup, okraj, oltář, pláž, ples, plot, pobřeží, přepážka, příjímání, pult, rande, schodiště, schůzka, setkání, sever, soud, stánek, studna, svět, večírek, veranda, vůz, zábradlí, zeď, zrcadlo</i>

13.1 jazyky, porozumění a vyjadřování	<i>zeptat se G na A, přeložit/překládat z G do G, vyslovit/vyslovovat (na)psát A, (po)rozumět D</i>	-
14.1 počasí a podnebí	-	-
2.1 typy ubytování	<i>být na prodej, být na pronajmutí pronajmout si, pronajímat si A stěhovat se kam</i>	-
2.5 (dům a domácnost) služby	<i>nefungovat (topení), zapnout/zapínat, vypnout/vypínat rozsvítit, zhasnout</i>	<i>baterka, lampička, vypínač, chodba, svíčka, lustr</i>
2.6 vybavení a úklid	<i>uklidit, uklízet mýt, prát</i>	<i>nádobí, dřez, prádlo</i>
2.7 okolí	-	-
2.8 flóra a fauna	-	-
3.1 každodenní život - doma	<i>vařit, připravit/připravovat, (u)dělat</i>	<i>sendviče, piknik</i>
3.3 každodenní život - příjem	<i>vydělat/vydělávat (si) A</i>	<i>živobytí, prachy, dolary, jmění</i>
3.4 každodenní život - vzdělání	viz specifické pojmy 8	viz specifické pojmy 8
3.5 každodenní život - plány	<i>stát se I, být N/I, (vy)studovat A, ucházet se o A</i>	<i>spisovatel, hrdina, mistr, šéf, malíř, majitel, společník, kněz, osud, nástupce, modelka, prezidentka, dědic, papež, úmysl, biskup, ministr, vzor, medicína, fakulta, Tóra, možnost, Harvard</i>
4.1 volný čas a zábava	<i>odpočinout si/odpočívat</i>	<i>klid</i>
4.2 koníčky a zájmy	<i>věnovat se D zajímat se o+A</i>	<i>výzkum, činnost, řízení, psaní, studium, případ</i>
4.7 sport	<i>sportovat, (za)hrát si, sledovat, závodit o, vyhrát/vyhrávat A, prohrát/prohrávat A</i>	<i>šachy, schovávaná, honěná, bitva, loterie, soutěž, zápas, sázka, boj, šance, spor, jackpot</i>
5.1 veřejná doprava obecně	<i>jet/jezdit I kam, cestovat I kam, zastavit A kde, nastoupit do, vystoupit z, přestoupit z/do, vyprovodit A kam</i>	<i>Londýn New York, Paříž, Skotsko, Řím, Moskva, Palestina, Evropa, Washington, Řecko, Indie, Kodaň, střed, sever, jih, západ, lov, nebe, klášter, synagoga, bazén, bod, brána, branka, budova, cíl, dvůr, lov, okraj, pláž, pobřeží, přijímání, sever, svět</i>
5.10 zavadla	<i>odnést/odnášet</i>	-
5.11 vstup do země a její opuštění	<i>dovézt/dovážet, proclít/proclívat</i>	-
5.12 cestovní doklady	-	-
5.2 letecká doprava	<i>letět/létat I kam, odletět, přiletět</i>	<i>výška, vesmír, pilot, vrtulník, sever, nebe, balon, svět, obloha, Měsíc, raketa, Venuše, křídla, jih, pondělí, Země, hladina, prostor, Videaň</i>
5.3 vlaková doprava	<i>zmeškat spojení do, viz též 5.1</i>	-
5.4 lodní doprava	-	-
5.5 taxislužba	-	-

5.6 soukromá doprava	<i>řídít A, jet/jezdit čím</i>	<i>vůz, dodávka, stop, kočár</i>
5.7 silniční provoz	-	-
5.8 prázdniny, dovolená	<i>navštívit/navštěvovat</i>	<i>příbuzní přednášky</i>
5.9 ubytování	<i>zamluvit/zamlouvat, rezervovat, nechat/nechávat, zapsat/zapisovat se kam</i>	-
8.1 škola, vzdělání, výchova	<i>(vy)studovat A, (na)učit A, (na)učit se A</i>	<i>medicína, fakulta, Tóra, možnost, Harvard, řemeslo, řeč, dětství</i>
8.2 vyučovací předměty	-	-
8.3 kvalifikace, zkoušky	<i>(u)dělat A</i>	<i>příliš velké množství dat</i>
9.1 druhy obchodů	-	-
9.2 nakupování	<i>koupit, kupovat jít nakoupit, nakupovat být na řadě přát si ukázat, ukazovat (za)balit (vy)měnit vybrat, vybírat (si) A</i>	<i>pozemek, stánek, vůz, knížka, domek, akcie, narodeniny, klobouk, barák, prsten, jachta, los, pláště, pokrývka</i>
9.3 potraviny	-	-
9.4 oblečení, móda	<i>obléknout, oblékat (si) svléknout, svlékat (si) převléknout, převlékat (si) mít na sobě nosit zkusit, zkoušet (si) obout, obouvat (si) zout, zouvat (si)</i>	<i>bradku, brejle, bunda, dres, džíny, kalhotky, kapsa, klobouk, maska, oděv, pláště, podpatky, podprsenka, prsten, punčochy, smutek, šortky, tričko, účes, uniforma, veřejnost, znamení, župan</i>
9.5 kuchyňské předměty	-	-
9.6 kouření	<i>kouřit</i>	<i>dýmka, doutník, tráva, marihuana</i>
9.7 placení	<i>(za)platit kolik stát</i>	<i>dolar, pokuta, dluh, libro, částka, výloha, kauce, služba, nájemné, poplatek, dvojnásobek</i>

2 Základní soubor a náhodný vzorek

2.1 Náhodný stratifikovaný vzorek

<i>advokát</i>	<i>houska</i>	<i>nebezpečí</i>	<i>přízemí</i>	<i>tělocvik</i>
<i>alkohol</i>	<i>hruška</i>	<i>Němka</i>	<i>prosinec</i>	<i>teta</i>
<i>Angličan</i>	<i>hudba</i>	<i>novina</i>	<i>průkaz</i>	<i>toaleta</i>
<i>automobil</i>	<i>jednolůžko-</i>	<i>novinář</i>	<i>průmysl</i>	<i>tramvaj</i>
<i>autobus</i>	<i>vý</i>	<i>nuda</i>	<i>půlhodina</i>	<i>treska</i>
<i>banán</i>	<i>pokoj</i>	<i>nůžky</i>	<i>půllitr</i>	<i>trh</i>
<i>barva</i>	<i>jídlo</i>	<i>obchodní</i>	<i>pult</i>	<i>třicítka</i>
<i>benzín</i>	<i>kalhoty</i>	<i>dům</i>	<i>Rakušan</i>	<i>ubytovna</i>
<i>biologie</i>	<i>karavan</i>	<i>oblečení</i>	<i>rande</i>	<i>učitel</i>
<i>bradka</i>	<i>keř</i>	<i>obraz</i>	<i>recepce</i>	<i>ulice</i>
<i>Bratislava</i>	<i>klobása</i>	<i>odpočinek</i>	<i>ředitel</i>	<i>umělec</i>
<i>Brno</i>	<i>kniha</i>	<i>okno</i>	<i>řidič</i>	<i>Ústí nad</i>
<i>bunda</i>	<i>koniček</i>	<i>opravna</i>	<i>řidičák</i>	<i>Labem</i>
<i>byt</i>	<i>konkurz</i>	<i>oslovení</i>	<i>říjen</i>	<i>vajíčko</i>
<i>čaj</i>	<i>krám</i>	<i>osmička</i>	<i>řízek</i>	<i>vdova</i>
<i>čárka</i>	<i>křeslo</i>	<i>ovzduší</i>	<i>rukavice</i>	<i>večeře</i>
<i>centimetr</i>	<i>křížovka</i>	<i>pacient</i>	<i>Rus</i>	<i>vedro</i>
<i>centrum</i>	<i>květina</i>	<i>panák</i>	<i>šála</i>	<i>velikost</i>
<i>čepice</i>	<i>lampička</i>	<i>panelák</i>	<i>šálek</i>	<i>venkov</i>
<i>čeština</i>	<i>lékař</i>	<i>párek</i>	<i>sanitka</i>	<i>vesnice</i>
<i>chaos</i>	<i>letecká linka</i>	<i>pasta na</i>	<i>sešit</i>	<i>videokazeta</i>
<i>chata</i>	<i>léto</i>	<i>zuby</i>	<i>sjezd</i>	<i>vila</i>
<i>chemie</i>	<i>Liberec</i>	<i>patro</i>	<i>sklenička</i>	<i>víno</i>
<i>chléb</i>	<i>lístek</i>	<i>písnička</i>	<i>sklenka</i>	<i>voják</i>
<i>chvilé</i>	<i>lokalita</i>	<i>pití</i>	<i>školné</i>	<i>virtulník</i>
<i>chyba</i>	<i>Londýn</i>	<i>pivo</i>	<i>slečna</i>	<i>vůz</i>
<i>Čiňanka</i>	<i>losos</i>	<i>plášť</i>	<i>smlouva</i>	<i>vysvětlení</i>
<i>čtvrť</i>	<i>louka</i>	<i>podlaha</i>	<i>smysl</i>	<i>výtvarná</i>
<i>domek</i>	<i>lžička</i>	<i>podnik</i>	<i>sobotu</i>	<i>výchova</i>
<i>dům</i>	<i>manažer</i>	<i>podnikatel</i>	<i>součást</i>	<i>vzdělání</i>
<i>ekonom</i>	<i>mapa</i>	<i>pohyb</i>	<i>sousedství</i>	<i>vzkaz</i>
<i>ekonomie</i>	<i>maso</i>	<i>pokoj</i>	<i>souvislost</i>	<i>Washington</i>
<i>fakt</i>	<i>město</i>	<i>policie</i>	<i>spojení</i>	<i>zájem</i>
<i>fakulta</i>	<i>milimetr</i>	<i>politik</i>	<i>spor</i>	<i>zaměstnání</i>
<i>flaška</i>	<i>ministr</i>	<i>požadavek</i>	<i>sprcha</i>	<i>zážitek</i>
<i>fyzika</i>	<i>mír</i>	<i>práce</i>	<i>stereotyp</i>	<i>Ženeva</i>
<i>gulášová</i>	<i>mluvčí (m)</i>	<i>Praha</i>	<i>stolní tenis</i>	<i>život</i>
<i>polévka</i>	<i>muzikál</i>	<i>pravice</i>	<i>suterén</i>	<i>zločin</i>
<i>gymnázium</i>	<i>náboženství</i>	<i>předpověď</i>	<i>táta</i>	<i>zmrzlina</i>
<i>herec</i>	<i>nádobí</i>	<i>počasí</i>	<i>tatarská</i>	<i>zpráva</i>
<i>heslo</i>	<i>nádrž</i>	<i>přejezd</i>	<i>omáčka</i>	<i>ztráta</i>
<i>hlava</i>	<i>náplast</i>	<i>přesčas</i>	<i>telefonní</i>	
<i>hotel</i>	<i>návrh</i>	<i>příběh</i>	<i>seznam</i>	

2.2 Základní soubor

<i>adresa</i>	<i>biftek</i>	<i>čas</i>	<i>Čína</i>
<i>áčko</i>	<i>biograf</i>	<i>časopis</i>	<i>Čínan</i>
<i>advokát</i>	<i>biologie</i>	<i>část</i>	<i>Číňanka</i>
<i>Afrika</i>	<i>biskup</i>	<i>částka</i>	<i>činnost</i>
<i>akce</i>	<i>bitva</i>	<i>CD disk</i>	<i>činžák</i>
<i>akcie</i>	<i>bižuterie</i>	<i>Čech</i>	<i>církev</i>
<i>aktivita</i>	<i>blahopřání</i>	<i>céčko</i>	<i>cirkus</i>
<i>alkohol</i>	<i>blinkr</i>	<i>cédéčko</i>	<i>číslice</i>
<i>Alpy</i>	<i>blízkost</i>	<i>čekárna</i>	<i>číslo</i>
<i>ambasáda</i>	<i>bod</i>	<i>celnice</i>	<i>číslovka</i>
<i>Američan</i>	<i>boj</i>	<i>čelo</i>	<i>čistírna</i>
<i>Američanka</i>	<i>bon</i>	<i>celozrnný chléb</i>	<i>citrón / citron</i>
<i>Amerika</i>	<i>bonboniéra</i>	<i>cena</i>	<i>cizí člověk</i>
<i>Angličan</i>	<i>bota</i>	<i>centimetr</i>	<i>cizinec</i>
<i>Angličanka</i>	<i>bouřka</i>	<i>centrum</i>	<i>cizinecká policie</i>
<i>angličtina</i>	<i>brácha</i>	<i>čepice</i>	<i>cizinka</i>
<i>Anglie</i>	<i>bradka</i>	<i>červen</i>	<i>článek</i>
<i>armáda</i>	<i>bramborová kaše</i>	<i>červenec</i>	<i>člen</i>
<i>Asie</i>	<i>bramborová polévka</i>	<i>Česka</i>	<i>členka</i>
<i>Austrálie</i>	<i>bramborový salát</i>	<i>Česká republika</i>	<i>clo</i>
<i>auto</i>	<i>brambory</i>	<i>České Budějovice</i>	<i>člověk</i>
<i>autobus</i>	<i>brána</i>	<i>český jazyk</i>	<i>čočka</i>
<i>autobusové nádraží</i>	<i>brandy</i>	<i>česnek</i>	<i>čtvereční kilometr</i>
<i>automobily</i>	<i>branka</i>	<i>cesta</i>	<i>čtvrt</i>
<i>autoopravna</i>	<i>Bratislava</i>	<i>čeština</i>	<i>čtvrť</i>
<i>autor</i>	<i>bratr</i>	<i>cestovatel</i>	<i>čtvrtek</i>
<i>babička</i>	<i>břeh</i>	<i>cestovní kancelář</i>	<i>čtvrthodina</i>
<i>bageta</i>	<i>brejle</i>	<i>cestovní šeky</i>	<i>čtvrťka</i>
<i>balet</i>	<i>březan</i>	<i>cestující</i>	<i>čtyřicítka</i>
<i>balíček</i>	<i>brigáda</i>	<i>chaos</i>	<i>čtyřka</i>
<i>balík</i>	<i>brigádník</i>	<i>charakter</i>	<i>cukr</i>
<i>balon</i>	<i>Brno</i>	<i>chata</i>	<i>cukrárna</i>
<i>banán</i>	<i>brokolice</i>	<i>chatař</i>	<i>cukroví</i>
<i>banka</i>	<i>broskev</i>	<i>chatařka</i>	<i>cvičení</i>
<i>bankovka</i>	<i>brýle</i>	<i>chatka</i>	<i>dav</i>
<i>bar</i>	<i>brzdy</i>	<i>chemie</i>	<i>dcera</i>
<i>barák</i>	<i>buchta</i>	<i>chlapec</i>	<i>děda</i>
<i>barva</i>	<i>Budapešť</i>	<i>chléb</i>	<i>dědeček</i>
<i>baterie</i>	<i>budoucnost</i>	<i>chleba</i>	<i>dědic</i>
<i>baterka</i>	<i>budova</i>	<i>chlebiček</i>	<i>dějepis</i>
<i>batož</i>	<i>bufet</i>	<i>chodba</i>	<i>dějiny</i>
<i>bavlna</i>	<i>Bůh</i>	<i>chodník</i>	<i>dělnice</i>
<i>bazén</i>	<i>bunda</i>	<i>chrám</i>	<i>dělník</i>
<i>Benátky</i>	<i>buřt</i>	<i>chudoba</i>	<i>deník</i>
<i>benzín</i>	<i>burza</i>	<i>chuť</i>	<i>desítka</i>
<i>benzinová nádrž</i>	<i>bydliště</i>	<i>chvilé</i>	<i>deštník</i>
<i>benzinová stanice</i>	<i>byt</i>	<i>chyba</i>	<i>dětství</i>
<i>Berlín</i>	<i>čaj</i>	<i>cigareta</i>	<i>devítka</i>
<i>bezpečnostní pás</i>	<i>čárka</i>	<i>cíl</i>	<i>diář</i>
		<i>čin</i>	

<i>dílo</i>	<i>elektrický spotře-</i>	<i>hindština</i>	<i>hřeben</i>
<i>diplomová práce</i>	<i>bič</i>	<i>historička</i>	<i>hřiště</i>
<i>diskotéka</i>	<i>e-mail</i>	<i>historie</i>	<i>hrnec</i>
<i>dítě</i>	<i>esemeska</i>	<i>historik</i>	<i>hrneček</i>
<i>divadlo</i>	<i>Evropa</i>	<i>hlad</i>	<i>hrnek</i>
<i>dluh</i>	<i>Evropan</i>	<i>hladina</i>	<i>hromada</i>
<i>Dobrá voda</i>	<i>Evropanka</i>	<i>hlas</i>	<i>hroznové víno</i>
<i>dobrodružství</i>	<i>fakt</i>	<i>hlášení</i>	<i>hruška</i>
<i>dobrota</i>	<i>fakulta</i>	<i>hlava</i>	<i>hudba</i>
<i>dodávka</i>	<i>farma</i>	<i>hlávkový salát</i>	<i>hudební nástroj</i>
<i>dojem</i>	<i>fazole</i>	<i>hlavní město</i>	<i>hudební výchova</i>
<i>doklad</i>	<i>film</i>	<i>hlavní silnice</i>	<i>hudebnice</i>
<i>doktor</i>	<i>filozofie</i>	<i>hlavní věc</i>	<i>hudebník</i>
<i>doktorka</i>	<i>firma</i>	<i>hledisko</i>	<i>husa</i>
<i>dokument</i>	<i>flaška</i>	<i>hlt</i>	<i>Indie</i>
<i>dolar</i>	<i>flek</i>	<i>hluk</i>	<i>informace</i>
<i>domácí úkol</i>	<i>fotbal</i>	<i>hodina</i>	<i>informační kan-</i>
<i>domácí zvíře</i>	<i>fotoaparát</i>	<i>hodinářství</i>	<i>celář</i>
<i>domácnost</i>	<i>fotograf</i>	<i>hodinky</i>	<i>internát</i>
<i>domek</i>	<i>fotografie</i>	<i>hodiny</i>	<i>internet</i>
<i>dopisní papír</i>	<i>fotografka</i>	<i>hodnota</i>	<i>inzerát</i>
<i>dopisní schránka</i>	<i>Francie</i>	<i>hokej</i>	<i>inzerce</i>
<i>dopravní pro-</i>	<i>Francouz</i>	<i>holící strojek</i>	<i>Ital</i>
<i>středky</i>	<i>Francouzka</i>	<i>holka</i>	<i>Itálie</i>
<i>dopravní spojení</i>	<i>francouzština</i>	<i>honěná</i>	<i>Italka</i>
<i>dopravní zácpa</i>	<i>fráze</i>	<i>hora</i>	<i>italština</i>
<i>dort</i>	<i>fronta</i>	<i>hořčice</i>	<i>jablko</i>
<i>dostih</i>	<i>funkce</i>	<i>horečka</i>	<i>jachta</i>
<i>doušek</i>	<i>fyzika</i>	<i>hory</i>	<i>jackpot</i>
<i>doutník</i>	<i>galerie</i>	<i>hosпода</i>	<i>jahoda</i>
<i>dovolená</i>	<i>garáž</i>	<i>hospodářství</i>	<i>jakost</i>
<i>Drážďany</i>	<i>golf</i>	<i>host</i>	<i>Japonec</i>
<i>dres</i>	<i>gram</i>	<i>hostitel</i>	<i>Japonka</i>
<i>dřez</i>	<i>grant</i>	<i>hostitelka</i>	<i>Japonsko</i>
<i>drogerie</i>	<i>grapefruit</i>	<i>hotel</i>	<i>japonština</i>
<i>duben</i>	<i>guláš</i>	<i>houba</i>	<i>jaro</i>
<i>dům</i>	<i>gulášová polévka</i>	<i>houska</i>	<i>jazyk</i>
<i>důsledek</i>	<i>guma</i>	<i>hovězí maso</i>	<i>jednání</i>
<i>dvacítká</i>	<i>gymnázium</i>	<i>hovězí polévka</i>	<i>jednička</i>
<i>dvojice</i>	<i>gynekolog</i>	<i>hovor</i>	<i>jednohubka</i>
<i>dvojka</i>	<i>hajzl</i>	<i>hra</i>	<i>jednolůžkový</i>
<i>dvojnásobek</i>	<i>hala</i>	<i>hráč</i>	<i>pokoj</i>
<i>dvůr</i>	<i>haléř</i>	<i>hráčka</i>	<i>jesle</i>
<i>dýmka</i>	<i>Harvard</i>	<i>hrad</i>	<i>jezero</i>
<i>džíny</i>	<i>hasič</i>	<i>Hradec Králové</i>	<i>jídlna</i>
<i>džob</i>	<i>hasička</i>	<i>hranice</i>	<i>jídelní lístek</i>
<i>džus</i>	<i>hedvábí</i>	<i>hranolky</i>	<i>jídlo</i>
<i>ekonom</i>	<i>herec</i>	<i>hrášek</i>	<i>jih</i>
<i>ekonomie</i>	<i>herečka</i>	<i>hřbitov</i>	<i>Jihlava</i>
<i>ekonomka</i>	<i>heslo</i>	<i>hrdina</i>	<i>jízdenka</i>

<i>jízdní řád</i>	<i>klobása</i>	<i>kout</i>	<i>les</i>
<i>jmění</i>	<i>klobouk</i>	<i>kov</i>	<i>let</i>
<i>jméno</i>	<i>klub</i>	<i>krabice</i>	<i>letadlo</i>
<i>jogurt</i>	<i>kluk</i>	<i>kraj</i>	<i>létání</i>
<i>kabaret</i>	<i>knedlík</i>	<i>krajina</i>	<i>letecká linka</i>
<i>kabát</i>	<i>kněz</i>	<i>krám</i>	<i>letecká pošta</i>
<i>kabelka</i>	<i>kniha</i>	<i>kravata</i>	<i>letenka</i>
<i>kabina</i>	<i>knihkupectví</i>	<i>krejčovství</i>	<i>letišťe</i>
<i>kachna</i>	<i>knihovna</i>	<i>křeslo</i>	<i>léto</i>
<i>kadeřnice</i>	<i>knížka</i>	<i>křestní jméno</i>	<i>levice</i>
<i>kadeřník</i>	<i>knoflík</i>	<i>krev</i>	<i>Liberec</i>
<i>kafe</i>	<i>koberec</i>	<i>křídla</i>	<i>libro</i>
<i>kakao</i>	<i>kočár</i>	<i>kriminální čin</i>	<i>linka</i>
<i>kalhotky</i>	<i>kočárek</i>	<i>křížovatka</i>	<i>Lipsko</i>
<i>kalhoty</i>	<i>Kodaň</i>	<i>křížovka</i>	<i>lístek</i>
<i>kamarád</i>	<i>kola</i>	<i>krok</i>	<i>listopad</i>
<i>kamarádka</i>	<i>koláč</i>	<i>kroketa</i>	<i>literatura</i>
<i>kamera</i>	<i>kolečkové brusle</i>	<i>krůta</i>	<i>litr</i>
<i>Kanada</i>	<i>koleda</i>	<i>kuchařka</i>	<i>loď</i>
<i>Kanad'an</i>	<i>kolega</i>	<i>kuchyně</i>	<i>lodžie</i>
<i>Kanad'anka</i>	<i>kolegyně</i>	<i>kuchyňská linka</i>	<i>loft</i>
<i>kancelář</i>	<i>kolej</i>	<i>kufř</i>	<i>lokalita</i>
<i>kantýna</i>	<i>kolo</i>	<i>kulečník</i>	<i>Londýn</i>
<i>kapr</i>	<i>komora</i>	<i>kultura</i>	<i>los</i>
<i>kapsa</i>	<i>koňak</i>	<i>kuřák</i>	<i>losos</i>
<i>karaván</i>	<i>koncert</i>	<i>kuře</i>	<i>loterie</i>
<i>Karlovy Vary</i>	<i>koncovka</i>	<i>kuřecí polévka</i>	<i>louka</i>
<i>karta</i>	<i>konec</i>	<i>kurz</i>	<i>lov</i>
<i>kartáček na zuby</i>	<i>konference</i>	<i>kurzíva</i>	<i>ložnice</i>
<i>karty</i>	<i>koniček</i>	<i>kus</i>	<i>lustr</i>
<i>kaše</i>	<i>konkurent</i>	<i>kůže</i>	<i>lyžování</i>
<i>kašel</i>	<i>konkurz</i>	<i>kvalita</i>	<i>lžíce</i>
<i>katedra</i>	<i>kontakt</i>	<i>květák</i>	<i>lžička</i>
<i>katedrála</i>	<i>kontrola</i>	<i>květen</i>	<i>Maďar</i>
<i>kauce</i>	<i>konvice na vodu</i>	<i>květina</i>	<i>Maďarka</i>
<i>káva</i>	<i>konzerva</i>	<i>láhev</i>	<i>Maďarsko</i>
<i>kavárna</i>	<i>konzulát</i>	<i>lampička</i>	<i>magnetofon</i>
<i>kazeta</i>	<i>konzultace</i>	<i>láska</i>	<i>majitel</i>
<i>kečup</i>	<i>kopec</i>	<i>lázně</i>	<i>majitelka</i>
<i>kelímek</i>	<i>kopírka</i>	<i>led</i>	<i>majonéza</i>
<i>kemp</i>	<i>koření</i>	<i>leden</i>	<i>makrela</i>
<i>keř</i>	<i>koruna</i>	<i>lednička</i>	<i>malba</i>
<i>kilogram</i>	<i>koš</i>	<i>legrace</i>	<i>malíř</i>
<i>kilometr</i>	<i>košík</i>	<i>lehátko</i>	<i>malířka</i>
<i>kino</i>	<i>košile</i>	<i>lék(-y)</i>	<i>máma</i>
<i>klášter</i>	<i>kostel</i>	<i>lékař</i>	<i>maminka</i>
<i>klávesnice</i>	<i>koupelna</i>	<i>lékařka</i>	<i>mamka</i>
<i>klíč</i>	<i>kouř</i>	<i>lékárna</i>	<i>manažer</i>
<i>klid</i>	<i>kouření</i>	<i>lékárníčka</i>	<i>manažerka</i>
<i>klinika</i>	<i>kousek</i>	<i>lekce</i>	<i>manžel</i>

<i>manželka</i>	<i>Mnichov</i>	<i>návštěva</i>	<i>obrana</i>
<i>mapa</i>	<i>množství</i>	<i>název</i>	<i>obraz</i>
<i>marihuana</i>	<i>mobil</i>	<i>názor</i>	<i>obrázek</i>
<i>maska</i>	<i>modelka</i>	<i>nebe</i>	<i>obrazovka</i>
<i>máslo</i>	<i>moře</i>	<i>nebezpečí</i>	<i>obsluha</i>
<i>maso</i>	<i>Moskva</i>	<i>neděle</i>	<i>obuv</i>
<i>matematika</i>	<i>most</i>	<i>nehoda</i>	<i>obývací pokoj</i>
<i>materiál</i>	<i>motor</i>	<i>němčina</i>	<i>obyvatel</i>
<i>mateřská škola</i>	<i>motorka</i>	<i>Němec</i>	<i>ocet</i>
<i>matka</i>	<i>moučník</i>	<i>Němka</i>	<i>ochrana</i>
<i>maturita</i>	<i>mouka</i>	<i>nemoc</i>	<i>odborník</i>
<i>maturitní zkouška</i>	<i>možnost</i>	<i>nemocnice</i>	<i>oděv</i>
<i>mechanik</i>	<i>mrakodrap</i>	<i>nepřítel</i>	<i>odlet</i>
<i>med</i>	<i>mrkev</i>	<i>neúspěch</i>	<i>odpočinek</i>
<i>medicína</i>	<i>mše</i>	<i>New York</i>	<i>odpoledne</i>
<i>menšina</i>	<i>muž</i>	<i>nezaměstnanost</i>	<i>odpověď</i>
<i>menza</i>	<i>muzeum</i>	<i>nezávislost</i>	<i>ohěň</i>
<i>meruňka</i>	<i>muzikál</i>	<i>noc</i>	<i>ohled</i>
<i>Měsíc</i>	<i>myčka</i>	<i>nocleh</i>	<i>okamžik</i>
<i>měsíc</i>	<i>myčka aut</i>	<i>noční klub</i>	<i>okno</i>
<i>město</i>	<i>mýdlo</i>	<i>noha</i>	<i>oko</i>
<i>metoda</i>	<i>myš</i>	<i>nouzový východ</i>	<i>okolí</i>
<i>metr</i>	<i>myšlenka</i>	<i>novina</i>	<i>okraj</i>
<i>metro</i>	<i>nabídka</i>	<i>novinář</i>	<i>okružní cesta</i>
<i>meziměstský ho-</i>	<i>náboženství</i>	<i>novinářka</i>	<i>okurka</i>
<i>vor</i>	<i>nábytek</i>	<i>noviny</i>	<i>olej</i>
<i>mezinárodní vlak</i>	<i>nádobí</i>	<i>nuda</i>	<i>Olomouc</i>
<i>míč</i>	<i>nádraží</i>	<i>nula</i>	<i>oltář</i>
<i>milenec</i>	<i>nádrž</i>	<i>nůž</i>	<i>omáčka</i>
<i>milenska</i>	<i>nafta</i>	<i>nůžky</i>	<i>omeleta</i>
<i>milimetr</i>	<i>nájem</i>	<i>obálka</i>	<i>opera</i>
<i>mince</i>	<i>nájemné</i>	<i>občan</i>	<i>opravna</i>
<i>minerálka</i>	<i>nákladní auto</i>	<i>občanka</i>	<i>optika</i>
<i>ministerstvo</i>	<i>nákup</i>	<i>občanský průkaz</i>	<i>ordinace</i>
<i>ministr</i>	<i>náledí</i>	<i>obchod</i>	<i>ořechy</i>
<i>minulost</i>	<i>náměstí</i>	<i>obchodní dům</i>	<i>organizace</i>
<i>minuta</i>	<i>nápad</i>	<i>obchodnice</i>	<i>orientace</i>
<i>mír</i>	<i>náplast</i>	<i>obchodník</i>	<i>oslava</i>
<i>míra</i>	<i>nápoj</i>	<i>období</i>	<i>oslovení</i>
<i>mísa</i>	<i>národnost</i>	<i>obec</i>	<i>osmička</i>
<i>místenka</i>	<i>nárok</i>	<i>obecná čeština</i>	<i>osoba</i>
<i>místní hovor</i>	<i>narození</i>	<i>oběd</i>	<i>osobní vlak</i>
<i>místnost</i>	<i>narozeniny</i>	<i>oběť</i>	<i>Ostrava</i>
<i>místo</i>	<i>nástroj</i>	<i>obilí</i>	<i>ostrov</i>
<i>mistr</i>	<i>nástupce</i>	<i>oblast</i>	<i>osud</i>
<i>mítink</i>	<i>nástupiště</i>	<i>oblečení</i>	<i>otázka</i>
<i>mléko</i>	<i>NATO</i>	<i>oblek</i>	<i>otec</i>
<i>mlha</i>	<i>náušnice</i>	<i>obličej</i>	<i>ovoce</i>
<i>mluvčí (f)</i>	<i>návrat</i>	<i>obloha</i>	<i>ovzduší</i>
<i>mluvčí (m)</i>	<i>návrh</i>	<i>obor</i>	<i>oznámení</i>

<i>pacient</i>	<i>plán</i>	<i>pokuta</i>	<i>pozdrav</i>
<i>palác</i>	<i>planeta</i>	<i>Polák</i>	<i>pozemek</i>
<i>palačinka</i>	<i>plast</i>	<i>pole</i>	<i>pozitíří</i>
<i>Palestina</i>	<i>plášť</i>	<i>poledne</i>	<i>pozor</i>
<i>památky</i>	<i>plat</i>	<i>polévka</i>	<i>pozornost</i>
<i>pan</i>	<i>platební karta</i>	<i>police</i>	<i>pozvání</i>
<i>panák</i>	<i>plátek</i>	<i>policejní stanice</i>	<i>pozvánka</i>
<i>panelák</i>	<i>plavání</i>	<i>policie</i>	<i>pracák</i>
<i>pánev</i>	<i>pláž</i>	<i>policista</i>	<i>práce</i>
<i>paní</i>	<i>ples</i>	<i>policistka</i>	<i>prachy</i>
<i>papež</i>	<i>plná penze</i>	<i>poliklinika</i>	<i>prací prášek</i>
<i>papír</i>	<i>plot</i>	<i>politická</i>	<i>pračka</i>
<i>papírnictví</i>	<i>plyn</i>	<i>politik</i>	<i>pracovní doba</i>
<i>paprika</i>	<i>Plzeň</i>	<i>politika</i>	<i>pracovní úřad</i>
<i>pár</i>	<i>pobřeží</i>	<i>Polka</i>	<i>pracovník</i>
<i>Pardubice</i>	<i>pobyt</i>	<i>poloha</i>	<i>prádlo</i>
<i>párek</i>	<i>počasí</i>	<i>polopenze</i>	<i>Praha</i>
<i>Paříž</i>	<i>počátek</i>	<i>polovina</i>	<i>pravda</i>
<i>park</i>	<i>počet</i>	<i>Polsko</i>	<i>pravice</i>
<i>parkoviště</i>	<i>pocit</i>	<i>polštář</i>	<i>pravidlo</i>
<i>parlament</i>	<i>počítač</i>	<i>polština</i>	<i>právníčka</i>
<i>partner</i>	<i>podkroví</i>	<i>pomeranč</i>	<i>právník</i>
<i>partnerka</i>	<i>podlaha</i>	<i>pomlázka</i>	<i>právo</i>
<i>party</i>	<i>podmínka</i>	<i>pomník</i>	<i>praxe</i>
<i>pas</i>	<i>podnájem</i>	<i>pomoc</i>	<i>prázdniny</i>
<i>pás</i>	<i>podnik</i>	<i>pondělí</i>	<i>přechod</i>
<i>pásek</i>	<i>podnikatel</i>	<i>ponožka</i>	<i>předměstí</i>
<i>pasová kontrola</i>	<i>podnikatelka</i>	<i>popelník</i>	<i>předmět</i>
<i>pasta na zuby</i>	<i>podoba</i>	<i>poplatek</i>	<i>přednáška</i>
<i>pastelka</i>	<i>podpatky</i>	<i>pořádek</i>	<i>předpis</i>
<i>pátek</i>	<i>podpis</i>	<i>porce</i>	<i>předpověď počasí</i>
<i>patro</i>	<i>podpora</i>	<i>pórek</i>	<i>předprodej</i>
<i>pečivo</i>	<i>podpora v neza-</i>	<i>portugalština</i>	<i>předseda</i>
<i>pekařství</i>	<i>městnanosti</i>	<i>porucha</i>	<i>předsín</i>
<i>peněženka</i>	<i>podprsenka</i>	<i>poschodí</i>	<i>představa</i>
<i>peníze</i>	<i>podstata</i>	<i>posilovna</i>	<i>představení</i>
<i>penze</i>	<i>podzim</i>	<i>pošta</i>	<i>přehled</i>
<i>penzion</i>	<i>poezie</i>	<i>postava</i>	<i>přehrávač</i>
<i>pepř</i>	<i>pohled</i>	<i>postel</i>	<i>přejezd</i>
<i>pero</i>	<i>pohlednice</i>	<i>postup</i>	<i>překlad</i>
<i>pětka</i>	<i>pohoda</i>	<i>potomek</i>	<i>premiér</i>
<i>pevná linka</i>	<i>pohotovost</i>	<i>potraviny</i>	<i>přepážka</i>
<i>piknik</i>	<i>pohovor</i>	<i>potřeba</i>	<i>přesčas</i>
<i>pilot</i>	<i>pohyb</i>	<i>poukázka</i>	<i>přestávka</i>
<i>píseň</i>	<i>pojištění</i>	<i>povaha</i>	<i>prezident</i>
<i>písnička</i>	<i>pokladna</i>	<i>povinnost</i>	<i>prezidentka</i>
<i>pití</i>	<i>pokoj</i>	<i>povlečení</i>	<i>příběh</i>
<i>pivo</i>	<i>pokračování</i>	<i>povolání</i>	<i>příbuzná</i>
<i>pizza</i>	<i>pokrývka</i>	<i>požadavek</i>	<i>příbuzný</i>
<i>plakát</i>	<i>pokus</i>	<i>požár</i>	<i>příchod</i>

<i>příčina</i>	<i>průmysl</i>	<i>režisérka</i>	<i>šance</i>
<i>příjem</i>	<i>průvodce</i>	<i>řeznictví</i>	<i>sanitka</i>
<i>přijímací zkouška</i>	<i>psaní</i>	<i>řidič</i>	<i>šátek</i>
<i>přijímání</i>	<i>PSC</i>	<i>řidič taxi</i>	<i>šatna</i>
<i>příjmení</i>	<i>pstruh</i>	<i>řidičák</i>	<i>šaty</i>
<i>příjmení za svo-</i>	<i>psychologie</i>	<i>řidička</i>	<i>sázka</i>
<i>bodna</i>	<i>půjčovní auto-</i>	<i>řidičský průkaz</i>	<i>scéna</i>
<i>příklad</i>	<i>mobilů</i>	<i>říjen</i>	<i>schod</i>
<i>přikrývka</i>	<i>půl</i>	<i>Řím</i>	<i>schodiště</i>
<i>přilet</i>	<i>půlhodina</i>	<i>řízek</i>	<i>schody</i>
<i>příležitost</i>	<i>půlka</i>	<i>řízení</i>	<i>schopnost</i>
<i>příloha</i>	<i>půllitr</i>	<i>rod</i>	<i>schovávaná</i>
<i>případ</i>	<i>pult</i>	<i>rodeo</i>	<i>schránka</i>
<i>příplatek</i>	<i>pumpa</i>	<i>rodiče</i>	<i>schůze</i>
<i>příprava</i>	<i>punčochy</i>	<i>rodina</i>	<i>schůzka</i>
<i>příroda</i>	<i>rada</i>	<i>rodné příjmení</i>	<i>sedací souprava</i>
<i>přírodopis</i>	<i>řada</i>	<i>roh</i>	<i>sedáčka</i>
<i>přísada</i>	<i>rádio</i>	<i>rohlík</i>	<i>sedadlo</i>
<i>přistání</i>	<i>radost</i>	<i>rok</i>	<i>sedmička</i>
<i>přístav</i>	<i>rajčata</i>	<i>roláda</i>	<i>šéf</i>
<i>přístroj</i>	<i>raketa</i>	<i>role</i>	<i>ségra</i>
<i>přístup</i>	<i>Rakousko</i>	<i>rostlina</i>	<i>šek</i>
<i>přistýlka</i>	<i>Rakušan / Raku-</i>	<i>rovina</i>	<i>sendvič</i>
<i>přítel</i>	<i>šák</i>	<i>rozdíl</i>	<i>sešit</i>
<i>přítelkyně</i>	<i>Rakušanka</i>	<i>rozhlas</i>	<i>šestka</i>
<i>přízemí</i>	<i>rámec</i>	<i>rozhodnutí</i>	<i>sestra</i>
<i>problém</i>	<i>rameno</i>	<i>rozhovor</i>	<i>setkání</i>
<i>procento</i>	<i>rámus</i>	<i>rozvoj</i>	<i>sever</i>
<i>proces</i>	<i>rána</i>	<i>ručník</i>	<i>seznam</i>
<i>procházka</i>	<i>rande</i>	<i>ruka</i>	<i>sídliště</i>
<i>profese</i>	<i>ráno</i>	<i>rukavice</i>	<i>síla</i>
<i>program</i>	<i>řeč</i>	<i>Rus</i>	<i>silnice</i>
<i>prohlídka</i>	<i>recepce</i>	<i>Ruska</i>	<i>Silvestr</i>
<i>projekt</i>	<i>recept na léky</i>	<i>Rusko</i>	<i>síň</i>
<i>promoce</i>	<i>Řecko</i>	<i>ruština</i>	<i>šipky</i>
<i>propiska</i>	<i>ředitel</i>	<i>ryba</i>	<i>sirup</i>
<i>prosba</i>	<i>ředitelka</i>	<i>rybník</i>	<i>síť</i>
<i>prosinec</i>	<i>Řek</i>	<i>rychlost</i>	<i>situace</i>
<i>prostěradlo</i>	<i>řeka</i>	<i>rýma</i>	<i>sjezd</i>
<i>prostor</i>	<i>reklama</i>	<i>rýže</i>	<i>sjezdovka</i>
<i>prostředek</i>	<i>rekonstrukce</i>	<i>sáček</i>	<i>skála</i>
<i>prostředí</i>	<i>Řekyně</i>	<i>šachy</i>	<i>skanzen</i>
<i>protisměr</i>	<i>řemeslo</i>	<i>sako</i>	<i>sklad</i>
<i>proud</i>	<i>reproduktor</i>	<i>šála</i>	<i>skladatel</i>
<i>provoz</i>	<i>řešení</i>	<i>salám</i>	<i>skladatelka</i>
<i>próza</i>	<i>restaurace</i>	<i>salát</i>	<i>sklenice</i>
<i>prst</i>	<i>revue</i>	<i>šálek</i>	<i>sklenička</i>
<i>prsten</i>	<i>rezervace</i>	<i>salon</i>	<i>sklenka</i>
<i>průběh</i>	<i>režim</i>	<i>samoobsluha</i>	<i>sklep</i>
<i>průkaz</i>	<i>režisér</i>	<i>šampaňské</i>	<i>sklo</i>

škoda	soused	štěstí	svíčková omáčka
škola	sousedka	stezka	Švýcar
školení	sousedství	stín	Švýcarka
školné	soustrast	stipendium	Švýcarsko
školní rok	soutěž	století	syn
skopové maso	souvislost	stolní tenis	synagoga
Skotsko	Španěl	stop	sýr
skříň	Španělka	stopař	systém
skupina	Španělsko	stopařka	tabák
skutečnost	španělština	strach	talíř
skútr	špenát	strana	taška
slanina	šperky	stránka	táta
sláva	spěšný vlak	strava	tatarská omáčka
slečna	spisovatel	stravenka	tatínek
šlehačka	spisovná čeština	stravování	taťka
sleva	Spojené státy	střecha	taxík
Slovák	americké / USA	střed	technika
Slovenka	spojení	středa	telecí maso
Slovensko	spojka	středisko	telefonní seznam
slovní druh	společenský pro-	střední škola	tělesná výchova
slovní zásoba	blém	strejda	tělo
slovník	společník	stříbro	tělocvik
slovo	společnost	strojek	tenis
sluchátko	Spolková repub-	strom	teplo
šluk	lika Německo	strýc	teplota
slunce	spolupráce	studená voda	terasa
služba	spolužák	student	termín
služební cesta	spor	studentka	test
směr	sporák	studium	těstoviny
smetana	sport	studna	teta
smlouva	sprcha	stůl	text
smrt	srážky	stupeň	tisícovka
smůla	srdce	sukně	toaleta
smutek	srpen	sůl	toastový chléb
mysl	stadion	šunka	Tóra
snaha	stan	supermarket	továrna
snídaně	stánek	sušenky	trafika
sníh	stanice	suterén	tramvaj
sobota	stanoviště taxiků	svačina	tráva
socha	stát	svatba	treska
sociální dávky	statek	Švéd	třetina
sociální zařízení	státní občanství	svědek	trh
sociologie	státní úřad	Švédka	třicítka
sortiment	státní zkouška	svědkyně	tričko
šortky	státnice	Švédsko	třída
soubor	stav	švédský stůl	trojka
součást	stavba	svět	trouba
soud	stáž	světlo	tuk
soudce	stěna	světlo	tužka
soudkyně	stereotyp	svíčka	tvář

<i>tvaroh</i>	<i>vánočka</i>	<i>vlak</i>	<i>vývoj</i>
<i>tykáni</i>	<i>vaříč</i>	<i>vlakové nádraží</i>	<i>výživa</i>
<i>typ</i>	<i>varieté</i>	<i>vlas</i>	<i>výzkum</i>
<i>ubytování</i>	<i>Varšava</i>	<i>vliv</i>	<i>význam</i>
<i>ubytovna</i>	<i>vchod</i>	<i>vlna</i>	<i>vyznání</i>
<i>účast</i>	<i>vdova</i>	<i>voda</i>	<i>vzdálenost</i>
<i>učebna</i>	<i>vdovec</i>	<i>vojačka</i>	<i>vzdělání</i>
<i>učení</i>	<i>věc</i>	<i>voják</i>	<i>vzduch</i>
<i>účes</i>	<i>večer</i>	<i>volant</i>	<i>vzkaz</i>
<i>účet</i>	<i>večeře</i>	<i>volba</i>	<i>vzor</i>
<i>účetní (f)</i>	<i>večírek</i>	<i>volkmen</i>	<i>vztah</i>
<i>účetní (m)</i>	<i>věda</i>	<i>volno</i>	<i>Washington</i>
<i>účetnictví</i>	<i>vedení</i>	<i>volný čas</i>	<i>WC</i>
<i>uchazeč</i>	<i>vedro</i>	<i>vrah</i>	<i>whisky</i>
<i>učiliště</i>	<i>vegetarián</i>	<i>vrtulník</i>	<i>zábradlí</i>
<i>učitel</i>	<i>vegetariánka</i>	<i>všední den</i>	<i>začátek</i>
<i>učitelka</i>	<i>vejce</i>	<i>vstup</i>	<i>záchod</i>
<i>událost</i>	<i>věk</i>	<i>vstupenka</i>	<i>záda</i>
<i>údolí</i>	<i>veka</i>	<i>vteřina</i>	<i>žádost</i>
<i>ulice</i>	<i>veletrh</i>	<i>vůně</i>	<i>zahrada</i>
<i>umělec</i>	<i>Velikonoce</i>	<i>vůz</i>	<i>zahraničí</i>
<i>umělkyně</i>	<i>velikost</i>	<i>vybavení</i>	<i>zájem</i>
<i>umění</i>	<i>Velká Británie</i>	<i>výběr</i>	<i>zájezd</i>
<i>úmysl</i>	<i>velvyslanec</i>	<i>východ</i>	<i>zájezdní hospoda</i>
<i>umyvadlo</i>	<i>velvyslanectví</i>	<i>výchova</i>	<i>žák</i>
<i>uniforma</i>	<i>venkov</i>	<i>vydavatelství</i>	<i>zákaz</i>
<i>univerzita</i>	<i>Venuše</i>	<i>výhled</i>	<i>zákaz kouření</i>
<i>únor</i>	<i>vepřové maso</i>	<i>vyjádření</i>	<i>zákaz stání</i>
<i>úřad</i>	<i>veranda</i>	<i>výjimka</i>	<i>zákaz zastavení</i>
<i>úraz</i>	<i>veřejnost</i>	<i>vykání</i>	<i>zákaznice</i>
<i>úřední formulář</i>	<i>věšák</i>	<i>výkon</i>	<i>zákazník</i>
<i>úřední hodiny</i>	<i>vesmír</i>	<i>výlet</i>	<i>základ</i>
<i>úřednice</i>	<i>vesnice</i>	<i>výloha</i>	<i>základní škola</i>
<i>úředník</i>	<i>většina</i>	<i>vypínač</i>	<i>zákon</i>
<i>úroveň</i>	<i>věž</i>	<i>výpis z účtu</i>	<i>zákusek</i>
<i>úschovna</i>	<i>vězení</i>	<i>výprodej</i>	<i>žákyně</i>
<i>zadel</i>	<i>vid</i>	<i>výraz</i>	<i>záležitost</i>
<i>úspěch</i>	<i>Vídeň</i>	<i>vyřízení</i>	<i>zámek</i>
<i>Ústí nad Labem</i>	<i>video</i>	<i>výroba</i>	<i>zaměstnanec</i>
<i>útok</i>	<i>videokazeta</i>	<i>výrobek</i>	<i>zaměstnání</i>
<i>úvazek</i>	<i>videopřehrávač</i>	<i>výška</i>	<i>zaměstnavatel</i>
<i>území</i>	<i>videorekordér</i>	<i>výsledek</i>	<i>západ</i>
<i>uzenářství</i>	<i>vidlička</i>	<i>vysoká škola</i>	<i>zápalky</i>
<i>uzeniny</i>	<i>víkend</i>	<i>výstava</i>	<i>zápas</i>
<i>vagon</i>	<i>vila</i>	<i>vysvětlení</i>	<i>září</i>
<i>váha</i>	<i>víno</i>	<i>výtah</i>	<i>zařízení</i>
<i>vajíčko</i>	<i>Vinohrady</i>	<i>výtvarná výchova</i>	<i>zásilkový obchod</i>
<i>válka</i>	<i>vítr</i>	<i>výtvarné umění</i>	<i>zastávka</i>
<i>vana</i>	<i>vízum</i>	<i>výuční list</i>	<i>zástupce</i>
<i>Vánoce</i>	<i>vláda</i>	<i>vyučování</i>	<i>zavazadlo</i>

<i>závěr</i>	<i>zemědělec</i>	<i>zkušenost</i>	<i>zpáteční jízdenka</i>
<i>závěrečná zkouška</i>	<i>zemědělkyně</i>	<i>zlato</i>	<i>zpěvačka</i>
<i>zavírací doba</i>	<i>zemědělství</i>	<i>Zlín</i>	<i>zpěvák</i>
<i>závod</i>	<i>zeměpis</i>	<i>zločin</i>	<i>zpoždění</i>
<i>závorka</i>	<i>žena</i>	<i>zločinec</i>	<i>zpráva</i>
<i>zážitek</i>	<i>Ženeva</i>	<i>zloděj</i>	<i>způsob</i>
<i>záznamník</i>	<i>židle</i>	<i>změna</i>	<i>zranění</i>
<i>zbraň</i>	<i>zima</i>	<i>zmrzlina</i>	<i>zrcadlo</i>
<i>zbytek</i>	<i>živnostenský list</i>	<i>značka</i>	<i>ztráta</i>
<i>zdraví</i>	<i>živobytí</i>	<i>znalost</i>	<i>zubní pasta</i>
<i>zdroj</i>	<i>život</i>	<i>známá</i>	<i>zuby</i>
<i>zed'</i>	<i>životní styl</i>	<i>znamení</i>	<i>župan</i>
<i>zelenina</i>	<i>životní úroveň</i>	<i>známka</i>	<i>zvěst</i>
<i>zelí</i>	<i>životopis</i>	<i>známý</i>	<i>zvíře</i>
<i>Země</i>	<i>žízeň</i>	<i>znečištěné ovzduš- ší</i>	<i>zvuk</i>
<i>země</i>	<i>zkouška</i>	<i>zoo</i>	
	<i>zkratka</i>		

3. Vyhodnocení dat

3.1 Abstrakta a konkréta

Morfologické profily abstrakt

Typ	Substantiva z náhodného vzorku	abs. F	rel. F
P2-P4	<i>centimetr, požadavek, přesčas</i>	3	4,9 %
S1-S2	<i>biologie, chaos, chemie, náboženství, nuda, průmysl, spor, velikost</i>	8	13,1 %
S1-S4	<i>chyba, fakt, letecká linka, návrh, nebezpečí, oslovení, předpověď počasí, příběh, rande, školné, sobota, spojení, stolní tenis, vysvětlení, vzkaz, zájem, zážitek, zpráva, ztráta</i>	19	31,1 %
S1-S7	<i>heslo</i>	1	1,6 %
S2-P2	<i>milimetr, zločin</i>	2	3,3 %
S2-P4	<i>stereotyp</i>	1	1,6 %
S2-S4	<i>barva, ekonomie, fyzika, konkurz/konkurs, mír, odpočinek, osmička, pohyb, práce, tělocvik, výtvarná výchova, vzdělání, zaměstnání, život</i>	14	23,0 %
S2-S6	<i>čeština, sousedství, třicítka prosinec, říjen</i>	5	8,2 %
S4-P2	<i>půllitr</i>	1	1,6 %
S4-S6	<i>chvilu, půlhodina, smysl, souvislost</i>	4	6,6 %
S4-S7	<i>součást</i>	1	1,6 %
S6-P4	<i>léto</i>	1	1,6 %
S7-P4	<i>koniček</i>	1	1,6 %
	celkem	62	100 %

Morfologické profily konkrét

Typ	Substantiva z náhodného vzorku	abs. F	rel. F
P1-P2	<i>banán, hruška, keř</i>	3	4,1 %
P1-P4	<i>nůžky, řízek, rukavice</i>	3	4,1 %
P2-P4	<i>kalhoty, květina, vajíčko</i>	3	4,1 %
S1-S2	<i>autobus, hudba, chléb, karavan, kniha, nádrž, tramvaj, vrtulník</i>	8	11,0 %
S1-S4	<i>čárka, polévka, jednolůžkový pokoj, klobása, lampička, lžička, náplast, obraz, párek, šála, sanitka</i>	11	15,1 %
S1-S7	<i>tatarská omáčka</i>	1	1,4 %
S2-P2	<i>vůz</i>	1	1,4 %
S2-S4	<i>alkohol, auto, benzin, bunda, čaj, flaška, jídlo, mapa, maso, muzikál, nádoby, oblečení, zubní pasta, pivo, plášť, průkaz, pult, řidičák, šálek, sešit, sklenka, sprcha, smetana, víno, zmrzlina</i>	25	34,2 %
S2-S6	<i>křeslo</i>	1	1,4 %
S2-S7	<i>okno</i>	1	1,4 %
S3-S4	<i>pití</i>	1	1,4 %
S4-P2	<i>sklenička, videokazeta</i>	2	2,7 %
S4-S6	<i>louka, podlaha, přejezd, telefon, večere</i>	5	6,8 %
S4-S7	<i>bradka, čepice, hlava</i>	3	4,1 %

S4-P4	<i>houska, křížovka, lístek, panák, písnička</i>	5	6,8 %
	celkem	73	100 %

3.2 Frekventované kolokáty substantivních tvarů typu <S4-S2, S1>

Kolokáty byly vyhledávány v subkorpusu SYNEK2010 v rozsahu <-3 do 3> od dotyčného genitivního nebo akuzativního tvaru.

Z výsledných kolokátů byly vybrány ty, které svou vazbou určí pád substantiva. Jedná se především o předložkové a slovesné výrazy, a také o substantiva. Z konkordancí vybraných kolokací byly potom vyřazeny kolokáty, které jsou součástí jiného větného členu (např. výskyty předložek po hledaném tvaru).

Kolokáty se v následujících přehledech uvádějí ve své základní podobě (lemma).

Frekventované kolokace genitivních tvarů abstrakt

ekonomie (314)	<i>profesor (+Ad) ekonomie (59)</i> <i>z (oboru) ekonomie (20)</i> <i>student (+Ad) ekonomie (18)</i>
fyzika/fysika (680)	<i>z (83), např.: z (+Ad) fyziky (70), z oboru/ze světa/z oblasti (7)</i> <i>profesor (+Ad) fyziky (62)</i> <i>katedra (+Ad) fyziky (27)</i> <i>do (+Ad) fyziky (27)</i> <i>oblast (+Ad) fyziky (25)</i> <i>ústav (+Ad) fyziky (23)</i> <i>učitel (+Ad) fyziky (20)</i> <i>do (14), např.: Úvod do teoretické fyziky, nepatří do oblasti fyziky</i>
mír (1904)	<i>náměstí Míru (217)</i> <i>třída Míru (117)</i> <i>Nobelova cena míru (103)</i> <i>uzavření míru (66)</i> <i>v době/v dobách (klidu i) míru (59)</i>
muzikál (397)	<i>z muzikálu (53), např.: písně z (15), herci z, číslo z, hvězdy z</i> <i>do muzikálu (26), např.: obsadit do, nastoupit do, konkurz do muzikálu</i> <i>verze (+Ad) muzikálu (21)</i>
osmička (129)	<i>do (+Ad) osmičky (54), např.: postup do, posunout se do, dostat se do</i> <i>z (+Ad) osmičky (27)</i> <i>tvar osmičky (15)</i>
pohyb (3123)	<i>do (+Ad) pohybu (840), např.: dát se /dávat se do pohybu (500), uvést/uvádět do pohybu (162), dostat se do pohybu (18)</i> <i>bez (+Ad) pohybu (133), např.: ležet bez pohybu, sedět bez pohybu</i> <i>(ne)schopen/(ne)schopný (+Ad) pohybu (127)</i>

	<p><i>z (+Ad) pohybu (105), např.: radost z, zážitek z, pocit z, těšit se z</i></p> <p><i>volnost pohybu (68)</i></p> <p><i>svoboda pohybu (67)</i></p> <p><i>nedostatek pohybu (60)</i></p> <p><i>dostatek pohybu (59)</i></p>
tělocvik (139)	<p><i>hodina tělocviku (39)</i></p> <p><i>učitel/učitelka tělocviku (43)</i></p> <p><i>z tělocviku (15), např.: známka z, jednička z, osvobozená z</i></p>
výtvarná výchova (57)	<p><i>učitel/učitelka výtvarné výchovy (15)</i></p> <p><i>katedra/fakulta výtvarné výchovy (11)</i></p> <p><i>hodina výtvarné výchovy (10), např.: v hodině/hodinách (8)</i></p>
vzdělání (1111)	<p><i>do (+Ad) vzdělání (91), např.: investice / investovat do (70)</i></p> <p><i>bez (+Ad) vzdělání (90), např.: člověk bez (18)</i></p> <p><i>úroveň (+Ad/P) vzdělání (74)</i></p> <p><i>stupeň (+Ad/P) vzdělání (56)</i></p> <p><i>dostat se někomu (+Ad) vzdělání (54)</i></p> <p><i>investice do vzdělání (38)</i></p> <p><i>dosáhnout (+A/P) vzdělání (36)</i></p> <p><i>možnost (+Ad) vzdělání (15)</i></p> <p><i>ze vzdělání (14), např.: radost ze, vycházet ze, vyplývat ze vzdělání</i></p>

Frekventované kolokace genitivních tvarů konkrét

alkohol (3233)	<p><i>promile alkoholu (547)</i></p> <p><i>vliv alkoholu (436), např.: pod vlivem alkoholu (421)</i></p> <p><i>obsah alkoholu (149)</i></p> <p><i>pití (+Ad) alkoholu (144)</i></p> <p><i>konzumace (+Ad) alkoholu (133)</i></p> <p><i>množství (+Ad) alkoholu (112)</i></p> <p><i>naměřit (číslovka) promile alkoholu (93)</i></p> <p><i>požití (+Ad) alkoholu (84)</i></p> <p><i>láhev (+Ad) alkoholu (75)</i></p> <p><i>bez alkoholu (67)</i></p>
benzín (1202)	<p><i>cena benzínu (354)</i></p> <p><i>litr benzínu (178); např.: za litr benzín (30)</i></p> <p><i>spotřeba benzínu (36)</i></p> <p><i>do benzínu (32), např.: přimíchávat, přidávat</i></p> <p><i>z benzínu (27), např.: daň z benzínu (16)</i></p>
čaj (1949)	<p><i>šálek čaje (392)</i></p> <p><i>trochu (+Ad) čaje (80)</i></p> <p><i>u (+Ad) čaje (74)</i></p> <p><i>do (+Ad) čaje (71): přidávat do, dát si do, rum do čaje</i></p> <p><i>pití (+Ad) čaje (67)</i></p>

	<p><i>hrnek (+Ad) čaje (67)</i> <i>z (+Ad) čaje (48)</i> <i>napít se (+Ad) čaje (47)</i> <i>příprava (+Ad) čaje (44)</i> <i>od (+Ad) čaje (35): bedna od, šálek od, hrnek od, plechovka od, pytlíky od, krabičky od čaje</i> <i>druh čaje (34)</i> <i>konvice (+Ad) čaje (31)</i> <i>sklenice (+Ad) čaje (30)</i> <i>kvalita (+Ad) čaje (28)</i> <i>účinek (+Ad) čaje (26)</i> <i>sklenka (+Ad) čaje (23)</i></p>
maso (3429)	<p><i>z (+Ad) masa (473): z mletého (42), z vepřového (34), z hovězího (23) masa</i> <i>g (+Ad) masa (167): g mletého, g vepřového, g hovězího, g kuřecího</i> <i>do (+Ad) masa (218): např. zařezávat do, hluboko do masa, do živého, do měkkého</i> <i>kus (+Ad) masa (239)</i> <i>kousek (+Ad) masa (169)</i> <i>plátek (+Ad) masa (83)</i> <i>druh masa (76)</i> <i>bez (+Ad) masa (46)</i> <i>porce (+Ad) masa (39)</i> <i>zbytek (+Ad) masa (38)</i> <i>prodej (+Ad) masa (37)</i> <i>kilo (+Ad) masa (36)</i> <i>od (+Ad) masa (19)</i></p>
sprcha (290)	<p><i>do (té/tý) sprchy (90), např.: vlézt do, jít do sprchy</i> <i>ze sprchy (86), např.: vylézt z, vyjít z, voda ze sprchy</i> <i>od sprchy (11), např.: hadice od, závěs od, odtok od sprchy</i></p>
zmrzlina (273)	<p><i>kopeček (+Ad) zmrzliny (28)</i> <i>od zmrzliny (12), např.: kornout od, ulepený od zmrzliny</i> <i>porce (+Ad) zmrzliny (12)</i></p>
zubní pasta (38)	<p><i>tuba zubní pasty (8)</i> <i>od zubní pasty (7), např.: tuba od zubní pasty</i></p>

Frekventované kolokace akuzativních tvarů abstrakt

ekonomie (209)	<p><i>za ekonomii (55), např.: cena za ekonomii (51),</i> <i>z toho: Nobelova cena za ekonomii (47)</i> <i>studovat/vystudovat ekonomii (65)</i> <i>na (14), např.: odborník na, zapsat se na, přihlásit se na, dostat se na ekonomii</i> <i>o ekonomii (11), např.: zájem o/zajímat se o, jít o ekonomii</i> <i>pro (7): význam pro, nástroje pro, peníze pro</i></p>
---------------------------------	---

fyzika/fysika (284)	<i>studovat/vystudovat fyziku (44)</i> <i>pro fyziku (43), např.: Ústav pro teoretickou fyziku/fysiku (14); znamená pro / mít význam pro / mít nadání pro / časopis pro</i> <i>za (+Ad) fyziku, např.: cena za fyziku (34), z toho: Nobelova cena za fyziku (27)</i> <i>učit (+Ad) fyziku (16)</i>
mír (1218)	<i>za mír (195), např.: cena za mír (55), boj za mír (22), bojovník za mír (15)</i> <i>pro mír (90), např.: partnerství pro mír (13) být pro mír (8)</i> <i>uzavřít mír (91)</i> <i>mít mír (43)</i> <i>přinést mír (32)</i> <i>chtít mír (31)</i>
muzikál (389)	<i>na (+Ad) muzikál (59), např.: lákat, zvát, konkurz na, práva na, výlet na muzikál</i> <i>pro (+Ad) muzikál (20), např.: zpěvák / zpěváčku pro, kostým pro, hodí se pro, je stvořený pro muzikál</i>
osmička (107)	<i>mezi (nejlepší/elitní) osmičku (25), např.: postoupit mezi, dostat se mezi, patřit mezi (nejlepší/elitní) osmičku</i> <i>na osmičku (23), např.: čekat na, sázet na osmičku</i>
pohyb (2801)	<i>na (177) (+Ad/P) pohyb, např.: reagovat na, čekat, na, soustředit se na pohyb</i> <i>sledovat (+Ad/P) pohyb (177), např.: sledovat každý můj/svůj/jeho/její pohyb (69)</i> <i>udělat (+Ad/P) pohyb (119), např.: udělat rychlý pohyb, udělat pohyb rukou</i> <i>mít (+Ad) pohyb (85), např.: mít pohyb, mít volný pohyb, mít rád pohyb</i> <i>o (+Ad/P) pohyb (51), např.: jít o pohyb (24)</i> <i>vidět (+Ad/P) pohyb (44)</i> <i>zahlédnout (+Ad/P) pohyb (44)</i> <i>umožňovat (+Ad) pohyb (32)</i>
tělocvik (80)	<i>na (školní) tělocvik (22): mít někoho na, plavky/věci na, jít na (školní) tělocvik</i> <i>mít tělocvik (11), např.: dopoledne jsme neměli tělocvik, (ne)mít rád, tělocvik</i>
výtvarná výchova (21)	<i>učit/vyučovat výtvarnou výchovu (7)</i> <i>na výtvarnou výchovu (6)</i>
vzdělání (1691)	<i>(ne)mít (+Ad/P) vzdělání (417): mít potřebné/formální/vysokoškolské vzdělání</i> <i>(ne)získat (+Ad/P) vzdělání (123)</i> <i>o (+Ad/P) vzdělání (72): mít zájem o, starat se o</i> <i>pro (+Ad/P) vzdělání (55): ministerstvo/ministr pro, ústav pro, výbor pro, svoboda pro, motivace pro, udělat pro, sloužit pro vzdělání</i> <i>poskytovat vzdělání (37)</i> <i>dostat vzdělání (29)</i> <i>dokončit vzdělání (23)</i> <i>zajistit vzdělání (18)</i>

Frekventované kolokace akuzativních tvarů konkrét

alkohol (1443)	<i>na alkohol (248), např.: zkouška na (59), test na (21), reklama na (14) pít (+Ad) alkohol (273) mít (+Ad) alkohol (28) za alkohol (42) popíjet (+Ad) alkohol (45) požít alkohol (36)</i>
benzín (737)	<i>na (+Ad) benzín (155), např.: daň na, poukaz na, příspěvky na, mít na, peníze na benzín za (+Ad) benzín (57), např.: zaplatit za, výdaje za, poplatky za benzín pro (+Ad) benzín (26), např.: přijet/jet pro, zastavit pro benzín tankovat/natankovat benzín (47) zaplatit benzín (13)</i>
čaj (2613)	<i>na čaj (388), např. postavit na čaj (28), voda na čaj (28), zvát/pozvat na čaj (33), jít/přijít na čaj (25), konvice na čaj (13), pozvání na čaj (10), (ne)mít chuť na čaj (4) vařit/uvařit (si/někomu) (+Ad) čaj (224) dát si (+Ad) čaj (84) dát někomu čaj (48) mít (+Ad) čaj (24), např. mít rád (+Ad) čaj (8) udělat (si/někomu) (+Ad) čaj (48) připravit/připravovat (si) (+Ad) čaj (68) popíjet (+Ad) čaj (70) vypít čaj (66) přinést čaj (65) chtít čaj (30) nalít (si/někomu) čaj (54) nabídnout čaj (44) dopít čaj (41)</i>
maso (2816)	<i>jíst (+Ad) maso (174) nakrájet (+Ad) maso (147) přidat (+Ad) maso (98) osolit (+Ad) maso (60) vložit (+Ad) maso (57) o (+Ad) maso (43): jít o, dělit se o dát maso (43), např.: dát maso někam/někomu (33), dát si maso (10) vyjmout maso (42) opéci maso (38) promíchat (33) prodávat (31) osmahnout (29) za maso (21), např.: zaplatit maso, peníze za maso, ryba za maso</i>

sprcha (346)	<i>dát si sprchu (74)</i> <i>pod sprchu (58), např.: vlézt pod, stoupat pod, dát někoho pod sprchu</i> <i>mít (+Ad) sprchu (17)</i> <i>pustit sprchu (16)</i>
zmrzlina (443)	<i>na zmrzlinu (90), např.: jít/zajít na, pozvat na, reklama na, strojek na</i> <i>koupit (někomu) zmrzlinu (34)</i> <i>dát si zmrzlinu (22)</i> <i>jíst (+Ad) zmrzlinu (22)</i>
zubní pasta (56)	<i>na zubní pastu (10), např.: reklama na zubní pastu</i>

3.3 Náplň typů středního frekvenčního pásma

TYP	SUBSTANTIVA Z NÁHODNÉHO VZORKU
P2-P4	<i>centimetr, kalhoty, květina, požadavky, přesčas, vajíčko</i>
P2-P1	<i>banán, hruška, keř, politik, Rus, voják</i>
S4-P4	<i>houska, křížovka, lístek, panák, písnička, toaleta</i>
P2-S2	<i>byt, domek, milimetr, vůz, zločin</i>
S4-S7	<i>bradka, čepice, hlava, součást</i>
P1-P4	<i>nůžky, řízek, rukavice</i>
S4-P2	<i>půllitr, sklenička, videokazeta</i>
S1-S7	<i>heslo, tatarská omáčka, vdova</i>