

Univerzita Karlova v Praze, Filosofická fakulta

Ústav teoretické a komputační lingvistiky

Valenční rámce substantiv v češtině na základě korpusových dat

Valency Frames of Czech Nouns:
Corpus-driven Study

Anna Čermáková

školitel: Doc. RNDr. Vladimír Petkevič, CSc.

rok podání: 2008

studijní program: Matematická lingvistika

studijní obor: Korpusová lingvistika

Prohlašuji, že jsem práci vypracovala samostatně za použití uvedených pramenů.

Velké díky patří všem, kteří se mnou trpělivě diskutovali nad problémy a mnohými otázkami této práce a dále všem, kteří mi byli velkou inspirací a pomohli mi tak práci dokončit. Explicitně mé díky patří mému laskavému vedoucímu práce doc. Vladimíru Petkevičovi, dále mému otci prof. Františku Čermákovi, který mě vlastně k tématu práce přivedl. Velkým zdrojem inspirace a podporou mi byli i prof. Wolfgang Teubert a prof. John Sinclair.

Nemenší díky patří i mé rodině. Díky jejich bezmezné trpělivosti a podpoře je tento rukopis nyní dokončený. Explicitně bych ráda poděkovala všem, kteří statečně „hlídali“ Honzíka, aby maminka mohla pracovat – tedy především jeho tatínkovi, babičce a kmotrovi.

OBSAH

1	ÚVOD	1
2	TEORETICKÝ RÁMEC	4
2.1	POJEM VALENCE	4
2.2	JAZYK Z POHLEDU KORPUSOVÉ LINGVISTIKY	10
2.3	KORPUSOVÁ LINGVISTIKA: TEORIE A METODOLOGIE	16
3	METODOLOGIE A DATA	20
3.1	ČESKÝ NÁRODNÍ KORPUS	20
3.2	VÝBĚR SUBSTANTIV A JEJICH ANALÝZA	23
3.2.1	<i>Pilotní studie</i>	25
3.2.2	<i>Vlastní analýza</i>	27
3.3	ZÁVĚRY PILOTNÍ STUDIE	29
3.4	ROZŠÍŘENÉ HLEDÁNÍ VALENCE SUBSTANTIV	30
4	SUBSTANTIVUM A JEHO VALENCE Z POHLEDU ČNK	31
4.1	GENITIVNÍ VALENCE	32
4.1.1	<i>Substantiva s převažující genitivní valencí</i>	36
4.1.1.1	<i>Jednotlivá substantiva</i>	37
4.1.1.2	<i>Shrnutí</i>	44
4.1.2	<i>Ostatní substantiva s genitivním doplněním</i>	48
4.1.2.1	<i>Substantiva s jedním sémantickým typem genitivního doplnění</i>	49
4.1.2.2	<i>Substantiva s více sémantickými typy genitivní valence</i>	51
4.1.2.3	<i>Shrnutí</i>	55
4.1.3	<i>Sémantické typy genitivní valence</i>	57
4.1.3.1	<i>Substantiva s převažující genitivní valencí</i>	57
4.1.3.2	<i>Ostatní substantiva s genitivní valencí</i>	59
4.1.4	<i>Několikanásobná genitivní valence</i>	61
4.1.5	<i>Konkurence genitivu a atributivního adjektiva</i>	65
4.1.5.1	<i>Shrnutí</i>	74
4.1.6	<i>Alternativní vyjádření genitivní valence u některých substantiv</i>	76
4.1.7	<i>Závěr</i>	78
4.2	PŘEDLOŽKOVÁ VALENCE	81
4.2.1	<i>Předložka „do“</i>	85
4.2.2	<i>Předložka „k/ke“</i>	88
4.2.3	<i>Předložka „mezi“</i>	91
4.2.4	<i>Předložka „na“</i>	93
4.2.5	<i>Předložka „o“</i>	99
4.2.6	<i>Předložka „od“</i>	102
4.2.7	<i>Předložka „po“</i>	104
4.2.8	<i>Předložka „pro“ a „proti“</i>	106
4.2.9	<i>Předložka „s/se“</i>	109
4.2.10	<i>Předložka „u“</i>	112
4.2.11	<i>Předložka „v/ve“</i>	113
4.2.12	<i>Předložka „z/ze“</i>	114
4.2.13	<i>Předložka „za“</i>	116
4.2.14	<i>Ostatní předložky</i>	117
4.2.14.1	<i>Kolem</i>	117
4.2.14.2	<i>Nad</i>	117
4.2.14.3	<i>Před</i>	118
4.2.14.4	<i>Při</i>	119
4.2.14.5	<i>Vůči</i>	119

4.2.15	Alternující předložky	120
4.2.16	Závěr	122
4.3	OSTATNÍ PÁDY	124
4.3.1	Instrumentál	124
4.3.2	Dativ	125
4.4	ADVERBIUM/ADVERBIÁLE	126
4.5	KOORDINOVANÁ SPOJENÍ	128
4.6	NOMINATIVNÍ VALENCE	130
4.7	DALŠÍ KONSTRUKCE.....	131
4.7.1	Infinitiv.....	131
4.7.2	Vedlejší věta	134
4.7.3	Další výrazy.....	136
4.8	VÍCEMÍSTNÁ VALENCE	137
4.9	SUBSTANTIVA S VYSOKOU FREKVENCÍ VÝSKYTU	139
4.9.1	Skupina ČAS.....	141
4.9.2	Skupina ČLOVĚK.....	144
4.9.3	Závěr.....	145
4.10	ABSTRAKTA A KONKRÉTA	147
4.10.1	Substantiva abstraktní.....	148
4.10.2	Substantiva konkrétní.....	149
4.10.3	Substantiva s abstraktním i konkrétním významem.....	149
4.10.4	Substantiva s konkrétním denotátem a abstraktní složkou významu	150
4.10.5	Závěr.....	150
5	ZÁVĚR.....	152
5.1	DATA A METODOLOGIE.....	152
5.2	VALENCE SUBSTANTIV V ČEŠTINĚ.....	155
5.3	OTEVŘENÉ OTÁZKY.....	160
	ABSTRAKT.....	163
	ABSTRACT.....	165
	LITERATURA	167
	JAZYKOVÉ KORPUSY	173
	PŘÍLOHA A. - SUBSTANTIVA VYBRANÁ PRO PILOTNÍ ANALÝZU VALENCE SUBSTANTIV V ČEŠTINĚ.....	174
	PŘÍLOHA B. - SÉMANTICKÉ TYPY GENITIVŮ.....	175
	PŘÍLOHA C. - PŘEHLED BEZPŘEDLOŽKOVÉ GENITIVNÍ VALENCE PODLE FREKVENCE.....	180
	PŘÍLOHA D. - INFINITIVNÍ VALENCE	181

1 ÚVOD

Cílem této práce je vytvoření lexikologického rámce pro systematický popis valence českých substantiv. Práce se pokusí o formální klasifikaci typů substantivní valence podle valenčních exponentů zjištěných na základě analýzy dat z Českého národního korpusu. Valence je v této práci chápána jako lexikologická systémová vlastnost slov, která by jako taková měla být součástí slovníkových hesel.

Pojem valence je v lingvistické literatuře obvykle kategorií syntaktickou, která bývá spojována především se slovesy. Valence substantiv bývá pojímána jako odvozená od sloves a obvyklá při jejich postavení v predikátu. Tato práce však předpokládá existenci valence substantivní, která může mít (a často má) s valencí slovesnou souvislost, nicméně není od ní výlučně odvozená. Navíc řada substantiv nejsou deverbativa, a takto se omezovat tedy nestačí. Valenci jako lexikologickou vlastnost slov tedy v této práci chápeme jako součást lexikální (významové) jednotky a způsob její syntaktické realizace není pro její popis zásadní. Při takto pojatém popisu valence je předpokládána souvislost mezi valenčním exponentem slovesa a deverbativního substantiva jejím zajímavým aspektem, který však přesahuje rozsah této práce.

Korpusově pojatý lingvistický popis zdůrazňuje závislost jednotlivých slov na kontextu, ve kterém se nacházejí. Textové (/ortografické¹) slovo tedy nemusí nutně být vždy totožné s významovou (lexikální) jednotkou. Chceme-li se věnovat korpusově pojatému popisu významu, musíme vycházet z kontextů (a tedy i typických slovních vzorců), ve kterých se dané slovo opakovaně vyskytuje. Výsledky analýz rozsáhlých korpusových dat zpochybňují tradiční úhledné dělení

¹ Tedy „slovo“ ve zcela tradičním a běžném významu – slovo v textu oddělené mezerami, viz dále.

jazyka na lexikon a syntax a ukazují, že pro popis kontextu přirozeného jazyka, ve kterém se slova vyskytují, není toto tradiční dělení vyhovující.

Popis valence v této práci vychází z metodologie korpusové lingvistiky. Přístup k práci s korpusovými daty ve vztahu k teoretickým tvrzením může být na korpusu založený nebo korpusem řízený. Přístup korpusem řízený odvozuje deskriptivní kategorie na základě korpusových dat, přístup na korpusu založený obvykle vychází z teorií existujících vně a na korpusu ověřuje jejich relevanci. Přístup v této práci je korpusem řízený a omezuje se tak na minimální výchozí teoretický rámec, který je pro popis dat nutný a který si data sama vynucují.

Substantiva mohou mít jako valenční exponent jméno v určitém pádě, předložku, infinitiv, adverbium/adverbiále a vedlejší větu. Nejfrekventovanější, nejtypičtější a nejkompexnější valencí substantiv je valence genitivní. Tato práce se pokusí o klasifikaci genitivního valenčního exponentu vyskytujícího se v souvislosti se zkoumanými substantivy. Genitivní valenční exponent se u substantiv jeví jako produktivní rys, práce se pokusí tuto tendenci blíže popsat.

Práce dále nabízí popis dalších pádů vyskytujících se jako substantivní valence, nejběžněji se vyskytujících předložek, infinitivu a vedlejších vět. Dále zde pracujeme s rozšířeným pojmem valence, do kterého je zahrnuto např. i příslovecné určení. V těchto případech je kritériem pro určení „valenčnosti“ frekvence výskytu daného jevu a s tím související otázka sémantické a syntaktické fakultativnosti či obligatornosti. Budeme se zde zabývat i některými dalšími jevy systematicky se vyskytujícími na pozici valenčního exponentu nebo s ním jinak souvisejícími.

Tato korpusová studie si vzhledem k rozsáhlosti tématu nečiní nárok na vyčerpávající popis substantivní valence. Na daném vzorku substantiv se budeme snažit ukázat nové možnosti, které metodologie vycházející ze studia korpusů

může při popisu tradičních kategorií nabídnout. Kapitola 2 představí pojem valence. Dále se tato kapitola v odd. 2.1 věnuje jazyku z pohledu korpusové lingvistiky a nastíní její stručný vývoj. V odd. 2.2 představíme teorii a metodologii korpusové lingvistiky a s nimi související otázky a problémy. Kapitola 3 je věnována metodologii této práce a datům, z kterých práce vychází. Oddíl 3.1 nás stručně seznámí s Českým národním korpusem, odd. 3.2 se pak podrobně věnuje metodologii této práce.

Kapitola 4 tvoří jádro této práce. Odd. 4.1 se věnuje bezpředložkové genitivní valenci, odd. 4.2 pak valenci předložkové, odd. 4.3 ostatním pádům a odd. 4.4. až 4.7 dalším doplněním valenčního charakteru nebo jevům s valencí substantiv souvisejícím. Odd. 4.8 se stručně věnuje valenci vícemístné, odd. 4.9 stručně nastiňuje problematiku valence substantiv s vysokou frekvencí výskytu a odd. 4.10 se na valenci podívá z hlediska rozdělení substantiv na abstrakta a konkréta. Kapitola 5 se pak pokusí o shrnutí dané problematiky.

2 TEORETICKÝ RÁMEC

2.1 Pojem valence

Pojem valence je inspirován chemií a pravděpodobně i Saussurovým pojmem hodnoty (*valeur*), s kterým je tento pojem spjat i etymologicky (Čermák 1991:16). Duchovní otcovství pojmu valence v lingvistice je připisováno Lucienu Tesnièreovi (1969). Valence je však jen součástí jeho rozsáhlé koncepce strukturální syntaxe. U Tesnière je valence pojem syntaktický a spojuje se výlučně s verbem². Jedná se tedy zčásti o tradiční pojem slovesné reky a rozšíření pojmu tranzitivity. V současné lingvistice je valenční teorie etablovanou teorií především v lingvistice německé, vyšel zde i první významný valenční slovník (Helbig & Schenkel 1973). Později vyšly valenční slovníky i v řadě dalších jazyků.

Téměř všechny současné přístupy k valenci jsou rovněž orientovány syntakticky a o aspektech lexikálně-sémantických najdeme jen náhodné a periferní poznámky (Čermák 1991:15). Navíc, valence je stále primárně jako kategorie spojována se slovesy. Valence je zde součástí syntaktických teorií, které považují za základ věty sloveso. V české lingvistice nacházíme pojem valence např. v tzv. dvourovinné valenční syntaxi vypracované v 60. letech F. Danešem. Z této koncepce vychází i *Mluvnice češtiny* (1986), která valenci definuje následujícím způsobem:

Valenci rozumíme schopnost dominujícího členu vyžadovat jistý počet dominovaných, valenčních členů, zpravidla též v jistých tvarech, a je projevem jeho intence na rovině gramatické formy. Valence je především vlastností slovesných predikátů, u ostatních

² Le nombre de crochets que présente un verbe et par conséquent le nombre d'actants qu'il est susceptible de régir, constitue ce que nous appellerons la **valence** du verbe (Tesnière 1969:238).

slovních druhů můžeme sice rovněž s valencí počítat, ale má poněkud jiný charakter a vedle toho je ve většině případů ze slovesné valence odvozena.

(*Mluvnice češtiny* /3/ 1986:18)

V dalších českých mluvnicích najdeme valenci pojatou velmi podobně. Grepl a Karlík chápou pojem valence jako

schopnost určitých lexikálních jednotek (zejména sloves – včetně dějových a stavových substantiv – a adjektiv) vázat na sebe jistý počet syntaktických pozic, determinovaný počtem sémantických aktantů, a to v jistých morfologických formách obsaditelných primárně jistými výrazy, buď nevětnými (substantivy a výrazy adverbiálními), nebo větnými (vedlejšími větami)

(Grepl & Karlík 1998:45)

Karlík (2000) později pracuje s tzv. modifikovanou valenční teorií, která se pokouší najít pravidla pro rozlišení intence jako sémantické valence a způsobu její realizace jako syntaktické valence (Kopřivová 2006). Popisem valence se zabývá i tzv. funkční generativní popis³, který uspořádává jazykové výrazy a významy do jazykových rovin. Nejvyšší rovinou je zde rovina významu vět, tzv. tektogramatická rovina. Tektogramatický zápis, který jednoznačně charakterizuje význam věty, je zachycen formou tzv. závislostního stromu. Pojetí valence je v rámci tohoto popisu následující:

Souhrn všech možných typů doplnění daného slova označujeme jako jeho valenci (v širším smyslu)... Určitý druh doplnění je aktantem nebo volným doplněním ve všech svých výskytech. Naproti tomu ve vztahu k určitému řídicímu slovu může být dané doplnění buď obligatorní (nezbytně přítomné), nebo fakultativní. Tato opozice se tedy netýká přímo typů doplnění, ale jejich vztahu k jednotlivým lexikálním jednotkám (nebo jejich skupinám) vystupujícím jako jejich slova řídicí... Rozlišujeme tedy dva pojmy valenčního rámce: v užším smyslu do něho patří obligatorní i fakultativní aktanty daného slovesa a jeho obligatorní volná doplnění, v širším smyslu všechna doplnění, která mohou na daném slovesu záviset. Ačkoli tedy valenci (v širším smyslu slova) obecně vymezujeme u sloves i u jiných slovních druhů na základě toho, které typy doplnění na něm mohou nebo musejí záviset, v užším smyslu mluvíme o doplněních valenčních a nevalenčních.

(Hajičová et al. 2002:94–95)

³ Podle funkčního generativního popisu vzniká na MFF UK Pražský závislostní korpus, který má v současné verzi PDT 2.0 zhruba 2 mil. slov a je anotován nejen morfologicky, ale i syntakticky (1,5 mil. slov) a na rovině tektogramatické (informace na <http://ufal.mff.cuni.cz/pdt2.0/index-cz.html> v listopadu 2007).

Vztahem valence v rámci funkčního generativního popisu k dalším teoriím se zabývá např. Sgall (1998:368):

V přístupu nazvaném funkční generativní popis (FGP) se valenční pozice (závislostní relace) dělí na dvě skupiny, v základě odpovídající Tesnièreovým 'actants' a 'circonstants'. Jsou to jednak aktanty čili doplnění vnitřní..., jednak doplnění volná (u slovesa adverbialní). K operativním kritériím... patří zejména to, že volná doplnění mohou rozvíjet kterékoli řídicí slovo, a to i víc než jednou (při jeho jediném výskytu). Rozlišení jednotlivých aktantů je u Tesnière dáno čistě syntakticky a tak to zůstává ve FGP při prvních dvou: má-li sloveso ve svém rámci jediný aktant, je to agens, má-li dva, je tím druhým patiens jako typický hloubkový protějšek povrchového přímého předmětu. U dalších aktantů se přikláníme k Fillmoreovu třídění, založenému sémanticky (i pak jsou ovšem nutná testovatelná kritéria). Třetí a další aktant tedy může být adresát, původ nebo výsledek. Existují i slovesa, v jejichž rámci není žádný aktant, která tedy mohou mít jen volná doplnění.

Specificky valenci substantiv se věnuje Jiří Novotný (1980). Jeho *Valence dějových substantiv v češtině* se však, jak již název napovídá, věnuje pouze substantivům odvozeným od sloves a sleduje především, „k jakým změnám ve valenci při transformaci sloves v dějová substantiva dochází“ (1980:67). Popis valence, kterému se Novotný věnuje, je umístěn do roviny gramaticko-syntaktické. Specificky deverbativním substantivům a jejich valenci v češtině se věnuje Veronika Kolářová (2004, 2005).

Většina valenčních teorií (i těch syntaktických) se shodne na tom, že valence se týká počtu pozic, které to které slovo vyžaduje, aby bylo plně určeno. Ne vždy je však patrné, že jsou případy, kdy může valence např. sloužit i k rozlišení významu polysémických lexémů (např. Filipec & Čermák 1985:63 a 145)⁴. Z tohoto hlediska je však třeba valenci vidět ne pouze primárně jako kategorii syntaktickou, ale

⁴ Filipec a Čermák zde uvádějí mezi dalšími příklady adjektiva *bohatý*: *bohatý něčím; na něco; v něčem; bohatý člověk - bohatá úroda; život b. událostmi; látka bohatá na kyslík; jazyk bohatý ve výrazových prostředcích* (1985:63).

především lexikologickou a systémovou (tedy danou v jazykové struktuře) a až pak syntaktickou. Je to vlastnost slov, která by měla být systematicky uváděna ve slovnících, a to nejen ve specializovaných slovnících valenčních. Filipec a Čermák (1985:63) tedy např. valenci chápou jako „schopnost slova vázat na sebe další slova nutná k jeho syntaktické a významové realizaci“. Valence, podobně jako rekce či intence je gramatickým vyjádřením obecné schopnosti slova se v textu spojovat s jinými slovy, jde tedy o kombinovatelnost paradigmatu nebo jeho členu vůči jinému paradigmatu/členu. V systémovém přístupu (v Saussurově pojetí v *langue*) je valence chápána jako jedna z inherentních lexikálních vlastností slova. Valence má tedy paradigmatickou povahu, můžeme si např. v rámci téhož slova všimnout jedné valence v určitém protikladu k jiné valenci⁵ (Čermák 1991:18–19).

S valencí bývá dále spojována otázka obligatornosti a fakultativnosti (viz i výše uvedené definice), a to především v pohledu syntaktickém. V interpretaci je pak často problémem vymezení hranice mezi aktantem a doplněním volným. Otázku tzv. obligatornosti, fakultativnosti a potenciálnosti se snaží řešit více autorů. Např. Sgall v *Úvodu do syntaxe a sémantiky* (1986:129) upozorňuje na nutnost rozlišovat mezi obligatorností sémantickou a gramatickou. Nutnost doplnění se u některých sloves tak týká např. i příslovečných určení (*octnout se* „kde“, *chovat se* „jak“), která jsou tradičně považována za fakultativní větné členy. Autoři připomínají, že o „doplňujících“, tj. nutných příslovečných určeních mluví však např. už i Vladimír Šmilauer (1947).

⁵ např. *vzpomínat něčeho* – *vzpomínat na něco*

Při valenčních popisech tedy stojí v popředí zájmu tradičně verbum. Je však zřejmé, že po stránce valenční žádná ostrá hranice mezi slovesem a jinými slovními druhy nevede. V této souvislosti vyvstává otázka, které slovní druhy valenci mají (Čermák 1991:25). Chápeme-li valenci jako systémový syntagmatický (ve slovníku potenciální) vztah slova ke slovům za ním i před ním, jejichž interpretaci a formu predeterminuje, často i pomocí morfologického či lexikálního signálu, můžeme říci, že valenci má každé sloveso stejně jako i každé slovo libovolného slovníkového okruhu (s výjimkou interjekcí). Valence slova je dána již v systému a má povahu lexikální (v syntaxi je realizována), nemůže být ovlivněna volbou mluvčího ani potřebami komunikace (textu), může být jen jako možnost realizována či nikoliv (Čermák 2001:134).

Ačkoliv je tedy valence obvykle chápána jako kategorie gramatická, je třeba si klást otázku, zda není i do určité míry vázána na přítomnost určitého sémantického rysu v sémantice samotného slova. Valence by pak byla zdvojeným explicitním vyjádřením toho, co už je obsaženo v samém významu slova implicitně (Čermák 1991:21). Poznatků o sémantické kombinatorice slov a jejich pravidlech je však doposud málo. To, co se v povrchové syntaxi jeví jako podřízené (tedy i valence), může však být v sémantickém smyslu determinující (Čermák 1991:25).

Valence je pouze jednou ze syntagmatických vlastností slova, která má bezprostřední význam pro konstrukci gramatické stavby věty. Vedle ní je třeba rozlišovat kolokabilitu, která má sémantickou povahu. Valenci je možné chápat jako schopnost slova vyžadovat, aby např. člen jeho kolokačního paradigmatu byl určité gramatické povahy, tedy např. substantivum v nějakém určitém pádě.

„Přídorys“ valence tak tvoří kolokabilita slova, tedy úhrn jeho kolokačních paradigmat (Čermák 1991:27–28).

Chceme-li pro valenci najít místo v širším kontextu současné lingvistiky, můžeme se na ni např. dívat podobně jako Wolfgang Teubert (2007:223–249), který valenci vidí jako součást evropské kontinentální⁶ tradice – valence v tomto pojetí spojuje lokální gramatiku s lexikální gramatikou a obecnou syntaxí. Ve svém článku Teubert navrhuje třístupňový přístup. Nejvyšší rovinu zde představuje rovina sémantická, která zahrnuje vztahy mezi řídicím členem a členy závislými. Prostřední rovina je místem střetu sémantiky a syntaxe. Rozlišuje se zde mezi aktanty (v Teubertově terminologii *complement*) a doplnění volnými (*adjunct*). Tyto skupiny doplnění jsou shora definovány svou souvislostí se sémanticky danými hloubkovými pády (ve Fillmorově pojetí) a zdola způsoby svých existujících syntaktických realizací. Tato rovina je potřebná pro rozlišení mezi syntakticky identickými povrchovými strukturami, jejichž sémantická interpretace není jednoznačná. Třetí rovinu tvoří rovina syntaktické realizace, která specifikuje možnosti syntaktického realizování jednotlivých doplnění.

⁶ Tj. v protikladu k britské nebo šířeji pojato i angloamerické.

2.2 Jazyk z pohledu korpusové lingvistiky

Jazyk se nám ... zcela bezprostředně jeví jako systém znaků... jazyk je ve skutečnosti především něco jiného, totiž systém prvků, které jsou určeny k tomu zaujímat určitá místa v řetězu, vstupovat v určitá spojení, zatímco do jiných vstupovat nemohou. ... Každý prvek v jazyce náleží k určité kategorii, která je definována jistými možnostmi spojení a vyloučením jiných možností spojení.

(Hjelmslev 1971:35–36, 34)

Korpusová lingvistika vychází z tvrzení, že neadekvátnější popis jazyka může být zprostředkován pouze popisem jazyka přirozeného. Lingvisté se vždy zajímali o přirozený jazyk, nikdy však neměli možnost jej pozorovat tak systematickým způsobem jako prostřednictvím korpusů. Jazykový korpus⁷ je elektronickou kolekcí textů přirozeného jazyka a v dnešní době jsou tyto kolekce značně rozsáhlé. Příchod počítačů a vznik korpusů způsobily v lingvistice revoluci. Počítačová technologie totiž umožňuje uchovávání a zpracovávání obrovského množství textů.

Počátky korpusové lingvistiky sahají do poloviny minulého století, kdy Randolph Quirk začal sbírat data pro *Survey of English Usage*, ze které později vznikla mluvnice angličtiny *A Comprehensive Grammar of the English Language* (Quirk et al. 1985). Data sbíraná Quirkem však získala elektronickou podobu až na konci osmdesátých v rámci projektu International Corpus of English⁸. *Survey* obsahuje nejen texty psané, ale i mluvené. Mluvená část získala elektronickou podobu díky

⁷ Existuje samozřejmě několik různých více či méně přesných definicí korpusu. V dnešní době se většina z nich shodne na několika základních vlastnostech korpusu – je to kolekce textů 1) ve strojově čitelné podobě, 2) texty jsou autentické (mohou zahrnovat přepisy mluvené řeči), 3) jedná se o vzorky diskursu, které byly vybrány tak, aby 4) reprezentovaly určitý jazyk nebo jeho variantu. Zatímco první dvě vlastnosti nejsou nijak kontroverzní, názory na otázku reprezentativnosti se liší podobně jako názory na techniky vzorkování pro dosažení reprezentativnosti (McEnery et al. 2006:5).

⁸ Více informací lze získat na <http://www.ucl.ac.uk/english-usage/ice/>

Janu Svartvikovi v sedmdesátých letech a vznikl tak London Lund Corpus⁹. Druhým velkým korpusovým projektem byl v šedesátých letech minulého století korpus Brown Corpus¹⁰ (1 milión slov psané americké angličtiny) sestavený Nelsonem Francisem a Henrym Kučerou. V letech sedmdesátých pak vznikl podobný projekt pro angličtinu britskou Lancaster-Oslo-Bergen Corpus (LOB)¹¹.

Dalším raným korpusovým projektem byl projekt *English Lexical Studies*¹², který byl zahájen v roce 1963 pod vedením Johna Sinclaira. John Sinclair byl první, kdo využil korpusu především pro výzkum lexikonu a znovu oživil pojem kolokace, který ve třicátých letech zavedli H. Palmer a A. S. Hornby v *Second Interim Report on English Collocation* (Palmer 1933) a který později vlivně použil J. R. Firth ve svém článku *Modes of Meaning* (1951)¹³. Sinclairův korpus se ve spolupráci Birminghamské univerzity a nakladatelství Collins významně rozrostl a z projektu vznikl první, dnes již legendární slovník vycházející z korpusových dat *Collins Cobuild English Language Dictionary* (1987). S přibývajícemi technickými možnostmi v následujících letech se korpusový výzkum začal potýkat s problémy, které s sebou sestavování objemných korpusů přináší. Kromě problémů technických, ale i finančních to byla (a do určité míry stále ještě je) především otázka reprezentativnosti, standardizace, anotování, ale i copyrightu apod.

Možnosti, které moderní technologie lingvistiky nabízejí, však nebyly zpočátku mnohdy doceněny a význam korpusů pro lingvistický výzkum býval zlehčován. Význam korpusů je však nezastupitelný jako zdroj dat a poznání

⁹ Manuál vydaný k tomuto korpusu je v elektronické podobě k dispozici na <http://khnt.hit.uib.no/icame/manuals/LONDLUND/INDEX.HTM>

¹⁰ Manuál tohoto korpusu je k dispozici na <http://icame.uib.no/brown/bcm.html>

¹¹ Manuál tohoto korpusu lze najít na <http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM>

¹² Závěrečná zpráva z tohoto projektu byla v nedávné době publikována pod názvem *English Collocation Studies* (2004, ed. R. Krishnamurthy).

¹³ Jeden ze slavných a neustále opakovaných citátů J. R. Firtha ve vztahu k pojmu kolokace „You shall know a word by the company it keeps!“ najdeme v jeho článku *A Synopsis of linguistic theory, 1930–55* (Palmer 1968:179).

kombinatorických možností slov. Je nezastupitelný i např. při zkoumání frekvence určitých jazykových jevů. Korpusy nám mj. mohou ukázat i jevy, kterých si nejsme za normálních okolností vůbec vědomi. Jak Biber et al. (1998:3) podotýká, lidé si spíše všimnou jevů neobvyklých než typických. V této souvislosti je však třeba si uvědomit, že i korpus je jako zdroj dat omezený. Existence velkých korpusů tak otevírá především možnost zkoumání jevů vyskytujících se v jazykovém centru. Jako zdroj dat pro popis jazykové periferie jsou však existující korpusy stále nedostačující.

Vraťme se však k úvodnímu citátu z Hjelmslevova *Jazyka* (1971). Hjelmslev zde popisuje jazyk jako systém prvků, jejichž možnosti spojení jsou dány vnitřní organizací jazyka. Hjelmslev zde užívá neutrálního termínu prvek. Chápání toho, co si dnes lingvisté pod tímto pojmem představují, se však jistě značně liší – a to především v pojetí lingvistiky tradiční a lingvistiky korpusové.

Tradiční lingvistika totiž studuje lexikon a syntax jako dvě zcela odlišné a spolu nesouvisející části jazyka. Mezi lexikonem a syntaxí však žádná ostrá hranice nevede. Řez mezi lexikologií a syntaxí odmítal již Saussure, který uznával pouze jazykový systém a integrálně chápanou gramatiku. Saussure dokonce popírá obecnou existenci syntaxe jako takové, tj. syntaxe vně konkrétní kombinace slov (Čermák 1991:16–17; dále viz např. pozn. 251 v de Maurových poznámkách k textu *Kursu* (de Saussure 1996:415–419)).

Studium rozsáhlých korpusových materiálů přináší nové pohledy na jazyk. Jedním z centrálních problémů se jeví právě chápání onoho výše uvedeného prvku. V pojetí tradiční lingvistiky by se tento prvek jistě dal ztotožnit převážně se slovy¹⁴, nositeli významu. Avšak právě korpusová lingvistika přináší další materiál, který potvrzuje, že slovo jako jednotka k adekvátnímu popisu jazyka není nejvhodnější

¹⁴ „Slovo“ zde chápeme v jeho nejtradičnějším pojetí, tedy jako ortografické slovo. V tomto pojetí má slovo i svůj samostatný definovatelný význam.

a zdůrazňuje jeho závislost na okolním kontextu. Jednotka nesoucí lingvistický význam není tedy nutně ztotožnitelná se slovem.

Koncept slova jako základní jednotky popisu jazyka narušují především tři skutečnosti: polysémie a homonymie, existence několikaslovných jednotek a tzv. slov gramatických (Sinclair 2004a:1–2). Polysémie (a homonymie), tak jak ji chápe tradiční lingvistika, vlastně však z pohledu korpusové lingvistiky problémem není. Korpusový výzkum ukázal a potvrdil, že význam nevzniká z jednotlivých slov, ale je dán až souhrou slov v textu¹⁵ – v přirozených textech totiž velice zřídka nacházíme nejednoznačnost významu, kterou bychom museli jako mluvčí jazyka (na rozdíl od počítačů) nějak složitě řešit. Jde tedy vlastně o problém potenciální, v praxi jej téměř vždy vyřeší okolní slova, tedy textový kontext (nebo kontext celé situace). Mluvčí harmonizují výběr slov v rámci krátkých textových úseků a přirozeně tak řeší možný problém interpretace významu. Tento zásadní fakt však za běžných okolností součástí popisu jazyka nebývá a nepočítají s ním ani gramatiky ani slovníky (Sinclair 2004a:2).

Hovoříme-li o několikaslovných jednotkách, většinou si představíme idiomy a frazémy. Avšak význam je ve velké části textů, ve více než 50 % (a u některých typů textů je toto procento mnohem vyšší) tvořen slovy v kombinacích (Sinclair 2004a:3). U tzv. gramatických slov je význam tradičně těžko definovatelný. Lingvisté se sice shodují na existenci této skupiny slov, nejsou však schopni ji přesně definovat (Sinclair 2004a:2). Tato slova však hrají významnou roli v několikaslovných lexikálních jednotkách a nemohou být ignorována (Sinclair 2004a:4).

¹⁵ Máme zde na mysli textový význam, ani korpusová lingvistika se dosud neumí vyrovnat s langueovým významem lexému. Hranice lexému tedy nejsou jasné, jak nám ostatně dokazují i existující slovníky.

Slova nevytvářejí význam samostatně. Některé specializované termíny se mohou ubránit procesům, do kterých slova vstupují v textu, ale jakmile se stanou součástí běžného jazyka, podléhají pohybům významu a vstupují do kolokací, koligací¹⁶ atd. Při lexikálním popisu je tradičně pozornost věnována dvěma extrémům – izolovaným slovům a více či méně pevným několikaslovným jednotkám, např. idiomům a frazémům. Většina textu je však tvořena kolokacemi a víceslovnými slovními spojeními (Sinclair 2004b:421–422), která jsou více či méně pevná. Jejich větší či menší ustálenost je dána dvěma základními jazykovými principy kombinace slov v jazyce, tzv. „open-choice principle“ (princip otevřené volby) a „idiom principle“ (princip idiomu) (Sinclair 1991:109–115). „Open-choice principle“ je tradičním popisům jazyka nejbližší, v jeho rámci je volba slov v daném textovém úseku volná a jejím jediným kritériem je gramatická správnost a sémantická kompatibilita. „Idiom principle“ stojí na druhém konci, realizace výběru slov v daném textovém úseku je díky němu velmi omezená.¹⁷

Pojem kolokace se stal ústředním pojmem korpusové lingvistiky. V 50. letech jej užíval J. R. Firth a na jeho dílo později navázal především Michael Halliday a dnes již legendární postava korpusové lingvistiky John Sinclair. Kolokace je různými autory definována různě. Sinclair (1991:170) ji definuje jako „výskyt dvou nebo více slov v textu nedaleko od sebe. Obvyklou mírou blízkosti jsou maximálně čtyři intervenující slova. Kolokace mohou být dramatické a zajímavé, protože jsou neočekávané, nebo mohou být důležité pro lexikální strukturu jazyka, protože se často opakují. V lingvistických pojednáních obvykle máme na mysli

¹⁶ V kontextu současné korpusové lingvistiky hojně používaný pojem koligace definuje Firth jako „the interrelation of grammatical categories in syntactical structure“ (Palmer 1968:169).

¹⁷ Sinclair definuje *idiom principle* následovně (1991:173): „One of the main principles of the organization of language is that the choice of one word affects the choice of others in its vicinity. Collocation is one of the patterns of mutual choice, and idiom is another. The name given to this principle of organization which contrasts with the idiom principle is the open-choice principle.“ *Open-choice principle* pak definuje takto (ibid.:175): „In many descriptions of language-grammars and dictionaries especially words are treated as independent items of meaning. Each of them represents a separate choice. Collocations, idioms, and other exceptions to this principle are given lower status in the description.“

tento druhý typ kolokací, který je často spojovaný s měřením statistické významnosti“.

Čermák (2001:254) kolokaci definuje jako „(smysluplné) spojení lexémů/slov, lexikální syntagma, zvl. v podobě víceslovného pojmenování, jehož vznik je podmíněn jejich vzájemnou kolokabilitou, a tedy i kompatibilitou“. Ačkoliv se v lingvistice (a korpusové zvláště) s kolokací hojně pracuje, její definice a různé kolokační analýzy přinášejí především mnoho otázek. Tyto otázky jsou důležité nejen pro teorii jazyka zabývající se významem, ale mají i praktický důsledek. Objevují se specializované kolokační slovníky (kolokace by však jistě měly být systematicky zahrnuty i ve slovnících výkladových, alespoň ty statisticky významné¹⁸), které musí tyto a další otázky řešit.

Valenci korpusová lingvistika obecně, poněkud neprávem¹⁹, tolik pozornosti nevěnuje. Valence navíc jako pojem nemá ani v lingvistice tradiční jednotnou definici (viz kapitola 2.1). Valence je však spíše chápána jako pojem formální (morfologicko-syntaktický), avšak většina autorů připouští i existenci určitých sémantických aspektů. Chápeme-li však valenci jako systémovou lexikologickou kategorii, je v tomto smyslu v některých případech hranice mezi valencí a širší kolokabilitou (v nejširším významu ji chápeme jako spojovatelnost lexikálních jednotek) velice neostrá. Kolokabilita se v některých případech může jevit jako (nutná) sémantická specifikace valence. Korpusová analýza nám tak v tomto smyslu může přinést kromě dokonalejšího popisu i nový pohled na valenci.

¹⁸ Korpusová lingvistika používá několik různých statistických měř pro určení statistické významnosti kolokací. Názory na tyto výpočty se v komunitě korpusových lingvistů značně liší.

¹⁹ Korpusová lingvistika je v současné době z nejrůznějších důvodů stále ještě především doménou v anglicky hovořícím světě (angličtina také má k dispozici nejvíce korpusů a některé z nich mají co do velikosti opravdu úctyhodné parametry) a valence nebývá součástí tradičních gramatických popisů angličtiny (viz např. Teubert 2007:223-248).

2.3 Korpusová lingvistika: teorie a metodologie

Korpusová lingvistika je mnohými stále nazírána především jako metodologie. Korpusová lingvistika však v posledním desetiletí prošla značným vývojem. Zatímco zpočátku se korpusová lingvistika zdála být metodologií především pro aplikovanou lingvistiku (hlavně lexikografii, ale i např. pro výuku angličtiny), nabízí dnes pohled na jazyk, který ji odlišuje od lingvistiky tradiční i kognitivní (Teubert 2005).

Na rozdíl od ostatních lingvistických přístupů, které se spoléhají především na jazykovou introspekci, korpusová lingvistika vychází z jazykových dat. Elena Tognini-Bonelli (2001) tento vývoj dokládá na tzv. *corpus-driven* přístupu, tedy analýze pouze těch dat, které nabízí korpus. Vztah korpusové lingvistiky k existujícím lingvistickým teoriím Tognini-Bonelli (2001) shrnula do dvou základních přístupů: přístupu založeného na korpusu (*corpus-based*)²⁰ a přístupu korpusem řízeného (*corpus-driven*)²¹. U korpusem řízeného přístupu by měla teoretická tvrzení být odrazem závěrů učiněných na základě analýzy dat a metodologický postup je charakterizován jako postupující od pozorování přes hypotézu k závěrům a teoretickému tvrzení (Tognini-Bonelli 2001:85). Co to však znamená v praxi?

Představa jazykové teorie opravdu vycházející z jazykových dat je jedna věc, avšak praktická proveditelnost postupů navržených Tognini-Bonelli (2001) je věc druhá. Dnešní korpusy nabízejí natolik rozsáhlé množství dat, že technické

²⁰ „...*corpus-based* is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study“ (Tognini-Bonelli 2001:65)

²¹ „...*corpus-driven* approach... aims to derive linguistic categories systematically from the recurrent patterns and the frequency distributions that emerge from language in context“ (Tognini-Bonelli 2001:87)

možnosti jejich zpracování nevyhnutelně ovlivňují i z nich odvozené lingvistické závěry.²² V současném korpusovém výzkumu tedy ve skutečnosti nevede žádná ostrá hranice mezi přístupem „corpus-based“ a „corpus-driven“. Jde spíše o to více či méně.²³

V komunitě korpusových lingvistů však v poslední době vzniká snaha vyprofilovat teoretický status korpusové lingvistiky. Někdejší šéfredaktor odborného časopisu *International Journal of Corpus Linguistics*, mezinárodního fóra korpusových lingvistů, Wolfgang Teubert např. v prvním čísle jubilejního desátého ročníku publikoval 25 tezí „své verze korpusové lingvistiky“ (Teubert 2005). V tomto pojetí je korpusová lingvistika zaměřena především na význam. Na každou víceslovnou jednotku, frázi, část textu atd. je nutno nahlížet ze dvou na sobě závislých úhlů pohledu – na jejich formu a význam. Na jedné straně forma reprezentuje význam a na druhé význam nemůže existovat bez své formy. Význam nacházíme v diskursu. Korpusová lingvistika je empirickou disciplínou, která pracuje s daty přirozeného jazyka, využívá kategorií tradiční lingvistiky, avšak nebere je jako samozřejmost. Korpusová lingvistika jako taková není metodou, ale využívá při zpracování a analýze dat mnoha různých metod. Je pro ni charakteristická práce pouze s daty přirozeného jazyka, která jsou systematickým způsobem získaná z diskursu a která tvoří jazykový korpus. Korpusová data umožňují pohledy na jazyk, který nebyl dříve možný. Teorie a kategorie odvozené z introspektivního studia jazyka nebo převzaté z jiných oborů (např. programování) nemusí být pro popis přirozeného jazyka nejvhodnější. Korpusová lingvistika zpochybňuje tradiční postavení slova jako základní lingvistické jednotky nesoucí význam a pracuje s jazykovými jednotkami, které jsou formálně spojeny lokální gramatikou, a užívá termínů jako lexikální prvek

²² Více viz např. Mahlberg 2005:18–22.

²³ Avšak na toto téma se vedou debaty a někteří korpusoví lingvisté se striktně hlásí do jednoho nebo druhého z táborů. O sporu více např. v McEnery et al. (2006:6–11).

nebo významová jednotka.²⁴ Tyto prvky nebo jednotky mohou být jednotlivá slova, víceslovné jednotky, fráze nebo idiomy. Podobně jako se v diskursu opakují jednotlivá slova, opakují se i tyto jednotky. Frekvence výskytu je tak důležitým parametrem při zjišťování opakujících se slovních vzorců. Ať již se jedná o jednotlivá slova nebo složené významové jednotky, jsou lexikální prvky, na rozdíl od slov, v zásadě monosémické. (Teubert 2005:1–14) Korpusová lingvistika je tak v současné době v tomto pojetí především *parole*-ovou lingvistikou, je tak mnohem méně zaměřena na jazykový systém, který byl a je tradičně v popředí zájmu lingvistů.

Elektronické korpusy nám nabízejí přístup k nevídanému množství dat. Ve skutečnosti je tak problémem současné korpusové lingvistiky otázka, jak s takovým množstvím dat naložit. Jsme omezeni technickými možnostmi. Počítače nám nabízejí možnosti zpracovat frekvenční údaje různými statistickými výpočty. V současnosti existuje několik statistických měr standardně užívaných v korpusové lingvistice. Statistika tak mnohdy nabízí velmi elegantní způsob zpracování a seřazení nepřehledných několikamístných čísel. Úskalím plynoucím ze statistického zpracování však může být, že budou tradiční lingvistické kategorie pouze kvantifikovány (více viz. např. Mahlberg 2005:16). Jakkoliv nám tedy statistické míry mnohdy nabízejí zajímavé výsledky, pomohou určit centrum a periférii jazykových jevů nebo dokonce v některých případech i některé jevy a kombinace přímo objevit, mohou být nové pohledy na jazyk i zmařeny či zkresleny takovouto filtrací dat. Navíc např. žádná z existujících statistických měr pro výpočet kolokací není schopná všechny kolokace objevit opravdu spolehlivě. Korpus je tedy cenným a zatím nepřekonatelným zdrojem poznání jazyka, avšak je třeba si na druhou stranu připustit i skutečnost, že pokud něco není v korpusu (být obrovským), neznamená to, že to neexistuje.

²⁴ lexical item / unit of meaning

Korpusově pojatý přístup k jazyku, který chce být otevřený novým pohledům, musí tedy nalézt takový způsob popisu jazyka, který je dostatečně pružný, aby mohl do sebe zahrnout nová poznání vycházející z korpusu. K tradiční gramatické terminologii je někdy třeba přistupovat s rezervami. Analýza korpusových dat by tak měla ideálně vycházet z minimálních teoretických východisek (srov. mj. např. Mahlberg 2005:16 a 29 a Tognini-Bonelli 2001) a hypotézy založené na pozorování dat.

3 METODOLOGIE A DATA

3.1 Český národní korpus

Pro vyhledávání valence pro tuto práci byl použit korpus SYN2000, který je součástí projektu Český národní korpus²⁵. SYN2000 je lematizovaný²⁶ a taggovaný²⁷ korpus, jehož rozsah je 100 miliónů slov. Jedná se o reprezentativní²⁸ synchronní korpus vytvořený v letech 1996–2000 obsahující české texty do roku 1999. Tyto texty jsou rozděleny v následujícím poměru: publicistika 60 %, odborné a tematické texty (25 %) a beletrie 15 %. Mezi odbornými a publicistickými texty jsou texty vydané po roce 1990, zatímco do beletrie byla zahrnuta díla v současnosti vydávaná a čtená, včetně překladů. S korpusem lze pracovat pomocí vyhledávacího programu Bonito. Vyhledávat lze na základě různých uživatelem zadaných kritérií a parametrů, včetně vyhledávání tzv. náhodných vzorků. Výstupy jsou v tradičním, uživatelem specifikovaném, formátu KWIC²⁹. Ústav Českého národního korpusu vydal pro uživatele korpusu podrobnou příručku (*Úvod a příručka uživatele* (2000)), informace lze najít i na webových stránkách Českého národního korpusu.

O složení SYN2000 bylo rozhodnuto na základě výzkumů recepce psaného jazyka. Největší zastoupení tak zde má publicistika. Publicistika jako taková je z hlediska jazykového do značné míry specifická a poměrně velké proporční zastoupení publicistiky v SYN2000 má tak menší či větší vliv na výsledky analýz

²⁵ <http://ucnk.ff.cuni.cz/>

²⁶ Každému tvaru je přiřazeno jeho lemma, na jehož základě lze v korpusu vyhledávat.

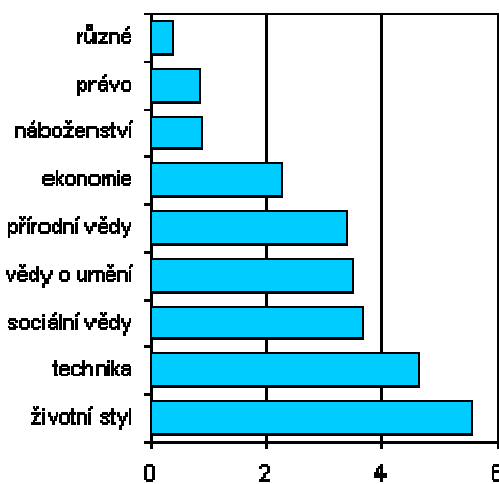
²⁷ Každý tvar má svou morfologickou značku, která zachycuje souhrn morfologických vlastností daného slovního tvaru, v SYN2000 tedy lze vyhledávat i pomocí tagu.

²⁸ O otázce reprezentativnosti více viz např. McEnery et al. (2006:13–16).

²⁹ KWIC=Key Word In Context

dat tohoto korpusu. Čtvrtinu textů dále tvoří texty odborné, jejichž jazyk rovněž nese určité charakteristiky. Na výsledky zkoumání dat v SYN2000 je tedy nutno nahlížet v tomto kontextu, a v tomto kontextu je tedy definována i reprezentativnost pro SYN2000. Termín 'reprezentativnost' se v korpusové lingvistice obvykle vztahuje k otázce, do jaké míry dovolují výsledky analýzy dat daného korpusu zobecnění.

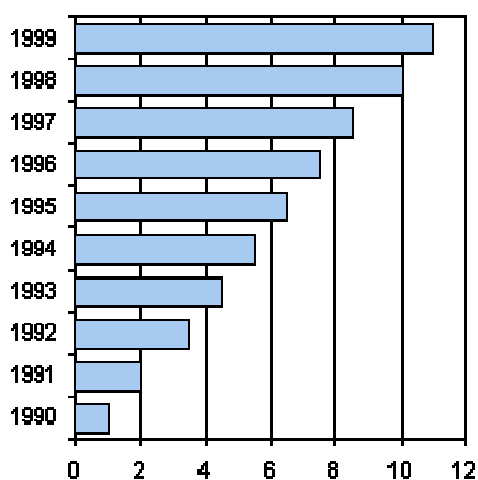
Následující obrázek převzatý z webových stránek Českého národního korpusu ukazuje rozložení odborné a další specializované literatury v SYN2000 podle odborného zaměření. Horizontální osa představuje počet slov v miliónech. Odborná literatura tvoří 25 % textů.



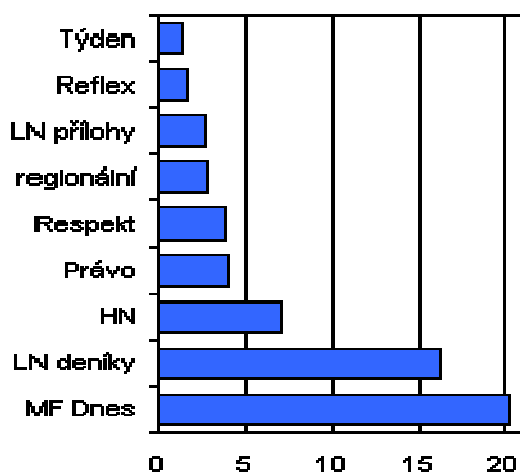
Obr. 1. Rozložení odborné a další specializované literatury v SYN2000 podle odborného zaměření.

Následující dva obrázky rovněž převzaté z webových stránek Českého národního korpusu nám názorně poskytují podrobnější informace o největším komponentu korpusu SYN2000 – publicistice, která zaujímá 60 %. Horizontální osa představuje u obou obrázků počet slov v mil. První obrázek zobrazuje proporční

zastoupení publicistiky podle data vydání a druhý obrázek složení publicistiky podle zastoupených titulů.



Obr. 2. Rozložení publicistiky v SYN2000 podle data vydání.



Obr. 3. Rozložení publicistiky v SYN2000 podle jednotlivých titulů.

3.2 Výběr substantiv a jejich analýza

SYN2000 svým rozsahem nabízí jedinečný zdroj dat, která reprezentují současný český jazyk³⁰. Takto veliký rozsah dat však má při svém zpracování určitá úskalí. Je evidentní, že při takovémto rozsahu nelze manuálně projít a spolehlivě zpracovat všechna relevantní data, které nám korpus nabízí. Pro jazykovou analýzu je třeba z tohoto množství učinit pokud možno reprezentativní výběr.

Jak bylo řečeno výše, tato práce pojímá valenci substantiv jako lexikální vlastnost jednotlivých slov a popis valence substantiv vychází z korpusu (viz 2.3). Tato práce se pokouší o systematický popis všech důležitých aspektů substantivní valence. U systematického popisu je důležitá otázka velikosti zkoumaného vzorku a jeho reprezentativnost. Vzorek musí být dostatečně veliký a reprezentativní, aby bylo zachyceno nejen to, co je pro valenci substantiv typické, ale i většina problémů dané oblasti. Při práci s takto velikým zdrojem dat je však třeba zohlednit i praktické možnosti analýzy a popisu.

Morfologická analýza i lemmatizace korpusu probíhá automaticky a díky vysoké tvarové homonymii češtiny jsou tak zde mnohé chyby (na některé z nich upozorníme dále). S tímto problémem pak souvisí i přesnost/nepřesnost získaných frekvenčních údajů. Pro práci využíváme lemmatizace jak při vyhledávání, tak i při samotném počátečním výběru substantiv. Předpokládáme, jak už bylo řečeno výše, že valence je aspekt lexému, a nikoliv pouze jeho tvaru. V této práci pracujeme většinou pouze s jednoduchými statistickými údaji (procentuální zastoupení), abychom se vyvarovali větších chyb a zkreslení, které v některých

³⁰ Tedy jazyk do roku 2000. Ústav Českého národního korpusu nedávno zpřístupnil další 100milionový korpus současné psané češtiny SYN2005, který se od SYN2000 liší v několika ohledech, mj. publicistika zde má nižší zastoupení (pouze 33 %) a jde o tituly vydané po roce 2000.

případech statistické metody v korpusové lingvistice jinak obvyklé (např. MI-score, t-score apod.) mohou ne zcela průhlednou filtrací dat způsobit (více viz kapitola 2.3). Analýza dat probíhala na zmenšených vzorcích a frekvenční údaje, které tato práce nabízí, se vztahují ke zkoumaným vzorkům. Jde tedy spíše o orientační údaje, které nám pomohou oddělit úzus typický od okrajových výskytů.

Za účelem vyhledávání valence substantiv byla valence jako kategorie pojímána poměrně volně. Valenci substantiv předpokládáme napravo od substantiva a ponecháváme zde možnost ovlivnění rámce až daty samými. Valence byla vyhledávána na náhodných vzorcích dat (viz dále odd. 3.2.2), jejichž zpracování bylo převážně manuální. V souladu s metodou *corpus-driven* (analýza korpusem řízená), tak jak ji charakterizuje Tognini-Bonelli (2001:85 a viz odd. 2.3), jsou závěry v této práci činěny pouze na základě analýzy dat a metodologický postup začíná od pozorování a pracovní hypotézy. Za tímto účelem byla nejprve provedena malá pilotní studie. Pro tuto studii byla provedena analýza substantiv v nižších frekvenčních pásmech. Výhodu pohybu v nižších frekvenčních pásmech představuje skutečnost, že máme možnost prozkoumat daný vzorek v jeho úplnosti a nemusíme se omezovat na náhodný výběr tak, jako je tomu u velkého množství dat (a později u vlastní analýzy).

3.2.1 Pilotní studie

Substantiva pro analýzu v rámci pilotní studie valence substantiv byla vybrána náhodným výběrem z nižších frekvenčních pásem. Zcela v souladu se Zipfovým zákonem (Zipf 1935) je i na konci frekvenčního slovníku češtiny³¹ na jednotlivých pozicích podstatně více slov než na jeho začátku³². Zkušební manuální výběr ve frekvenčním pásmu 230–240 výskytů představoval 195 substantiv. Z této skupiny bylo vybráno každé osmé³³ substantivum a byl tak získán vzorek 25 lemmat. Podobným způsobem byly vybrány další dva vzorky po 25 substantivech ze dvou dalších frekvenčních pásem. Druhým pásmem byla frekvence výskytu 430–450, ve kterém se nacházelo 154 substantiv a bylo vybráno každé šesté substantivum. Posledním frekvenčním pásmem byla frekvence výskytu 630–660, kde bylo 138 substantiv a bylo tak vybráno každé páté substantivum. Substantiva vybraná pro analýzu v rámci pilotní studie jsou uvedena v příloze A.

Jak již bylo řečeno, výhodou práce s daty v nízkých frekvenčních pásmech je možnost projít všechny výskyty daného slova v korpuse. Další parametr, který je nutno při práci s korpusem stanovit, je délka kontextu. Ta se liší jazyk od jazyka a je samozřejmě dána i povahou zkoumaného problému. Pro češtinu vzhledem k jejímu volnějšímu slovosledu musí toto rozpětí být samozřejmě větší než např. pro

³¹ V době výběru substantiv jak pro pilotní studii, tak pro práci samou ještě nebyl k dispozici nový frekvenční slovník českého jazyka, proto zde pracujeme s hrubým frekvenčním seznamem lemmat, který byl k dispozici na podzim roku 2001.

³² Pro srovnání *Frekvenční slovník češtiny* (2004) uvádí pouze 75 lemmat s vyšší frekvencí než 100 000 výskytů.

³³ Substantiva jsou vybírána zcela náhodně, jako každé n -té. Velikost n závisí na konečné velikosti vzorku, který chceme analyzovat, a na množství substantiv v daném frekvenčním pásmu.

angličtinu³⁴. U valence se však dá očekávat, že se bude nacházet bezprostředně v okolí daného slova a to především na pravé straně.

Pro pilotní studii bylo využito především materiálu na CD-ROM vydaného Ústavem Českého národního korpusu (2001), který by měl představovat reprezentativní vzorek z ČNK³⁵. Na tomto CD jsou k dispozici dva korpusy: SYNEK (desetimilionová „zmenšenina“ velkého korpusu SYN2000, u kterého jsou zachovány proporce stylového rozložení textů: 60 % publicistiky, 25 % odborné literatury a 15 % beletrie) a LITERA (vzorek české prózy 20. století). Pilotní studie využívá dat z korpusu SYNEK. Výsledky této pilotní studie (viz dále 3.3) naznačily, jakým způsobem je valence substantiv v češtině realizována a na co bychom se měli ve vlastní analýze zaměřit.

³⁴ Ideální délku kontextu pro angličtinu již v 70. letech vypočítal B. Daley ze Sinclairova týmu na 4 slova nalevo i napravo od KWIC. Později byl tento výpočet znovu ověřován na větším korpuse angličtiny se zjištěním, že ze sémantického hlediska lepší výsledky může přinést kontext pěti slov nalevo a čtyř slov napravo, tedy téměř stejný výsledek (Krishnamurthy 2004:xix).

³⁵ CD Český národní korpus, Vzorek českého jazyka a literatury, 2000

3.2.2 Vlastní analýza

Substantiva pro vlastní analýzu byla vybrána podobně jako pro pilotní studii na základě náhodného výběru z frekvenčního seznamu lemmat ze SYN2000. Jednalo se o pracovní verzi pro pozdější frekvenční slovník, ve které se bohužel ještě vyskytovaly chyby způsobené nesprávnou lemmatizací (např. *soudnost/nesoudnost* tvoří jedno lemma; lemma *nora* v sobě zahrnuje anglické *nor*, ženské jméno *Nora*, občany Norska i zvířecí *noru*; *saň* a *saně* jsou lemmatizovány dohromady; *alkoholik* a *alkoholička* tvoří jedno lemma; *baba* v sobě zahrnuje i anglické *baby* apod.). Jelikož však je náš konečný výběr substantiv náhodný, tyto chyby pro nás nejsou nijak závažné.

Vycházíme z předpokladu, že pozice slova ve frekvenčním slovníku není náhodná (např. Zipf 1935). Slova na vrcholu pomyslného frekvenčního žebříčku se mohou chovat jinak než slova na jeho konci. Lingvisté se u obou skupin, tedy skupiny velice frekventovaných versus velice řídkých slov, potýkají s praktickými problémy při pozorování jejich chování. U slov frekventovaných je to problém obrovského množství dat a jeho velice obtížného zpracování a u slov řídkých je to problém přesně opačný – nedostatečné množství dokladů o výskytu daných slov, a tudíž nemožnosti přesvědčivého popisu a lingvistických závěrů. Stejný problém se samozřejmě týká i této práce. Částečně jsme se tomuto problému vyhnuli tím, že jsme analýzu zaměřili spíše na střední frekvenční pásmo.

Pro základní analýzu byla substantiva vybrána na čistě statistickém základě a výběr se snaží pokrýt střední a vyšší frekvenční pásmo výskytu. Z frekvenčního seznamu lemmat s frekvencí nad 1000 výskytů v ČNK jsme vybrali všechna substantiva (výběr byl proveden manuálně a vynechána byla vlastní jména, zkratky, měny, míry, měsíce, dny v týdnu a cizí slova) a získali jsme tak 3 494

substantiv. Z této skupiny bylo dále vybráno každé dvacáté substantivum, čímž jsme získali skupinu 173 substantiv. Jádrem analýzy této práce pak tvoří 99 substantiv od konce tohoto seznamu, pohybujeme se tak ve frekvenčním pásmu 1005 až 3339.

Skupinu vybraných substantiv tvoří následující lemmata (číslo před slovem udává celkový počet výskytů v SYN2000³⁶):

3339 zájezd, 3291 civilizace, 3227 vyhlášení, 3158 sval, 3113 funkcionář, 3079 nález, 3025 plátno, 2990 pohoda, 2934 mítink, 2889 pojistné, 2846 vila, 2800 ateliér, 2759 odstranění, 2714 stanovení, 2684 pasáž, 2640 chata, 2596 poskytování, 2556 udržení, 2520 vzhled, 2479 příznak, 2455 stáří, 2421 pec, 2395 zodpovědnost, 2357 recenze, 2328 monopol, 2308 obliba, 2279 soustředění, 2252 korupce, 2220 odjezd, 2191 ves, 2151 pleť, 2124 rozpaky, 2087 výdělek, 2068 porost, 2040 dlužník, 2019 stanoviště, 1990 pomník, 1960 pero, 1942 konání, 1922 krb, 1899 deprese, 1873 vaření, 1843 samec, 1822 obilí, 1793 sonda, 1773 odpočinek, 1748 tolerance, 1729 housle, 1712 dostih, 1689 směrnice, 1665 křivka, 1652 kuře, 1638 pomocník, 1623 policajt, 1604 viník, 1583 koridor, 1561 letectvo, 1544 vlček, 1533 školka, 1520 zavádění, 1509 důležitost, 1488 vdova, 1470 pocta, 1450 chudák, 1437 povstání, 1417 smetana, 1401 recese, 1389 tribuna, 1375 vzpoura, 1362 operátor, 1344 sklonek, 1325 šátek, 1308 maličkost, 1294 rukavice, 1277 dodatek, 1266 žíla, 1249 mor, 1234 bankovníctví, 1224 klika, 1215 kompresor, 1205 plemeno, 1196 disident, 1187 hádka, 1179 ubytovna, 1166 volnost, 1152 plavec, 1140 citát, 1130 partnerka, 1121 zástupkyně, 1112 žlutek, 1102 moucha, 1092 kolaps, 1082 tušení, 1071 překročení, 1062 zamyšlení, 1050 respektování, 1038 klávesnice, 1027 baba, 1020 ambasáda, 1005 podzemí.

Při analýze jsme pak pracovali s náhodnými vzorky, tak jak je nabízel program Bonito. Mělo-li substantivum frekvenci výskytu nižší než 1500, byla velikost vzorku 300 náhodných konkordančních řádků, u frekvence do 2000 byla velikost vzorku 400 a u frekvence vyšší než 2000 je pak velikost vzorku 500 konkordančních řádků.³⁷

³⁶ Frekvenční údaje mohou být ve vztahu k *Frekvenčnímu slovníku češtiny* (2004) jiné.

³⁷ Skutečné číslo mohlo být mírně nižší následkem špatné lemmatizace.

3.3 Závěry pilotní studie

Na základě předběžné analýzy má valence českých substantiv následující podobu. Nejčastějším valenčním exponentem je (1) bezpředložkový genitiv. Genitiv pak může mít i (1a) více významů, jeho sémantice je proto třeba věnovat zvýšenou pozornost. Např. substantivum *chorál* má dva sémantické typy genitivních exponentů: genitiv může označovat, kdo je původcem/autorem chorálu (*chorál domácího publika*) nebo, co je jeho předmětem (*chorál výčitek*). Podobně např. genitiv u substantiva *podloží* může označovat materiál podloží nebo i místo, kde se podloží nachází atd. Genitivní valenční exponent může u některých substantiv následovat i nějaká (1b) další valence, např. jiný pád nebo předložka – např. substantivum *ohledání* bývá následováno bezpředložkovým genitivem a dále instrumentálem (*po ohledání místa policisty byly nalezeny...*).

Další (2) pády nacházíme na pozici valenčních exponentů méně – např. dativ u substantiva *léčka*. Valence je dále u substantiv zprostředkována různými (3) předložkami (např. *léčka na, zádrhel v*). Z analýzy dále vyplývá, že u některých substantiv je vhodné si všimnout i zvýšené frekvence výskytu určitého typu okolí – např. častější (ze statistického hlediska) výskyt (4) příslovečného určení místa. Príslovečné určení místa je syntaktickou kategorií a bývá považováno za fakultativní větný člen, avšak v souvislosti s valencí bývá někdy uváděno jako člen obligatorní (více viz 2.1). Ačkoliv se tedy jedná primárně o kategorii syntaktickou³⁸, vypůjčíme si dále toto označení a budeme se příslovečnému určení věnovat v kapitole 4.4.

V souvislosti s valencí méně zřejmé, avšak u některých substantiv frekvenčně velice výrazné jsou dva další jevy, kterých si budeme při další analýze všimnout.

³⁸ Adverbium však je v některých případech součástí systému, např. *dopadnout výborně*.

Jsou to dvě substantiva, která jsou nápadně často (5) spojená spojkou *a*, přičemž tato dvě substantiva spolu vždy nějakým způsobem sémanticky souvisejí – např. *soudnost a rozum/opatrnost, romantika a něžnost/dobrodružství* (více odd. 4.5). A dále jsou to (6) dvě substantiva ve stejném pádě, přičemž druhé substantivum je často proprium, např. *cestovatel* + vlastní jméno (více viz 4.6).

Pozdější vlastní analýza ukázala, že pilotní studie naznačila problematiku dané oblasti poměrně spolehlivým způsobem. V pilotní studii jsme pouze nenarazili na (7) infinitiv jako valenční doplnění, tento jev je méně frekventovaný a i vlastní studie odhalila pouze několik málo výskytů (odd. 4.7.1). K valenci dále patří i (8) vedlejší věta (odd. 4.7.2).

Struktura čtvrté kapitoly, která představuje výsledky korpusové analýzy substantiv, zohledňuje frekvenci výskytu jednotlivých valenčních exponentů. Odd. 4.1 se zabývá bezpředložkovou genitivní valenci, odd. 4.2 se věnuje předložkové valenci, odd. 4.3 ostatním pádům, tedy instrumentálu a dativu, odd. 4.4. se věnuje adverbciále, odd. 4.5 koordinovaným spojením se spojkou *a*, odd. 4.6 nominativní valenci a odd. 4.7 dalším konstrukcím. Najdeme zde valenci infinitivní, vedlejší větu a některé další prostředky mající valenční charakter.

3.4 Rozšířené hledání valence substantiv

Závěry a data v této práci jsou, pokud není uvedeno jinak, činěny na základě výsledků analýzy skupiny substantiv uvedené v odd. 3.2.2. V některých případech jsme považovali za vhodné tuto základní analýzu doplnit vyhledáním dalších dat pomocí specializovaných dotazů do korpusu. Tato další dodatečná (v některých případech poměrně rozsáhlá) vyhledávání jsou vždy specifikována dále u konkrétních příkladů.

4 SUBSTANTIVUM A JEHO VALENCE Z POHLEDU ČNK

Jádro analýzy této práce vychází z podrobné analýzy náhodných vzorků náhodně vybraných 99 substantiv ze středního frekvenčního pásma (viz odd. 3.2 a 3.2.2). Na základě výsledků analýzy provedené v rámci pilotní studie jsme se zaměřili na bezpředložkovou genitivní valenci (odd. 4.1) a její sémantickou analýzu (odd. 4.1.3). V souvislosti s genitivním valenčním exponentem jsme se podrobněji věnovali i jeho řetězení, tedy výskytu několika genitivů za sebou (odd. 4.1.4). Odd. 4.1.5 je věnován konkurenci genitivu a atributivního adjektiva a odd. 4.1.6 alternativnímu vyjádření genitivního doplnění.

Odd. 4.2 popisuje předložkovou valenci a je rozdělen podle jednotlivých předložek. Odd. 4.3 se věnuje ostatním pádům vyskytujícím se v našich datech na pozici valenčního exponentu, tedy instrumentálu a dativu. Odd. 4.4 se zabývá otázkou „valenčnosti“ příslovečného určení, odd. 4.5 se věnuje koordinovaným spojením, tedy dvěma substantivům spojeným spojkou *a* vyskytujícím se ve frekvenčně významnější míře a odd. 4.6 popisuje valenci nominativní. Odd. 4.7 nabízí pohled na některé další konstrukce a výrazy, které mají valenční charakter, a to především infinitiv (4.7.1) a vedlejší větu (4.7.2).

Odd. 4.8 krátce pojedná vícemístnou valenci, odd. 4.9 se pak stručně věnuje některým charakteristikám substantiv s vysokou frekvencí výskytu. Závěrečný odd. 4.10 se pokusí shrnout rozdíly mezi abstrakty a konkréty.

4.1 Genitivní valence

V této kapitole se pokusíme na základě korpusové analýzy substantiv ze SYN2000 popsat bezpředložkovou genitivní valenci. Bepředložkový genitivní valenční exponent se jeví jako vůbec nejběžnější valence českých substantiv. Genitiv jako doplnění na pozici valenčního exponentu se ve větší či menší míře vyskytuje dokonce u více než 90 % substantiv (podrobnější data viz odd. 4.1.7), a je tedy na místě si položit otázku, zda jsou všechny tyto genitivy valencí tak, jak ji chápeme v této práci – tedy, zda jsou součástí lexikálního profilu daných substantiv. Některé z těchto genitivů nabízejí i interpretaci jako volné syntaktické realizace atributu podobně jako např. většina atributivních adjektiv.

Stanovení kritérií pro stanovení míry „valenčnosti“ genitivního doplnění není jednoduché. Pro stanovení významnosti určitého jevu se v korpusové lingvistice často používá údajů o frekvenci výskytu (o frekvenci více viz i 2.3), která může být pro začátek určitým vodítkem i zde.³⁹ Zvýšená frekvence výskytu nám signalizuje, co je typické. Přehled genitivního doplnění jsme tedy rozdělili do dvou hlavních částí. Odd. 4.1.1 pojednává substantiva, u kterých se genitiv jako valenční doplnění vyskytuje ve více než 50 % všech výskytů daného substantiva. Odd. 4.1.2 pak pojednává ostatní substantiva, u kterých se vyskytuje genitivní doplnění, byť i třeba jen okrajově.

Genitiv je pád sémanticky polyfunkční a tomuto aspektu budeme u jednotlivých případech věnovat další pozornost. Substantiva pojednávána v kapitole 4.1.2 jsou z tohoto hlediska rozdělena do dvou částí. Odd. 4.1.2.1 uvádí substantiva s jedním sémantickým typem genitivního doplnění a odd. 4.1.2.2 uvádí substantiva s více

³⁹ Zvýšená frekvence výskytu určitého jevu potvrzuje jeho významnost ze statistického hlediska a v našem případě může být vodítkem pro stanovení valence. Na základě analýzy dat je však patrné, že valence ne vždy musí nutně mít zvýšenou frekvenci výskytu.

sémantickými typy genitivního doplnění. Odd. 4.1.3 se pak pokusí shrnout vyskytující se sémantické typy genitivního valenčního doplnění.

Dalším zajímavým aspektem genitivu jako valenčního doplnění substantiv je s významem daných substantiv související řetězení genitivů, jejichž prostřednictvím je sémantika těchto substantiv upřesňována. V povrchové syntaxi jsou tyto genitivy členy podřízené, ale z hlediska sémantického může být vztah podřízenosti–nadřízenosti i opačný nebo nejednoznačný. Několikanásobné genitivní valenci se věnuje odd. 4.1.4.

V povrchové syntaxi je genitivní doplnění podobně jako atributivní rozvíjející adjektivum přívlastkem substantiva. Valenční povaha genitivního doplnění je u některých substantiv zcela zřejmá a je zřejmá např. i z výrazné frekvence výskytu, avšak u některých substantiv je oslabená a genitivní doplnění zde má produktivní povahu a má tak určitou podobnost s atributivním adjektivem. V této souvislosti je tak zajímavé se podívat na konkurenci genitivního doplnění a atributivního adjektiva. Této problematice se věnuje odd. 4.1.5. Odd. 4.1.6 pak upozorní na některá alternativní vyjádření genitivního valenčního doplnění.

Typickou základní vlastností genitivu je adnominálnost (*Mluvnice češtiny 2*, s. 56). U adnominálního genitivu *Mluvnice češtiny 2* uvádí jako jeho nejčastější větněsémantickou funkci význam kvalifikující, a to v širokém slova smyslu. „Kvalifikátorem je tu entita vyjádřená právě adnominálním G⁴⁰, přičemž dalším sémantickým rysem je tu zužující, specifikující charakter kvalifikačního vztahu. Další sémantické třídění je pak závislé na povaze vztahů mezi kvalifikovanou a kvalifikující entitou“ (*Mluvnice češtiny 2* 1986:56). Jako příklady těchto vztahů *Mluvnice češtiny 2* (s. 56) uvádí např. vztah přívlastňovací, vztah části a celku, vztah mezi původcem a výsledkem, vztah, v jehož rámci je kvalifikovaná entita

⁴⁰ G=genitiv

pojmenováním děje zaměřeného na entitu kvalifikující (*lov makrel*), dále vztah míry, množství, úhrnu (tzv. genitiv partitivní). *Mluvnice češtiny 2* dále zmiňuje (s. 56), že se zde však uplatňují i další sémantické vztahy. Jejich podrobnější analýzou se však nezabývá a neuvádí ani jejich výčet. *Mluvnice češtiny 2* se dále sémantice genitivu stručně věnuje na str. 58–59, kde sémantiku genitivu rozděluje na základě jeho výskytu v subjektové nebo objektové pozici.

Vladimír Šmilauer (1966:174) dělí genitivity na: (1) genitiv přivlastňovací, posesivní (příslušnost k vlastnosti na otázku *jaký?*), (2) genitiv vlastnosti, kvalitativní (příslušnost podle obsahu), (3) genitiv vysvětlovací, explikativní, (4) genitiv celkový, partitivní, (5) genitiv původce děje, subjektový a (6) genitiv cíle děje (zasaženého předmětu), objektový.

Genitiv posesivní Šmilauer charakterizuje (1966:174–175) jako genitiv označující příslušnost k substanti (v nejširším smyslu), který vyjadřuje nejrůznější těsné vztahy: vlastnictví, příslušnost k osobě, místu, části k celku, vlastnosti k substanti. Zvláštním druhem tohoto genitivu je genitiv původce (*auctoris*).

Genitiv kvalitativní, tedy genitiv vlastnosti, označuje příslušnost k vlastnosti na otázku *jaký?* Vyskytuje se u jmen lidí (u obecných substantiv), u jmen částí těla, jiných konkrétních i abstraktních a je to vždy výraz rozvitý, tedy skládající se ze substantiva a rozvíjejícího adjektiva (více viz Šmilauer 1966:175–176). Genitiv explikativní označuje příslušnost podle obsahu připojením pojmu užšího, druhového k pojmu širšímu, rodovému. Bývá součástí ustálených spojení (více viz Šmilauer 1966:176). Genitiv partitivní k udání míry, množství nebo části připojuje udání předmětu (více viz Šmilauer 1966:176–179).

Genitiv původce děje (subjektový) stojí u dějových substantiv a označuje podmět děje (Šmilauer 1966:179). Genitiv cíle děje, zasaženého předmětu (objektový) stojí u substantiv odvozených z přechodných sloves a ze sloves s genitivem a

označuje předmět děje vyjádřeného základním slovesem (Šmilauer 1966:179). Může se jednat o dějové jméno, činitelské jméno nebo místní jméno (více viz Šmilauer 1966:179–180).

Sémantika genitivu pojednávána v těchto dílech má syntaktické pojetí. Jak však bylo již dříve uvedeno (odd. 2.1), není syntaktické pojetí pro popis valence prvořadé ani vždy vhodné (např. odd. 4.1.4). Podrobný sémantický popis genitivu jako valenčního exponentu však přesahuje rámec této práce. Chceme zde tedy pouze načrtnout pohled na genitiv jako valenční exponent substantiv tak, jak vyplývá z analýzy korpusových dat.⁴¹ V této práci se tedy v zásadě budeme dále držet terminologie podle Vladimíra Šmilauera (1966). V případě tzv. kvalitativního genitivu (Šmilauer 1966:174) je však naše pojetí širší a volnější, nemusí se jednat o výraz rozvitý, jde obecně o vyjádření nějaké vlastnosti, kvality.

V této souvislosti je zde třeba připomenout, že sémantika patří k lingvistickým oblastem, ve kterých je často obecně obtížné stanovit zcela jednoznačná kritéria interpretace a klasifikace. Sémantická interpretace a klasifikace genitivů do výše uvedených kategorií je tedy diskutabilní a subjektivní. Pojetí této práce jako korpusem řízené vybízí k vlastní klasifikaci a stanovení kritérií, nicméně takovýto krok přesahuje rámec této práce. Jakkoliv se tedy dá o jednotlivých interpretacích genitivů dále diskutovat, je např. v případě substantiv s více sémantickými typy genitivů důležitější než samotné zařazení k jednotlivým sémantickým typům fakt, že genitivní valence daných substantiv má různou sémantickou charakteristiku. Tento fakt je u většiny příkladů zcela zřejmý.

⁴¹ Sémantické dělení je však vždy nutně do značné míry subjektivní.

4.1.1 Substantiva s převažující genitivní valencí

Substantiva s převažující genitivní valencí jsou pro tuto práci definována jako substantiva, která se ve více než 50 % ze všech svých výskytů (v našem zkoumaném vzorku) používají s bezpředložkovým genitivním doplněním. Takovýchto substantiv bylo v námi zkoumaném vzorku 12: *vyhlášení, odstranění, stanovení, poskytování, udržení, příznak, konání, zavádění, sklonek, zástupkyně, překročení, respektování*. Již na první pohled je zřejmé, že ve většině případů jde o deverbativní substantiva odvozená pomocí přípon *-ení* nebo *-ání*. Jde tedy o specifickou skupinu substantiv. Pro slova v této skupině je charakteristická i určitá sémantická neúplnost. Význam většiny těchto slov je velmi abstraktní a vágní, a vyžaduje tak další kontextové doplnění, které může být (a bývá) realizováno právě prostřednictvím valence.

Stručně si připomeňme frekvenční údaje vztahující se k těmto substantivům. Následující tabulka pracuje s údaji z *Frekvenčního slovníku češtiny* (2004) (dále FSC) a uvádí v prvním sloupci frekvenci výskytu, v druhém sloupci pořadí frekvence a další tři sloupce uvádějí v procentech žánrové rozložení výskytu daného substantiva.

Tabulka 1. Substantiva s převažující genitivní valencí z pohledu FSC

Substantivum	Frekvence	Pořadí frekv.	Beletrie	Odborná	Publicistika
<i>vyhlášení</i>	3202	3222	10 %	27 %	64 %
<i>odstranění</i>	2741	3639	5 %	63 %	32 %
<i>stanovení</i>	2657	3706	1 %	79 %	20 %
<i>poskytování</i>	2592	3786	1 %	58 %	42 %
<i>udržení</i>	2423	3961	7 %	48 %	46 %
<i>příznak</i>	2453	3934	14 %	70 %	16 %
<i>konání</i>	1888	4786	15 %	49 %	36 %
<i>zavádění</i>	1502	5633	2 %	69 %	30 %
<i>sklonek*</i>	-	-	-	-	-
<i>zástupkyně</i>	1122	6921	17 %	7 %	76 %
<i>překročení</i>	1071	7122	6 %	59 %	35 %
<i>respektování</i>	1020	7363	5 %	56 %	39 %

* Srovnatelné údaje pro tvar *sklonku+* jsou – frekv.: 1335, poř. frekv.: 6101, beletrie: 25 %, odb.: 29 % a publ. 46 %.

Jak je patrné z tabulky výše tato substantiva jsou rozložena ve frekvenčním pásmu 1020 až 3202, což odpovídá zkoumanému vzorku substantiv. Pro tato substantiva je však charakteristické specifické žánrové rozložení. V beletrii se vyskytují v malé míře a jsou typická pro styl odborný nebo publicistický. Specifičnost těchto stylů a důsledek pro reprezentativnost výsledků analýzy jsme již zmiňovali v odd. 3.1. Pro srovnání, např. v Pražském mluveném korpusu se většina těchto substantiv nevyskytuje vůbec nebo se vyskytují naprosto okrajově⁴². Při interpretaci výsledků analýzy je tedy potřeba mít tuto skutečnost na zřeteli.

4.1.1.1 Jednotlivá substantiva

VYHLÁŠENÍ

Substantivum *vyhlášení* má silnou genitivní valenci (v 88 % ze všech výskytů), a to dvou sémantických typů. Převládající je genitiv objektový vyjadřující, co je předmětem *vyhlášení* (*vyhlášení samostatnosti*). V malé míře se objevuje genitiv

⁴² Data výskytu v PMK: *odstranění* 2x (oba výskyty se stejným gen. doplněním: *odstranění evidenčního nesouladu*); *poskytování* 1x; *udržení* 1x (s gen. doplněním: *udržení standardní náky úrovně*); *příznak* 1x, *příznaků* 1x; *konání* 5x, *konáním* 1x (s gen. doplněním: *konáním druhéjch lidí*); *zástupkyni* 1x (s gen. doplněním: *dělá zástupkyni primářky*); *překročení* 1x; *respektování* 1x.

subjektivý vyjadřující, kdo *vyhlášení* činí (necelé 2 % z celkového počtu genitivů). Nejčastějším příkladem tohoto druhého typu je *vyhlášení vlády*.

Za genitivním valenčním exponentem v mnoha případech následuje další valenční exponent. Jedná se především o následující typy:

- *vyhlášení* + gen. + instrum. (*Korejského poloostrova pásmem bez jaderných zbraní*)
- gen. + gen. (viz dále)
- gen. + adv. času / místa (*vyhlášení stávkové pohotovosti od příštího týdne; vyhlášení příměří v Ulsteru*)
- gen. + dat. (*vyhlášení války Srbsku*)⁴³
- gen. + předložka *o*⁴⁴ (*vyhlášení příměří a o výměně zajatců; vyhlášení vládních představitelů zemí Středoevropské dohody o zóně volného obchodu*)
- gen. + další předložky plynoucí ze sémantiky slova v gen. (*vyhlášení nucené správy nad špatně hospodařícím fondem; vyhlášení konkursu na dlužníka; vyhlášení hospodářských sankcí proti Pchjongjangu*)

Pro genitivní doplnění substantiva *vyhlášení* je navíc typické, že většina slov na pozici jeho genitivního valenčního exponentu má určitou sémantickou charakteristiku. Některá z nich se navíc často opakují. Mezi těmito nejčastěji se opakujícími kolokáty jsou např. následující substantiva: *konkurs, nezávislost, válka, výsledek* (převážně v plurálové podobě), *referendum, samostatnost, soutěž, vítěz* (častěji v pl.), *výjimečný stav*. Typicky se na této pozici vyskytuje i adjektivum v superlativu následované substantivem jako např. *vyhlášení nejlepšího fotbalisty*. Pro genitivní valenční exponent stojící u substantiva *vyhlášení* je typické i jeho řetězení, tedy to, že genitiv je následován dalším substantivem v genitivu. Konkrétně můžeme toto pozorovat např. ve spojení se

⁴³ omezeno na slova *válka* a *ultimátum*

⁴⁴ Předložka *o* se zdá být i samostatným valenčním exponentem substantiva *vyhlášení*, avšak velice řídkým a pravděpodobně omezeným na určitý žánr (*vyhlášení o navyšování základního jmění*). V mnoha dalších výskytech lze výskyt této předložky interpretovat jako valenci slova v genitivu (např. *anketa, zákon, dohoda* apod. (*vyhlášení válečného zákona o cenzuře*)).

slovy *nezávislost, výsledek* (převážně v pl.), *samostatnost, svrchovanost, stav*⁴⁵ (např. *vyhlášení výsledků soutěže*).

ODSTRANĚNÍ

Substantivum *odstranění* je následováno bezpředložkovým genitivem v asi 88 % svých výskytů. Většinu slov na pozici genitivního exponentu substantiva *odstranění* spojuje negativní sémantická preference⁴⁶, tedy sémantika těchto slov v sobě zahrnuje určitý negativní sémantický rys (např. *havárie, hrozba, nedostatky, překážky, škoda, vada, závada*). Pokud jsou to slova sémanticky neutrální, představují v daném kontextu něco negativního, nežádoucího – např. v případě *odstranění vaječnicků* je z kontextu zřejmé, že se jedná o nemocné vaječníky. Genitiv je v necelých 3,5 % všech výskytů následován ještě předložkou *z* (*odstranění Senátu z právního řádu*).

STANOVENÍ

Substantivum *stanovení* je následováno genitivní valencí v asi 94 % svých výskytů. Toto substantivum patří svým významem především do slovníku odborného, čemuž odpovídá i jeho rozložení v korpuse – v beletrii je zastoupeno 1 %, zatímco v odborné literatuře 79 % a v publicistice 20 %. Tomuto charakteru samozřejmě odpovídá i kolokační profil tohoto substantiva. Genitivní valenční exponent je u substantiva *stanovení* poměrně často následován dalšími valenčními exponenty. Jsou to především předložky (*pro, z, na, za, při, u*) (např. *stanovení hranice pro odchod do důchodu*), další genitiv (*stanovení koncentrace kovů*) a instrumentál (*stanovení železa metodou RFS*).

Pro substantivum *stanovení* je tedy také charakteristické řetězení genitivů. Mezi substantiva, která se vyskytují jako genitivní valence substantiva *stanovení* a která

⁴⁵ Podrobnější analýzu typických kolokátů lze najít v Čermáková (2006).

⁴⁶ Sémantickou preferenci definuje Sinclair např. ve svém článku *The lexical item* následovně: „Semantic preference is the restriction of regular co-occurrence to items which share a semantic feature...“ (Sinclair 2004c:142).

jsou dále následována dalším genitivem patří např. *cena, cíl, hodnota, kurs, výše, limit, obsah, počet, pravidlo, hranice, datum, dávka, doba, sazba, koncentrace, termín, kvóta, míra, pořadí, priorita, rozsah, stupeň* a *základ*. Některé z těchto exponentů jsou dalším genitivem následovány ve vysoké míře: *míra* a *rozsah* (93 %), *stupeň* a *výše* (91, resp. 90 %), *počet, obsah* a *koncentrace* (89, 88 a 87 %).

Podíváme-li se na frekvenci výskytu výše uvedených kolokátů substantiva *stanovení*, je patrné, že některé z kolokátů s podobnou nebo stejnou frekvencí výskytu tvoří určité sémantické skupiny, např. *kurs* a *hodnota* se stejnou frekvencí; *hranice, limit, počet* s podobnou frekvencí; *doba* a *termín* se stejnou frekvencí; nebo *rozsah, dávka, míra, pořadí* s podobnou frekvencí. Následující tabulka uvádí pro větší názornost konkrétní čísla.

Tabulka 2. Kolokace substantiva *stanovení* v SYN2000 (interval 0/3)*

Pořadí	Kolokát	Počet výskytů	Frekv. v SYN2000
1	<i>cena</i>	172	81 879
2	<i>Výše</i>	68	22 902
3	<i>Cíl</i>	56	27 662
4	<i>hodnota</i>	51	32 203
4	<i>Kurs</i>	51	14 774
5	<i>základ</i>	50	27 657
6	<i>pravidlo</i>	49	12 559
7	<i>hranice</i>	41	26 237
8	<i>limit</i>	39	2 901
9	<i>počet</i>	37	39 564
10	<i>obsah</i>	34	11 800
11	<i>priorita</i>	29	11 800
12	<i>sazba</i>	25	9 638
13	<i>koncentrace</i>	23	3 516
14	<i>doba</i>	20	111 053
14	<i>termín</i>	20	12 009
15	<i>kvóta</i>	19	1 993
16	<i>rozsah</i>	15	8 601
16	<i>dávka</i>	15	7 535
17	<i>míra</i>	14	22 885
17	<i>pořadí</i>	14	8 402
17	<i>datum</i>	14	5 532
18	<i>stupeň</i>	11	11 716

* Interval vyjadřuje v počtu slov rozsah kontextu, ve kterém jsou kolokáty vyhledávány. 0 představuje pozici hledaného slova (tedy v tomto případě *stanovení*) a kladné hodnoty představují kontext napravo, hodnoty záporné kontext nalevo (tedy zde jsou hledány kolokáty v kontextu 3 slov napravo od KWIC).

Vzhledem k již zmíněné sémantické povaze substantiva *stanovení* je třeba uvést, že nejčastěji na pozici genitivního exponentu najdeme obecně nebo konkrétně nějakou chemickou či fyzikální látku, event. jednotku (např. *Principem analytické metody je stanovení uhličitaniů jako CO₂ metodou headspace*)⁴⁷.

POSKYTOVÁNÍ

Substantivum *poskytování* je následováno genitivem až v 96 % všech svých výskytů. I u jeho genitivních doplnění můžeme pozorovat jisté charakteristické sémantické okruhy. Nejčastěji se jedná o slova sémanticky spadající do okruhu financí (*půjčky, dotace* atd.), dále se často poskytují např. *informace, služby, pomoc* nebo *zdravotní péče*. Genitivní valence je u substantiva *poskytování* v necelých 5 % všech výskytů dále následována předložkou *na* vyjadřující účel a to především ve spojení s výrazy ze sémantického okruhu financí (*poskytování subvencí na vývoz whisky*). V asi 11 % všech výskytů je genitivní výraz dále následován výrazem dativním (*poskytování dotací soukromým školám*) a paralelně zde i najdeme vyjádření pomocí předložky *pro* (+ akuz.) (*poskytování daňových úlev pro malé a střední podnikatele*).

UDRŽENÍ

Substantivum *udržení* se s genitivem vyskytuje v přibližně 71 % svých výskytů. Na místě genitivního valenčního exponentu se s vyšší frekvencí vyskytují např. slova *příjem* (především v pl.), *úroveň, rovnováha, konkurenceschopnost, růst, stabilita*.

⁴⁷ Podrobnější analýza kolokátů tohoto substantiva viz Čermáková (2006).

PŘÍZNAK

Substantivum *příznak* se s genitivem vyskytuje v asi 48 % svých výskytů, což je těsně pod výše stanovenou padesátiprocentní hranicí. Avšak vzhledem k sémantické povaze tohoto substantiva, tedy jeho sémantické neúplnosti vyžadující další kontextové doplnění a vzhledem k tomu, že jsou, jak již bylo výše řečeno, procentuální údaje přibližné a vztahují se pouze k analyzovaným vzorkům, bylo toto substantivum přesto zařazeno do skupiny substantiv s převažující genitivní valencí. Slova vyskytující se na pozici genitivního valenčního exponentu jsou v souladu s očekáváním nejčastěji z oblasti lékařské: *choroba, nemoc, onemocnění, otrava* atd. (*viditelnější příznaky rozrušení; nezjistila příznaky arytmie*). Substantivum *příznak* má nejvyšší zastoupení – 70 % – v textech odborných.

KONÁNÍ

Substantivum *konání* se s genitivní valencí v našem korpuse vyskytuje přibližně v 57 % svých výskytů. Většinou se jedná o genitiv objektový, kdy na pozici genitivní valence nacházíme výrazy jako *valná hromada, volby, výstava, festival* nebo *akce*. Velice okrajově zde najdeme i genitiv subjektový (např. *konání obžalované*).

ZAVÁDĚNÍ

Substantivum *zavádění* se s genitivní valencí vyskytuje v asi 89 % svých výskytů. Mezi typickými genitivními kolokáty najdeme slova *technologie* (v plurálové i singulárové podobě), *reforma* (sg. i pl.), *systém* (sg. i pl.), *norma* (sg. i pl.), *technika*. Velice typickými kolokáty rozvíjejícími substantivum na pozici genitivního valenčního exponentu najdeme adjektiva *nový* a *informační*.

SKLONEK

Substantivum *sklonek* je velice specifické, protože je ve většině svých výskytů součástí několikaslovné předložky následované genitivem. Takto jej nacházíme ve více než 99 % svých výskytů. V naprosté většině případů je součástí víceslovné

frazeologické předložky „*na sklonku* + gen.“.⁴⁸ Následující substantivum nebo výraz v genitivu má téměř vždy určitou časovou charakteristiku, můžeme zde tedy mluvit o sémantické preferenci časových a obdobných výrazů: *měsíc, rok, století, život, kariéra* atd.

ZÁSTUPKYNĚ

Substantivum *zástupkyně* se vyskytuje asi v 57 % svých výskytů jako následované genitivem. Je to jediné substantivum z této skupiny, které označuje osobu, a je tak konkrétní povahy. Podle očekávání se kolokáty tohoto substantiva nejčastěji vztahují k různým institucím (*bývalá zástupkyně ředitele Dívčí školy*).

PŘEKROČENÍ

Substantivum *překročení* se s genitivním valenčním exponentem vyskytuje asi v 88 % svých výskytů. I zde se na pozici genitivního valenčního exponentu objevuje sémanticky charakteristická skupina slov, např. *limit, mez, rychlost, zákon, kompetence, hodnota, míra, úroveň, norma, plán, pravomoc, práh, dávka, teplota, zákaz, řeka*. Některé z těchto typických kolokátů bývají následovány dalším genitivem – *hranice, limit, mez, hodnota, úroveň, pravomoc, norma, práh, dávka, zákaz (překroční míry inflace)*.⁴⁹ I zde, podobně jako u substantiva *stanovení* výše, můžeme pozorovat, že frekvence jejich výskytu koreluje se sémantikou, např. prvních pět kolokátů lze v určitých kontextech považovat do určité míry za synonymní: *hranice, limit, mez, hodnota, úroveň*. Výrazy *limit* a *mez* pak k sobě mají sémanticky nejbližší a rovněž jejich frekvence výskytu jsou téměř stejné. Následující tabulka uvádí podrobnější číselné údaje.⁵⁰

⁴⁸ Objevuje se však i s jinými předložkami: *od, do, ze, a ke* (již *od sklonku* pravěku začali lidé záplavy regulovat; *dotáhl svůj výklad až do sklonku 30. let; veršovaných českých knížek ze sklonku 16. a začátku 17. století; Vincent van Gogh si ke sklonku života*).

⁴⁹ Dále za genitivem v okrajové míře následuje předložka *o* (*překročení plánu o 31 procent*).

⁵⁰ Podrobnější analýza kolokátů tohoto substantiva viz Čermáková (2006).

Tabulka 3. Vybrané kolokace substantiva *překročení* v SYN2000 (interval 0/3)

překročení (0/3) +	počet / z toho následuje gen.
<i>hranice</i>	253 / 55
<i>limit</i>	78 / 15
<i>mez</i>	52 / 14
<i>hodnota</i>	30 / 12
<i>úroveň</i>	21 / 6
<i>pravomoc</i>	21 / 6 (<i>veřejného činitele</i>)
<i>norma</i>	13 / 3
<i>práh</i>	14 / 8
<i>dávka</i>	11 / 3
<i>zákaz</i>	8 / 3

RESPEKTOVÁNÍ

Substantivum *respektování* se s genitivní valencí vyskytuje asi v 85 % svých výskytů. Mezi jeho typické genitivní kolokáty patří např. *právo* (především v plurálové podobě a často s atributivním adjektivem *lidský*), *zásada*, *zájem*, *požadavek* a *pravidlo* (všechna tato substantiva především ve své plurálové podobě).

4.1.1.2 Shrnutí

V našem analyzovaném vzorku 99 substantiv (viz. 3.2.2) bylo 12 substantiv s převažující genitivní valencí, tedy výskytem genitivního exponentu ve více než 50 % všech výskytů daného substantiva. Většina těchto substantiv je deverbativního původu odvozená pomocí přípon *-ení* nebo *-ání*. Pro celou skupinu těchto substantiv je charakteristická sémantická neúplnost, která vyžaduje další kontextové doplnění, které může být (a bývá) zprostředkováno např. právě valencí. U všech těchto substantiv má valence formální podobu bezpředložkového genitivu, takováto specifikace je však, jak je patrné z analýzy dat výše, pro popis jejich valence neúplná. Většinu výrazů na pozici valenčního genitivu totiž u jednotlivých substantiv spojuje určitá sémantická charakteristika, byť v některých případech těžko uchopitelná.

Tato sémantická charakteristika jejich kolokability by vyžadovala mnohem podrobnější, rámec této práce přesahující, sémantickou analýzu genitivních kolokátů těchto substantiv. U některých substantiv je jednoznačnější, např. u substantiva *odstranění* můžeme říci, že většinu kolokátů po sémantické stránce spojuje určitá negativnost (negativní sémantická preference, viz výše), u substantiva *příznak* se obvykle jedná o kolokáty spadající do sémantické oblasti různých nemocí apod.

Specifickým substantivem v této skupině je slovo *skloněk*, které je téměř bezvýhradně součástí víceslovné předložky (a tak je chápáno např. i ve *Frekvenčním slovníku*), i u něj jsou však genitivní kolokáty jednoznačně sémanticky charakterizovány, jedná se o různé výrazy časové v nejširším slova smyslu.

U tří z těchto substantiv (*vyhlášení, stanovení, překročení*) je navíc frekvenčně výraznější i další rys – a to řetězení genitivů. Tedy genitivní valenční exponent je následován dalším/(i) genitivním/(i) výrazem/(-zy). Podrobnější analýza genitivních kolokátů těchto tří substantiv je nastíněna v článku *Kolokace a valence některých případů substantiv* (Čermáková 2006). Tomuto jevu se budeme dále věnovat podrobněji i v této práci (odd. 4.1.4).

Ze sémantického hlediska byl genitiv vyskytující se u většiny substantiv jednoho sémantického typu, a to genitiv objektový. U substantiv *příznak* a *zástupkyně* se jednalo o genitiv posesivní. Specifickým případem je substantivum *skloněk*, jehož genitivní doplnění má posesivní charakter, je však součástí celého předložkového spojení. U dvou substantiv (*vyhlášení, konání*) se vyskytovaly dva sémantické typy genitivů. V převažující míře se podobně jako u ostatních substantiv jednalo o genitiv objektový a v okrajové míře pak o genitiv subjektový.

Následující tabulka uvádí substantiva z této skupiny v pořadí podle četnosti jejich genitivního doplnění. S výjimkou substantiva *sklonek* si zde můžeme všimnout, že nejsilněji se genitivní valence pojí právě se skupinou deverbativních sloves odvozených příponami *-ení* a *-ání*. V povrchové syntaxi tak genitiv u těchto substantiv odpovídá akuzativní valenci sloves, z kterých jsou tato substantiva derivována.

Tabulka 4. Substantiva s převažující bezpředložkovou genitivní valencí.

Substantivum	% výskytu s gen. doplněním
<i>sklonek</i>	99 %
<i>poskytování</i>	96 %
<i>stanovení</i>	94 %
<i>zavádění</i>	89 %
<i>odstranění</i>	88 %
<i>překročení</i>	88 %
<i>vyhlášení</i>	88 %
<i>respektování</i>	85 %
<i>udržení</i>	71 %
<i>konání</i>	57 %
<i>zástupkyně</i>	57 %
<i>příznak</i>	48 %

Na základě výše uvedené charakteristiky této skupiny substantiv se pokusíme stanovit kritéria pro určení míry valenčnosti genitivního doplnění. Prvním kritériem by tedy mohla být frekvence výskytu tohoto typu doplnění u daného substantiva. Jako druhé kritérium se nabízí určitá společná sémantická charakteristika genitivních kolokátů vyskytujících se na pozici valenčního exponentu, jejichž prostřednictvím je upřesňován či zužován význam tohoto substantiva. Kritériem třetím, avšak velmi obecným, by pak mohla být určitá sémantická „nesamostatnost“ substantiva, která vyžaduje další kontextové zpřesnění, tedy substantivum se nevyskytuje (nebo prakticky nevyskytuje) jako stojící samostatně bez dalšího zpřesnění. Tato otázka je dosud neprozkoumaná a velmi obecná, avšak data na ni nepřímo odkazují, a otevírá se zde tedy velký prostor pro další zkoumání. Doprovodným (ale fakultativním) čtvrtým kritériem

posilujícím „valenčnost“ (a ve většině případů i všechny další výše uvedené charakteristiky) by mohl být výskyt „řetězení“ genitivních doplnění (viz dále 4.1.4).

4.1.2 Ostatní substantiva s genitivním doplněním

Pro kategorizaci genitivu jako valenčního exponentu se na základě dosavadní analýzy jeví jako důležitá především dvě kritéria: (1) frekvence výskytu genitivního doplnění u jednotlivých substantiv a (2) sémantická charakteristika genitivních kolokátů a s ní související stupeň sémantické samostatnosti substantiva (viz 4.1). Některá substantiva se pojí s jedním sémantickým typem genitivu, některá s více a v podstatě se tak jedná o různé valence, chápeme-li valenci jako kategorii formálně sémantickou.

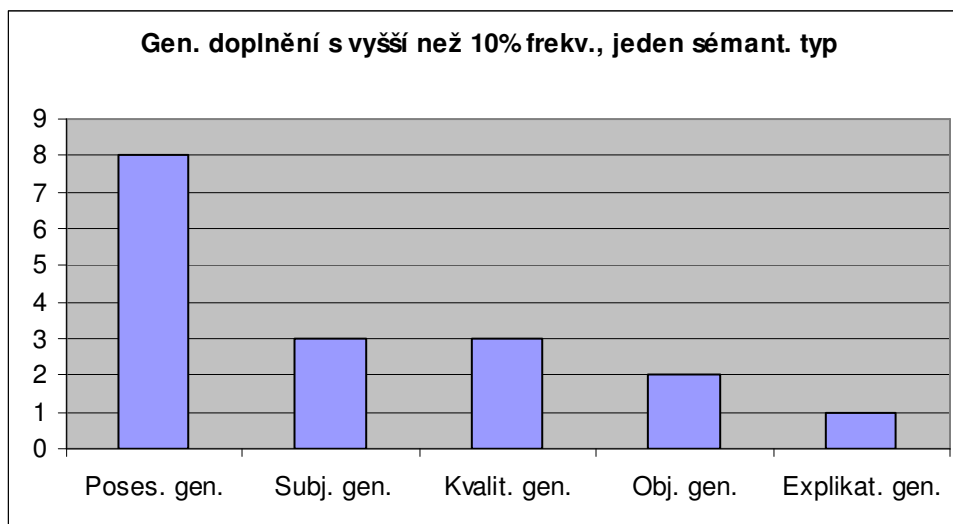
Analyzovaná substantiva, s nimiž se vyskytovalo genitivní doplnění s frekvencí nižší než v 50 % výskytů, jsme v této kapitole rozdělili do dvou skupin. Část 4.1.2.1 pojednává o substantiva s jedním sémantickým typem genitivní valence a část 4.1.2.2 o substantiva se dvěma nebo více sémantickými typy. Následující tabulka (srov. i tabulka v 4.1.7) v přehledu ukazuje rozložení genitivního doplnění u substantiv vzhledem k frekvenci jeho výskytu a sémantické složitosti. Jak je z tabulky patrné, rozdíl v počtu substantiv s jedním a více sémantickými typy genitivního doplnění není velký.

Tabulka 5. Ostatní substantiva s genitivním doplněním

Frekvence genitivní doplnění	Počet substantiv s jedním sémantickým typem genitivního doplnění	Počet substantiv s více sémantickými typy genitivního doplnění
více než 10 % výskytů	17	18
5 – 10 % výskytů	7	8
3 – 5 % výskytů	6	2
méně než 3 % výskytů	13	7
Celkem	43	35

4.1.2.1 Substantiva s jedním sémantickým typem genitivního doplnění

Substantiv s jedním sémantickým typem genitivního valenčního doplnění bylo mezi analyzovanými substantivy v této skupině 43. Mezi těmito substantivy bylo 17 substantiv s vyšším výskytem genitivního doplnění než 10 % ze všech výskytů substantiva⁵¹. Mezi těmito 17 substantivy jich bylo 8 takových, jejichž genitivní doplnění mělo povahu posesivního genitivu (*funkcionář; pomník; křivka; pero*⁵²; *vzhled; pomocník; tribuna; ubytovna*), se 3 se vázal genitiv kvalitativní (*důležitost; volnost; podzemí*), se 3 subjektivý (*stáří; vzpoura; kolaps*), se 2 objektový (*odjezd; viník*) a s 1 explikativní (*plemeno*). Sémantické rozložení jednotlivých typů genitivů mezi těmito substantivy zobrazuje následující graf.



Graf 1. Sémantické typy genitivního doplnění s frekvencí výskytu vyšší než 10%.

⁵¹ Avšak nižším než 50 %. Tato substantiva viz 4.1.1.

⁵² Zde se ve většině případů jednalo o zvláštní případ posesivního genitivu, tzv. genitiv auctoris (z *pera Hanse Christiana Andersena*).

Ve frekvenčním pásmu výskytu genitivů 5–10 % bylo 7 substantiv s jedním sémantickým typem genitivního doplnění. U pěti z těchto substantiv měl genitiv posesivní povahu (*rozpaky*⁵³, *partnerka*, *klávesnice*, *výdělek*), jedním z nich byl genitiv auctoris (*vila*), u substantiva *vaření* má genitiv charakter genitivu objektového a u substantiva *korupce* genitivu subjektivního. Šest dalších substantiv (*ambasáda*, *plet'*, *odpočinek*, *pocta*, *hádky*, *plavec*) mělo genitivní doplnění v ještě okrajovější míře (3–5 %). Podrobnější data a příklady jsou uvedeny v odd. 4.1.7 a příloze B.

Zcela okrajově (v méně než 3 % výskytů) se genitivní doplnění vyskytovalo s 13 substantivy. Většina těchto genitivů měla posesivní charakter (pec (0,5 %) *pecí koncentračního tábora*; sonda (0,75 %) *sonda Evropské kosmické agentury*; policajt (1x) *policajt demokratického státu*; pojistné (1 %) *pojistné České kooperativy*; šátek (2x) *šátků firmy Mileta Hořice*; rukavice (1 %) *rukavice*; bankovníctví (1,7 %) *ředitel odboru investičního bankovníctví České spořitelny*; kompresor (2,6 %) *kompresor chladničky*; vdova (1x) *nadací pro vdovy policistů*). Explikativní charakter mělo genitivní doplnění u dvou z těchto substantiv (civilizace (1,6 %) *civilizace starověku*⁵⁴; samec (2,75 %) *samců druhu Felis*), u jednoho ze substantiv mělo doplnění charakter subjektivní (zamyšlení (2,3 %) *novoročního zamyšlení prezidenta Havla*) a u jednoho partitivní (maličkost (2x) *maličkosti každodenního života*).

Genitiv jako valenční exponent má větší sémantickou různorodost u těch substantiv, kde se vyskytuje s vyšší mírou (v 10–50 % z výskytů substantiva)⁵⁵. Genitiv posesivní je nejfrekventovanější a nejméně se vyskytuje genitiv

⁵³ Substantivum *rozpaky* má výrazný kolokační profil a často bývá součástí např. následujících spojení: *bez rozpaků* 8,8 %; *do rozpaků* 12,6 %; *na rozpacích* 8,2 %; *s rozpaky* 6,6 %; *v rozpacích* 14,4 %. Genitiv by mohl být interpretován i jako subjektivní.

⁵⁴ Více viz 4.1.5.

⁵⁵ U substantiv, kde se však genitiv vyskytuje ve vysoké míře (viz 4.1.1), převládá pouze jeden sémantický typ genitivu, a to genitiv objektový.

explikativní. U substantiv, kde má genitivní valence slabší charakter (výskyt méně než 10 %) se nejméně uplatňuje opět genitiv posesivní.

4.1.2.2 Substantiva s více sémantickými typy genitivní valence

Substantiv s více sémantickými typy genitivního valenčního doplnění bylo mezi analyzovanými substantivy 35. Mezi těmito substantivy bylo 18 substantiv s celkovým výskytem genitivního doplnění vyšším než 10 % ze všech výskytů substantiva. Srovnávání dat frekvence výskytu se substantivy s jedním sémantickým typem genitivního doplnění však může být v těchto případech poněkud zavádějící. Formálně se sice jedná o stejný valenční exponent (doplnění), tedy bezpředložkový genitiv, avšak vnitřní sémantické členění nabízí interpretaci, že se jedná o různé valence nebo dokonce v některých případech valence různých slov (lexémů). U některých zdánlivě polysémických⁵⁶ substantiv se totiž různé sémantické typy genitivního doplnění pojí s různými „významy“ daného substantiva, a mohou tak vlastně lišit mezi těmito významy podobně jako širší textový kontext nebo kolokace. I v ostatních případech mohou různé sémantické typy genitivního doplnění upřesňovat význam slova (substantiva) podobně jako širší kontext⁵⁷. V těchto případech je sémantická specifikace formy, tedy genitivu, velice důležitá informace (viz i 4.1.3).

Zdá se, že sémantický typ genitivního valenčního doplnění tak i může např. lišit mezi abstraktním a konkrétním významem slova. U substantiva *pasáž* partitivní genitiv odkazuje na abstraktní význam tohoto substantiva (*pasáž americké verze dokumentu*), zatímco genitiv posesivní odkazuje na pasáž konkrétní budovy (*nad vchodem do pasáže divadla Čas*). Podobně u substantiva *mor* objektový genitiv

⁵⁶ Otázka polysémie, více viz 2.2.

⁵⁷ Slovo jako základní jednotka lingvistické analýzy, více viz 2.2.

odkazuje na konkrétnější podobu moru (*mor prasat*)⁵⁸ a subjektový genitiv odkazuje na jeho abstraktnější podobu (*mor komunismu*). Ve spojení se substantivem *mor* se vyskytnul i časově specifikující genitiv kvalitativní (*mor minulého měsíce*).

Substantivum *soustředění* ve spojení s genitivem posesivním má význam sportovního soustředění (*soustředění národního týmu*), jinak se v abstraktním i konkrétním významu váže s genitivem objektovým (*soustředění podnikatelských aktivit; soustředění jejich pozornosti na strategii*) nebo subjektovým (*soustředění sil poblíž hranic Belgie a Lucemburska; soustředění motoristy*).

Většina z těchto analyzovaných substantiv se vyskytla se dvěma sémantickými typy genitivního valenčního doplnění, a to nejčastěji s genitivem posesivním (i auctoris) a kvalitativním (*mítink, ateliér, stanoviště, dodatek, recenze, citát*), přičemž posesivní genitiv byl vždy výraznější, zatímco genitiv kvalitativní měl okrajový charakter. S genitivem posesivním (*monopolu České pojišťovny*) a explikativním (*monopol moci*) se vyskytovalo substantivum *monopol*, substantivum *směrnice* se pojilo s genitivem auctoris (*směrnice EU*) a okrajově s genitivem objektovým (*směrnice pořádku*). Substantivum *povstání* se 3krát vyskytlo s posesivním genitivem, který označoval vůdce povstání (*Bedri Chánovo ve 40. letech*), avšak v 10,7 % výskytů se vyskytoval s gen. subjektovým (*povstání sedláků*).

Substantivum *obliba* se vyskytovalo s gen. objektovým (*obliba značky IBM*) a subjektovým (*ztratila oblibu voličů*). S genitivem objektovým (*s nulovou tolerancí zločinu*) a subjektovým (*tolerance veřejnosti k ...*) se vyskytovalo i substantivum *tolerance*. U tohoto substantiva byl genitiv objektový dále členěn na dva typy: první byl obecnějšího rázu (viz příklad výše) a druhý se vyskytoval v technickém kontextu (*tolerance jednoho bodu; tolerance vnitřního radiusu*). S

⁵⁸ Zde se nabízí i interpretace tohoto genitivu jako genitivu posesivního.

genitivem kvalitativním (*porosty trávy* – tedy tráva tvoří porost) a okrajově gen. objektovým (*porost houby* – tedy porost na houbě) se vyskytovalo substantivum *porost*.

Substantivum *operátor* se vyskytovalo pouze s genitivem kvalitativním, avšak tento genitiv byl dále sémanticky členěn a specifikoval místo (*operátor Baltiku*), zařízení (*operátor Titaniku*) a předmět (*operátor mobilních telefonů*). Sémanticky členěn byl i kvalitativní genitiv u substantiva *nález*. Kromě časového označení období, ze kterého nález pochází (*nález miocénu u Hluchova*), se však jiné významy vyskytovaly pouze okrajově. Substantivum *nález* se dále pojilo s genitivem objektovým (*nález maligního nádoru*) a subjektivním (*nález soudu*).

Podobně se pouze s genitivem posesivním vyskytovalo substantivum *klika*. Ve většině případů se jednalo o konkrétní kliku (*klika dveří*), avšak toto substantivum se homonymně vyskytovalo s genitivem i v abstraktním významu politickém (*klika Chruščova a Brežněva*). Podrobnější data a příklady jsou uvedeny v příloze B a C.

S frekvencí genitivního valenčního doplnění v pásmu 5–10 % bylo 8 substantiv. Mezi těmito substantivy byla nejčastější kombinací kombinace genitivu subjektivního a objektového (*zájezd, zodpovědnost, tušení*). Pouze s genitivem posesivním se vyskytovalo substantivum *sval*. Zde však byl tento genitiv dále členěn, ve většině případů genitiv označoval specifickou část těla (*svaly břišní stěny*), několikrát však byl jeho význam obecnější (*svaly koně*). Substantivum *letectvo* se ve většině případů vyskytovalo s genitivem posesivním (*letectvo Spojenců*), jeden z jeho genitivů však byl kvalitativní povahy (*nejznámější stroj námořního letectva II. světové války*). Substantivum *dostih* se vyskytovalo s genitivem objektovým (*dostih klusáků*) a kvalitativním (*dostih letošního roku*), kvalitativní gen. v polovině případů udával časovou specifikaci. S genitivem

objektovým (*největším dlužníkem Fondu národního majetku*) a kvalitativním (*dlužníků nájemného*) se vyskytovalo i substantivum *dlužník*.

Substantivum *plátno* se ve významu uměleckého díla vyskytovalo s genitivem auctoris a genitivem kvalitativním (*plátna velkých rozměrů*). S posesivním genitivem se vyskytovalo i ve významu filmového plátna (*na plátna kin*), v tomto významu se okrajově vyskytovalo i s genitivem kvalitativním. S genitivem kvalitativním se toto substantivum vyskytovalo i ve svém metaforickém významu (*na plátně světové literatury*).

S celkovou frekvencí v pásmu 3–5 % se s více sémantickými typy genitivního doplnění vyskytovala substantiva *pohoda* a *žíla*. Podrobnější data a příklady jsou uvedeny v odd. 4.1.7 a příloze B a C. S frekvencí výskytu nižší než 3 % se vyskytovalo 7 substantiv (*chata, housle, koridor, školka, deprese, kuře, recese*).

Podobně jako u substantiv s jedním sémantickým typem genitivního doplnění má genitiv jako valenční exponent větší sémantickou různorodost u substantiv, kde se vyskytuje s vyšší mírou (v 10–50 % z výskytů substantiva), je si zde však potřeba uvědomit, že se jedná o součet všech výskytů genitivu, jednotlivé sémantické typy se tedy mohou vyskytovat i ve velice malé míře. Genitiv posesivní je opět nejfrekventovanější a nejméně se vyskytuje genitiv explikativní a partitivní. Výrazný je zde i genitiv kvalitativní, který se u mnoha substantiv vyznačuje ještě dalším vnitřním sémantickým členěním (specifikuje např. čas, místo apod.). Se substantivy, kde má genitivní valence slabší charakter (výskyt méně než 10 %), se vyskytují genitivy posesivní, kvalitativní, objektový a subjektový.

4.1.2.3 Shrnutí

Výše naznačená typová různorodost genitivních doplnění a z toho plynoucí malé množství dat u jednotlivých typů neumožňují učinit jakékoliv formální závěry o míře „valenčnosti“ těchto genitivních doplnění. Ani kritéria pro její stanovení navržená v odd. 4.1.1.2 nejsou pro tuto skupinu zcela vhodná. Prvním kritériem zde byla výrazná frekvence výskytu – toto kritérium většina substantiv v této skupině nesplňuje. Genitiv se vyskytoval ve výraznější míře⁵⁹ pouze u pěti substantiv v této skupině: *důležitost* (46,2 %); *funkcionář* (36,4 %); *kolaps* (34 %); *křivka* (32 %) a *viník* (27,25 %).

Druhým navrženým kritériem je určitá významová nesamostatnost daného substantiva, která vyžaduje další specifikaci. Toto kritérium, jakkoliv je problematické jej přesně definovat, se však zde zdá být pro určení míry valenčnosti zásadní. Vyjdeme-li z tohoto kritéria a na valenci se nejprve díváme jako na vlastnost, kterou slova mají nebo nemají, a pak teprve zkoumáme její konkrétní realizace, je třeba se dívat i na frekvenční údaje jako na celek, tedy v jaké míře se dané substantivum s nějakým doplněním vyskytuje. V tomto světle se frekvenční údaje stávají i u dalších substantiv výraznějšími, tedy z hlediska frekvence svých výskytů se tato substantiva častěji vyskytují s realizovanou valencí nebo doplněním valenčního charakteru (např. *ateliér*; *citát*; *dodatek*; *korupce*; *letectvo*; *mítink*; *monopol*; *nález*; *odjezd*; *pocta*; *pohoda*; *pomník*; *porost*; *povstání*; *recenze*; *směrnice*, *soustředění*; *stanoviště*; *stáří*; *tolerance*; *ubytovna*; *vila*; *vzpouora*, *zodpovědnost*. Podrobnější data poskytuje příloha B a odd. 4.10).

⁵⁹ Alespoň ve třetině svých výskytů.

Takto pojatá valence úzce souvisí se sémantickou charakteristikou daného slova, tedy mírou jeho sémantické samostatnosti a tím tendence se vyskytovat např. v kolokacích. Vyšší míře sémantické nesamostatnosti odpovídá i vyšší míra valenčnosti; existence valence tedy ukazuje na nutnost dále dané slovo specifikovat. Valence samozřejmě není jediným způsobem vyjádření této specifikace, tuto funkci může plnit i textový (či situační) kontext. Nicméně právě hraniční případy toho, co do valence spadá a co ne (některé genitivity, příslovečná určení a další), otevírají prostor pro tuto interpretaci valenčnosti.

4.1.3 Sémantické typy genitivní valence

Jak již bylo dříve řečeno (odd. 4.1), je sémantická interpretace a klasifikace genitivů diskutabilní a značně subjektivní. Sémantická interpretace jednotlivých genitivů na pozici valenčního exponentu substantiv v této práci je tedy pouze orientační. Snažíme se zde ukázat především komplexnost situace. Sémantická klasifikace genitivů mohla být provedena podle jiných kritérií a získali bychom odlišný obrázek, konkrétní počty genitivů uváděné v jednotlivých grafech i textu tedy nejsou nikterak absolutní. Nicméně i klasifikace podle jiných sémantických kritérií by ukázala odlišnosti mezi sémantickými typy genitivní valence.

4.1.3.1 Substantiva s převažující genitivní valencí

U většiny (9 substantiv) substantiv s převažující genitivní valencí se jednalo o genitiv objektový. U dvou substantiv (*příznak* a *zástupkyně*) se jednalo o genitiv posesivní. Specifickým případem je substantivum *sklonek*, jehož genitivní doplnění má posesivní charakter, je však součástí celého předložkového spojení. Většina zkoumaných genitivů byla jednoho sémantického typu, dva sémantické typy se vyskytovaly pouze u dvou substantiv (*vyhlášení*, *konání*). I zde však jednoznačně vždy převládal jeden sémantický typ (genitiv objektový) a v okrajové míře se vyskytoval typ druhý (genitiv subjektový).

Podívejme se však nyní na druhý aspekt spadající do oblasti sémantiky – tedy konkrétní textovou realizaci valenčního exponentu. Jak již bylo uvedeno, charakteristickým znakem substantiv spadajících do této skupiny je jejich sémantická neúplnost vyžadující další specifikaci kontextem. Podíváme-li se blíže na skupiny typických genitivních kolokátů těchto substantiv, je patrné, že většina těchto slov má specifičtější význam než jejich ze syntaktického hlediska člen

„řídící“. Podobně je pak i význam kolokátů, které bývají frekventovaně následované dalším genitivem, neúplný. Ze sémantického hlediska je tedy vztah podřízenosti a nadřízenosti buď nejednoznačný, nebo opačný, než by byl v syntaktické interpretaci. Navíc se mnohdy některá spojení jeví spíše jako ze sémantického hlediska nedělitelné víceslovné jednotky, kde vlastně nemá o vztahu podřízenosti–nadřízenosti smysl uvažovat. Jednotlivá slova tvořící tyto lexikální jednotky jsou na sobě sice závislá, nikoliv však ve smyslu hierarchickém, ale spíše by se zde dalo hovořit o úplnosti či neúplnosti těchto jednotek.⁶⁰

Následující tabulka uvádí v přehledu vybraná substantiva s převažující genitivní valencí (odd. 4.1.1) s jejich typickými genitivními kolokáty. Tučně jsou označeny ty kolokáty, které jsou frekventovaně následovány dalším genitivem.

Tabulka 6. Vybraná substantiva s převažující genitivní valencí a jejich typické kolokáty v genitivu.

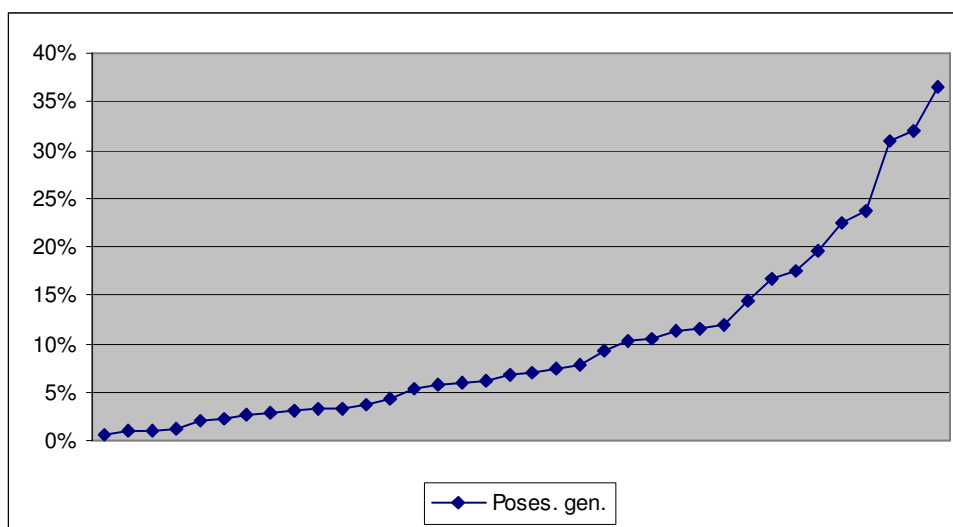
Substantivum	Typické genitivní kolokáty
<i>vyhlášení</i>	<i>konkurs; nezávislost; válka; výsledek; referendum; samostatnost; soutěž; svrchovanost; vítěz; výjimečný stav</i>
<i>odstranění</i>	<i>havárie; hrozba; nedostatky; překážky; škoda; vada; závada</i>
<i>stanovení</i>	<i>cena; cíl; datum; dávka; doba; hodnota; koncentrace; kurs; kvóta; limit; míra; obsah; počet; pořadí; pravidlo; priorita; rozsah; sazba; stupeň; termín; výše</i>
<i>poskytování</i>	<i>dotace; informace; pomoc; půjčky; služby; zdravotní péče</i>
<i>udržení</i>	<i>konkurenceschopnost; příjem; rovnováha; růst; stabilita; úroveň</i>
<i>příznak</i>	<i>choroba; nemoc; onemocnění; otrava</i>
<i>konání</i>	<i>akce; festival; valná hromada; volby; výstava</i>
<i>zavádění</i>	<i>norma; reforma; systém; technika; technologie</i>
<i>překročení</i>	<i>dávka; hodnota; hranice; kompetence; limit; mez; míra; norma; plán; práh; pravomoc; rychlost; řeka; teplota; úroveň; zákaz; zákon</i>
<i>respektování</i>	<i>požadavek; pravidlo; právo; zájem; zásada</i>

⁶⁰ Viz třeba již klasický text Johna Sinclaire (1991:81–98), kde se zabývá podrobnou analýzou slova *of* (které v určitých konstrukcích může odpovídat českému genitivu) a mimo jiné upozorňuje na podobný aspekt. V této analýze pak dále rozvíjí Hallidayovu myšlenku (1966), že každé slovo vlastně má svou vlastní gramatiku.

4.1.3.2 Ostatní substantiva s genitivní valencí

Nejčastějším sémantickým typem genitivního doplnění je v této skupině genitiv posesivní (u substantiv s převažující genitivní valencí byl nejčastějším sémantickým typem genitiv objektový). Ačkoliv je to nejčastější sémantický typ, má z frekvenčního hlediska slabší míru valenčnosti. Posesivní genitiv se vyskytoval pouze u 3 substantiv s vyšší frekvencí než 30 % (*funkcionář; křivka; vzhled*) a s frekvencí mezi 20 a 30 % se vyskytoval u dvou substantiv (*pomník; směrnice*). S frekvencí mezi 10 a 20 % se vyskytoval s 11 substantivy (*ateliér; citát; dodatek; klika; mítink; monopol; pero; pomocník; stanoviště; ubytovna; tribuna*) a více než polovina výskytů tak spadala pod frekvenci nižší než 10 %.

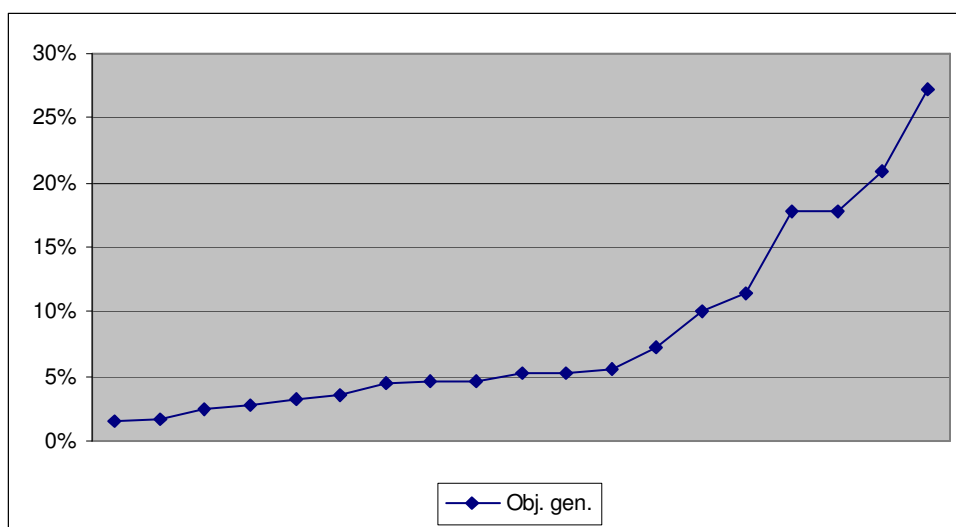
Rozložení valenčnosti posesivního genitivu přibližuje následující graf.



Graf 2. Posesivní genitiv

Genitiv objektový má u této skupiny substantiv ještě slabší míru valenčnosti než genitiv posesivní. Se substantivem *kolaps* překročil hranici 30 % (frekvence

výskytu 34 %) a dále se vyskytoval pouze se 2 substantivy s vyšší frekvencí než 20 % (*nález*; *viník*). Mezi 10 a 20 % se vyskytoval se 3 substantivy (*obliba*; *odjezd*; *recenze*). Téměř tři čtvrtiny jeho výskytů tak nepřesáhly hranici 10 %. Tendenci přibližuje následující graf.



Graf 3. Objektový genitiv

Z ostatních genitivních doplnění v této skupině substantiv měl výraznější frekvenci výskytu (přes 10 %) genitiv subjektového charakteru vyskytující se se substantivem *důležitost* (46,2 %), dále genitiv kvalitativní u substantiva *operátor* (12,2 %), partitivní genitiv u substantiva *pasáž* (10,4 %), explikativní genitiv u substantiva *plemeno* (17,3 %), kvalitativní genitiv u substantiv *podzemí* (21 %) a *porost* (13,2 %), subjektový genitiv u substantiv *povstání* (10,7 %) a *stáří* (13,8 %), kvalitativní genitiv u substantiva *volnost* (17,3%) a subjektový genitiv u substantiva *vzpoura* (18 %).

4.1.4 Několikanásobná genitivní valence

Jak jsme výše již několikrát zmínili, je jedním z typických rysů genitivní valence v případě některých substantiv její řetězení. Jedná se především o substantiva, která nejsou sémanticky zcela samostatná (vyžadují další specifikaci) a jejich význam bývá upřesňován právě prostřednictvím genitivní valence. V mnoha případech se jedná o substantiva s natolik obecným významem, že i následující substantivum v genitivu, které je jeho valencí, má poměrně obecný význam a vyžaduje další specifikaci. O takto vázaných substantivech se tedy dá říci, že jsou spolu výskytově spojená⁶¹. Při popisu jejich valenčního rámce se tudíž nestačí omezit na konstatování, že se vyskytují s genitivním valenčním exponentem, protože tento exponent bývá sémanticky specifický. Ve vzorku námi analyzovaných substantiv bylo řetězení genitivů výrazné u tří substantiv: *vyhlášení, stanovení, překročení* (více viz i Čermáková 2006).

Analyzovaný vzorek nám tedy tak naznačil oblast, které bychom měli věnovat více pozornosti. Pomocí specializovaného dotazu⁶² jsme v korpuse vyhledali další typická substantiva, která se vyskytují následována dalšími dvěma substantivy v genitivu. Analýza výsledků nám naznačila několik sémantických skupin⁶³, která tato substantiva tvoří. Substantiva jsou často typická pro určitý žánr nebo žargon. Pro ilustraci uvádíme u každé skupiny i příklady ze SYN2000.

⁶¹ Zajímavý pohled zde nabízí Hoeyho teorie „lexical priming“ (Hoey 2005).

⁶² Dotazem do SYN2000 [tag="NN.*"] [tag="NN..2.*"] [tag="NN..2.*"] jsme získali 383 029 konkordančních řádků. Z nich jsme provedli náhodný výběr 5000 řádků, které byly podrobeny manuální analýze.

⁶³ Sémantické skupiny jsou utvořeny na základě analýzy výše uvedeného vzorku, nejedná se tedy o nijak vyčerpávající seznam sémantických skupin.

(1) Skupina PROCES (jeho začátek, konec nebo průběh):

cíl; dokončení; jednání; obnovení; odstranění; omezení; postup; proces; průběh; regulace; rozšiřování; stanovení; studium; udržení; ukončení; vyšetřování; využití; vývoj; zjištění; způsob

Příklady:

... se neustále stává cílem útoků fundamentalistů...

...předpokladem úspěšného dokončení procesu transformace...

(2) Skupina ZVĚTŠOVÁNÍ/ZMENŠOVÁNÍ (změna velikosti, délky, objemu atd.):

nárůst; pokles; růst; snížení; zmenšení; zvýšení

Příklad:

Meziroční nárůst indexu objemu dovozu se zvýšil...

(3) Skupina KAUZÁLNÍCH vztahů (velmi obecně):

důsledek; důvod; nutnost; podmínka; příčina

Příklad:

Důsledky oteplování klimatu sahají daleko za hranice antarktické oblasti...

(4) Skupina nějakého SESKUPENÍ (velmi obecně):

koncentrace; setkání; síť; skupina

Příklad:

...zmenšuje koncentrace inhibitorů koroze...

(5) Skupina ČASových výrazů:

doba; období; termín

Příklad:

... byl určen termín konání konference.

(6) Skupina výrazů označujících ZMĚNU (obecně):

zlepšení; změna

Příklad:

... nevedlo ke zlepšení prognózy onemocnění...

(7) Skupina PARTITIVNÍCH výrazů:

část; míra

Příklad:

A právě v tom okamžiku se propadla část střechy pavilonu

(8) Skupina výrazů pro JEDNOTKY míry:

tempo; rychlost

Příklad:

...s všeobecně akcelerujícím tempem života společnosti...

(9) Skupina AKCE (obecně):

aktivita; kontrola; ochrana; zákaz

Příklad:

Například kontrola znečištění ovzduší představuje...

(10) Skupina EVALUATIVNÍCH výrazů:

*bezpečnost; hodnocení; konflikt; možnost; nebezpečí; potřeba; požadavek;
problém; respektování*

Příklady:

... podporuje bezpečnost přenosu dat po Internetu...

Při hodnocení vlivu rodiny na spotřební chování...

(11) Výrazy součástí TERMÍNŮ, jako např.:

trest odnětí svobody; čin

Příklad:

trestný čin podvracení republiky

(12) TECHNICKÝ žargon:

*funkce; metoda; princip; problematika; program; systém; technologie; výpočet;
výsledky*

Příklad:

Při použití dálkového ovládní pak funkce rychlosti svařování slouží současně jako...

...typicky mužskou metodu řešení problémů jejich komplikováním muži používají v práci a ve veřejném životě...

Specializovaný dotaz do SYN2000 jsme zadali i na substantiva následovaná 3 substantivy v genitivu (*pustil se sám do líčení detailů případu muže bez tváře*) a získali jsme 47 729 konkordančních řádků (srov. 383 029 konkordančních řádků při dotazu na substantiva následovaná 2 gen.). Letmá manuální analýza vzorku náhodných 5000 řádků ukázala, že se zde částečně vyskytují stejná substantiva jako výše a dále mnoho substantiv následovaných *propriem* (*učinil ho zpovědníkem císařovny Marie Terezie*). Při dalších dotazech na řetězcí se genitivu se číslo úměrně snižuje a ve výsledcích stoupá počet *proprií*. Dotaz na substantivum následované 4 genitivu nám dal 6327 konkordančních řádků, dotaz na substantivum následované 5 genitivu 830 konkordančních řádků a dotaz na substantivum následované 6 genitivu 144 konkordančních řádků (zde již šlo téměř výhradně o vlastní jména nebo názvy). Níže uvádíme 3 příklady z výsledků na dotaz na substantivum následované 7 genitivu:

- (1) Znájména zrádných generálů Velké vzrušení v sále Nejvyššího sovětu SSSR vyvolalo včerejší **vystoupení poslance předsedy Svazu důstojníků občanů Ukrajiny plukovníka Vilena Martyrosjana**. (zde je genitivů dokonce 9)
- (2) Americký týdeník se odvolává na **prohlášení zástupce předsedy sboru náčelníků štábů admirála Davida Jeremiaha**, který...
- (3) Podle **sdělení vedoucího odboru Správy majetku města Miroslava Grepla** projevila o tento druh spolupráce zájem ještě další firma, která ...

Obecně se tedy dá říci, že řetězení genitivů je typické pro psaný jazyk, pro publicistiku i odborný styl. Vyskytuje se se substantivy, jejichž sémantika není dostatečně specifická a genitiv upřesňuje a zužuje jejich význam (*vyšetřování zločinů komunismu*). Frekventovaně se tato konstrukce objevuje v názvech (institucí apod.) a bývá součástí titulů, kde na genitivní pozici bývá vlastní jméno (*prezident Václav Havel*), dále je typická pro právnícký jazyk (*trest odnětí svobody; návrh novely zákona* atd.) a v nezanedbatelné míře se objevuje v textech se sportovní tematikou (*titul mistra světa*).

4.1.5 Konkurence genitivu a atributivního adjektiva

Jak již bylo uvedeno, nejběžnějším valenčním exponentem substantiv je bezpředložkový genitiv. Z 99 analyzovaných substantiv se pouze 8 s genitivem nevyskytovalo (*ves; obilí; chudák; smetana; disident; žloutek; moucha; baba*). Genitiv, zejména posesivní a kvalitativní, je však natolik produktivní, že není vyloučené, že by se s některými z těchto substantiv mohl pojít. Tato tendence přibližuje některé typy genitivního doplnění charakteru atributivního adjektiva. V povrchové syntaxi atributivní adjektivum i genitivní doplnění představují stejný větný člen – přívlástek (shodný, resp. neshodný). Tato shoda jistě není náhodná, a nejedná se tak pouze o syntaktickou a formální charakteristiku.

Při bližším zkoumání si u některých substantiv všimneme toho, že stejné nebo podobné významy vyjadřované u daného substantiva prostřednictvím bezpředložkové genitivní valence, např. *porost trávy*, mohou být vyjádřeny (a bývají vyjadřovány) i atributivním adjektivem, např. *travní porost*. Jde především o substantiva, která jsou hojně rozvíjena i jinými atributivními adjektivy a jejich genitivní valence je spíše okrajová. Kdy tedy dává substantivum přednost genitivu a kdy adjektivu? Důvody by mohly být např. právě kolokační (frekvence výskytu ustálila jistá spojení) a/nebo stylistické. Některá substantiva navíc snáze tvoří adjektiva, některá však méně často nebo vůbec ne. Důvody by pak mohly být i syntaktické – adjektiva se např. užívá tam, kde za rozvíjeným substantivem následuje jiný valenční exponent (více viz Čermáková 2006:111–116).

Čermák a Holub (1988:155–156) zmiňují dvě hlavní kombinatorické role adjektiva: strukturu A–S a (S)–V–A, které lze vyjádřit synonymním víceslovným korelátem. Autoři upozorňují, že strukturně zde však jde o dva případy souvztažnosti: (1) adjektivum se nemění a druhé substantivum je jeho určitý

interpretovaný rys (*zelené auto* vs. *auto je zelené barvy*), (2) adjektivum se mění v druhé substantivum a většinou se zde dodává i druhé, pomocné adjektivum (*spokojený přítel* vs. *přítel je plný spokojenosti*; sem patří i následující autory uváděné příklady: *cenná sbírka* vs. *sbírka velké ceny*; *výrobní program* vs. *program výroby*).

Podíváme se blíže na některé příklady tohoto typu vyskytující se v našich datech. Příklady uvádíme v abecedním pořadí.

AMBASÁDA

Substantivum *ambasáda* je ve více než polovině svých výskytů modifikováno atributivním adjektivem. V náhodném vzorku 300 konkordančních řádků stálo před tímto substantivem 148krát adjektivum specifikující stát, jemuž ambasáda přísluší (toto číslo zahrnuje i výskyt dvou obecnějších adjektiv *zahraniční* a *západní*). Další 8 adjektiv označovalo město, kde se ambasáda nachází. 16krát se před substantivem *ambasáda* vyskytlo přivlastňovací zájmeno (*jejich, naše*).

Genitivní doplnění specifikující stát, kterému ambasáda patří, se ve vzorku objevilo pouze 15krát. V bezprostřední blízkosti tohoto substantiva se však 17krát nacházelo i příslovečné určení místa označující, kde se ambasáda nachází. V dalších 20 případech bylo z blízkého kontextu (tedy kontext jedné konkordanční řádky) patrné, o jakou ambasádu jde. Z výše uvedeného tedy vyplývá, že substantivum *ambasáda* se v textu v převážné většině (75 %) vyskytuje blíže určeno.

Podíváme-li se na žánrové rozložení slova *ambasáda* do *Frekvenčního slovníku češtiny* (2004), zjistíme, že se nejčastěji vyskytuje v publicistice (74 %) a nejméně v literatuře odborné (5 %). Výsledky analýzy uvedené v odd. 4.1.2 naznačují, že bezpředložková genitivní valence je typická u slov používaných v odborném

stylu. Volba mezi genitivem a adjektivem tedy může být daná žánrem, a je tudíž stylistická.

CITÁT

Ve zkoumaném vzorku se genitivní doplnění u substantiva *citát* vyskytovalo ve 14 % a na sémantické rovině vyjadřovalo tři různé významy. Nejčastějším významem (téměř 79 % genitivů) genitiv vyjadřoval autora *citátu* (konkrétně: *Aristotela* nebo obecněji (okrajově): *klasiků*). V menší míře genitiv specifikoval časové zařazení *citátu* (*citát týdne/měsíce*) a ještě méně text/dílo, z kterého *citát* pochází (*kanonických spisů; židovských žalmů*).

První a třetí výše uvedený význam bývá u tohoto substantiva častěji vyjadřován prostřednictvím předložky *z/ze* (26,6 % z výskytů substantiva), a to opět častěji konkrétně a méně často na obecné rovině (*z klasiků*). Atributivní adjektivum těmito dvěma způsobům vyjadřování těchto významů konkuruje v okrajovější míře (3 % výskytů). I zde adjektivum častěji vyjadřuje konkrétního autora nebo text (*biblický, Masarykův*) a okrajově obecný původ *citátu* (*klasický, marxistický*).

CIVILIZACE

Substantivum *civilizace* nemá výrazné doplnění genitivního typu (v necelých 2 % svých výskytů), je však hojně rozvíjeno atributivním adjektivem (v asi polovině svých výskytů). S některými adjektivy tvoří ustálenější kolokace – *západní* (5 %); *evropská* (4 %); *lidská* (3 %); *moderní* (3 %); *technická* (2 %). Atributivní adjektiva vyskytující se se substantivem *civilizace* lze typově rozdělit především na adjektiva časoprostorová (*evropská; americká; dnešní; dávná*) a dále skupinu různých hodnotících adjektiv (*choromyslná; industriální; konzumní; kosmopolitní*).

Genitivní doplnění vyskytující se zde okrajově spadá ze sémantického hlediska do časoprostorové oblasti (*civilizaci celé Evropy; byzantské civilizace přelomu 11.–*

12. století), dále genitivy označují kulturní prostor (*civilizaci křesťanské Evropy*), etnický, národnostní nebo historický původ civilizace (*civilizace Mayů*) a dále zde najdeme samozřejmě různorodou skupinu specifických atributů (*civilizace Gulagu*).

Ze sémantického hlediska jsou atributivní adjektiva a genitivy u tohoto substantiva podobné, substantivum však dává přednost atributům v adjektivní podobě. V některých případech se jedná o konkurenci těchto typů vyjádření, kde není zcela zřejmý důvod pro to či ono vyjádření, v jiných je vyjádření jednou nebo druhou formou elegantnější a jednodušší a v některých si lze převod na druhou konkurenční formu představit jen stěží (podrobnější analýza viz Čermáková (2006)).

FUNKCIONÁŘ

V náhodném vzorku 500 řádků bylo substantivum *funkcionář* modifikované atributivním adjektivem ve více než polovině svých výskytů. Po vyloučení různých hodnotících adjektiv typu *bývalý, odpovědný, přední, vedoucí, vysoký*, kterých bylo 119, nám zůstává skupina adjektiv v zásadě dvou sémantických typů. První, menší skupina (27krát), označuje místní příslušnost (např. *hradecký*) a zbývající adjektiva označují formální zařazení *funkcionáře*, tedy např. nejčastěji *akademický, fotbalový, hokejový, komunistický, odborový, stranický, svazový* apod.

Genitivní valenci mělo toto substantivum ve 36,4 % svých výskytů, jde tedy o poměrně výraznou genitivní valenci. Při bližším pohledu na genitivní kolokáty se ukazuje, že nejčastěji se ze sémantického hlediska jedná o užší specifikaci obsahu vyjadřovaného největší skupinou atributivních adjektiv (*funkcionáře sociální demokracie; f. fotbalového klubu Baník Havířov*). Jsou zde však i genitivy vyjadřující místní příslušnost (*f. Hradce a Plzně*) a ojediněle i genitivy odpovídající hodnotícím adjektivům první skupiny (*f. nižších stranických orgánů*)

– i zde je však patrná užší specifikace. V případě tohoto substantiva lze tedy konstatovat, že ačkoliv jsou atributivní adjektiva a genitivní valenční kolokáty ze stejných sémantických oblastí, atributivní adjektiva určují substantivum ve většině případů obecněji a genitivní kolokáty pak specifitěji.

HÁDKA

Substantivum *hádky* se s genitivem (vyjadřujícím účastníky *hádky*) nevyskytuje nijak výrazně (3 %). Atributivní adjektivum vyjadřující účastníky *hádky* se vyskytuje o něco častěji – ve 4,6 % příkladů analyzovaného vzorku. Ve vzorku se však jednalo pouze o tři adjektiva – *manželská*, *rodinná* a *sousedská*. Genitiv je v tomto smyslu rozmanitější. Stejně sémantické významy jsou u tohoto substantiva častěji vyjadřovány předložkou (*mezi* (3,3 %), *s/se* (8,3 %)).

KLIKA

Substantivum *klika* nacházíme ve třech základních významech: v konkrétním (nejčastěji) a abstraktním (ve smyslu neformálního mocenského uskupení) a jako součást různých frazémů (nejčastěji *mít kliku*). Ve vzorku 300 konkordančních řádků⁶⁴ se substantivum *klika* vyskytovalo ve svém abstraktním významu v téměř 10 % výskytů. Z toho bylo substantivum ve 44 % užito s atributivním adjektivem (bez genitivního doplnění) (*mocenská*, *vládnoucí*, *vojenská*, *liberální*...) a ve 14 % s genitivním doplněním (bez atributivního adjektiva) (*s klikou zabijáků*). V 17 % u substantiva stálo atributivní adjektivum i genitivní doplnění (*revizionistická klika Dubčeka*; *početná klika příznivců*). Třikrát bylo substantivum následováno příslovecným určením (*kolem arcivévodý*; *uvnitř BIS*; *v zemi*), dvakrát zde bylo atributivní adjektivum a další specifikace v podobě nominativu (*vládnoucí klika Masaryk – Beneš*), v jednom případě jsme narazili na pouze nominativní doplnění (*s chybnými praktikami kliky Göring – Hitler*). Jak vyplývá z výše uvedeného,

⁶⁴ U tohoto substantiva bohužel narážíme na špatnou lemmatizaci – ve vzorku se nacházejí i příklady *kliky* (ve smyslu posilovacího cvičení), *Klika* jako příjmení a v jednom z příkladů jde dokonce o *klíče* (chybí diakritika).

atributivní adjektivum si s genitivem (event. nominativem) konkuruje pouze v některých případech. Adjektivum v těchto případech častěji charakterizuje *kliku* obecněji (*vojenská, vládnoucí*), zatímco genitiv (či nominativ a někdy i příslovečné určení) je mnohem konkrétnější.

KOLAPS

Ve zkoumaném vzorku bylo substantivum *kolaps* rozvíjeno atributivním adjektivem vyjadřujícím „co kolabuje“ ve 31 % výskytů (nejsou zde tedy zahrnuta adjektiva jako *hrozící, faktický, totální...*). Třetina těchto adjektiv bylo adjektivum *dopravní*. Mezi další frekventovanější adjektiva patřila adjektiva *ekonomický, finanční, hospodářský* a *srdeční*. V 9 příkladech bylo substantivum *kolaps* rozvíjené atributivním adjektivem následováno ještě příslovečným určením místa (*Dopravní kolaps v Karmelitské ulici...*) a v 11 případech bylo následováno genitivním doplněním (*hlavní příčinou finančního kolapsu amerického zdravotnictví; dopravní kolaps velkoměstské civilizace*).

Genitivní doplnění mělo toto substantivum ve 34 % svých výskytů, tedy frekvence je podobná jako u atributivního adjektiva. Na rozdíl od výskytů s atributivním adjektivem, kde tvoří frekvenčně výrazné kolokace, však nebylo mezi genitivními doplněními žádné s natolik výraznou frekvencí jako některá adjektiva. Mezi genitivními doplněními se opakovaně objevily následující kolokáty: *dopravy, trhu, Sovětského svazu, systému, banky*. Tedy slova ze stejné sémantické oblasti jako nejčastější adjektiva. Adjektiva specifikují význam tohoto substantiva obecněji než jeho genitivní valence, která bývá po sémantické stránce většinou velice konkrétní (*hrozil úplný kolaps slovenských elektráren*).

MOR

Substantivum *mor* se vyskytuje s genitivním doplněním v asi 10 % svých výskytů. Asi polovina těchto genitivních kolokátů odkazuje na abstraktní význam *moru* (*komunismu, nacismu*) a polovina na konkrétní výskyt *moru* jako choroby (např.

prasad). V tomto smyslu je dvakrát častější vyjádření pomocí atributivního adjektiva (*prasečí*). Podobně často se vyskytují atributy specifikující typ *moru* jako choroby – *plicní, dýmějový* apod. Adjektiva typu *prasečí* (*včelí* a další) se zdají být typičtější pro odbornější styl a většinou se jedná i o název choroby.

PLÁTNO

U substantiva *plátno* je třeba upozornit na tři sémantické kontexty, ve kterých se objevuje. Nejméně je užíváno jako abstraktum (*velké plátno dějin*), nejčastěji je užíváno v kontextu uměleckém, a to malířském a filmovém. Ve zkoumaném vzorku bylo substantivum modifikováno atributivním adjektivem ve 42 % svých výskytech a genitivní valenční doplnění mělo v 7,8 %.

Atributivní adjektiva patřila do několika hlavních sémantických oblastí, označují např. barvu a materiál (12 % adjektiv), rozměry (12 % adjektiv) včetně specifického adjektiva z oblasti kinematografie *širokouhlý*, 7 % adjektiv označovalo umělecký směr nebo žánr (*barokní, monumentální*), necelá 3 % adjektiv označovala konkrétního autora (*giottovské, Picassova*) a 10 % adjektiv charakterizovalo kvalitu *plátna* (*jemné, voskované* apod.). Samotné adjektivum *malířské* se objevilo v 1 % příkladů.

Nejčastějším adjektivem bylo adjektivum *filmové* (17,5 %) a druhé nejčastější *stříbrné* (9,5 %). Adjektiva *projekční* a *promítací* se vyskytla v 3,8 % příkladů. Specifickým adjektivem pak bylo adjektivum *turínské* (necelé procento výskytů). Dále se zde vyskytovala různorodá skupina adjektiv bez výraznější společné sémantické charakteristiky (*staré, krásné, zrestaurované...*).

Většina genitivních kolokátů byla ze stejných sémantických oblastí. 23 % z genitivů označovalo konkrétního malíře (*Picassa*) – tedy zde je tendence ve srovnání s adjektivy jednoznačná a 18 % označovalo uměleckou skupinu

(*impresionistů*)⁶⁵. Dalších 23 % (tedy 9 výskytů) genitivních kolokátů označovalo „kino“ (tedy *kin, biografu, Febiofestu*). 4krát se objevila slova jako *malíř, malířka, student akademie*. 5krát genitivní kolokát specifikoval, na výrobu čeho bylo plátno použito (*obrazu, stanu, slunečníku*).

POROST

Genitiv u substantiva *porost* vyjadřuje dvě hlavní specifikace porostu, 1) čím je porost tvořen (85 % všech genitivů) a 2), kde se porost nachází (13,3 % genitivů). Celkově se genitiv vyskytuje se substantivem *porost* ve 14,5 % všech výskytů analyzovaného vzorku.

Dvě nejvýraznější sémantické skupiny atributivních adjektiv rozvíjejících substantivum *porost* odpovídají výše naznačeným skupinám charakterizujícím genitivity, tedy první nejvýraznější skupina označuje, čím je *porost* tvořen (nejčastějším adjektivem je zde *lesní*), a druhá skupina vyjadřuje, kde se *porost* nachází (*břehový, polní* apod.). Atributivní adjektivum je však častější a vyskytují se zde z hlediska frekvenčního i poměrně ustálené kolokace (*lesní, smrkový, listnatý*).

Srovnáme-li genitivní kolokáty a atributivní adjektiva z hlediska jejich případné konkurence, zjišťujeme, že u asi 52 % genitivů si lze poměrně snadno představit „převod“ na atributivní adjektivum, tedy lze hovořit o konkurenci na sémantické rovině. U zbývajících genitivů je však tento převod obtížnější nebo nemožný. Jsou to především případy, kdy je typ *porostu* modifikován adjektivem (v *porostech divokého konopí*), genitiv je několikanásobný (v *porostech dubů, habru, babyky, jasanu a jiných vyšších dřevin*) a genitiv má přívlástek neshodný (*porost dřevin stejného druhu a přibližně i věku*). V těchto případech se tedy domněnka, že

⁶⁵ Zde je třeba připomenout, že procentuální údaje jsou uváděny v poměru k celkovému počtu atributivních adjektiv, resp. genitivních kolokátů. Procentuální údaj je tak např. zde vyšší než u adjektiv, avšak u absolutních čísel tomu tak samozřejmě není. Atributivních adjektiv označujících umělecký směr nebo žánr bylo 15, zatímco genitivních kolokátů bylo 7.

důvod volby jedné nebo druhé alternativy (adjektiva nebo genitivu) je gramatický, zdá být na místě. Podrobnější analýza a další příklady jsou uvedeny v Čermákové (2006:112–114).

POVSTÁNÍ

Substantivum *povstání* je rozvíjeno atributivním adjektivem ve více než polovině svých výskytů (54 %). Mezi těmito atributivními adjektivy je několik výrazných sémantických skupin. Největší skupinu tvoří adjektiva označující konkrétní národ (event. náboženství): *arabské, kurdské, slovenské národní* apod. (téměř 36 % z adjektiv); další dvě velké skupiny (25 % resp. 24 %) tvoří adjektiva specifikující typ povstání (*antikoloniální, lidové, selské, stavovské* apod.) a místo jeho konání (*pražské*). Menší ucelenější skupinu pak ještě tvoří adjektiva charakterizující průběh povstání: *ozbrojené, pokojné, živelné, nezdařené* (téměř 12 %).

Největší skupina genitivních kolokátů (10,7 % z výskytů substantiva) odpovídá skupině adjektiv označujících národ (*p. Arabů*) a částečně i skupině specifikující typ povstání (*sedláků* vs. *selské*). Avšak substantivum *povstání* má další výraznou valence v podobě předložky *proti*, která se projevuje např. i mezi atributivní adjektivy typu *protibritské, protivládní* apod. a velice často se v bezprostředním kontextu tohoto substantiva vyskytuje i příslovecné určení místa.

SVAL

V náhodném vzorku 500 konkordančních řádků stálo substantivum *sval* s atributivním adjektivem 216krát. Z toho asi polovina adjektiv⁶⁶ (110) měla takový charakter, že umožňovala paralelní vyjádření pomocí genitivu. Nejčastějšími adjektivy byla adjektiva (ne nutně v těchto tvarech a častěji v pl.): *srdeční, stehenní, zádový, břišní, lýtkový, prsní, hýžděový*.

⁶⁶ Do tohoto čísla nebyla zahrnuta taková adjektiva jako *dýchací*, u kterého by převod musel být zprostředkován ještě pomocí dalšího slova (např. *svaly dýchacího ústrojí*).

Genitiv se se substantivem *sval* vyskytl 32krát. Z toho pouze 10krát se opravdu potenciálně jednalo o konkurenci s adjektivem, např. *paže, stehna, zad, bérce*. Ostatní genitivy tento převod neumožňovaly (nebo pouze obtížně), jednalo se často o spojení A+S, např. *svaly pánevního dna, dolních končetin, trávící trubice*. Dále zde byl typ genitivů, kde sice převod byl gramaticky možný, nicméně nebyl vhodný ze stylistických nebo jiných důvodů.

UBYTOVNA

Ve vzorku 300 konkordančních řádků se substantivum *ubytovna* vyskytlo ve 13,3 % výskytů s atributivním adjektivem vyjadřujícím, pro koho je *ubytovna* určena (*cizinecká, turistická*), nebo kdo je jejím zřizovatelem/majitelem (*nemocniční*). V některých případech adjektivum vyjadřovalo oba významy (*vojenská*). První z významů, tedy pro koho je *ubytovna* určena, bývá u tohoto substantiva však nejčastěji vyjadřován prostřednictvím předložky *pro* (téměř 11 % výskytů ve vzorku). Genitivní doplnění, které se vyskytuje stejně často jako předložka *pro*, však vyjadřuje, podobně jako adjektivum, oba výše uvedené významy (*ubytovny studentů; před ubytovnamí poslanců; v ubytovnách pařížské radnice*). V prvním významu se často jedná o konkurenci adjektiva a genitivu (*studentská* vs. *studentů*), ve druhém se však ve většině případů o sémantickou konkurenci nejedná.

4.1.5.1 Shrnutí

Pro zjištění, zda a případně jaká existuje konkurence mezi genitivním doplněním a atributivním adjektivem, jsme podrobněji analyzovali 13 substantiv. U pěti z těchto substantiv (*ambasáda; citát; hádka; povstání; ubytovna*) k vzájemné sémantické „konkurenci“ mezi adjektivem a genitivem existovala výrazněji i konkurence v podobě předložky nebo příslovečného určení.

Ve většině analyzovaných příkladů se ukázalo, že se nejedná o opravdovou konkurenci, byť byla adjektiva i genitivy ze stejné nebo podobné sémantické oblasti. Adjektivum většinou vyjadřovalo obecnější specifikaci, zatímco genitiv byl konkrétnější. V případech, kde by se dalo hovořit o opravdové konkurenci, se nabízí například hypotéza, že genitiv je příznačnější pro odborný styl a adjektiva pro styl publicistický. Volba mezi genitivem a adjektivem může také být dána gramatickými a stylistickými důvody. Tato problematika by vyžadovala rozsáhlejší analýzu, která přesahuje rámec této práce.

4.1.6 Alternativní vyjádření genitivní valence u některých substantiv

Alternativním vyjádřením genitivní valence zde máme na mysli vyjádření stejného sémantického významu prostřednictvím jiného valenčního exponentu (jiný pád nebo předložka), event. prostřednictvím příslovečného určení.

V odd. 4.1.5 výše jsme na několik takovýchto případů upozornili. Jednalo se o substantiva *citát*, *hádka*, *povstání* a *ubytovna*. U substantiva *citát* se genitivní doplnění vyskytovalo ve 14 % případů a na sémantické rovině vyjadřovalo tři různé významy: (1) nejčastěji genitiv vyjadřoval autora *citátu* (konkrétně: *Aristotela* nebo obecněji (okrajově): *klasiků*), (2) dále (méně často) genitiv specifikoval časové zařazení *citátu* (*citát týdne/měsíce*) a (3) okrajově text/dílo, z kterého *citát* pochází (*kanonických spisů; židovských žalmů*). První a třetí z těchto významů bývá u tohoto substantiva častěji vyjadřován prostřednictvím předložky *z/ze* (26,6 % z výskytů substantiva), a to opět častěji konkrétně a méně často na obecné rovině (*z klasiků*). Substantivum *hádka* se s genitivem, který označuje účastníky *hádky*, vyskytuje pouze ve 3 % výskytů. Stejný sémantický rys je u tohoto substantiva častěji vyjadřován pomocí předložek: *mezi* (3,3%) a *s/se* (8,3%). U substantiva *ubytovna* se vyskytuje genitivní doplnění přibližně v 11 % výskytů a vyjadřuje, pro koho je *ubytovna* určena, nebo kdo je jejím zřizovatelem/majitelem (*ubytovny studentů; před ubytovny poslanců; v ubytovnách pařížské radnice*). První z významů, tedy pro koho je *ubytovna* určena (genitiv jej vyjadřuje v 2,66 % výskytů substantiva), bývá u tohoto substantiva však nejčastěji vyjadřován prostřednictvím předložky *pro* (téměř 11 % výskytů ve vzorku). Následující tabulka shrnuje alternativní vyjádření genitivního doplnění u těchto substantiv.

Tabulka 7a. Příklady alternativního vyjádření bezpředložkového G předložkami

Substantivum	Frekv. gen. doplnění	Předložka a její frekvence
<i>citát</i>	14 %	<i>z/ze</i> 26,6 %
<i>hádky</i>	3 %	<i>mezi</i> 3,3 %; <i>s/se</i> 8,3 %
<i>ubytovna</i>	2,66 %*	<i>pro</i> 11 %

* Procentuální údaj zahrnuje pouze genitivy vyjadřující stejný význam jako předložka *pro*.

Podívejme se na několik dalších příkladů v následující tabulce (příklady jsou uvedeny v abecedním pořadí).

Tabulka 7b. Příklady alternativního vyjádření bezpředložkového G předložkovým spojením/ jiným pádem

Substantivum	Genitivní doplnění	Alternující předložka/pád
<i>Dodatek</i>	12 % – <i>dodatek ústavy</i>	<i>k/ke</i> 25 % – <i>dodatek k ústavě</i>
<i>pasáž</i> **	12,2 % – <i>pasáž Vieweghovy knihy</i>	<i>v/ve</i> 1,6 % – <i>pasáže v Jamesi Joyceovi</i> ; <i>z/ze</i> 4,8 % – <i>pasáž z Máchova Máje</i>
<i>Pomník</i>	24,75 % – <i>pomník padlých</i>	dativ 5,75 % – <i>pomník padlým</i>
<i>recenze</i> **	9,2 % – <i>dvě recenze románu Jana Pelce</i>	<i>na</i> 3,6 % – <i>napsal recenzi na sbírku Pavla Kohouta</i>
<i>Tolerance</i>	6,25 %*** – <i>tolerance různých konfesí</i>	<i>k/ke</i> 4 % – <i>tolerance k jinému názoru</i> ; <i>vůči</i> 2 % – <i>tolerance vůči tomuto předsudku</i>
<i>Volnost</i>	21,33 % – <i>volnost pohybu</i>	<i>v/ve</i> **** 5,66 % – <i>naprosto volnost v rozhodování</i>

* ve svém partitivním významu

** Zkoumaný vzorek měl 500 konkordančních řádků, avšak substantivum *recenze* se často vyskytuje jako název rubriky v novinách/časopisech, např. rubrika *Recenze a glosy* se v daném vzorku vyskytovala 79krát.

*** Toto číslo zahrnuje pouze genitivy vyjadřující "co je tolerováno", aby byl údaj srovnatelný s procenty uváděnými u přeložek.

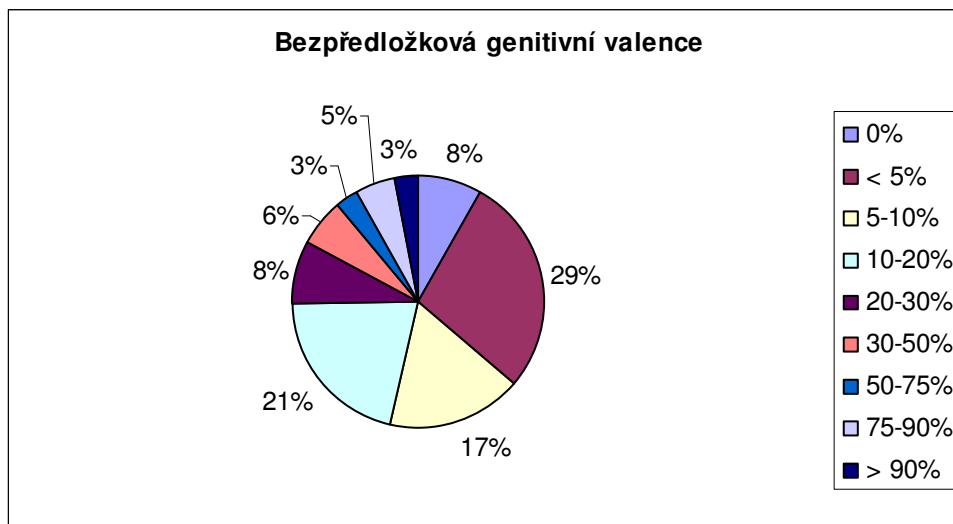
**** pouze v tomto smyslu, předložka *v/ve* je samozřejmě i součástí jiných významových konstrukcí

Z výše uvedených tabulek vyplývá, že existují určité alternativní způsoby vyjádření genitivní valence a z hlediska frekvenčního není toto vyjádření vždy opominutelné: předložka *z/ze* u substantiva *citát*; *s/se* u substantiva *hádky*; *proti* u substantiva *povstání*; *pro* u substantiva *ubytovna*; *k/ke* u substantiva *dodatek* a dativ u substantiva *pomník*. Pro učinění obecnějších závěrů nám však naše analýza neposkytla dostatečné množství dat.

4.1.7 Závěr

Bezpředložkový genitiv je nejčastějším valenčním exponentem substantiv v češtině. V námi analyzovaném vzorku 99 náhodně vybraných substantiv pouze osm nemělo genitivní valenční exponent vůbec a dalších devět substantiv mělo genitivní doplnění v naprosto okrajové míře (v méně než 1 % výskytů). Na druhou stranu pouze pět substantiv mělo genitivní valenci ve více než 80 % svých výskytů.

Následující graf ukazuje rozložení genitivní valence podle četnosti jejího výskytu u jednotlivých substantiv, tedy např. 6 % substantiv mělo genitivní doplnění s frekvencí výskytu mezi 30 až 50 %.



Graf 4. Procentuální rozložení výskytu bezpředložkové genitivní valence

Následující tabulka pak přehledně shrnuje výsledky analýzy. Jednotlivá analyzovaná substantiva jsou uvedena v prvním sloupci v abecedním pořadí. Druhý sloupec uvádí frekvenci výskytu genitivního valenčního doplnění u těchto substantiv a třetí sloupec uvádí, zda se jednalo o genitiv jednoho nebo více

sémantických typů. Frekvenční údaj ve druhém sloupci je třeba vnímat právě s ohledem na případnou sémantickou různorodost genitivního doplnění.

Tabulka 8. Bezpředložkový genitiv jako valenční exponent

Substantivum	Genitiv	Počet sémant. typů	Substantivum	Genitiv	Počet sémant. typů	Substantivum	Genitiv	Počet sémant. typů
<i>ambasáda</i>	4,30%	1	<i>nález</i>	28%	5	<i>sklonek</i>	99%	1
<i>ateliér</i>	16,80%	2	<i>obilí</i>	0		<i>směrnice</i>	23,80%	2
<i>baba</i>	0		<i>oblíba</i>	18,80%	2	<i>smetana</i>	0	
<i>bankovnictví</i>	1,70%	1	<i>odjezd</i>	17,80%	1 (2)	<i>sonda</i>	<1 %	1
<i>citát</i>	14%	3	<i>odpočinek</i>	4,50%	1	<i>soustředění</i>	13,60%	3
<i>civilizace</i>	1,60%	1	<i>odstranění</i>	88%	1	<i>stanovení</i>	94%	1
<i>deprese</i>	2,50%	3	<i>operátor</i>	12,30%	3	<i>stanoviště</i>	18%	2
<i>disident</i>	0		<i>partnerka</i>	7%	1	<i>stáří</i>	13,80%	1
<i>dlužník</i>	6%	2	<i>pasáž</i>	12,20%	2	<i>sval</i>	6,80%	2
<i>dodatek</i>	12%	2	<i>pec</i>	<1 %	1	<i>šátek</i>	<1 %	1
<i>dostih</i>	10,30%	2	<i>pero</i>	18,30%	1	<i>školka</i>	2%	2
<i>důležitost</i>	46,20%	1	<i>plátno</i>	7,50%	3	<i>tolerance</i>	13%	2
<i>funkcionář</i>	36,40%	1	<i>plavec</i>	4%	1	<i>tribuna</i>	10,30%	1
<i>hádky</i>	3%	1	<i>plemeno</i>	17,30%	1	<i>tušení</i>	8,30%	2
<i>housle</i>	<1 %	2	<i>plet</i>	3,40%	1	<i>ubytovna</i>	11%	1
<i>chata</i>	2,80%	2	<i>pocta</i>	6,60%	1 (2)	<i>udržení</i>	71%	1
<i>chudák</i>	0		<i>podzemí</i>	21%	1	<i>vaření</i>	10%	1
<i>klávesnice</i>	6%	1	<i>pohoda</i>	4,20%	3	<i>vdova</i>	<1 %	1
<i>klika</i>	14%	2	<i>pojistné</i>	1%	1	<i>ves</i>	0	
<i>kolaps</i>	34%	1	<i>policajt</i>	<1 %	1	<i>vila</i>	7,80%	1
<i>kompresor</i>	2,60%	1	<i>pomník</i>	24,80%	1	<i>viník</i>	27,30%	1
<i>konání</i>	57%	2	<i>pomocník</i>	12%	1	<i>volnost</i>	21,30%	1
<i>koridor</i>	2,30%	3	<i>porost</i>	14,50%	1	<i>výdělek</i>	6,20%	1
<i>korupce</i>	5,20%	1	<i>poskytování</i>	96%	1	<i>vyhlášení</i>	88%	2
<i>krb</i>	<1 %	1	<i>povstání</i>	10,70%	2	<i>vzhled</i>	31%	1
<i>křivka</i>	32%	1	<i>překročení</i>	89%	1	<i>vzpoura</i>	18%	1
<i>kuře</i>	<1 %	2	<i>příznak</i>	48%	1	<i>zájezd</i>	5,60%	2
<i>letectvo</i>	7,50%	2	<i>recenze</i>	23,50%	2	<i>zamýšlení</i>	2,30%	1
<i>maličkost</i>	<1 %	1	<i>recese</i>	3%	2	<i>zástupkyně</i>	57%	1
<i>mitink</i>	9,40%	3	<i>respektování</i>	85%	1	<i>zavádění</i>	89%	1
<i>monopol</i>	23,40%	2	<i>rozpaky</i>	5,40%	1	<i>zodpovědnost</i>	8,30%	2
<i>mor</i>	10%	2	<i>rukavice</i>	1%	1	<i>žila</i>	4,60%	3
<i>moucha</i>	0		<i>samec</i>	2,80%	1	<i>žlutek</i>	0	

Při analýze genitivní valence u substantiv je třeba brát v potaz především dva faktory. Prvním je frekvence výskytu – čím vyšší frekvence výskytu genitivního doplnění u substantiva, tím je valence jednoznačnější. Při velice nízkých

frekvencích se nabízí otázka, zda se vůbec o valenci jedná, zda nejde spíše o volný atribut. V této souvislosti si však na druhou stranu musíme položit i otázku, zda vůbec existují substantiva, u kterých genitivní doplnění (nějakého sémantického typu) nemůže stát. Na základě naší analýzy nemělo genitivní doplnění osm substantiv, avšak analýza byla provedena na relativně malém vzorku, větší vzorek by mohl dát samozřejmě jiné výsledky (buť pravděpodobně ne dramaticky jiné). Frekvence výskytu je tedy důležitým, ne však jediným zásadním faktorem při popisu genitivní valence. Pro určení valenčnosti se zdá být v některých případech zásadní sémantický charakter daného substantiva, tedy jeho určitá významová nesamostatnost (viz odd. 4.1.2.3). Jakkoliv je problematické ji jako kritérium přesně definovat, představuje tu složku významu substantiva, která si při jeho realizaci v textu vyžádá jeho další zpřesnění prostřednictvím valence nebo jiných prostředků (viz např. atributivní adjektiva, odd. 4.1.5). Takovéto pojetí valence je v souladu s jejím lexikálním pojetím v této práci.

Druhým důležitým faktorem při analýze genitivní valence u substantiv je polyfunkční charakter tohoto pádu. Při analýze se ukázalo, že tradiční sémantické kategorie genitivu nejsou v našem případě vždy nejvhodnější a že by vyžadovaly určitou modifikaci. Navržení sémantické klasifikace genitivních valenčních exponentů by však vyžadovalo další rozsáhlejší analýzu, která přesahuje rámec této práce.

4.2 Předložková valence

Předložku je třeba většinou vnímat jako součást valence slova. Ve velké většině případů tedy nejde o její samostatnou existenci a autorskou volbu mluvčího (Čermák 1996:30). Ve spojení se substantivou může předložka být součástí valence obligatorní nebo fakultativní. Obecně tu jde však vlastně někdy o valenci dvojí – na jedné straně valenci slova (verba / substantiva / adjektiva) a na straně druhé má svou valenci i samotná předložka (Čermák 1996:33).

Čermák (1996:32) uvádí tři základní funkce předložek. Nejčastější funkcí je (1) funkce adverbální, kde předložka vyjadřuje relaci mezi verbem a substantivem (V – S), další funkcí je (2) funkce adnominální (typ S – S), kde prepozice vyjadřuje relaci mezi dvěma nominálními výrazy. Poslední funkcí je funkce adverbiální (typ PROP – S) sloužící k vyjádření relace adverbiále k větě. Pro nás je důležitá funkce adnominální, kde předložky specifikují a doplňují význam substantiva, mnohdy analogicky k funkci rozvíjejícího adjektiva (Čermák 1996:33).

Sémantika předložky souvisí se sémantikou slova, se kterým se předložka kombinuje, a slova z podobného sémantického okruhu tak mnohdy mívají stejnou předložkovou valenci. V souvislosti s předložkovou valencí vyvstává i výše uvedená otázka její obligatornosti či fakultativnosti. Není vždy snadné rozlišit, kdy se jedná o pouhou kombinaci substantiva a předložky a kdy o valenci daného substantiva.

Ve zkoumaném vzorku substantiv (viz 3.2) se objevily pouze některé předložky a mnohdy se nevyskytly ve statisticky významné míře. Zkoumaný vzorek substantiv tedy zde nebyl dostačující pro reprezentativní popis předložkové valence substantiv. Předložková valence tak byla dohledána speciálním dotazem do

korpusu na <substantivum + předložka>. Výsledky tohoto hledání pojednáváme u jednotlivých předložek. U hodnocení těchto výsledků, zejména pak frekvenčních údajů, je třeba mít na paměti, že předložka za substantivem v některých případech s tímto substantivem vůbec nemusí valenčně souviset. Zatímco např. předložka *s* ve spojení *problémy s vlhkostí* je valenčním exponentem substantiva *problém*, ve větě *...druhý problém s ní úzce souvisí...* je předložka *s* valencí slovesa *souviset*. Díky svému postavení za substantivem se však i takovýto výskyt dostává do celkového počtu hledání <problém + s>. U těchto výsledků si tedy obvykle všímáme substantiv s předložkou s vyšší frekvencí výskytu (pokud není uvedeno jinak, tedy frekvenci alespoň 300 výskytů). Následující tabulka uvádí padesát nejfrekventovanějších spojení substantiva s předložkou v SYN2000.

Tabulka 9. Nejfrekventovanější spojení substantiva s předložkou v SYN2000.

Pořadí	Substantivum s předložkou	Frekvence	Pořadí	Substantivum s předložkou	Frekvence
1.	zákon o	14 356	26.	návrh na	4 639
2.	souvislost s	9 159	27.	pokus o	4 627
3.	rozdíl od	8 983	28.	práce v	4 625
4.	spolupráce s	8 599	29.	problém s	4 588
5.	informace o	8 417	30.	podmínka pro	4 553
6.	ohled na	7 995	31.	jednání s	4 501
7.	zájem o	7 794	32.	daň z	4 395
8.	vztah k	7 583	33.	setkání s	4 351
9.	rok v	7 363	34.	obchod s	4 337
10.	vstup do	7 239	35.	vztah mezi	4 337
11.	místo v	6 483	36.	právo na	4 289
12.	člověk v	6 323	37.	úřad v	4 261
13.	pohled na	6 283	38.	podíl na	4 190
14.	náklad na	6 075	39.	soulad s	4 155
15.	situace v	5 985	40.	smlouva o	4 128
16.	změna v	5 914	41.	rozdíl mezi	4 111
17.	srovnání s	5 780	42.	boj o	4 027
18.	zpráva o	5 713	43.	svět v	3 990
19.	cesta k	5 638	44.	soud v	3 987
20.	přístup k	5 613	45.	den v	3 983
21.	Ústí nad	5 275	46.	rozhovor s	3 924
22.	cesta do	5 111	47.	jednání o	3 865
23.	nárok na	5 104	48.	škola v	3 773
24.	dohoda o	5 068	49.	vliv na	3 738
25.	člověk z	4 992	50.	žádost o	3 679

Jak si můžeme na základě tabulky všimnout, mezi těmito padesáti nejfrekventovanějšími spojeními substantiva s předložkou se vyskytují pouze některé předložky: *o; s; od; na; k; v; do; nad; z; pro; mezi*. Na tomto vzorku jsou také na první pohled patrné možné problémy v souvislosti s určováním předložkové valence. Jsou zde např. slova s velmi vysokou frekvencí výskytu, která žádnou pevnou valenci nemají (*člověk; rok; den*) (viz také 4.9). Jde o hodně vysoká a obecná hyperonyma, která svou valenci dostávají, resp. specifikují v kontextu. Vzhledem k povaze korpusu (resp. jeho složení, viz 3.1) je na 21. místě i část vlastního jména (*Ústí nad ...*). Na první pohled je rovněž patrné, že některé kombinace nemají nutně valenční charakter, a to zejména ve spojení s předložkou *v*.

Předložky jako takové pak také mají velmi vysokou frekvenci výskytu. Mezi 10 nejfrekventovanějšími českými slovy jsou hned čtyři předložky (*v/ve; na; s/se; z/ze*) (*Frekvenční slovník češtiny 2004*). Následující tabulka uvádí v přehledu frekvenci výskytu předložek v češtině. Prostřední sloupec uvádí, kolikrát v pořadí na pomyslném frekvenčním žebříčku daná předložka je, a třetí sloupec pak uvádí její absolutní frekvenci v korpusu. V tabulce uvádíme předložky, které dále pojednáváme v souvislosti s valencí substantiv. Můžeme si zde všimnout, že i předložka s nejnižší frekvencí výskytu patří mezi nejfrekventovanější část slovní zásoby.

Tabulka 10. Frekvence výskytu českých předložek
(na základě *Frekvenčního slovníku češtiny* (2004))

Předložka	Frekvence - pořadí	Celková frekvence
<i>v / ve</i>	2	2 487 420
<i>na</i>	5	1 643 934
<i>s / se</i>	7	879 319
<i>z / ze</i>	9	844 271
<i>o</i>	11	656 915
<i>do</i>	13	612 558
<i>k / ke</i>	16	519 661
<i>pro</i>	18	393 986
<i>za</i>	21	387 212
<i>po</i>	26	311 686
<i>od</i>	30	245 516
<i>u</i>	45	182 277
<i>při</i>	48	167 659
<i>před</i>	57	142 297
<i>mezi</i>	70	115 815
<i>proti</i>	105	76 278
<i>nad</i>	111	73 019
<i>kolem</i>	208	42 927
<i>vůči</i>	720	15 534

Struktura kapitoly 4.2 je následující. Kapitoly 4.2.1 až 4.2.14 se pokusí popsat předložkovou valenci substantiv tak, jak ji naznačují výsledky analýzy založené na korpuse SYN2000. Jednotlivé podkapitoly (4.2.1 až 4.2.13) jsou věnovány jednotlivým předložkám. Kapitola nejprve uvádí popis předložky založený na analýze zkoumaného vzorku substantiv (viz 3.2.2) a dále výsledky rozšířeného hledání (viz výše a 3.4). Na závěr kapitoly jsme se pokusili substantiva vyskytující se s danou předložkou shrnout do sémantických skupin. Kapitola 4.2.14 se věnuje předložkám vyskytujícím se s nižší frekvencí. Kapitola 4.2.15 se věnuje alternacím mezi předložkami.

4.2.1 Předložka „do“

Ve zkoumaném vzorku substantiv se předložka <do + gen.> objevila ve spojení se substantivou především jako součást příslovečného určení místa (více viz kapitola 4.4). I při rozšířeném hledání <substantivum + předložka> se předložka *do* objevuje především jako součást adverbialu. Mezi 50 nejčastějšími kombinacemi (viz tabulka výše) substantiva následovaného předložkou se předložka *do* objevila se substantivy *vstup* a *cesta*. V rozšířeném hledání se s frekvencí vyšší než 400 předložka *do* objevila ve spojení s následujícími substantivy: *volba; investice; návrat; postup; odchod; zásah; vchod; nástup; přístup; výlet; zápis; směr; zařazení; přechod; kandidát; okno*.

Následující tabulka v přehledu uvádí výše uvedená substantiva a jejich nejčastější kolokáty. Ve druhém sloupci je celková frekvence výskytu spojení <substantivum + do>⁶⁷ a v závorce je velikost vzorku, ze kterého jsou vypočítávány kolokáty v rozsahu 0 až 1.

⁶⁷ Znovu zde připomínáme orientační charakter frekvenčních údajů daný tím, že se jedná o automatické výstupy hledání bez následné podrobné manuální analýzy.

Tabulka 11. Vybraná substantiva s předložkou *do* a jejich kolokáty.

Substantivum (lemma)	Frekvence výskytu s předložkou <i>do</i> (vel. vzorku)	Nejčastější kolokáty (počet ve vzorku)
<i>vstup</i>	7 239 (500)	<i>NATO</i> (70); <i>EU</i> (51)
<i>cesta</i>	5 111 (500)	<i>zahraničí</i> (27); <i>Evropy</i> (15); <i>pekél</i> (13); <i>USA</i> (11) <i>Evropského parlamentu</i> (10); <i>parlamentu</i> (11);
<i>volba</i>	2 331 (300)	<i>poslanecké sněmovny</i> (47); <i>Senátu</i> (62)
<i>investice</i>	2 202 (300)	<i>akcií</i> (8); <i>infrastruktury</i> (14); <i>modernizace</i> (7); <i>nemovitostí</i> (5); <i>vzdělání</i> (6)
<i>návrat</i>	2 178 (300)	<i>vlasti</i> (29); <i>Prahy</i> (12); <i>Evropy</i> (10); <i>Čech</i> (9)
<i>postup</i>	2 085 (300)	<i>semifinále</i> (36); <i>finále</i> (34); <i>čtvrtfinále</i> (19); <i>play off</i> (24)
<i>odchod</i>	1 260 (300)	<i>důchodu</i> (144); <i>penze</i> (19); <i>kabin</i> (7); <i>opozice</i> (8); <i>zahraničí</i> (7)
<i>zásah</i>	1 212 (300)	<i>krajiny</i> (9); <i>přírody</i> (8); <i>struktury</i> (7)
<i>vchod</i>	909 (200)	<i>budovy</i> (16); <i>domu</i> (11); <i>jeskyně</i> (8)
<i>nástup</i>	880 (200)	<i>funkce</i> (60); <i>zaměstnání</i> (14); <i>úřadu</i> (11)
<i>přístup</i>	735 (200)	<i>paměti</i> (7); <i>Internetu</i> (6); <i>sítě</i> (5)
<i>výlet</i>	604 (200)	<i>okolí</i> (10); <i>přírody</i> (9)
<i>zápis</i>	522 (200)	<i>Guinnessovy knihy</i> (23); <i>katastru</i> (13); <i>obchodního rejstříku</i> (30)
<i>směr</i>	512 (z toho 386 směrem do)	<i>centra</i> (56)
<i>zařazení</i>	490 (150)	<i>sítě</i> (8); <i>programu</i> (8); <i>seznamu</i> (5)
<i>přechod</i>	479 (150)	<i>útoku</i> (13); <i>Německa</i> (7)
<i>kandidát</i>	419 (150)	<i>Senátu</i> (49); <i>voleb</i> (9)
<i>okno</i>	412 (150)	<i>dvora</i> (16); <i>budoucnosti</i> (10)

Jak je z tabulky patrné, substantiva zde intuitivně tvoří jakési volné sémantické skupiny (1. *vstup*; *vchod*; *přístup*; *přechod*; *okno*, 2. *cesta*; *výlet*; *směr*, 3. *volba*; *kandidát*, 3. *investice*; *zásah*, 4. *návrat*; *odchod*, 5. *postup*; *nástup*, 6. *zápis*; *zařazení*). Nicméně, v některých případech je předložka *do* spíše součástí příslovečného určení místa, které má valenční charakter.

V rozšířeném hledání se objevila i substantiva *ruka* a *hlava*. Zde byla předložka *do* součástí frazémů (z *ruky do huby*; *ruka do ohně*; z *ruky do ruky*; z *ruky do úst*; *hlavu do oprátky*; *hlavu do písku*; *hlavou do zdi*) nebo častých kolokací (*ruce do kapes*; *ruku/rukou do kapsy*; *ruce do klína*; *hlavu do dlaní*, *hlavu do dveří/klína*), nebyla však jejich valencí.

Dále zde byla i skupinka časových substantiv *rok*, *den*, *hodina*, *leden* a *ráno*, ale ani zde se nejednalo o valenci. Předložka *do* byla součástí ustálenějších kolokací. Mezi nejčastějšími kolokátů⁶⁸ substantiva *rok* byla slova *vězení* a *konce*, nejčastějším kolokátem substantiva *den* bylo substantivum *práce*, u substantiva *hodina* to byly slovní formy *rána* a *hodin*, mezi nejčastější kolokátů u měsíce *leden* byly tvary *června*, *konce*, *února*, *roku*. Nejčastější kolokací substantiva *ráno* bylo spojení *od rána do večera*, dále *od rána do noci* a *ráno do práce*.

⁶⁸ Výpočet kolokátů v rozsahu 0 až 3.

4.2.2 Předložka „k/ke“

Ve zkoumaném vzorku substantiv se předložka <k/ke⁶⁹ + dat.> objevila ve spojení s následujícími substantivy:

zodpovědnost (*k budoucím generacím*); **tolerance** (5,5 %) (*k extremismu*); **směrnice** (*k organizování vyšších odborných mysliveckých zkoušek*); **smetana** (*ke šlehání*); **dodatek** (25,3 %) (*dotatku č. 1 k Platební smlouvě*); **volnost** (*k improvizaci*).

U těchto substantiv tvoří předložka jejich valenci. Podíváme-li se na níže uvedené sémantické skupiny, vidíme, že výše uvedená substantiva lze do těchto skupin bez problémů zařadit. Výjimkou je substantivum *smetana*, u kterého lze předložku interpretovat jako součást příslovečného určení účelu (*smetana ke kávě / smetana ke šlehání*), jako součást ustálené kolokace nebo jako fakultativní valenci, kde předložkové spojení upřesňuje význam substantiva.

Při rozšířeném hledání předložkové valence se mezi 50 nejfrekventovanějšími kombinacemi <substantivum+předložka> objevila předložka *k* ve spojení se třemi substantivy: *vztah*; *cesta*; *přístup*. Následující tabulka uvádí procentuální výskyt⁷⁰ těchto substantiv s předložkou *k/ke* z jejich celkového výskytu a příklad užití.

⁶⁹ Mohla by zde být uvažována i předložka *ku*, která se však ve zkoumaném materiálu nevyskytovala. Letmý pohled na tuto předložku do ČNK nám říká, že se nejčastěji vyskytuje v konstrukci „X ku Y“, kde X a Y jsou číslovky, dále ve frazeologických spojení typu *být ku prospěchu*; *sportem ku zdraví*. Je i součástí víceslovné předložky *směrem ku*, kde by mohla být interpretována jako valence substantiva *směr*. Vyskytuje se i ve specifickém spojení se substantivem *hlas* v plurálu, kde toto substantivum předchází číslovka: *X hlasů ku*.

⁷⁰ Uváděné procento je pouze velmi orientační (tedy výsledky zde nebyly manuálně tříděny).

Tabulka 12. Vybraná substantiva s předložkou *k/ke*

Substantivum	Výskyt s předložkou <i>k/ke</i> v %	Příklad
<i>vztah k</i>	20,6 %	<i>Máte vztah k hudbě?</i>
<i>cesta k*</i>	10 %	<i>na cestě k moudrosti</i>
<i>přístup k**</i>	36,4 %	<i>o novém přístupu k umění</i>

* Pro srovnání s předložkou *do*, se kterou se toto substantivum také objevuje (viz 4.2.1), je frekvence výskytu 8,9 %.

** Frekvence výskytu s předložkou *do* je 4,8 %.

Dále jsme se pokusili rozdělit substantiva vyskytující se v rozšířeném hledání s předložkou *k/ke* s frekvencí vyšší než 300 do několika sémantických skupin.⁷¹ U těchto substantiv lze předložku *k* považovat za součást jejich valence, ať již obligatorní nebo fakultativní.

skupina <i>láska</i>
chuť k; láska k; nedůvěra k; nechuť k; nenávisť k; obdiv k; odpor k; sympatie k; úcta k

skupina <i>dodatek</i>
dodatek k; komentář k; poznámka k; připomínka k; příspěvek k

skupina <i>možnost</i>
možnost k; nabídka k

skupina <i>klíč</i>
klíč k; nástroj k; prostředek k; opatření k

skupina <i>podmínka</i>
podmínka k; povolení k; předpoklad k; výhrada k

skupina <i>podnět</i>
Podnět k; příkaz k; příležitost k; výzva k

skupina <i>vztah</i>
poměr k; postoj k; protiklad k; přihlídnutí k; přístup k; stanovisko k; vztah k; souhlas k; příprava k

skupina <i>síla</i>
síla k; vůle k; důvod k

⁷¹ Dělení na sémantické skupiny je vždy nutně subjektivní.

skupina <i>sklon</i>

sklon k; tendence k; směr k; příslušnost k
--

skupina <i>cesta</i>

cesta k; krok k; návrat k; přechod k; připojení k

V rozšířeném hledání se předložka *k* vyskytovala i ve spojení se substantivem *záda*, ve kterém se jedná spíše o víceslovnou předložku typicky se vyskytující se slovesy *otočit* a *obrátit* ale i *sedět*, *stát*, *být* (*zády k*) apod.

Předložka *k* se v rozšířeném hledání (s frekvencí vyšší než 300) vyskytovala i se skupinou spaciotemporálních substantiv.

skupina <i>čas</i>

čas k; doba k; rok k

skupina <i>místo</i>

místo k; prostor k

4.2.3 Předložka „mezi“

Ve zkoumaném vzorku se předložka <mezi + instrum.> objevila ve spojení s následujícími substantivy:

udržení (*bojovali s týmem Zimbabwe o udržení mezi elitní šestnáctkou*); **oblíba** (*Klaus má více oblíby přímo mezi voliči ODS*); **hádky** (3,3 %) (*hádky mezi mámou a tátou*); **koridor** (2,2 %) (*otevření humanitárního koridoru mezi Mostarem a Sarajevem*).

U těchto substantiv tvoří předložka *mezi* jejich valenci, u substantiva *koridor* může být interpretována i jako součást adverbialie, které má valenční charakter. Při rozšířeném hledání <substantivum + předložka> se mezi padesáti nejfrekventovanějšími spojeními předložka *mezi* se substantivem objevila u dvou substantiv: *vztah* a *rozdíl*. Níže uvedená tabulka ukazuje procentuální výskyt těchto substantiv s předložkou *mezi* z jejich celkového výskytu.

Tabulka 13. Vybraná substantiva s předložkou *mezi*

Substantivum	Výskyt s předložkou <i>mezi</i> v %	Příklad
<i>vztah mezi</i>	11,8 %	<i>ve vztazích mezi muži a ženami</i>
<i>rozdíl mezi</i>	17,8 %	<i>známý rozdíl mezi životem a literaturou</i>

Substantiva s předložkou *mezi* nalezená mezi výsledky rozšířeného hledání s vyšší frekvencí než 300 jsou dále rozdělena do dvou sémantických skupin („konflikt“ a „jednání“).

skupina konflikt
boj mezi; konflikt mezi; rozpor mezi; spor mezi; napětí mezi; kompromis mezi

skupina jednání

jednání mezi; komunikace mezi; kontakt mezi; poměr mezi; rozhovor mezi; smlouva mezi; souvislost mezi; spojení mezi; spolupráce mezi; vazba mezi

S předložkou *mezi* se mezi výsledky rozšířeného hledání objevila i substantiva tvořící sémantickou skupinu „cesta“. U těchto substantiv je však předložka *mezi* součástí příslovečného určení, které má valenční charakter, podobně jako substantivum *obchod*, které se zde rovněž vyskytuje s předložkou *mezi*.

skupina cesta

cesta mezi; hranice mezi; prostor mezi; vzdálenost mezi

4.2.4 Předložka „na“

Uvažujeme zde především předložku <na + akuz.>. Předložka <na + lok.> se ve spojení se substantivou objevuje často jako součást příslovečných určení, ne však výlučně. Přííslovečná určení však mají v souvislosti s některými substantivy valenční charakter (více viz 4.4). Ve zkoumaném vzorku se předložka *na* jako valence objevila u následujících substantiv:

recenze (9,3 %) (*recenze na knihu*); **monopol** (+ **na** + akuz.) (9,6 %) (*monopol na zpracování české plazmy / nikdo nemá monopol na pravdu*); **monopol** (+ **na** + lok.) (0,8 %) (*z dob jejího monopolu na trhu*); **soustředění** (6,8 %) (*na sebe sama*); **soustředění se** (1,4 %) (*soustředění se na barvu tónu*); **recese**⁷² (*Je to recese na stoletý uherskohradišský smíšený sbor Svatopluk*); **klika** (*kliku na vrátka*).

Předložka *na* se dále objevila i ve spojení se substantivem *zájezd*, kde je součástí ustálené kolokace *zájezd na poslední chvíli*.

Dále se předložka *na* objevuje se substantivy, u kterých vyjadřuje účel: *sval* (*svaly na rychlost a svaly na výdrž*); *mítink* (*mítinky na její podporu / mítinků na počest výročí / mítink na protest proti*); *pec* (*pec na chleba*); *krb* (*na dřevo / plyn*); *tribuna* (*na sezení / stání*); *šátek* (*na utírání potu / na hlavu*); *kompresor* (*na plnění náhradních lahví*).

Předložka *na* (+ akuz.) se často objevuje i v kontextu slova *pojistné*. Vyjadřuje účel pojištění, typ nebo i období, na které se pojištění sjednává (*pojistné na*

⁷² Zde valence rozlišuje význam slova. Slovo *recese* se vyskytuje ve dvou hlavních kontextech: ekonomická *recese* a *recese* ve smyslu legrace nebo parodie. Předložka *na* ve spojení se slovem *recese* je valenční především ve druhém významu, v prvním bývá součástí příslovečných určení, kde je interpretace nejednoznačná. V prvním významu, pokud se vyskytuje, je součástí příslovečného určení místa.

nemocenské pojištění; pojistné na jeho auto / na odcizení; pojistné na dobu jednoho roku). V případě spojení *pojistné na důchodové zabezpečení*, které se ve vzorku vyskytovalo s 4,7 % frekvencí, a spojení *pojistné na sociální zabezpečení*, které se vyskytovalo s 7,9 %, se jedná spíše o administrativní (právní) termíny.

Předložka *na* (+ lok.) se objevila i jako součást příslovečného určení způsobu ve spojení *kuře na* (*kuře na paprice*), které se vyskytovalo s frekvencí 2,5 % . Ve sportovním žargonu se předložka *na* (+ akuz.) používá k vyjádření vzdáleností u jednotlivých sportovních disciplín: *dostih na* (*1660 m / jednu míli*); *plavec na* (*25 km / 5 km*).

Substantivum *nález* s atributivním adjektivem *pozitivní*⁷³ má také valenci v podobě předložky *na* (+ akuz.) (*pozitivní nález na kokain*), i když její výskyt není nikterak frekventovaný. O něco frekventovanější je spojení tohoto substantiva s předložkou *na* + „místo nálezu na těle“ (tedy <*na* + lok.>) (*nález na plicích*). Výskyty obou typů jsou však nižší než 1 % z celkového výskytu tohoto substantiva.

Při rozšířeném hledání předložkové valence bylo součástí nejfrekventovanějšího spojení s předložkou *na* substantivum *ohled* (+ akuz.). Mezi 50 dalšími nejfrekventovanějšími kombinacemi substantiva a předložky se s předložkou *na* objevilo i dalších 7 substantiv (*pohled, náklad, nárok, návrh, právo, podíl, vliv*). Následující tabulka ukazuje jejich procentuální výskyt s předložkou *na* z celkového výskytu těchto substantiv v SYN2000 a příklad užití.

⁷³ V korpuse nacházíme pouze adjektivum *pozitivní*, ale dá se zde předpokládat i adjektivum *negativní*.

Tabulka 14. Vybraná substantiva s předložkou *na*

Substantivum	Výskyt s předložkou <i>na</i> v %	Příklad
<i>ohled na</i>	72 %*	<i>bere ohled na místní tradice</i> (akuz.)
<i>pohled na</i>	26 %	<i>vrhl zběžný pohled na hodinky</i> (akuz.)
<i>náklad** na</i>	27 %	<i>vyšší náklady na cestování</i> (akuz.)
<i>nárok na</i>	50 %	<i>má nárok na penzi</i> (akuz.)
<i>návrh na</i>	13,6 %	<i>návrh na rozvod</i> (akuz.)
<i>právo na</i>	9 %	<i>právo na svůj názor</i> (akuz.)
<i>podíl na</i>	21 %	<i>podíl na zisku firmu</i> (lok.)
<i>vliv na</i>	21 %	<i>dobry vliv na mravy</i> (akuz.)

* Z toho téměř 55 % tvoří spojení *bez ohledu na* a dalších 38 % spojení *s ohledem na*.

** Většinou v pl.

Na základě dalších substantiv, která se v rozšířeném hledání s předložkou *na* vyskytují (s frekvencí vyšší než 400), jsme se pokusili vytvořit několik sémantických skupin substantiv, jejichž valenci (obligatorní nebo fakultativní) tato předložka představuje.

skupina *chut'*

chut' na; nabídka na; naděje na; šance na; tip na; recept na

skupina *útok*

atentát na; hon na; lov na; střelba na; útok na

skupina *program*

příprava na; program na; projekt na

skupina *expert*

expert na; odborník na; specialista na

skupina *vstupenka*

lístek na; pozvánka na; vstupenka na; reklama na; fronta na; účast na

skupina *návrh*

návrh na; nominace na; kandidát na

skupina *vliv*

dopad na; důraz na; tlak na; vliv na; závislost na; opatření na

skupina orientace
směr na; zaměření na; orientace na

skupina návaznost
napojení na; návaznost na; působení na; reakce na; vazba na; ohled na

skupina myšlenka
myšlenka na; vzpomínka na; názor na; odkaz na; narážka na; poukaz na; podezření na; odpověď na

skupina právo
požadavek na; právo na; nárok na; stížnost na; smlouva na; zakázka na; kontrakt na; kvalifikace na; licence na; žaloba na

U dalších tří sémantických skupin je předložka *na* spíše součástí adverbíále (adverbíále místa u skupin „výhled“ a „vstup“ a adverbíále účelu u skupiny „továrna“), které však má valenční charakter.

skupina výhled
výhled na; vyhlídka na; pohled na

skupina vstup
postup na; přechod na; návrat na; přístup na; vstup na; konkurs na; cesta na

skupina továrna
stroj na; továrna na; olej na; lék na ; materiál na

Předložka *na* se frekventovaně vyskytuje i ve spojení se substantivou ze sémantické oblasti FINANCE (viz následující tabulka). U některých z těchto substantiv se jedná o valenci, u některých jde o součást adverbíále a u některých o pevné kolokace či termíny. Podrobnější analýza však přesahuje rámec této práce.

skupina FINANCE

rozpočet na; sazba na; úvěr na; výdaj na; záloha na; podíl na; příspěvek na; půjčka na; prostředek na; náklad na; clo na; daň na; dotace na; přírůstek na; fond na; přepočítání na; pojistné na

Další sémantickou skupinu, se kterou se předložka *na* objevuje, je skupina slov ze sémantické oblasti ČAS (viz tabulka dále). V tomto případě se vesměs jedná o substantiva s vysokou frekvencí výskytu, jejichž podrobnější analýza přesahuje rámec této práce.

skupina ČAS

čas na; den na; doba na; hodina na; chvíle na; měsíc na; minuta na; týden na
--

Substantiva z oblasti sportovního žargonu (viz výše) byla zastoupena i v rozšířeném hledání.

skupina SPORT

běh na; závod na

Některá slova ve spojení s předložkou *na* jsou součástí ustálených kolokací a nedají se zařadit do výše navržených sémantických skupin: **ubližení na** (*duši* (1) / *těle* (48) / *zdraví* (657) / + lok.); **dopis na** (136krát *dopis na rozloučenou*); **krok na** (*k. na cestě* (*k*) / + lok.); **ztráta na** (*životech* / + lok.; *čelo závodu* / + akuz.).

Předložka *na* se vyskytuje frekventovaně i za substantivem *ruka*, nejedná se však o valenci, i když zde nacházíme některé frazémy (*ruku na srdce* (120); *ruku na to* (13)). Substantivum *ruka* však patří mezi velice frekventovaná substantiva (celkový výskyt v SYN2000 je 53 446), u kterých můžeme očekávat odlišné

chování (viz 4.9). Podobná je situace s vysoce frekventovaným substantivem *slovo* (celkový výskyt v SYN2000 je 54 389), i zde nacházíme ustálená spojení a nejedná se o valenci (*slovo na adresu* („někoho“) (43); *s. na rozloučenou* (16); *s. na závěr* (10)).

Účel předložka *na* vyjadřuje i u podstatného jména *akce* (ve vzorku 200 řádků to bylo ve 33 % případů). Nejčastěji v následujících spojení: *a. na podporu/záchranu/pomoc/ochranu...* Další sémantickou skupinku tvoří substantiva *hra* a *variace*, ke kterým by patřilo i např. výše zmíněné substantivum *recese*. Předložka *na* se typicky vyskytuje i se substantivem *řada*. Zde je součástí frazému <*být/přijít/dojít*> *řada na* <někoho> a „(*stát*) v řadě na...“.

4.2.5 Předložka „o“

Ve zkoumaném vzorku substantiv se předložka <o + lok.> (a méně často <o + akuz.> objevila s vyšší frekvencí se substantivy *pasáž* (5,6 %) (v *pasážích o evropské integraci, pasáž o třech až čtyřech větách*) a *tušení* (7,3 %) (*ani tušení o nebezpečí*). Dále se objevila ve spojení s následujícími substantivy:

dodatek (o *ústavním dodatku o povinnosti občanů*); *dostih* (*dostih o Pohár Nadace*) (akuz.); *hádká* (4,6 %) (*hádká o odměnu*) (akuz.); *nález* (*nález o nesouladu s novelizací paragrafu; nálezy o zpracování železa*); *recenze* (*Recenzi o divadelní inscenaci...*); *směrnice* (3,5 %) (*směrnice o obalech a obalovém odpadu*); *vyhlášení* (*vyhlášení o navyšování základního jmění*); *překročení* (*při překročení o 20 km*) (akuz.); *zamyšlení* (3 %) (*zamyšlení o světě*).

Při rozšířeném hledání předložkové valence se jako nejfrekventovanější kombinace <substantivum + předložka> objevilo spojení *zákon o* (14 356krát). Mezi 50 dalšími nejfrekventovanějšími kombinacemi substantiva a předložky se s předložkou <o + lok./akuz.> objevilo i dalších 9 substantiv (*informace, zájem, zpráva, dohoda, pokus, smlouva, boj, jednání, žádost*). Následující tabulka ukazuje jejich procentuální výskyt s předložkou *o* a příklad užití.

Tabulka 15. Vybraná substantiva s předložkou *o*

Substantivum	Výskyt s předložkou <i>o</i> v %	Příklad
zákon o	20 %	<i>zákon o občanství</i> (lok.)
informace o	19 %	<i>informace o stavu ovzduší</i> (lok.)
zájem o	18 %	<i>zájem o spolupráci</i> (akuz.)
zpráva o	17 %	<i>zprávy o hospodaření</i> (lok.)
dohoda o	20 %	<i>dohoda o bezvízovém styku</i> (lok.)
pokus o	36 %	<i>pokus o atentát</i> (akuz.)
smlouva o	12 %	<i>smlouva o dílo</i> (akuz.)
boj o	19 %	<i>boj o bronz</i> (akuz.)
jednání o	11 %	<i>jednání o členství</i> (lok.)
žádost o	37 %	<i>žádost o azyl</i> (akuz.)

Na základě dalších substantiv, která se v rozšířeném hledání s předložkou *o* vyskytují, jsme vytvořili několik sémantických skupin substantiv, u nichž předložka *o* tvoří valenci nebo má valenční charakter.

skupina <i>zákon</i>
zákon o; dohoda o; smlouva o; úmluva o; protokol o; rozhodnutí o; referendum o

skupina <i>diskuse</i>
diskuse o; konference o; jednání o; debata o; seminář o; přednáška o; řeč o; rozhovor o

skupina <i>potvrzení</i>
žádost o; doklad o; potvrzení o; osvědčení o

skupina <i>kniha</i>
kniha o; příběh o; film o; dokument o; komedie o; pohádka o; legenda o; vyprávění o; pověst o; píseň o; kapitola o; studie o; článek o; teorie o; nauka o; poznatek o

skupina <i>růst</i> <o + akuz.>
pokles o; nárůst o; růst o

skupina <i>tvrzení</i>
slovo o; pravda o; výrok o; důkaz o; tvrzení o

skupina <i>informace</i>
informace o; údaj o; zmínka o; záznam o; poznámka o; oznámení o; zpráva o; materiál o

skupina <i>zájem</i>
zájem / nezájem o; zájemce o; uchazeč o; žadatel o (+akuz.)

skupina <i>představa</i>
představa o; úvaha o; přehled o; pochybnost o; spekulace o; iluze o; sen o

skupina <i>pokus</i> <o + akuz.>

péče o; pokus o; snaha o; úsilí o; obava o; starost o; pochyba o; strach o
--

skupina <i>boj</i> <o + akuz.>

boj o; spor o; soutěž o; hlasování o; rozhodování o; zápas o; hra o; bitva o; souboj o
--

4.2.6 Předložka „od“

Předložka <od + gen.> se ve spojení se substantivou na pozici valenčního exponentu objevuje méně.⁷⁴ Ve zkoumaném vzorku se několikrát objevila se substantivem *klika* (např. *klika od flašinetu*) a dále se substantivy, u kterých vyjadřuje původ nebo autora, např. *šátek od Hermése* nebo *dotatky od jednoho z našich předních odborníků*. V těchto případech se však nejedná o valenci těchto substantiv jako takovou. Předložka zde vyjadřuje jeden ze svých základních významů, označuje původce nebo autora.

Podíváme-li se však, s jakými substantivy se vyskytuje ve zvýšené míře (více než 300 výskytů při rozšířeném hledání), můžeme si všimnout, že i tato substantiva lze rozdělit do dvou hlavních sémantických skupin ČAS a VZDÁLENOST. Vyjadřování spaciotemporálních souvislostí je jednou ze základních funkcí této předložky.

skupina ČAS
rok od; čas od; den od; doba od; období od; týden od; měsíc od; sobota od pondělí od; pátek od; čtvrtek od

skupina VZDÁLENOST
kilometr od; metr od; kousek od; směr od; vzdálenost od; krok od

Dále zde najdeme substantiva vyjadřující nějaké rozmezí (*rozmezí od; platnost od; cena od; věk od*⁷⁵), v tomto kontextu můžeme v blízkosti očekávat i předložku

⁷⁴ Při rozšířeném hledání třetí nejčastější kombinací <subst. + předl.> bylo spojení *rozdíl od*, které se v korpuse vyskytuje 8 983krát (z toho však 5 177krát je součástí výrazu *na rozdíl od*).

⁷⁵ V tomto smyslu do této skupiny bude patřit i spojení *dítě od*, které bývá součástí širších spojení typu *každé dítě od 6 do 15 let*.

do, podobně jako i u některých substantiv spadajících do skupiny ČAS a VZDÁLENOST.

Podobný typ jako výše uvedené substantivum *klika* představuje i substantivum *klíč*, kde předložka *od* vyjadřuje „patřící k“, tedy původ. Původce vyjadřuje i ve spojení se substantivem *dopis* a *peníze*. Obecněji se tak původ vyjadřuje i ve spojení typu *lidé od novin*.

Za valenční lze považovat předložku *od* ve spojení se substantivem *osvobození* (331 výskytů v SYN2000). „Osvobozuje“ se od *daní* ale i např. *strachu*. Zde valence přímo souvisí s valencí slovesnou.

Jak bylo řečeno výše předložka *od* bývá součástí různých spaciotemporálních výrazů a doplňuje se s předložkou *do*, je však výrazně méně frekventovaná. Pro srovnání na dotaz do SYN2000 na substantivum *cesta* plus předložka *do* získáme 5 111 konkordančních řádků. Výskytů substantiva *cesta* plus předložka *od* je podstatně méně, pouhých 380.

Předložka *od* se vyskytuje i ve spojení se substantivem *přechod*. Zde bývá celkem pravidelně součástí spojení „*přechod od něčeho k něčemu*“ (*přechod od věcné literatura k beletrii*) s možnými variacemi s předložkou *do* a *na* (*přechodu od dětství do dospělosti; přechod od vytápění uhlím na plynové a elektrické systémy*).

Zvýšená frekvence výskytu substantiva *strana* a předložky *od* je dána tím, že spolu často tvoří součást víceslovné předložky (*stranou od ruchu města; na pravé straně od Havelského trhu*).

4.2.7 Předložka „po“

Předložka <po + lok.> se ve zkoumaném vzorku objevila ve spojení s následujícími substantivy:

zájezd (*po stopách Neila Armstronga*); vdova po (26,3 %); kuře po (*cikánsku*).

Při rozšířeném hledání se s frekvencí vyšší 2 500 výskytů předložka *po* vyskytla pouze se substantivem *rok*. Zde se však můžeme domnívat, že je celé spojení součástí časového adverbialu.

Z dalších substantiv z rozšířeného hledání vyskytujících se s frekvencí nad 300 jsme vytvořili tři následující sémantické skupiny. Valenci předložka tvoří u skupiny „*pátrání*“ a „*vdova*“. U první skupiny ČAS se jedná o specifická substantiva, jejichž další analýza přesahuje rámec této práce.

skupina pátrání
pátrání po; poptávka po; snaha po; touha po; volání po

skupina vdova
stopa po; vdova po

skupina ČAS
den po; doba po; hodina po; měsíc po; minuta po; rok po; týden po

Na příkladu substantiva *otázka* si pak můžeme opět názorně všimnout, že ne vždy je zvýšená frekvence nejlepším vodítkem pro určení valence. Jednu z valencí tohoto velmi frekventovaného substantiva tvoří právě předložka *po* (celkem 318

výskytů v SYN2000, typicky představují kolokáty něco abstraktního, např. *otázka po smyslu života*). Tento počet však představuje pouhých 0,65 % z celkového výskytu substantiva *otázka* (frekvence v SYN2000: 48 635).

Specifickou valenci předložka *po* představuje u substantiva *výhra*. Předložka *po* je vždy následována udáním výše výhry (*613 výher po 1000 Kč*) a má zde tak distributivní funkci. Tento typ spojení se však téměř nevyskytuje v normálním textu, ale patří výhradně do výsledkových sázkových listin.

4.2.8 Předložka „pro“ a „proti“

Ve zkoumaném vzorku se předložka <pro + akuz.> objevuje se substantivy s nízkou frekvencí. Kromě substantiva *směrnice* (se kterým se vyskytuje v 7 % všech jeho výskytů) a substantiva *ubytovna* (se kterým se vyskytuje v 9 % všech jeho výskytů) se okrajově (více však než jeden výskyt) objevila u následujících substantiv:

zájezd (*zájezd pro náročné turisty*); *pojistné* (*pro letošní rok / pro muže*); *vila* (*pro Štrougalovu sekretářku*); *ateliér* (*pro knižní vazbu / výrobu divadelních dekorací a loutek*); *pec* 3,7 % (*pro tavení oceli*); *sonda* (*snímací sonda pro měření kyslíku*); *dostih* (*pro špičkové koně*); *směrnice* (7 %) (*pro případ ozbrojeného střetnutí*); *pomocník* (*pomocníkem pro všechny milovníky historie*); *koridor* (*pro průjezd mezinárodní humanitární pomoci*); *pocta* (*pro příští vítězný film*); *tribuna* (*pro čestné hosty*); *operátor* (*pro kontrolu sítí*); *kompresor* (*pro speciální aplikace*); *volnost* (*pro improvizaci*); *ubytovna* (9 %) (*pro ...*); *klávesnice* (*pro češtinu*); *stanoviště* (*pro měření emisí*); *důležitost* (*pro...*).

Z výše uvedených substantiv má předložka *pro* valenční charakter u substantiv *směrnice*; *pocta*; *volnost*; *důležitost*. U ostatních bychom ji spíše interpretovali jako součást příslovečného určení účelu. Můžeme zde pozorovat určitou sémantickou podobnost s předložkou *na* (viz 4.2.4).

Při rozšířeném hledání předložkové valence se mezi 50 nejfrekventovanějšími kombinacemi substantiva a předložky s předložkou *pro* objevilo pouze jedno substantivum (*podmínka*). Další substantiva s frekvencí výskytu vyšší než 300 jsme opět rozdělili do několika sémantických skupin.

skupina cit
cit pro; pochopení pro

skupina důvod
důvod pro; kritérium pro; podmínka pro; pravidlo pro; nebezpečí pro
skupina podklad
podklad pro; předpoklad pro; základ pro signál pro
skupina agentura
agentura pro; centrum pro; institut pro; komise pro; ministerstvo pro; ministr pro; nadace pro; organizace pro; sdružení pro; rada pro; služba pro; společnost pro; středisko pro; úřad pro; ústav pro; výbor pro
skupina čas
čas pro; doba pro; lhůta pro; limit pro; rok pro; termín pro
skupina místo
místo pro; prostor pro; prostředí pro
skupina příležitost
příležitost pro; možnost pro; šance pro; cesta pro
skupina přínos
přínos pro; význam pro; cena pro
skupina nástroj
nástroj pro; program pro; prostředek pro; přístroj pro; řešení pro; zařízení pro
skupina partner
partner pro; partnerství pro
skupina informace
informace pro; materiál pro; zpráva pro; otázka pro

Předložka <proti + dat.> se ve zkoumaném vzorku objevila ve spojení s následujícími dvěma substantivy:

povstání (6,3 %) (*proti Britům*); vzpoura (11,7 %).

Specifická a poměrně jednoznačná sémantika předložky *proti* určuje, že je tato předložka valencí sémanticky specifických substantiv. I v našem vzorku se jedná o valenci těchto substantiv. Při rozšířeném hledání předložkové valence <substantivum + předložka> se mezi 50 nejfrekventovanějšími kombinacemi substantiva a předložky předložka *proti* neobjevila. Ze substantiv s frekvencí vyšší než 300 jsme vytvořili následující sémantické skupiny. Dvě výše uvedená substantiva *povstání* a *vzpouora* by patřila do skupiny „*boj*“.

skupina akce
akce proti; kampaň proti

skupina obrana
obrana proti; odpor proti; protest proti; námitka proti

skupina lék
lék proti; odolnost proti; ochrana proti; opatření proti

skupina boj
boj proti; utkání proti; útok proti; válka proti; zápas proti

skupina sankce
sankce proti; zásah proti

4.2.9 Předložka „s/se“

Ve zkoumaném vzorku se předložka <s/se + instrum.> objevila u následujících substantiv:

zájezd (s Čedokem / s polopenzí); mítink (2,4 %) (mítink s občany); vila (vila s bazénem); pec (3,9 %) (pece s otevřeným ohništěm); porost (listnaté porosty s převahou lípy a klenu); krb (krb s předhřívací komorou); dostih (dostihy s časem); šátek (šátek s modrými růžemi); kompresor (kompresor s vodním chlazením); hádka (8,3 %) (hádka se sousedkou); klávesnice (klávesnice s velmi vysokou citlivostí)

Pětiprocentní hranici výskytu předložka *s* překročila pouze ve spojení se substantivem *hádka*.

U této skupiny je zajímavé, že více než polovina substantiv je konkrétní povahy (*vila, pec, krb, šátek, kompresor, klávesnice*) a předložku *s* u těchto substantiv stěží lze považovat za jejich valenci. Podobnou sémantickou funkci jako u těchto substantiv předložka *s* plní např. i u substantiva *zájezd* ve spojení typu *zájezd s polopenzí*, kde vyjadřuje, co *zájezd* obsahuje. Ve spojení se substantivem *zájezd* má však předložka *s* i další sémantickou funkci – vyjadřuje, kým je *zájezd* pořádán (*zájezd s Čedokem*).

Při rozšířeném hledání předložkové valence <substantivum + předložka> se mezi 50 nejfrekventovanějšími kombinacemi substantiva a předložky *s* předložkou *s* objevilo 9 substantiv (*souvislost, spolupráce, srovnání, problém, jednání, setkání, obchod, soulad, rozhovor*), přičemž druhou nejfrekventovanější kombinací vůbec (po „zákon o“) bylo spojení *souvislost s*, které se vyskytlo celkem 9 159krát.

Následující tabulka ukazuje procentuální výskyt nejfrekventovanějších spojení s předložkou *s/se* z celkového výskytu substantiva v korpuse a příklad užití.

Tabulka 16. Vybraná substantiva s předložkou *s/se*

Substantivum	Výskyt s předložkou <i>s/se</i> v %	Příklad
souvislost	52 %	<i>co mělo nějakou souvislost se zločinem</i>
spolupráce	40 %	<i>kolaborace je spolupráce s nepřitelem</i>
srovnání	67 %	<i>moře je ve srovnání se* zemí ohromně plodné</i>
problém	7,4 %	<i>měla problémy s dechem</i>
jednání	32 %	<i>podal zprávu o jednání s premiérem</i>
setkání	31 %	<i>možnost setkání s mimozemským životem</i>
obchod	14 %	<i>proti takzvanému obchodu s bílým masem</i>
soulad	85 %	<i>v souladu s* podmínkami závěti</i>
rozhovor	22 %	<i>první rozhovor s Havlem</i>

* zde by se však mohlo jednat i o interpretaci tohoto spojení jako víceslovné předložky

Na základě dalších substantiv vyskytujících se v rozšířeném hledání s předložkou *s/se* jsme vytvořili následující sémantické skupiny.

skupina souvislost
spojitost s (47 %); vztah s (7,6 %); spojení s (25 %)

skupina srovnání
porovnání s

skupina soulad
shoda s; dohoda s; smlouva s; souhlas s

skupina spolupráce
jednání s; setkání s; rozhovor s; dialog s; schůzka s; beseda s; diskuse s

skupina kontakt
styk s; hranice s

skupina problém
potíž s; starost s; rozpor s; konfrontace s; zápas s; boj s; utkání s; souboj s; spor s; konflikt s; střet s; střetnutí s; duel s; válka s; prohra s; srážka s

skupina hra
proces s manipulace s soud s zkušenost s

skupina trh
společnost s; obchod s / obchodník s

4.2.10 Předložka „u“

Předložka <u + gen.> se ve zkoumaném vzorku okrajově objevila ve spojení s následujícími substantivy:

nález (*nálezy u těchto pacientů*); pojistné (*u České pojišťovny*), příznak (*příznaky u dalších dvou osob*); obliba (*Jak si vysvětlujete oblibu u mladých?*); deprese (*výskyt deprese u sledovaných osob*); klika (*u dveří*).

Valenční charakter má předložka *u* pouze u substantiva *oblíba*, v ostatních případech bychom ji interpretovali spíše jako součást ustálených kolokací. Při rozšířeném hledání (s frekvencí výskytu vyšší než 300) se předložka *u* typicky objevila jako součást místního určení (*místo u; bitva u; dům u; návštěva u*), a to i jako součást názvu obce (*Brodek u ...*).

4.2.11 Předložka „v/ve“

Předložka <v/ve + lok.> stojící za substantivem (nebo v jeho bezprostřední blízkosti) je ve většině svých výskytů součástí adverbialé (místa, času i způsobu) (viz dále kapitola 4.4). Ve zkoumaném vzorku se však vyskytla i ve spojeních, ve kterých má valenční charakter:

zodpovědnost (*ve věcech veřejných*); monopol (4 %) (*v distribuci tisku*); obliba (*Nezapře oblibu v přírodních materiálech*); recese (*v automobilovém průmyslu*); volnost (*v rozhodování*); pomocník (*v nouzi*).

Předložka *v* se objevila i jako součást ustálených kolokací substantiva *citát*: *citát v plném rozsahu*; *citáty v jejich přesném znění*.

Při rozšířeném hledání <substantivum + předložka> se předložka *v/ve* vyskytla mezi padesáti nejfrekventovanějšími kombinacemi 11krát, avšak u všech těchto výskytů se dá předpokládat, že tvoří součást adverbialé. Všechna tato substantiva patří do vyšších frekvenčních pásem (*rok* (frekv. 399 340), *místo* (84 617), *člověk* (184 099), *situace* (50 722), *změna* (40 140), *práce* (82 919), *úřad* (43 425), *svět* (81 504), *soud* (30 329), *den* (110 321), *škola* (48 599)), kde se dají očekávat určité specifické charakteristiky jejich chování.

S frekvencí výskytu vyšší než 2 500 se s předložkou *v* vyskytlo i dalších 17 substantiv: *život, dům, muž, účast, hodina, pobyt, divadlo, volba, žena, členství, válka, týden, univerzita, republika, vláda, dítě, turnaj*.⁷⁶

⁷⁶ Zde by byla jistě na místě podrobnější analýza situace, většina těchto substantiv spadá do vyššího frekvenčního pásma, což vždy představuje komplexnější situaci i z hlediska valence. Nicméně příklady jako *Osmnáct let poskytoval své ženě život v blahobytu*. (SYN2000) jsou jistě relevantní.

4.2.12 Předložka „z/ze“

Ve zkoumaném vzorku se předložka <z/ze + gen.> objevila u následujících substantiv:

funkcionář (*funkcionáře z ODS*); pojistné (*pojistné z čistého příjmu*); odstranění (*odstranění z funkce*); pasáž (7 %) (*pasáž z Homéra*); rozpaky (*Rozpaky z jednání Havla i Bendy*); výdělek (*výdělek z chmelové brigády*); rukavice (*rukavice z jelenice*); citát (29,6 %) (*citát z Komenského díla*); deprese (*z toho, že...*); žlutek (*z čerstvých vajec*).

U dvou z těchto substantiv se předložka z vyskytovala s vyšší frekvencí: *pasáž* a *citát*. Obě tato substantiva by patřila do sémantické skupiny **ukázka** (viz níže). Při rozšířeném hledání předložkové valence <substantivum+předložka> se mezi 50 nejfrekventovanějšími kombinacemi substantiva a předložky s předložkou z objevilo pouze jedno substantivum (*daň*), které se vyskytlo v této kombinaci celkem 4 395krát.

Na základě rozšířeného hledání jsme ze substantiv vyskytujících se s předložkou z s vyšší frekvencí než 300 vytvořili následující sémantické skupiny.

skupina strach
strach z; obava z; radost z; hrůza z pocit z; zkušenost z; dojem z; zážitek z podezření z nemoc z

skupina daň
daň z; příjem z; výnos z; zisk z; úrok z; ztráta z; výtěžek z

skupina většina
většina z; procento z; část z; polovina z

skupina návrat

návrat z
cesta z
útěk z
přechod z
propuštění z

skupina dopis

dopis z; pohled z

skupina ukázka

ukázka z; výpis z; výběr z; příklad z; úryvek z; scéna z; výbor z; citát z; výjimka z; záběr z

skupina přenos

přenos z; záznam z

skupina medaile

medaile z
výsledek z
zkouška z

skupina původ

původ z
šťáva z
dovoz z

4.2.13 Předložka „za“

Ve zkoumaném vzorku se předložka <za + akuz.> vyskytla s následujícími substantivy:

pojistné za (období/co/koho); výdělek (za *Přelet nad kukaččím hnízdem*); pocta (za *své úspěchy*); dlužník (*dlužníků za zdravotní pojištění*); zodpovědnost (26,6 %) (*zodpovědnost za to, že/co/jestli...*).

Při rozšířeném hledání s frekvencí vyšší než 2 500 se předložka *za* objevila se třemi substantivy: *cena*; *koruna*; *odpovědnost*.

Další substantiva s frekvencí vyšší než 300 jsme rozdělili do následujících sémantických skupin.

skupina hodnota

akcie za; bod za; cena za; dolar za; Kč za; kilometr za; koruna za; metr za; obchod za; peníz za; škoda za; tržba za; USD za; zboží za; život za
--

skupina boj

boj za; bojovník za; hnutí za; honba za; odpovědnost za; poslanec za; zodpovědnost za; cesta za; krok za
--

skupina dík

daň za; dík za; náhrada za; odměna za; platba za; pokuta za; poplatek za; trest za; účet za; tečka za; výměna za
--

4.2.14 Ostatní předložky

4.2.14.1 Kolem

Předložka <kolem + gen.> nejčastěji vyjadřuje přibližnost nebo umístění. Takto se objevila i ve zkoumaném vzorku např. se substantivem *šátek*⁷⁷ (*šátek kolem krku*), kde se vyskytla i v modifikacích typu: *šátek ovinutý/uvázaný kolem* Při rozšířeném hledání se v SYN200 s frekvencí vyšší než 300 takto vyskytuje např. se substantivem *cesta*, kde přímo souvisí se sémantikou tohoto slova (*cesta kolem světa*).

Ve zkoumaném vzorku se však předložka *kolem* objevila i ve spojení se substantivy *rozpaky* (*rozpaky kolem jmenování Brita George Robertsona*), *klika* (*konzervativní kliky kolem arcivévodě Franze Ferdinanda d'Este*)⁷⁸ a *hádky* (*hádkách kolem rozsvícených světél*). Podíváme-li se v rozšířeném hledání na výsledky i s nižší frekvencí výskytu, najdeme zde s předložkou *kolem* sémanticky podobnou skupinu substantiv, ke kterým by patřila i výše uvedená substantiva *rozpaky* a *klika*. Jsou to následující substantiva (ve frekvenčním pořadí, frekvence 275 – 89): *aféra*; *situace*; *skandál*; *problém*; *dění*; *spor*; *věc*; *nejasnost*; *diskuse*; *událost*; *rozdělení*.

4.2.14.2 Nad

Ve zkoumaném vzorku se předložka <nad + instrum.> vyskytla pouze se 2 substantivy: *rozpaky* (*nad drahým jídlem*); *zamyšlení nad* (26,6 %). Při

⁷⁷ 6x šátek přes (*oči, tvář, obličej...*)

⁷⁸ Se substantivem *klika* v tomto významu se objevila i předložka *uvnitř*: *soupeřící klikou uvnitř BIS*

rozšířeném hledání se ukázalo, že předložka *nad* je typicky součástí místních jmen (*Veselí nad...; Kolín nad...; Žďár nad...* atd.). S frekvencí výskytu vyšší než 300 se předložka *nad* vyskytovala s následujícími substantivy: *vítězství; kontrola; výhra; dohled; dozor, střecha; metr*.

U substantiva *střecha* se však jedná o frazém – ze 462 výskytů byla *střecha* 411krát součástí výrazu (nebo jeho variace) *střecha nad hlavou*. Substantivum *metr* ve spojení s předložkou *nad* pak vždy předchází číselné (nebo obecně množstevní) udání, tedy o kolik metrů (přesně nebo obecně) se jedná.

Podíváme-li se i na spojení s nižší frekvencí (alespoň však 100 výskytů), získáváme následující skupinu substantiv: *zamyšlení; znepokojení; lítost; rozsudek; politování; údiv; diskuse; převaha; radost; uspokojení; moc; otazník; soud; zklamání*.

4.2.14.3 Před

Ve zkoumaném vzorku se předložka <*před* + instrum.> vyskytla pouze se substantivem *zodpovědnost* (*před Bohem a dějinami*). Při rozšířeném hledání se předložka *před* vyskytuje se substantivy označujícími časový údaj, jak i vyplývá ze sémantiky samotné předložky (*den; minuta; hodina; rok; týden; měsíc; doba*). Vyskytuje se však i se substantivy u nichž má valenční charakter, podobně jako výše uvedené substantivum *zodpovědnost*: *ochrana; přednost; strach, obava a náskok*.

4.2.14.4 Při

Předložka <při + lok.> se ve zkoumaném vzorku ve spojení se substantivou objevila pouze okrajově a jako součást příslovečného určení okolností (*nález při dopingové kontrole; pohody při práci*). S mírně vyšší frekvencí (59 výskytů v SYN2000) se objevila se substantivem *pomocník* (*pomocník při vyplňování formulářů*) a *volnost* (31 výskytů) (*volnost při rozhodování*). Při rozšířeném hledání se objevila se substantivy *pomoc, postup, role, spolupráce, problém, maximum*.

4.2.14.5 Vůči

Ve zkoumaném vzorku se předložka <vůči + dat.> objevila ve spojení se substantivou *zodpovědnost* a *tolerance*. Při rozšířeném hledání se s frekvencí vyšší než 300 vyskytuje pouze se substantivem *závazek* (frekvence 456). Podíváme-li se na spojení s nižší frekvencí (alespoň však 90 výskytů), předložku *vůči* najdeme u následujících substantiv: *dluh; politika; postoj; sankce; odpor; odolnost; výhrada; odpovědnost; povinnost; pohledávka; postup; embargo; nenávisť; opozice; nedůvěra*.⁷⁹

⁷⁹ Tato substantiva spojuje určitý negativní sémantický nádech. Tento fakt potvrzuje i letmý pohled do korpusu mluveného jazyka (oral 2006), kde mezi bezprostřední kolokáty této předložky patří např. slova *nezodpovědnej, kritický, blbý, laksnosti, námitky, nechápavá*.

4.2.15 Alternující předložky

Některá substantiva mají vícepředložkovou valenci, kde různé předložky vyjadřují podobnou (v některých případech stejnou) sémantickou funkci. Tuto míru podobnosti by odhalila podrobnější kolokační analýza těchto případů. Podrobná analýza však přesahuje rámec této práce, uvedeme zde však několik příkladů.

PEC

Substantivum *pec* (celková frekvence v SYN2000 je 2421 výskytů⁸⁰) vyjadřuje určení účelu pomocí dvou předložek – předložky *na* a *pro*. Výskyt těchto předložek se substantivem *pec* má i podobnou frekvenci. S předložkou *na* se substantivum *pec* vyskytuje v SYN2000 38krát⁸¹ (*pec na chleba*) a s předložkou *pro* 59krát (*pec pro tavení oceli/výrobu cementu*). Kolokáty následující předložku *pro* jsou však mnohem techničtější (viz příklady výše) a v podstatě se s kolokáty předložky *na* nekryjí.

OBLIBA

Valencí substantiva *obliba* (celková frekvence v SYN2000 je 2308) jsou předložky *u* (*Získal si oblibu u sportovních rybářů,..*) a *mezi* (*Parlament je frustrován stabilně nízkou oblíbou mezi naším obyvatelstvem...*). Předložka *u* je téměř dvakrát frekventovanější (74 výskytů v SYN2000 vs. 41 výskytů předložky *mezi*) a v této konstrukci jsou výraznými slovesnými kolokáty substantiva *obliba* slovesa *těšit se* a *získat*. Slovesný kolokační profil v konstrukci s předložkou *mezi* není tak výrazný, vyskytuje se zde také sloveso *získat*, ale i např. sloveso *být*.

⁸⁰ Včetně *Pece pod Sněžkou*.

⁸¹ Předložka *na* se substantivem *pec* specifikuje i typ paliva – 7 výskytů (např. *pece na generátorový plyn*).

ROZPAKY

Substantivum *rozpaky* (celková frekvence 2124) má též pro vyjadřování stejného (podobného) významu valenci v podobě dvou předložek: předložky *nad* a *kolem*. Předložka *nad* (*Jisté rozpaky **nad** výsledky vídeňské schůzky o českých a rakouských projektech v rámci programu...*) je výrazně frekventovanější, v SYN2000 se vyskytla 82krát zatímco předložka *kolem* (*Rozpaky **kolem** hledání ministra vnitra mohl sotva rozptýlit mluvčí ODS...*) pouze 10krát.

TOLERANCE

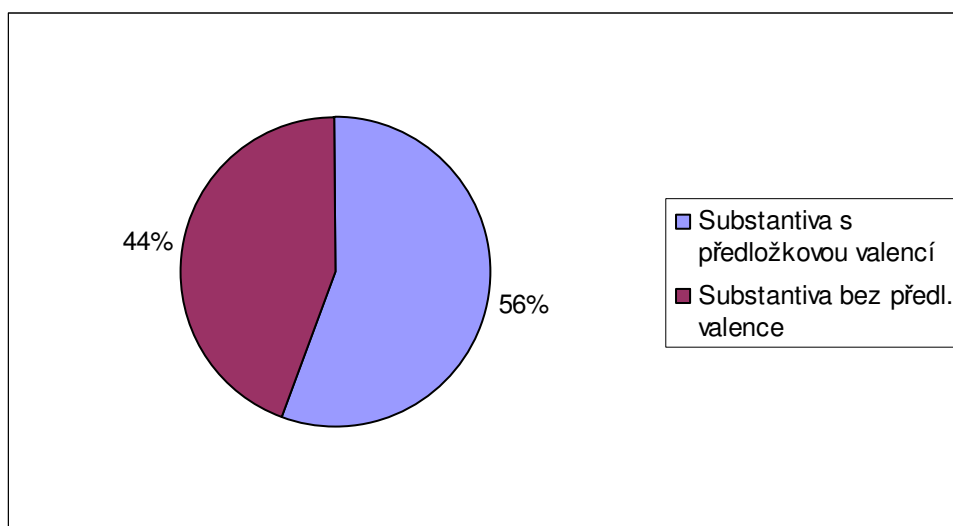
Obdobná situace je u substantiva *tolerance* (frekvence 1748), jehož alternativními předložkovými valencemi jsou předložky *k* a *vůči*. Ve srovnání s předložkou *vůči* (46 výskytů) je předložka *k* téměř dvakrát frekventovanější (88 výskytů), avšak jinak se jejich kolokáty vyznačují značnou podobností:

... *obdivuhodná tolerance k cizím zvykům...*
... *nedostatek tolerance vůči ostatním kulturám...*

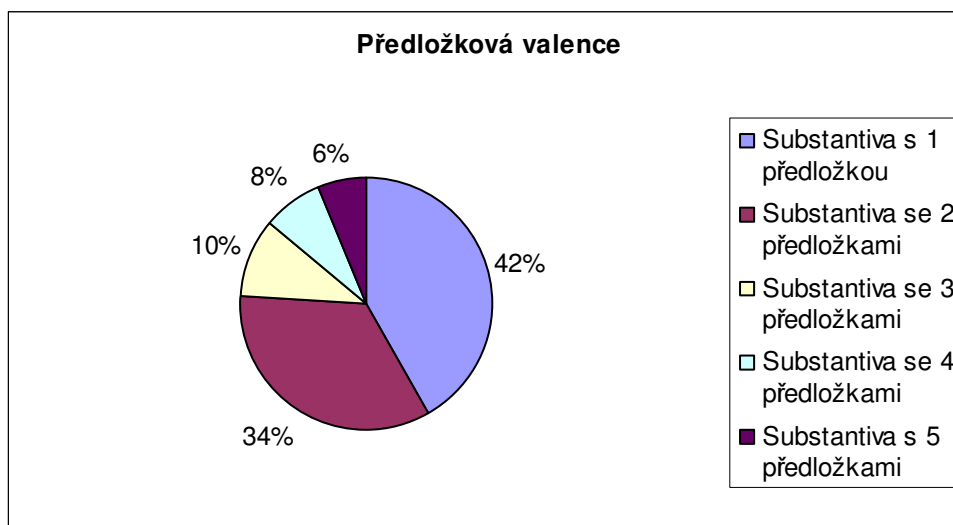
Alternativní, resp. synonymně konkurenční předložková valence by si zasloužila větší pozornost, nedostatečný rozsah dat této práce bohužel neumožňuje se této problematice věnovat podrobněji. Chceme se tedy pouze omezit na konstatování, že alternativní předložková valence (pro vyjadřování stejného nebo podobného významu) existuje. Kolokační profily s jednotlivými předložkami mohou být podobné (viz např. *tolerance* nebo *rozpaky*), mohou se výrazněji lišit např. v levé části kolokačního profilu (slovesa u substantiva *oblíba*) nebo mohou být kolokační profily značně odlišné (viz konkrétně *pec*).

4.2.16 Závěr

Pokusíme se předložkovou valenci pomocí následujících grafů shrnout obecněji. Na základě našeho vzorku substantiv můžeme říci, že předložkovou valenci má přibližně polovina substantiv. Ve většině případů předložkovou valenci těchto substantiv představuje pouze 1 předložka (42 %). Další velkou skupinu tvoří substantiva s dvoupředložkovou valencí (34 %). Valenci zbývajících substantiv představují 3 až 5 předložek.



Graf 5. Předložková valence z frekvenčního pohledu



Graf 6. Přehled předložkové valence

Škála předložek, které se substantivní valence účastní, je veliká a je třeba předpokládat, že přehled v této práci není vyčerpávající a naznačuje pouze tendence. Z hlediska frekvenčního se lze na předložkovou valenci dívat ze dvou hledisek a ptát se:

- (1) jak často se daná předložka se substantivy vyskytuje a
- (2) jaká je její míra valenčnosti. Příkladem může být předložka *pro*. Předložka *pro* je v námi zkoumaném vzorku nejčastěji se vyskytující předložkou (vyskytla se s 18 substantivy), avšak s žádným z těchto substantiv se nevyskytuje s velkou frekvencí (míra valenčnosti). Naopak, např. ze substantiv, která se objevila s předložkou *za* (ve zkoumaném vzorku celkem 5), se substantivum *zodpovědnost* vyskytuje s touto předložkou v téměř třetině svých výskytů. Míra valenčnosti předložek u substantiv je však obecně, např. ve srovnání s genitivem, nižší.

4.3 Ostatní pády

4.3.1 Instrumentál

Instrumentál se jako valenční exponent v našem materiálu objevil zcela okrajově. Objevil se se substantivem *zájezd*, kde následný instrumentál vyjadřoval typ dopravního prostředku (*zájezd autobusem*). Dále se objevil se substantivem *odstranění*, kde instrumentál vyjadřoval způsob provedení (*odstranění obyčejným chirurgickým zákrokem*). Instrumentál se vyskytoval i se substantivem *respektování* (*a její respektování ostatními*) a několikrát se substantivem *vyhlášení*, kde vyjadřoval dvě sémantické role: (1) *vyhlášení Korejského poloostrova pásmem bez jaderných zbraní* a (2) *vyhlášení radiačního nebezpečí sirénami*.

Ani specializovaný dotaz do korpusu na substantivum následované substantivem nepřinesl výrazné výsledky. Konkordančních řádků jsme sice získali 164 746, po zmenšení výsledku na náhodných 500 však již letný pohled ukazuje, že je zde příliš mnoho nerelevantních výsledků. Podrobnější manuální analýza vzorku tak v podstatě potvrdila výsledky analýzy původní (viz výše). Na základě analýzy relevantních výsledků z tohoto vzorku jsme získali tři substantiva s instrumentální valencí: *bodnutí* (z celkové frekvence v SYN2000 147 se 37krát toto substantivum objevuje s instrumentální valencí (*bodnutí hmyzem/nožem...*)); *inspirace* (celková frekvence 2127, s instrumentálem 62krát (*inspirace divadlem*)); *napadení* (celková frekvence 1274, s instrumentálem výskyt 223krát (*napadení mšicemi*)).

Mezi relevantními výsledky byla i další tři substantiva. *Cesta* a *jízda* spadají do stejné skupiny jako substantivum *zájezd* (viz výše), kde instrumentál obecně

vyjadřuje dopravní prostředek. Třetím substantivem bylo substantivum *rok*, kde následný instrumentál vyjadřoval povolání apod. (*rok poslancem*, typicky v konstrukci „X let kým/čím“).

4.3.2 Dativ

Ve zkoumaném vzorku substantiv mělo nejvýraznější dativní valenci substantivum *pocta* (31,3 %; *pocta Vladimíru Holanovi*). Dále mělo výraznější dativní valenci substantivum *pomník*, kde však dativ (5 % výskytů; *pomník bratrům Čapkům*) alternoval s častějším genitivem (21 %; *pomník Jana Husa*). V případě tohoto substantiva se zdá, že se jedná o valenční synonymii, která závisí na stylistických preferencích mluvčího a kolokační síle, obecně však i na různém ustálení úzu, např. výraz *pomník obětem*, tedy s dativní valencí se objevuje ve vzorku 5krát, zatímco *pomník obětí* se objevuje pouze jednou; *pomník padlých* se ve vzorku vyskytl 6krát a *pomník padlým* 11krát.

Dativ jako valenční exponent se objevil i v rámci širších kolokací, kde je třeba specifikovat i sloveso, např. ve 2 % výskytů se substantivem *důležitost* (kolokace *ne/přikládat důležitost* plus dativ) nebo se substantivem *volnost* (kolokace *ponechat/dát volnost...*).

Obdobně jako instrumentál není dativ příliš frekventovaným valenčním exponentem substantiv.

4.4 Adverbium/adverbiále

Status adverbium/adverbiále⁸² jako valenčního exponentu substantiv je problematický (viz 2.1), nicméně tuto problematiku nelze ignorovat nebo opomenout. Vzhledem k lexikologickému a frekvenčnímu nazírání na valenci v této práci je zde adverbium, příp. adverbiále (a to se v datech vyskytuje především), vyskytující se s určitým substantivem s výraznou frekvencí považované za valenci nebo mající valenční charakter. V této kapitole se v podstatě omezíme na výčet substantiv, do jejichž valence adverbium/adverbiále spadá (a to pouze na substantiva, u kterých frekvenční hranice výskytu adverbium/adverbiále překročila 10 % výskytu). Problematika si zasluhuje podrobnější zkoumání, takováto analýza však bohužel přesahuje rámec této práce. Povšimněme si, že naprostá většina adverbiálií ze zkoumaných dat specifikuje místo.

ODJEZD (29,6 %; *odjezd do zahraničí*)

MÍTINK (26,2 %; *mítink v Monte Carlu*)

SOUSTŘEDĚNÍ (sportovní) (24 %; *soustředění v Harrachově*)

AMBASÁDA (23 %; *Americká ambasáda na Haiti nejčastěji však s předložkou v*)

NÁLEZ (22 %; *nález na pobřeží Dánska; nálezy z období Velkomoravské říše; nález na plicích*⁸³)

ZÁJEZD (19,2 %; *zájezd do Paříže*)

POVSTÁNÍ (15 %; *povstání ve varšavském ghettu, povstání v prosinci 1989*)

VILA (16,6 %; *vila na Ořechovce*)

CHATA (13,2 %; *chata na horách*)

⁸² Tato práce neklasifikuje valenci z hlediska syntaktického, ovšem v tomto případě je třeba připustit i kategorii širší, než je lexikální adverbium, a vypůjčit si syntaktickou kategorii „adverbiále“ pro víceslovné výrazy.

⁸³ s předložkou *na* i typ: *nález na kokain, pozitivní nález na anabolický steroid*

SONDA (10,25 %; *sonda do žaludku, sonda k Marsu*)

KORUPCE (10 %; *korupce ve státní správě*)

STANOVIŠTĚ (10 %; *stanoviště na dálnici D8*)

Při bližším pohledu na tuto skupinu substantiv si můžeme všimnout, že jsou zde poněkud překvapivě především konkrétna (*ambasáda, vila, chata, sonda, stanoviště*) nebo substantiva mající v tomto významu konkrétní význam (*mítink, soustředění*). Dále jsou tu dvě substantiva ze sémantické oblasti cestování (*odjezd, zájezd*), kde je určení místa sémanticky velmi relevantní. Další dvě substantiva (*nález a povstání*) jsou příslovečným určením historicky specifikována a upřesněna. Posledním substantivem je *korupce*, kde vzhledem k povaze tohoto jevu je další specifikace také očekávána.

4.5 Koordinovaná spojení

Koordinovaným výrazem zde máme na mysli spojení dvou substantiv spojkou *a*⁸⁴, které má výraznější frekvenci výskytu a určitou sémantickou charakteristiku (viz také Čermáková (2006)). Můžeme zde pozorovat dva typy koordinovaných výrazů: ustálené kombinace (frazémy a ustálené kombinace dvou slov spojených předložkou *a*) a dále kombinace určitých sémantických skupin slov.

Ustálených kombinací tohoto typu existuje v češtině veliké množství. Příkladem může být např. frazém *ztráty a nálezy*, který tvoří téměř 3 % z výskytů hledaného lemmatu *nález*. Najdeme tu však i méně vzletná ustálená spojení typu *bankovníctví a pojišťovnictví* (tvoří 2,7 % z výskytů lemmatu *bankovníctví*). Tento typ spojení je však obtížné v korpuse automaticky vyhledávat, spojka *a* má v ČNK samozřejmě závratně vysokou frekvenci.

Slovní vzorec se spojkou *a* se častěji a výrazněji vyskytuje s abstrakty (konkrétně se s tímto vzorcem mezi analyzovanými substantivy objevilo pouze jednou – a to substantivum *sval*, které se frekventovaně vyskytuje ve spojeních *sval a kloub*, *klouby a svaly*, *svaly a kosti*, *kosti a svaly*, *svaly a vazy*, *svaly a šlachy*). Mezi abstrakty bylo substantiv v daném vzorku vyskytujících se se spojkou *a* pět: *civilizace*, *tolerance*, *pohoda*, *korupce*, *respektování*.

Valenční struktura S a S je velice obecná, avšak při pohledu na kolokační profily těchto substantiv v této konstrukci vidíme, že se zde rýsují jasné sémantické skupiny. U substantiva *civilizace* první skupinu tvoří abstrakta psycholingvisticky asociovaná s tímto substantivem (*kultura*, *demokracie*), druhou skupinu tvoří atributy *civilizace* (*křesťanství*, *materialismus*), třetí skupinu tvoří negativní

⁸⁴ Jedná se o tzv. binomiály, které jsou velmi obecné a běžné (*pepř a sůl* apod.).

atributy (*konzumní způsob života*) a čtvrtou skupinu slova, která jsou v sémantickém kontrastu (*příroda*). Substantivum *tolerance* se se spojkou *a* pojí se dvěma širokými sémantickými skupinami slov – pozitivními (*demokracie, dialog*) a negativními (*intolerance*). Substantivum *pohoda* se pojí se širokou škálou slov s pozitivními psychologickými konotacemi (*bezpečí, optimismus*). A konečně substantivum *korupce* je v tomto smyslu podobné, jeho negativní konotace jsou dalšími substantivy přes spojku *a* umocněny (*kriminality, podvody*). Substantivum *respektování* se častěji vyskytuje na druhé pozici (tj. za spojkou *a*), avšak i v tomto případě jde o sémanticky podobnou skupinu slov (*demokracie a respektování*) (více příkladů a podrobnější analýza viz Čermáková (2006:109–110)).

4.6 Nominativní valence

Nominativní valenci zde definujeme jako vztah substantiva a dalšího substantiva ve stejném pádě a čísle, přičemž obě se vztahují ke stejnému (konkrétnímu) denotátu. Druhé ze substantiv může být rozvíjeno atributivním adjektivem a poměrně často to bývá proprium. Kromě běžných případů jako např. <*ulice* + její název> se ve zkoumaném materiálu nominativní valence výrazně projevila u substantiva *vdova* (v 10,7 % výskytů; *stará vdova Abigail*), *disident* (v 15 % ze svých výskytů následováno propriem (*Sionistický disident Sacharov...*)) ale i u konkrétní *vila* (*Vila Tugendhat v Brně*), *ateliér* (*ateliéru A+G Flora*), *sonda* (*sonda Galileo*). Nejvýrazněji se objevila, a to ve 41,5 % výskytů, s hodnotícím substantivem *chudák* (*chudák Arnošt* ale i *chudák holka*) (viz také Čermáková (2006: 111)).

4.7 Další konstrukce

4.7.1 Infinitiv

Ve zkoumaném vzorku substantiv se infinitiv jako valenční doplnění objevuje zcela okrajově; a to pouze v několika málo příkladech ve spojení se substantivy *zodpovědnost* (*cítil obrovskou zodpovědnost nezklamat diváky*) a *volnost* (*Vy ani já nemáme volnost jednat podle vlastního přání.*)

Při rozšířeném hledání (substantivum následované slovesem v infinitivu) získáváme 335 061 konkordančních řádků, které bylo potřeba manuálně projít a najít zde případy infinitivní valence. Získali jsme takto 95 substantiv, u kterých se výrazněji⁸⁵ projevuje infinitivní valence. Abecední přehled těchto substantiv nabízí příloha D.

Některá substantiva zde tvoří výrazné sémantické okruhy, jiná je obtížnější zařadit. Substantiva, která měla tendenci se sémanticky seskupovat, jsme rozdělili do 13 skupin (53 z výše uvedených 95 substantiv):

(1) *ambice, ctižádost, motivace, touha, úsilí*

(2) *odhodlání, odvaha*

(3) *sklon, tendence, trend*

(4) *snaha (ve snaze), úsilí, vůle*

(5) *zásada, zvyk*

⁸⁵ Vzhledem k množství konkordančních řádků a z toho plynoucí náročnosti – a tedy i nepřesnosti manuálního zpracování, si zde naprosto nečiníme nárok na vyčerpávající přehled infinitivní valence substantiv v češtině.

(6) *čest, mandát, oprávnění, pocta, povolání⁸⁶, pravomoc, privilegium*

(7) *šance, příležitost, naděje*

(8) *chuť(/nechuť), ne/ochota*

(9) *možnost(/nemožnost), nezbytnost, nutnost, potřeba*

(10) *myšlenka, nápad, předsevzetí, záměr*

(11) *nabídka, návrh, povel, povolení, požadavek, příkaz, rada, rozkaz*

(12) *slib, souhlas, svolení, zákaz*

(13) *nařízení, pověření, povinnost*

U některých substantiv se infinitivní valence zdá být omezena pouze na některé tvary tohoto substantiva nebo substantiva bývají součástí širších slovních výrazů (frazémů⁸⁷), jejichž součástí bývá často např. sloveso *mít* (*nemít*) – např. substantiva *tendence, strach, chuť* (zde i sloveso *dostat*), *nechuť, důvod, plán, úkol* (zde i slovesa *dát, dostat za*), *trpělivost, možnost, smysl, nárok*.

U některých substantiv je patrná tendence k silným kolokacím a koligacím. Typickým příkladem infinitivní valence substantiva *hrůza* jsou dva následující příklady:

(a) *Ty bys dokázal celý svět obrátit vzhůru nohama – hrůza pomyslet.*

(b) *Je to určitě hrůza přijít o nohu.*

Situace však není tak přehledná u následujícího příkladu substantiva *cíl*. Substantivum *cíl* se vyskytuje ve spojení s předložkou *za* (*za cíl*) 511krát. V některých příkladech lze toto spojení vynechat, aniž by věta došla po stránce

⁸⁶ Zde infinitivní valence rozlišuje dva lexémy, *povolání* v tomto případě tedy ne ve smyslu „profese“.

⁸⁷ např. BÝT/NEBÝT *na škodu* + inf.

gramatické úhony (viz (a) níže), v některých to však není dost dobře možné (viz (b) níže).

(a) *Drobné úpravy v seznamech literatury měly **za cíl** dosáhnout určité proporcionality a ponechat převážně odkazy na základní prameny a souhrnné články.*

(b) *Ta si klade **za cíl** eliminovat carské metody řízení Vladimira Petrova, který je předsedou Ruské hokejové federace.*

Substantivum *cíl* se velice frekventovaně vyskytuje i s předložkou *s* (*s cílem* 1578krát). Ve většině případů nelze toto spojení z věty vynechat ani z gramatického hlediska, viz příklad (c) níže.

(c) *Ta je prodána **s cílem** dosáhnout uhrazení půjčky .*

Jak naznačují výše uvedená data, infinitivní valence substantiv má svá specifika, která by vyžadovala podrobnější, rámec této práce přesahující, korpusovou analýzu. Pokusili jsme tedy výše alespoň tuto problematiku naznačit.

4.7.2 Vedlejší věta

Vedlejší věta jako valenční exponent má, podobně jako adverbialé, volnější status. V této kapitole tedy pouze naznačíme výsledky získané na základě analýzy naší skupiny substantiv⁸⁸. S vyšší mírou se vedlejší věta uvozená spojkou *že* (18,3 %) vyskytovala pouze se substantivem *tušení* (*neměla jsem tušení, že takové chování je vůbec možné*). Substantivum *tušení* se frekventovaně vyskytuje i s dalšími typy vedlejších vět (spojky *jak, co* atd.). Frekvenční hranice 2 % (a vyšší) výskyt vedlejších vět v pozici valenčního atributu dosáhl u těchto substantiv: ***zamyšlení*** (*zda* – 2,3 % (*je to důvod k zamyšlení, zda by se problém nedal řešit úsporněji*)); ***volnost*** (*aby* – 2 % (*je třeba ponechat volnost, aby každá rodina „se zařídila podle svého“*)); ***dodatek*** (*že* – 4,3 % (*a to ještě s dodatkem, že malé selhání bylo zapláceno neskonale poctivou a rozsáhlou prací*)) a ***maličkost*** (*jako* – 3,3 % (*Každá maličkost, jako květiny, úsměv, křivý pohled,...*)).

Specializovaný dotaz do korpusu a další podrobná analýza by přinesly ucelenější obrázek statusu vedlejší věty jako valenčního exponentu substantiva. Již zběžný pohled na data však ukazuje, že by taková analýza byla velmi rozsáhlá. Dotaz do SYN2000 na substantivum následované podřadicí spojkou (která je oddělena čárkou) nám dává 753 904 konkordančních řádků (SYN2005 jich dává dokonce ještě více). Redukovali jsme vzorek na náhodných 1000 řádků a z výsledků analýzy tohoto vzorku vyplývá, že nejčastější spojkou uvozující vedlejší větu, která má valenční charakter, je spojka *že*. Ve vzorku se vyskytovalo 21 substantiv⁸⁹ následovaných vedlejší větou s valenčním charakterem. 19krát se jednalo o vedlejší větu uvozenou spojkou *že*, 1krát o vedlejší větu uvozenou spojkou *zda* a 1krát o vedlejší větu se spojkou *než*. Jednalo se o následující substantiva:

⁸⁸ Z analýzy byly vynechány vztažné věty se zájmenem *který, -á, -é* a pl. formy.

⁸⁹ Jsou zde započítána substantiva, která se ve vzorku takto vyskytla alespoň dvakrát.

div, že; doba, než; dojem, že; fakt, že; názor, že; obava, že; otázka, zda; pocit, že; podezření, že; pravda, že; předpoklad, že; představa, že; přesvědčení, že; případ, že; skutečnost, že; slib, že; škoda, že; štěstí, že; tvrzení, že; ujištění, že; úvaha, že.

I u těchto výše uvedených substantiv je patrné, že zde existují skupiny s určitou sémantickou podobností (*fakt, názor, otázka, pravda, předpoklad, představa, přesvědčení, skutečnost, tvrzení, ujištění, úvaha* nebo skupina *dojem, obava, pocit, podezření*).

4.7.3 Další výrazy

Zvláštním případem, který by zasloužil více pozornosti, je skupina substantiv (sémanticky vyhraněná a možná i uzavřená) následovaná ukazovacím zájmenem a spojkou *že*, typicky: *z toho, že...* Tato skupina substantiv v našem materiálu zahrnuje substantiva, jejichž hlavním sémantickým rysem jsou emoce: *deprese, šprýmy, obavy* (i v sg.), *strach, šok, úzkost, hrůza, radost, štěstí, rozpaky*. Např.

(1) *panický strach z toho, že člověk v životě vlastně nic neudělal*

(2) *paranoickou hrůzu z toho, že jsou její soukromé telefonické rozhovory odposlouchávány*

Tento typ konstrukce má větší škálu možných realizací, příkladem může být substantivum *zodpovědnost* následované svou předložkovou valencí: *zodpovědnost za to, že/co/jestli....*

4.8 Vícemístná valence

Substantivní valence může být i vícemístná (tedy substantivum současně (nebo potenciálně) realizuje více valenčních exponentů. U dvoumístné valence převažuje na první pozici genitiv a za ním pak obvykle stojí již valence vyjádřená explicitněji (předložkou apod.) (Čermák & Holub 1988:17). O některých případech jsme se již zmínili výše, především v souvislosti s genitivní valencí (odd. 4.1), patří sem i problematika několikanásobné genitivní valence (odd. 4.1.4). Interpretace vícemístné valence však není vždy jednoduchá, valence na druhé pozici může být mnohdy interpretována jako valence druhého substantiva, tedy substantiva na pozici prvního valenčního exponentu.

Vícemístná valence se nejvýrazněji projevila u substantiv s převažující genitivní valencí. Největší škála vícemístné valence se projevila u substantiva *vyhlášení*, připomeňme si několik příkladů: *vyhlášení* + gen. + instrum. (*Korejského poloostrova pásmem bez jaderných zbraní*), gen. + adv. času / místa (*vyhlášení stávkové pohotovosti od příštího týdne; vyhlášení příměří v Ulsteru*), gen. + dat. (*vyhlášení války Srbsku*)⁹⁰, gen. + předložka *o*⁹¹ (*vyhlášení příměří a o výměně zajatců; vyhlášení vládních představitelů zemí Středoevropské dohody o zóně volného obchodu*).

U substantiva *odstranění* byl genitiv v téměř 3,5 % všech výskytů následován předložkou *z* (*odstranění Senátu z právního řádu*), genitivní valenční exponent u substantiva *stanovení* bývá následován např. předložkami (*pro, z, na, za, při, u*), dalším genitivem nebo instrumentálem. V případě substantiva *poskytování* je v asi

⁹⁰ Omezeno na slova *válka* a *ultimátum*.

⁹¹ Předložka *o* se zdá být i samostatným valenčním exponentem substantiva *vyhlášení*, avšak velice řídkým (*vyhlášení o navyšování základního jmění*). V mnoha dalších výskytech lze výskyt této předložky interpretovat jako valenci slova v genitivu (např. *anketa, zákon, dohoda* apod. (*vyhlášení válečného zákona o cenzuře*)).

11 % všech výskytů genitivní výraz dále následován výrazem dativním (*poskytování dotací soukromým školám*) a paralelně zde i najdeme vyjádření pomocí předložky *pro* (+ akuz.) (*poskytování daňových úlev pro malé a střední podnikatele*). V případě substantiv s méně výraznou genitivní valencí jsme také narazili na několik příkladů, např. substantivum *soustředění* realizuje někdy současně obě své valence, tedy genitiv i předložku *na* (*soustředění jejich pozornosti na strategii*).

Některé z předložek realizujících substantivní valenci ve spojení s daným substantivem sémanticky inherentně v některých případech vyžadují další specifikaci, příkladem může být předložka *od*, která bývá valenčním exponentem u substantiv vyjadřujících nějaké rozmezí (*rozmezí od; platnost od; cena od; věk od*⁹²) a substantiv spadajících do sémantických skupin ČAS a VZDÁLENOST – v těchto případech bývá předložka *od* následována předložkou *do*.

V případě substantiva *přechod* bývá předložka *od* následována předložkami *k*, *na* a *do* (*přechod od věcné literatury k beletrii; přechodu od dětství do dospělosti; přechod od vytápění uhlím na plynové a elektrické systémy*).

Dvou či vícemístná valence tedy bývá realizována (1) jako genitiv následovaný dalším genitivem, (2) genitiv následovaný předložkovým spojením, (3) genitiv následovaný dalším pádem, (4) genitiv následovaný adverbíále nebo eventuálně jako (5) předložkové spojení následované dalším předložkovým spojením.

⁹² V tomto smyslu do této skupiny bude patřit i spojení *dítě od*, které bývá součástí širších spojení typu *každé dítě od 6 do 15 let* (i když se zde vlastně jedná o elipsu „*dítě ve věku od*“).

4.9 Substantiva s vysokou frekvencí výskytu

Závěry této práce vycházejí ze vzorku náhodně vybraných substantiv ze středního frekvenčního pásma (frekvence 1005 až 3339). Substantiva z vyšších frekvenčních pásem zde nebyla podrobněji analyzována ze dvou důvodů. První důvod je čistě praktický: u substantiv z vysokých frekvenčních pásem bychom museli pracovat s podstatně většími vzorky a analýza několika málo substantiv by tak svým objemem přesáhla praktické možnosti této práce. Druhým důvodem je hypotéza, že se substantiva s vysokou frekvencí chovají jinak. Často se jedná o velmi vysoká a obecná hyperonyma, která tudíž svou valenci specifikují až v kontextu. U takovýchto substantiv se také dají očekávat výrazné a ustálené kolokace, jejichž profil pak může někdy tvořit jejich valenci. Valence zde může využívat i jiných prostředků, než bylo popsáno výše. Substantiva s vysokou frekvencí tvoří jádro slovní zásoby a dá se tak očekávat, že se budou vyskytovat hojně ve všech žánrech a typech textů a v tomto smyslu jsou na ně kladeny i větší lexikálně-gramatické požadavky a z toho plynoucí pravděpodobně velmi roztržitý valenční profil. V této kapitole se pro ilustraci podíváme na několik málo příkladů.

Do seznamu lemmat ze SYN2000 (viz odd. 3.2) jsme se podívali na prvních 100 nejfrekventovanějších lexikálních slov (lemmat) (tedy ne spojky, předložky a číslovky) a z nich jsme vybrali 26 substantiv. Frekvenci udanou v našem seznamu lemmat jsme pak porovnali s frekvencí ve *Frekvenčním slovníku češtiny* (2004) (dále FSČ) a zároveň jsme se podívali na procentuální rozložení těchto substantiv v jednotlivých žánrech. Údaje přehledně shrnuje následující tabulka.

Tabulka 17. Nejfrekventovanější česká substantiva.

Substantivum	Frekvence v seznamu lemmat (SYN2000)	Rank FRQ v FSC	Frekvence uvedená v FSC	% zastoupení v beletrii	% zastoupení v odborném textu	% zastoupení v publicistice
<i>rok</i>	406 834	17	399 340	18%	32%	50%
<i>člověk</i>	183 509	42	184 099	41%	28%	30%
<i>strana</i>	130 835	63	128 158	18%	34%	38%
<i>doba</i>	111 053	72	110 150	24%	41%	45%
<i>den</i>	110 289	71	120 321	39%	26%	35%
<i>země</i>	103 622	75	102 186	26%	30%	44%
<i>místo</i>	103 493	85	84 617	28%	35%	38%
<i>Praha</i>	99 695	-	-	-	-	-
<i>společnost</i>	90 657	82	89 958	11 %*	40%	49%
<i>svět</i>	85 503	95	81 504	35%	33%	32%
<i>firma</i>	84 285	91	82 844	4 %*	38%	58%
<i>město</i>	82 728	97	80 493	28%	20%	52%
<i>vláda</i>	81 887	94	81 621	9 %*	15 %*	76 %**
<i>cena</i>	81 879	100	78 795	10 %*	30%	60%
<i>práce</i>	80 043	90	82 919	24%	41%	36%
<i>koruna</i>	80 002	98	79 709	7 %*	8 %*	85 %**
<i>život</i>	78 600	104	77 144	42%	36%	22%
<i>případ</i>	78 228	103	77 578	15 %*	44%	41%
<i>hodina</i>	77 122	149	58 063	39%	21%	41%
<i>dítě</i>	71 829	116	70 354	32%	32%	36%
<i>zákon</i>	68 750	126	67 467	12 %*	34%	54%
<i>část</i>	64 495	132	63 666	12 %*	52%	36%
<i>čas</i>	62 174	144	59 700	41%	30%	30%
<i>problém</i>	61 952	139	61 784	12 %*	42%	46%
<i>žena</i>	59 524	151	56 628	48%	26%	26%
<i>muž</i>	59 362	159	54 508	52%	17%	31%

*15 % a méně

**75 % a více

Jak je z tabulky patrné, frekvenční údaje v seznamu lemmat a FSC se výrazně neliší, všechna substantiva mají frekvenci vyšší než 54 000 výskytů, ve většině případů podstatně vyšší. Všechna tato substantiva také patří mezi 160 nejfrekventovanějších slov v češtině. Podíváme-li se na jejich procentuální žánrové rozložení, vidíme, že většina z nich je poměrně rovnoměrně rozložena mezi třemi srovnávanými žánry (beletrie, odborný text a publicistika), což odpovídá faktu, že patří do centra slovní zásoby. Je zde několik výjimek, které nám opět připomínají, že při analýze korpusových dat je vždy třeba mít na mysli

složení korpusu jako takové. Několik těchto substantiv má poměrně nízké zastoupení v beletrii (*společnost, firma, vláda, cena, koruna, případ, zákon, část, problém*) a odborném textu (*vláda, koruna*) a naopak vysoké v publicistice (*vláda, koruna*). Tato substantiva jsou v tabulce označena „*“.

U většiny výše uvedených substantiv se nabízí jejich rozdělení do 4 základních sémantických skupin: (1) ČAS (*rok, doba, den, hodina, čas*), (2) ČLOVĚK (*člověk, dítě, žena, muž*), (3) SVĚT (*země, místo, svět, město*), (4) PRÁCE A ŽIVOT (*práce, život*).

Abychom si udělali alespoň velmi obecnou představu o jejich chování z valenčního hlediska, podrobili jsme substantiva z prvních dvou skupin (ČAS a ČLOVĚK) analýze na základě dat ze SYN2000. Pokud není uvedeno jinak, pracovali jsme s náhodným vzorky 500 konkordančních řádků.

4.9.1 Skupina ČAS

Do sémantické skupiny ČAS patří 5 substantiv: *rok, doba, den, hodina, čas*. Substantivum **rok** (počet výskytů v SYN2000: 406 834) má výrazný kolokační profil nalevo (kolokace v intervalu tří slov nalevo od KWIC), najdeme zde typicky slova *příští, loňský, letošní, poslední, počátek, průběh, konec* apod. Napravo (kolokace v intervalu tří slov napravo od KWIC) je typická specifikace určitým letopočtem. Valenční profil tohoto substantiva není nikterak výrazný.

U substantiva **doba** (počet výskytů v SYN2000: 111 053) se nalevo (kolokace v intervalu tří slov nalevo) s významnou statistickou mírou vyskytují např. adjektiva *současná, delší, dohledná, krátká, pracovní, dnešní* apod. a mezi statisticky významnými kolokáty (MI-score) se napravo objevilo několik substantiv v

genitivu: *lovu, trvání, působení, růstu*. Při podrobnějším pohledu na konkordanční řádky zjistíme, že substantivum *doba* má kromě genitivní valence (ve vzorku genitiv představoval 20,2 % výskytů) i valenci v podobě věty vedlejší uvozené zájmenem *kdy* (9,2 % výskytů ve vzorku: *Stane se tak u nás poprvé od dosti dávných **dob, kdy** byla objevena televize.*).

Substantivum ***den*** (počet výskytů v SYN2000: 110 289) má opět výraznější kolokační profil nalevo (kolokace v intervalu tří slov nalevo), který tvoří taková adjektiva jako *osudný, šťastný, nejbližší, poslední, Štědrý, každý*. Napravo se ve vzorku s vysokým MI-score objevilo spojení v genitivu *otevřených dveří*. Při podrobnějším pohledu na konkordanční řádky vidíme, že v typické textové realizaci substantiva *den* je toto substantivum následováno udáním konkrétního data (8,4 % výskytů ve vzorku), ve 2,8 % výskytů následuje vedlejší věta uvozená zájmenem *kdy* (*Ve stejný **den, kdy** vypukly ničivé povodně...*), ve 2,4 % výskytů následuje předložka *po* (*Pátý **den po** operaci většinou umyjeme hlavu...*), dále specifictěji v 1,2 % výskytů následuje adverbium *poté*, ve většině případů je dále následováno vedlejší větou uvozenou zájmenem *co* (*Lety s humanitární pomocí do Sarajeva byly přerušeny před 23 **dny poté, co** „něčí“ řízená střela sestřelila italské letadlo...*), v 1 % výskytů následuje předložka *o* v konstrukci *o (X) den/(dní) později* (***O den později** v rozhovoru s italským listem La Repubblica řekl...*) a v 1 % výskytů následuje předložka *před* (*Zima letos přišla do pražských ulic pět **dní před** Vánocemi...*).

Den je také typicky následován jinou časovou specifikací, např. v podobě příslovce (*předem, ráno*). Nejvýraznější je zde však genitivní valence – 3,8 % výskytů, včetně výrazů jako výše uvedený atribut *otevřených dveří* (*Den nezávislosti, Dny otevřených dveří...*), avšak v nejčastější (4,6 %) konstrukci, do které genitiv vstupuje má substantivum *den* ještě nějaký specifický atribut napravo (typicky adjektivum nebo číslovku): *vyhlásil tři dny smutku; zbylé dny svého života; Závěrečný den pracovního semináře*.

Substantivum *hodina* (počet výskytů v SYN2000: 77 122) má méně výrazný kolokační profil, typicky se vyskytuje v konstrukcích jako: X *hodin denně* (1,8 %) (*mláďata musela v průměru osm hodin denně čekat na návrat a matčinu potravu*); X *hodin po* (1,8 %) (*V neděli, dvě hodiny po půlnoci, vedení mise OBSE v Sarajevu pravidla změnilo.*); X *hodin před* (1,4 %) (*kdyby lidé jezdili na stadion alespoň hodinu před začátkem promítání*); X *hodin ráno* (1,6 %) (*Je sedm hodin ráno, pane.*). Významným kolokátem substantiva *hodina* je adjektivum *ranní*, např. v *Xtá hodina ranní (do páté hodiny ranní inkriminované noci)*. I substantivum *hodina* má genitivní valenci (3 %), z toho téměř třetina výskytů je specifického sémantického typu, *hodina* ve významu „lekce“ (*Nestačí mu hodiny tělocviku ve škole?*). Zbývající genitivy mají obecný charakter, např. *...kterou jsme ani po hodině snahy nemohli otevřít.*

Substantivum *čas* (počet výskytů v SYN2000: 62 174) má opět výrazný kolokační profil nalevo (kolokace v intervalu tří slov na nalevo od KWIC) a najdeme zde adjektiva *reálný, volný, ztracený, krátký, nějaký, určitý, pravý, nejvyšší*. Při podrobnějším pohledu na vzorek konkordančních řádků si všimneme, že toto substantivum je součástí mnoha frazémů (*čas od času* – 3,4 %) a ustálených kolokací (typicky např. se slovesy *věnovat, ukázat*). Ve 2,2 % výskytů je substantivum *čas* následováno infinitivem, a to v typických konstrukcích: BÝT (*nejvyšší čas + inf.*; MÍT *čas + inf.*; NASTAT *čas + inf.* (...*je nejvyšší čas dát se do práce...*; *Takže máš dost času něco skládat.*); v 1,8 % výskytů následuje předložka *k* (*Dostatek času k projednání 165 změn v Zákoníku práce*); ve 2,8 % výskytů předložka *na* (*čas na generační změnu / rozmyšlenou*); v 1 % následuje vedlejší věta uvozená spojkou *aby* (*Je na čase⁹³, aby to vzali aspoň trochu vážně.*); v 1 % následuje vedlejší věta uvozená spojkou *kdy* (*A tak je zřejmě už jen*

⁹³ Výraz „*Je na čase*“ je sám o sobě frazémem.

otázkou času⁹⁴, *kdy tam kterási podnikavá česká cestovka začne pořádat pravidelné zájezdy.*); v 1,4 % výskytů následuje vztažná vedlejší věta uvozená zájmenem *který (...v čase, který takto získám, se věnuji svým zájmům...)*. Genitivní valence substantiva *čas* byla patrná v 8,8 % výskytů *Nyní je čas rozhodnutí...*). Substantivum *čas* vstupuje i do koordinovaných spojení (viz 4.5) se spojkou *a* (*čas a energie/klid/místo/námaha/peníze/prostor/smrť...*) a tato spojení nacházím i ve variantě se spojkou *ani* (*čas ani chuť/energie...*).

4.9.2 Skupina ČLOVĚK

Substantivum *člověk* (počet výskytů v SYN2000: 183 509) má natolik obecný význam, že i jeho valenční profil je nevýrazný. V náhodném vzorku 500 konkordančních řádků nenacházíme žádný frekvenčně významnější valenční vzorec. Ani při automatické analýze kolokací na vzorku 5 000 konkordančních řádků nenacházíme nic nečekaného, typickými kolokáty napravo (interval tří slov od KWIC) jsou slovesa *stvořit, prožít, naučit* a kolokáty nalevo (interval 3 slov) adjektiva *obyčejný, vzdělaný, slušný, uvažující* (s vysokým MI-score se tu objevuje i spojení *v tísní*).

Ani substantivum *dítě* (frekvence v SYN2000: 71 829) nemá žádnou výraznou valenci. Kolokační analýza vzorku nám říká, že významnými kolokáty nalevo jsou adjektiva *romské, zdravé, postižené*. Substantivum *dítě* v pl. vstupuje do koordinovaných spojení se spojkou *a* (3,2 %) (*děti a mládež / dospělí*). Substantivum *dítě* bývá (3 %) specifikováno určitým omezením, typicky např. *děti do 10 let; děti ve věku od šesti do patnácti*. Obecnou specifikaci realizuje substantivum např. předložkou *s* (1,2 %) (*děti s autismem*) a samozřejmě se vyskytuje i s genitivem (1,4 %) (*děti cizinců; děti tohoto věku; dítě štěstěny*).

⁹⁴ Je otázkou času je opět frazém.

Substantivum *žena* (počet výskytů v SYN2000: 59 524) rovněž nemá žádnou výraznou valenci. V omezené míře se vyskytuje v koordinovaných spojeních se spojkou *a* (a po této spojce typicky bývá substantivum *děti* – 17krát *žena a děti* v SYN2000), s vedlejší větou vztažnou realizovanou zájmenem *kteřá* a samozřejmě zde nacházíme i genitivní valenci. Substantivum *muž* (frekvence v SYN2000: 59 362) také nemá výraznou valenci. Vyskytuje se také do nějaké míry v koordinovaných spojeních se spojkou *a* (zde však po spojce *a* typicky nacházíme substantivum *žena*: MUŽ A ŽENA nacházíme v SYN2000 1425krát⁹⁵). Obecnou specifikaci realizuje substantivum *muž* (podobně jako *dítě*) předložkou *s* (*muž s černým obočím*), i zde nacházíme větu vztažnou uvozenou zájmenem *který* a několik příkladů genitivní valence.

4.9.3 Závěr

Analýza dvou výše uvedených sémantických skupin substantiv z vysokofrekvenčního pásma naznačila jejich značnou různorodost a potvrdila tak hypotézu o jejich specifitě a roztržitém valenčním profilu. Nicméně i zde můžeme vyzorovat určité tendence⁹⁶ a rozdíly mezi dvěma analyzovanými sémantickými skupinami. Zatímco v podstatě konkrétní substantiva ve skupině ČLOVĚK nemají výrazné valenční profily a valenci realizují konkrétně v textu (např. předložkou *s*), substantiva ve skupině ČAS jsou výrazně kolokační s mnoha idiomy a u některých nacházíme zcela zřetelnou valenci. Všechna substantiva však mají více valenčních exponentů (u všech nacházíme genitiv) a žádný z nich se nevyskytuje s výraznou frekvencí. Vysokofrekvenční substantiva by si

⁹⁵ Pro zajímavost ŽENA A MUŽ nacházíme v SYN2000 pouze 217krát.

⁹⁶ Které však mohou být i dílem náhodné vzhledem k malému množství analyzovaných substantiv i malým vzorkům podrobeným analýze.

zasloužila více pozornosti, jejich podrobnější analýza však bohužel přesahuje rámec této práce.

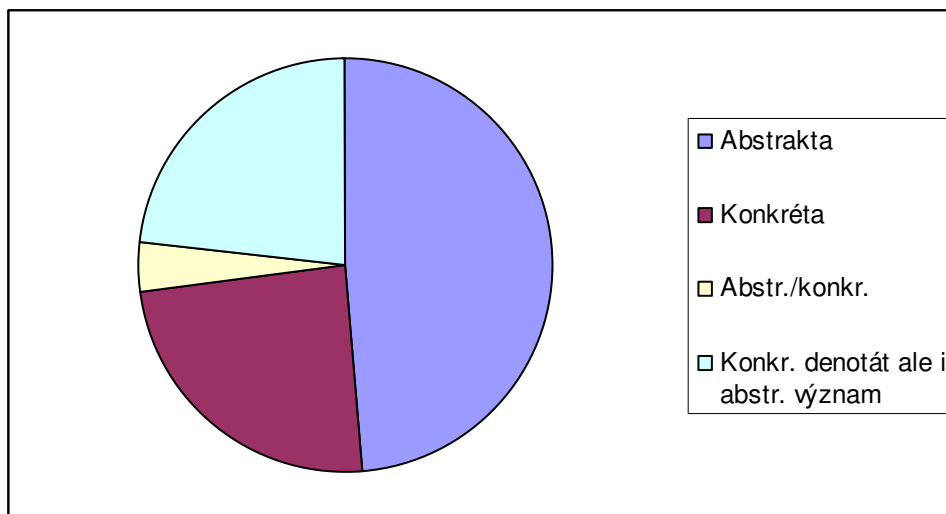
4.10 Abstrakta a konkréta

Při popisech substantivní valence někteří autoři rozlišují mezi substantivy derivovanými (od slovesa, s předpokladem, že i např. valence je od něj odvozená) a původními substantivy a dále mezi substantivy konkrétními a abstraktními. Rozdělení na substantiva konkrétní a abstraktní je podloženo hypotézou (a do jisté míry i různými příklady), že se konkréta a abstrakta chovají jinak (více např. Čermák 2005). Konkréta nemají stejné funkční, syntaktické a kolokační vzorce, zatímco abstrakta mají několik obecných vzorců, které jsou více strukturované a tím dovolují méně variací. Vzorce u abstrakt jsou pak dány jejich funkcí a významem (Čermák 2005:145). Čermák a Holub (1988:27) valenci konkrét charakterizují následně:

Valence konkrét (mají, zhruba řečeno, hmotný denotát) není tak výrazná a složitá jako valence abstrakt. Sémanticky v ní lze zjistit některé jasnější oblasti, třídy, jež jsou charakterizované stejnou valencí, u většiny případů je to však obtížné. Charakteristickým rysem velkého množství konkrét je nepevnost, resp. fakultativnost širokého spektra valence daná konkrétní specifikací podle potřeb textu. Lze tedy pochytit valenci jen poměrně pevnou, danou vyšší frekvencí výskytu a zaměřit se na povahu vztahu, aspektu, který je valencí vyjadřován, a to často bez jednoznačného vztahu k povaze substantiva.

V této kapitole se pokusíme námi analyzovaná data shrnout i z tohoto pohledu (konkrétní příklady pak nabízejí jednotlivé kapitoly a přílohy). Na vzorku námi analyzovaných substantiv jsme se pokusili najít a charakterizovat některé rozdíly mezi valencí abstrakt a konkrét. Náš vzorek substantiv jsme rozdělili do 4 skupin. Na (1) substantiva abstraktní, (2) substantiva konkrétní (tj. velmi obecně mající konkrétní denotát), (3) substantiva mající oba významy a (4) substantiva mající sice konkrétní denotát (osobu nebo věc), ale výraznější abstraktní složku významu. Největší skupinu tvoří substantiva abstraktní (48), dále jsou to konkréta (24), téměř stejně velká je skupina substantiv s konkrétním denotátem a s

abstraktní složkou významu (23) a nejmenší skupinu, 4 substantiva, představuje skupina substantiv s oběma významy, konkrétním i abstraktním. Následující graf ilustruje toto rozdělení.



Graf 7. Abstrakta vs. konkréta

4.10.1 Substantiva abstraktní

Skupinu abstraktních substantiv tvoří následující substantiva⁹⁷:

bankovnictví, citát, civilizace, deprese, dostih, důležitost, hádka, kolaps, konání, korupce, maličkost, mítink, monopol, mor, obliba, odjezd, odpočinek, odstranění, plemeno, pocta, pohoda, pojistné, poskytování, povstání, překročení, příznak, recenze, recese, respektování, rozpaky, sklonek, směrnice, soustředění, stanovení, stáří, tolerance, tušení, udržení, vaření, volnost, výdělek, vyhlášení, vzhled, vzpoura, zamyšlení, zájezd, zavádění, zodpovědnost.

Nejběžnější valencí abstraktních substantiv (tak jako i všech ostatních substantiv) je genitivní a předložková valence. Nacházíme zde však i dativ, instrumentál a

⁹⁷ Abstrakta jsou vybrána z původního vzorku substantiv, viz. 3.2

infinitiv. Mezi abstrakty jsou pouze 3 substantiva (6,25 %), která nemají genitivní valenci (nebo ji mají v naprosto omezené míře)⁹⁸ a nacházíme zde i poměrně málo příslovečných určení (18,75 %), která mají valenční charakter. Nominativní valenci zde nenacházíme žádnou.

4.10.2 Substantiva konkrétní

Skupinu konkrétních substantiv tvoří následující substantiva:

ateliér, housle, chata, klávesnice, kompresor, krb, kuře, moucha, obilí, pec, plátno, plavec, pleť, pero, porost, rukavice, smetana, sval, šátek, tribuna, ubytovna, vila, žíla, žloutek.

I zde je nejběžnější valencií genitiv a předložky, ve vzorku však nenacházíme jiné pády ani infinitiv. Je zde větší proporce substantiv bez (nebo téměř bez) genitivní valence (29,2 %), nacházíme zde nominativní valenci (12,5 %), je zde i vyšší zastoupení příslovečného určení, které má valenční charakter (29,2 %) a je zde i větší proporce substantiv, která nemají žádnou výraznou valenci (16,6 %).

4.10.3 Substantiva s abstraktním i konkrétním významem

Tuto malou skupinku tvoří pouze 4 substantiva: *klika, pasáž, podzemí, sonda*. Vzhledem k velikosti vzorku není možné činit jakékoliv závěry. Jako valenční exponenty zde nacházíme genitiv, předložky, nominativní valenci i příslovečné určení místa. Zatímco u substantiva *pasáž* se nominativní valence pojí s jeho

⁹⁸ Zde je však třeba připomenout, že povaha genitivní valence je natolik produktivní, že výše uvedená tvrzení je vždy třeba uvažovat ve vztahu k analyzovaným vzorkům.

konkrétním významem a příslovečné určení spíše s jeho abstraktním významem, u substantiva *sonda*, kde rovněž nacházíme nominativní valenci a příslovečné určení, se valenční exponenty pojí s oběma významy tohoto slova.

4.10.4 Substantiva s konkrétním denotátem a abstraktní složkou významu

Tuto skupinu substantiv jsme vytvořili, protože nebylo vždy snadné rozhodnout, zda se jedná o konkrétní, nebo abstraktní. Patří sem tedy substantiva mající konkrétní hmotný denotát, avšak významnou abstraktní složku významu. Téměř polovina z těchto substantiv má životný denotát. Tuto skupinu tvoří následující substantiva:

ambasáda, baba, disident, dlužník, dodatek, funkcionář, chudák, koridor, křivka, letectvo, nález, operátor, partnerka, policajt, pomník, pomocník, samec, stanoviště, školka, vdova, ves, viník, zástupkyně.

Kromě genitivní a předložkové valence zde nacházíme i jeden případ valence dativní (*pomník*). Je zde však vyšší proporční zastoupení substantiv bez genitivní valence (26 %) a s příslovečným určením majícím valenční charakter (21,7 %) a to především u substantiv neživotných. V této skupině nacházíme i nejvyšší proporční zastoupení valence nominativní (26 %). Bez valence (nebo téměř bez valence) bylo 8,7 % z těchto substantiv.

4.10.5 Závěr

Na základě výsledků námi provedené analýzy a námi sledovaných jevů se jednotlivé skupiny těchto substantiv ve vztahu k jejich valenčním exponentům liší

ve čtyřech rysech: (1) počtem substantiv nemajících genitivní valenci, (2) počtem substantiv bez výrazné valence, (3) počtem substantiv s nominativní valencí (viz 4.6) a (4) počtem substantiv s příslovečným určením majícím valenční charakter. Genitivní a předložková valence je společná všem skupinám. Ostatní exponenty též nacházíme více méně ve všech skupinách, liší se však frekvencí výskytu. Sledované jevy shrnuje následující tabulka. Vzhledem k velikosti, a tudíž nedostatečné reprezentativnosti vzorku není v tabulce uvažována třetí skupina, tedy substantiva s abstraktním i konkrétním významem.

Tabulka 18. Frekvence valence ve vztahu k rozdělení substantiv na abstrakta/konkréta

	ABSTRAKTA	KONKRÉTA	Substantiva s konkr. denotátem a abstr. složkou významu
Bez gen.	6,25 %	29,2 %	26 %
Bez valence	-	16,6 %	8,7 %
Nom. val.	-	12,5 %	26 %
Přísl. urč.	18,75 %	29,2 %	21,7 %

Vzhledem k povaze substantivní valence jako takové je nejzvláštnější skupinou skupina substantiv s konkrétním denotátem a abstraktní složkou významu. Je to skupina s nejvyšším zastoupením nominativní valence (nacházíme ji jak u životných i neživotných substantiv v této skupině). Je zde i poměrně velké zastoupení substantiv bez výrazné genitivní valence podobně jako u konkrét. Abstrakta se od zbývajících substantiv liší v tom, že všechny mají nějakou valenci, nenacházíme zde valenci nominativní a je zde malé zastoupení takových substantiv, která nemají genitivní valenci, což je naopak výraznějším rysem konkrét, kde téměř třetina substantiv nemá genitivní valenci. Mezi konkréty nacházíme i poměrně větší množství substantiv bez výrazného valenčního profilu.

5 ZÁVĚR

5.1 Data a metodologie

Základem práce je analýza náhodných vzorků 99 substantiv, která byla manuálně vybrána na čistě statistickém základě jako každé n -té ze substantiv ve středním a vyšším frekvenčním pásmu českého lexikonu. Střední a vyšší frekvenční pásmo zde představuje frekvenci mezi 1000 až 3339. U substantiv ve středním pásmu předpokládáme určité typické jazykové vzorce, které se v korpusovém materiálu vyskytují dostatečně často, aby umožňovaly určitá zobecnění. Tento přístup nám tak umožňuje poměrně spolehlivě popsat jazykové centrum této problematiky.

Naopak možnosti popisu jazykové periferie jsou dány velikostí a reprezentativností korpusového materiálu, který je k dispozici. Pro spolehlivý popis jazykové periferie však žádný národní korpus samozřejmě není dostatečně veliký.⁹⁹ Pro takovéto popisy se lépe hodí různé specializované korpusy. Nicméně i zde, tak jako v našem případě, může všeobecný korpus problematiku spolehlivě alespoň naznačit. Na základě výsledků pilotní studie pro tuto práci, která byla prováděna v nízkých frekvenčních pásmech, se dá předpokládat, že v jazykové periférii z hlediska valenčního nenajdeme nové typy valenčních exponentů nebo struktury mající valenční charakter.

Vysoké frekvenční pásmo výskytu pak představuje naopak zcela jiné problémy. Slovní zásoba je zde specifická, mezi nejfrekventovanější slova patří tzv. slova

⁹⁹ Rosamund Moon (2008:20) ve svém článku upozorňuje na problematiku řídké se vyskytujících dokladů v jazykovém korpusu a otázku míry jejich reprezentativnosti. Evert (2004) se ve své práci zabýval způsobem, jakým jsou statistické výpočty zkresleny v případě velice řídkých slovních kombinací ve velkých korpusech, které mohou být náhodné. Evert dochází k závěru, že kombinace, které se v korpusu vyskytují třikrát nebo čtyřikrát, pravděpodobně nejsou náhodné, avšak kombinace, které se vyskytují pětkrát a vícekrát, nejsou náhodné zcela určitě.

gramatická. V češtině je prvních 10 nejfrekventovanějších slov: *a, v/ve, se, být/bejt, na, ten, s/se, že, z/ze, který/kterej* (FSC), tedy samá slova gramatická. První substantivum se nachází na 17. pozici a je to substantivum *rok/léta*. Další substantivum – *člověk/lidé* – se nachází až na 42. pozici. Ve vysokém frekvenčním pásmu se tedy nemůžeme vyhnout problému popisu gramatických slov a jejich frekventovaných kombinací a kombinací se slovy negramatickými, tedy i substantivy (srov. např. Sinclair 1999). Kolokační profily vysokofrekvenčních substantiv mají svá výrazná specifika, která je třeba uchopit jiným způsobem. V neposlední řadě je pak problémem i praktická otázka, jak efektivně zpracovat tak velké množství dat.

Za účelem konkrétní analýzy byla valence jako kategorie v této práci pojímána poměrně volně a široce: v podstatě jako formálně vymežitelný jazykový vzorec vyskytující se s vyšší než náhodnou frekvencí napravo od substantiva. Takto pojatá analýza vyhovuje kritériím korpusově řízené analýzy („corpus-driven“, Tognini-Bonelli 2001), o kterou se tato práce pokusila. Na druhou stranu je však třeba připomenout, že toto pojetí nepostihne možné marginálněji se vyskytující konkrétní valenční exponenty některých substantiv a ani systematicky nevyhledává pouze určitá – z valenčního hlediska atraktivní – substantiva. Jak vyplývá z výše uvedeného, nevšímá si ani levého okolí substantiva (viz ale např. otázku atributivních adjektiv, odd. 4.1.5).

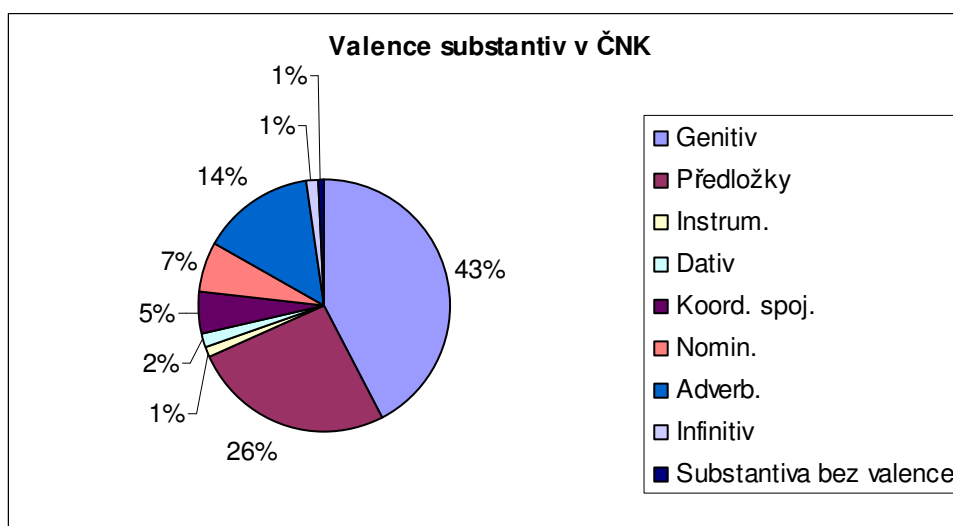
Korpusově řízená analýza formuluje teoretické závěry na základě analýzy dat a metodologický postup je charakterizován jako postupující od pozorování přes hypotézu k závěrům a teoretickému tvrzení (Tognini-Bonelli 2001:85). K analýze valence substantiv zde tedy přistupujeme s minimálními teoretickými východisky. Z tohoto důvodu byla také zvolena výše uvedená objektivní kritéria: výběr substantiv jako každé *n*-té, náhodné vzorky dat a jejich manuální analýza a frekvence výskytu jako kritérium pro oddělení typického úzu od okrajových výskytů. V některých případech nám však naše vzorky neposkytly dostatečné

množství dat pro objektivní zobecnění. V těchto případech byla analýza rozšířena o další vzorky dat.

U korpusově zpracovaných studií je zobecnitelnost výsledků dána i reprezentativností daného jazykového korpusu. V této práci jsme pracovali s daty korpusu SYN2000, jednoho z rozsáhlých korpusů projektu ČNK, který byl ve své době sice koncipován jako korpus reprezentativní (o složení korpusu bylo rozhodnuto na základě výzkumů recepce psaného jazyka), má však rozsáhlou (60 %) složku publicistických textů. Publicistika je žánr velmi specifický a nutně výsledky práce ovlivňuje. Na druhou stranu zde v podstatě chybí složka mluvená, která samozřejmě má také svá významná specifika (viz např. Sinclair & Mauranen 2006) .

5.2 Valence substantiv v češtině

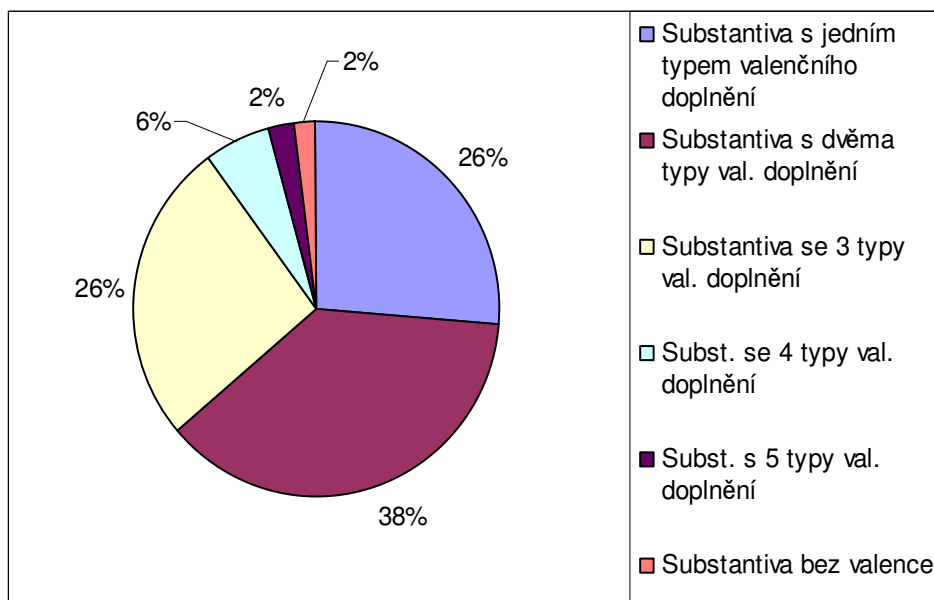
Následující graf (graf č. 8) nám nabízí celkový shrnující pohled na valenci¹⁰⁰ substantiv v češtině. Bezpředložkový genitiv (vyskytující se ve větší či menší míře u 91 substantiv z 99 analyzovaných) je valencí nejtypičtější. Druhou nejčastější valenci představuje předložkové spojení (55 substantiv z 99). V námi analyzovaném vzorku byla pouze 2 substantiva bez valence (*obilí, moucha*).



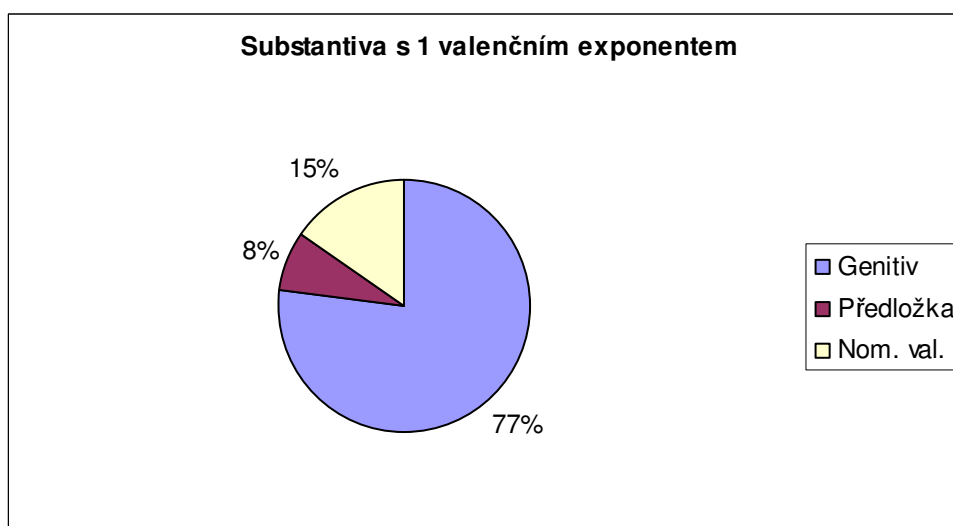
Graf 8. Valence substantiv v češtině

Graf č. 9 nám zprostředkovává na substantivní valenci jiný pohled. Ukazuje nám procentuální zastoupení substantiv podle počtu typů valenčních exponentů. Nejvíce substantiv má dva typy valenčního doplnění, o něco méně je pak substantiv s jedním nebo 3 typy valenčního doplnění. Grafy 10–12 nám pak dále ukazují, o které kombinace typů se jedná.

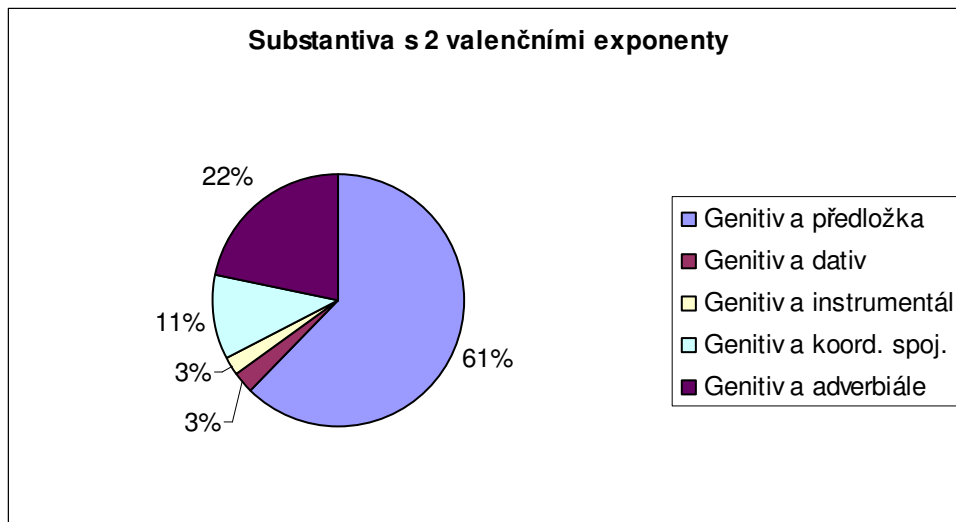
¹⁰⁰ Je zde třeba znovu připomenout, že valence jako taková je především abstraktní vztah něčím konkrétním realizovaný a v této práci se pro jednoduchost pojem „valence“ používá pro obojí, především však pro onu realizaci, tzn. že valencí zde nazýváme i konkrétní exponenty.



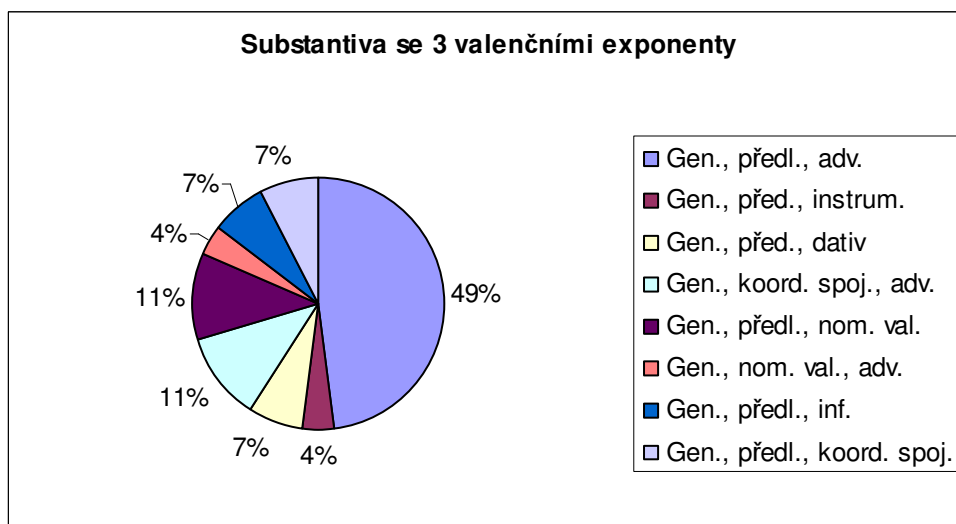
Graf 9. Počet typů valenčních exponentů u substantiv v %



Graf 10. Substantiva s jedním valenčním exponentem



Graf 11. Substantiva se 2 valenčními exponenty



Graf 12. Substantiva se 3 valenčními exponenty

Jak je z výše uvedených grafů patrné, nejčastějším valenčním exponentem substantiv v češtině je (1) bezpředložkový genitiv. U bezpředložkového genitivu jsme si všímali především jeho frekvence výskytu s jednotlivými substantivy a jeho sémantických typů. Ve vzorku bylo několik substantiv s velmi výraznou genitivní valencí, většinou se jednalo o deverbativa s abstraktním významem typická pro odborný a publicistický styl (např. *vyhlášení, odstranění, stanovení, poskytování*). Obecně se však dá říci, že bezpředložkový genitiv je téměř univerzální valencí českých substantiv. Bepředložkový genitiv jako valence přináší mnohé otazníky a tato problematika by si zasloužila více pozornosti. Existující popisy jsou zcela nedostatečné. Tato práce otevírá především prostor pro další zkoumání: otázku toho, kdy je bezpředložkový genitiv opravdu valencí a součástí lexikálního profilu daného substantiva a kdy se jedná spíše o volnou syntaktickou realizaci. Dále se otevírá prostor pro určení co nejobjektivnějších kritérií pro sémantickou klasifikaci genitivních valenčních exponentů. Dílčími otázkami pro další analýzu jsou řetězení genitivů a problematika korelace genitivního valenčního exponentu a atributivního adjektiva (viz dále 5.3).

Další z frekvenčního hlediska významnou valenci tvořila (2) předložková spojení. Předložkovou valenci má přibližně polovina substantiv. U předložkové valence jde obecně vlastně o valenci dvojí – substantiva a samotné předložky. Substantiva mohou mít jednopředložkovou, dvoupředložkovou a okrajově i vícepředložkovou valenci. Sémantika zde hraje nemalou roli, substantiva u jednotlivých předložek nabízejí rozdělení, řečeno terminologií *WordNetu*, do jakýchsi sémantických synsetů, a podíváme-li se na ně z opačné strany, tvoří vlastně kolokační paradigmata dané prepozice. Možnosti těchto kolokačních paradigmat jsou v této práci pouze naznačené, pro úplnější popis je nutná odlišná metodologie práce s daty korpusu.

Valence substantiv v češtině může být dále realizována (3) dalšími pády - a to dativem a instrumentálem. Valence může být realizována i (4) infinitivem nebo vedlejší větou.

Při analýze jsme dále identifikovali z frekvenčního hlediska významné struktury vyskytující se na pozici valenčního exponentu, které však v existujících popisech substantivní valence nebývají uvažovány. Patří sem (1) adverbialie, které někteří autoři při popisech valence slovesné připouštějí jako obligatorní valenci, dále (2) spojení dvou substantiv spojkou *a*, které jsme zde nazvali „koordinovaná spojení“, a (3) apozici dvou substantiv, kterou jsme zde nazvali „nominativní valence“.

V práci v podstatě zavádíme z deskriptivních potřeb pojem „valenčnost“ nebo „míra valenčnosti“. Obecně míra valenčnosti vyjadřuje, jak často se substantivum v konkrétních textových realizacích vyskytuje, a je tedy ve svém bezprostředním okolí sémanticky dále upřesňováno. Toto upřesnění může být realizováno jeho valencí nebo dalšími textovými prostředky (atributivní adjektivum, příslovecné určení). Zavedení tohoto pojmu si vynutila otázka tzv. obligatornosti či fakultativnosti valenčních exponentů. Určení není vždy jednoznačné a existuje mnoho mezních případů. Valenčnost bezprostředně souvisí se sémantikou daného lexému a je nejsilnější u lexémů, jejichž význam je abstraktní nebo vágní.

5.3 Otevřené otázky

Tato práce více otázek pokládá, než zodpovídá. Zvolená metodologie sice přinesla otevřený přístup a odkryla otázky, které je třeba pro popis substantivní valence dále řešit. Na druhou stranu nepřinesla dostatečné množství dat pro celistvější popisy jednotlivých typů valence. V některých případech jsme nedostatek dat kompenzovali dodatečnými hledáními. Obecně tato práce opět potvrdila náročnost spolehlivého zpracování dat velkého korpusu v realistickém čase.

Bezpředložkový genitivní valenční exponent je natolik běžný, že jsou spíše výjimkou ta substantiva, která se s bezpředložkovým genitivem nevyskytují. Otázkou tedy zůstává, zda jsou všechny tyto realizované a realizovatelné genitivy valencí, tedy součástí lexikálního profilu daných substantiv. Mnohdy se zde totiž nabízí i interpretace genitivního doplnění jako volné syntaktické realizace atributu podobně jako atributivní adjektivum. Nakolik se však skutečně jedná o „volnost“, se však nedá jednoznačně odpovědět.¹⁰¹ Tradiční pohled syntaktika nám pravděpodobně dá jinou odpověď než např. lingvista korpusový. Jak již bylo několikrát řečeno nebo naznačeno – ne vše má objektivní jazyková kritéria a distinkce fakultativní/obligatorní nemusí být vždy ta nejvhodnější.

V našem případě se jako vhodné kritérium pro rozlišení mezi statusem těchto genitivů jevila frekvence výskytu. Situaci u některých substantiv však komplikuje různá sémantika těchto genitivů. Bezpředložkový genitiv má sice identickou gramatickou realizaci – substantivum v genitivu –, ale může vyjadřovat různé sémantické role, a to i u jednoho substantiva (např. G objektový a G subjektový u substantiva *soustředění* – *soustředění jejich pozornosti* resp. *soustředění sil poblíž*

¹⁰¹ Uvážíme-li např. atributivní adjektivum *pouhý*, které stojí vždy jako premodifikující adjektivum v atributivní, tj. z hlediska syntaktického volné pozici před substantivem. Avšak když tak stojí vždy, nabízí se otázka, zda nelze tento vztah interpretovat jako valenci.

hranic). Frekvenční údaje výskytu bezpředložkového genitivu v pozici valenčního exponentu by měly být ideálně konfrontovány i s frekvencí výskytu ostatních valenčních exponentů, adverbíale a atributivních adjektiv a jejich případnou sémantickou korelací s genitivem. Takovýto komplexní pohled na valenci nám u některých substantiv přináší zajímavý pohled, např. substantivum *ambasáda* je v 75 % svých výskytů kontextem blíže určováno a to atributivním adjektivem, genitivem a příslovečným určením místa, substantivum *citát* je kontextem blíže určováno ve 44 % svých výskytů, a to genitivem, předložkou *z/ze* a atributivním adjektivem, nebo substantivum *kolaps*, které je kontextem blíže určováno v 65 % svých výskytů, a to atributivním adjektivem a genitivem.

U takového pojetí se nabízí hned několik otázek, na které nemáme jednoznačnou odpověď. Kde vede v těchto případech hranice, aby byla frekvence výskytu považována za významnou? Je to oněch 75 % u substantiva *ambasáda* nebo stačí 44 % u substantiva *citát*? Navíc zde uvažujeme i atributivní adjektivum a adverbíale, přitom obojí má z hlediska valence sporný charakter. U otázky „konkurence“ atributivního adjektiva a genitivu se navíc na několika příkladech ukázalo, že se vlastně o konkurenci v pravém slova smyslu vždy nejedná – adjektivum často specifikuje obecněji (*dopravní kolaps*), zatímco genitiv je konkrétnější (*kolaps slovenských elektráren*).

Pracujeme zde s pojmem „míra valenčnosti“, která může být vyjádřena prostým frekvenčním údajem. I zde však záleží na individuální interpretaci, zda je významné již číslo 44 %, nebo až 75 % (viz příklady výše). Při této příležitosti je třeba i zdůraznit, že ona výše zmiňovaná „sémantická neúplnost“ je především daná interpretativností v textu. „Valenčnost“ zde tedy chápeme jako obecnou vlastnost substantiva vyžadovat v konkrétních textových realizacích další formální specifikaci, a tedy i sémantické zpřesnění, forma konkrétní realizace pak může být různá. Při určení valenčnosti však frekvence výskytu není jediným kritériem, je třeba zohlednit sémantiku kolokátů a sémantiku samotného substantiva – jeho

sémantickou nesamostatnost či vágnost. Povaha valence je tak v zásadě dána povahou sémantiky substantiva. V řadě případů je pak zřejmé, že valence bývá přizpůsobena potřebám kontextu.

Abstrakt

Cílem práce je vytvoření lexikologického rámce pro systematický popis valence českých substantiv na základě analýzy dat z Českého národního korpusu (SYN2000). Valence je zde chápána jako systémová lexikologická vlastnost slov. Valence je abstraktním vztahem, který má konkrétní realizace, v této práci se pro zjednodušení pojem valence používá pro obojí. Pro účely analýzy je valence pojímána poměrně volně, protože není vhodné dopředu předpokládat, že tu jsou všechny její aspekty jasné. Předpokládá se jako platná napravo od substantiva a analýza si všímá jevů, které se zde systematicky (z hlediska frekvence výskytu) vyskytují. Práce nabízí korpusem řízený lingvistický popis, který se zaměřuje na slova v kontextu, ve kterém se obvykle nacházejí a vytvářejí tak typické slovní vzorce.

Základ analýzy tvoří náhodné vzorky 99 náhodně vybraných substantiv ze středního frekvenčního pásma. V případech, kdy vzorky neposkytly dostatečné množství dat, byla analýza rozšířena o hledání pomocí specializovaných dotazů do korpusu. U substantiv z vysokých frekvenčních pásem předpokládáme roztráštěnější valenční profily, pro potvrzení této hypotézy bylo několik substantiv z vysokých frekvenčních pásem podrobena dodatečné stručné analýze.

Nejběžnějším valenčním exponentem českých substantiv je bezpředložkový genitiv, který se vyskytuje s více než 90 % substantiv. Bepředložková genitivní valence je u některých substantiv velmi typická (i více než 90 % z jejich textových realizací), u jiných pak představuje pouze několik málo výskytů. Genitiv je pád sémanticky polyfunkční a to se projevuje i při jeho realizaci na pozici valenčního exponentu (i v rámci jednoho substantiva můžeme nacházet ze sémantického hlediska různé genitivy). Další typickou valencí je předložkové spojení. Předložkovou valenci měla přibližně polovina z analyzovaných

substantiv. Z ostatních pádů jsme dále narazili na instrumentál a dativ. Valence může být realizována i infinitivem a vedlejší větou.

Z hlediska valenčního má tradičně problematický status adverbium/adverbiále. Kritérium frekvence výskytu však potvrzuje, že s valenční povahou adverbia/adverbiále je třeba u některých substantiv počítat. Na pozici valenčního exponentu se u některých substantiv s vyšší frekvencí objevila i tzv. koordinovaná spojení (dvě substantiva s určitým společným sémantickým rysem spojená spojkou *a*) a nominativní valence (dvě substantiva ve stejném pádě a čísle vztahující se ke společnému denotátu).

Abstract

This thesis aims at providing a lexicological framework for systematic description of valency of Czech nouns. Valency is seen here as a lexicological property of words. Valency is an abstract relation with concrete textual realizations and the term „valency“ is used here for both: the abstract notion and the concrete valency exponents and realisations. The analysis is corpus-driven and as such it is based on a rather loose notion of valency, devoid of any pre-conceived ideas, concentrating on typical structural patterns of occurrence on the right side of the noun under investigation.

For the analysis the corpus SYN2000, a part of the Czech National Corpus has been used. The analysis is based on random selections of concordance lines of randomly chosen 99 nouns from the middle frequency range. In some cases, where the data proved insufficient, we have carried out additional specialized corpus queries. For high frequency nouns we assume highly differentiated valency profiles; to confirm this hypothesis we have carried out additional brief analysis of several high frequency nouns.

The most frequent valency of Czech nouns is genitive complementation, which we find as occurring with more than 90% of the analysed nouns. For some of the nouns, the genitive valency is a very dominant valency pattern (in some cases occurring in more than 90% of occurrences). With some other nouns, however, we find only a handful of examples. From the semantic point of view, genitive has several semantic functions and this is apparent also in its valency realizations (we may find semantically different genitives occurring with the same noun). The second most frequent and typical valency pattern of Czech nouns is a prepositional complementation. About half of the analysed nouns occurred with some prepositional complementations. Among other valency patterns we find

dative and instrumental complementations and infinitive and clause complementations.

The valency status of adverb/adverbial complementation is traditionally viewed as problematic. However, if we take into account the frequency of occurrence, it becomes apparent that with some nouns adverb/adverbial has a clear valency potential. With some nouns we also find, with a higher frequency of occurrence, two following structures: (1) two nouns with the conjunction „and“ (*a*), where the two nouns have some common semantic features and (2) two nouns in the same case and number with common denotation.

LITERATURA

- Atkins, S., Clear, J. & Ostler, N. (1992). Corpus Design Criteria. *Literary and Linguistic Computing*, 7 (1), 1–16.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8 (4), 243–257.
- Church, K. & Hanks, P. (1990). Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics*, 16 (1), 22–29.
- Collins COBUILD English Language Dictionary* (1987). Sinclair, J. et al. (Eds.). London/Glasgow: William Collins Sons & Co. Ltd.
- Čermák, F. (2005). Abstract Noun Collocations: Their Nature in a Parallel English-Czech Corpus. In G. Barnbrook, P. Danielsson & M. Mahlberg (Eds.), *Meaningful Texts. The Extraction of Semantic Information from Monolingual and Multilingual Corpora* (pp. 143–153). London/New York: Continuum.
- Čermák, F. (2001). *Jazyk a jazykověda*. Praha: UK Karolinum.
- Čermák, F. (1996). Systém, funkce, forma a sémantika českých předložek. *Slovo a slovesnost* 57, 30–46.
- Čermák, F. (1991). Podstata valence z hlediska lexikologického. In *Walencja czasownika a problemy leksykografii dwujęzycznej* (pp. 15–39). Wrocław.
- Čermák, F. & Holub, J. (1988). *Syntagmatika a paradigmatica českého slova I. Valence a kolokabilita*. Praha: Univerzita Karlova.
- Čermáková, A. (2006). Kolokace a valence některých případů substantiv. In F. Čermák & M. Šulc (Eds.), *Kolokace* (pp. 107–141). Praha: Nakladatelství Lidové noviny/Ústav Českého národního korpusu.
- Danielsson, P. (2001). *The Automatic Identification of Meaningful Units in Language*. PhD thesis, Språkdata, Department of Swedish, Göteborg University.

- Evert, S. (2004). *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. PhD. Institut für maschinelle Sprachverarbeitung, Universität Stuttgart. (Publikováno 2005)
- Filipec, J. & Čermák, F. (1985). *Česká lexikologie*. Praha: Academia.
- Fillmore, C. J. (1992). Corpus Linguistics or Computer-aided Armchair Linguistics. In J. Svartvik (Ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991* (pp. 35–60). Berlin/New York: Mouton de Gruyter.
- Firth, J. R. (1957). *Papers in Linguistics 1934–51*. London: Oxford University Press.
- Frekvenční slovník češtiny*. (2004). (Eds. F. Čermák et al.). Praha: Nakladatelství Lidové noviny.
- Greenbaum, S. (Ed.) (1996). *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.
- Grepl, M. & Karlík, P. (1998). *Skladba češtiny*. Olomouc.
- Hajičová, E., Panevová, J. & Sgall, P. (2002). *Úvod do teoretické a počítačové lingvistiky*. 1. svazek – *Teoretická lingvistika*. Praha.
- Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell et al. (Eds.), *In Memory of J. R. Firth* (pp. 148–162). London: Longman.
- Halliday, M. A. K., Teubert, W., Yallop, C. & Čermáková, A. (2004). *Lexicology and Corpus Linguistics*. London: Continuum.
- Helbig, G. & Schenkel, W. (1973). *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig: VEB Bibliographisches Institut.
- Hjelmslev, L. (1971). *Jazyk*. Praha: Academia.
- Hoey, M. (2005). *Lexical Priming. A new theory of words and language*. London/New York: Routledge.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: John Benjamins.

- Karlík, P. (2000). Hypotéza modifikované valenční teorie. *Slovo a slovesnost*, 61 (3), 170–189.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London/New York: Longman.
- Kolářová, V. (2005). *Valence deverbativních substantiv v češtině*. PhD. ÚFAL, MFF UK. Praha.
- Kolářová, V. (2004). Valence deverbálních substantiv: Některé specifické posuny v povrchových realizacích participantů. In P. Karlík (Ed.), *Korpus jako zdroj dat o češtině* (pp. 113–125). MU, Brno.
- Kopřivová, M. (2006). *Valence českých adjektiv*. Praha: Nakladatelství Lidové noviny/Ústav Českého národního korpusu.
- Králík, J. & Šulc, M. (2005). The Representativeness of Czech Corpora. *International Journal of Corpus Linguistics*, 10 (3), 357–366.
- Krishnamurthy, R. (Ed.) (2004). Sinclair, J., Jones, S. & Daley, R., *English Collocation Studies. The OSTI Report*. London/New York: Continuum.
- Leech, G. (1991). The state of art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics. Studies in Honour of Jan Svartvik* (pp.8–29). London/New York: Longman.
- Louw, W. E. (1993). Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In M. Baker et al. (Eds.), *Text and Technology: in honour of John Sinclair* (pp. 157–174). Amsterdam/Philadelphia: John Benjamins.
- Mahlberg, M. (2005). *English General Nouns. A corpus theoretical approach*. Amsterdam/Philadelphia: John Benjamins.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-Based Language Studies*. London/New York: Routledge.
- McEnery, T. & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Mluvnice češtiny /2/. Tvarosloví*. (1986). Praha: Academia.
- Moon, R. (2008). Conventionalized *as*-similes in English. A problem case. *International Journal of Corpus Linguistics*, 13 (1), 3–37.

- Moon, R. (1998). *Fixed Expressions and Idioms in English. A Corpus-based Approach*. Oxford: Clarendon Press.
- Novotný, J. (1980). *Valence dějových substantiv v češtině*. Sborník pedagogické fakulty v Ústí nad Labem. Praha: SPN.
- Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Owen, Ch. (2007). Notes on the ofness of *of* – Sinclair and grammar. *Words, grammar, text: revisiting the work of John Sinclair*, R. Moon (Ed.), Special issue of *International Journal of Corpus Linguistics*, 12 (2), 201–222.
- Palmer, F. R. (Ed.) (1968). *Selected Papers of J. R. Firth, 1952–1959*. Bloomington/London: Indiana University Press.
- Palmer, H. (1933). *Second Interim Report on English Collocations*. Tokyo: Institute for Research in English Teaching.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Sgall, P. (1998). Teorie valence a její formální zpracování. *Slovo a slovesnost* 59: 367–381.
- Sgall, P. (1986). *Úvod do syntaxe a sémantiky*. Praha: Academia.
- de Saussure, F. (1996). *Kurs obecné lingvistiky*. (S komentářem Tullio de Maura, český překlad F. Čermák). Praha: Academia.
- de Saussure, F. (1972). *Cours de linguistique générale, édition critique préparée par T. de Mauro*. Paris: Payot.
- Sinclair, J. & Mauranen, A. (2006). *Linear Unit Grammar. Integrating speech and writing*. Amsterdam: John Benjamins.
- Sinclair, J. (2004a). In praise of the dictionary. In Williams, G. & Vessier, S. (Eds.), *Proceedings of the Eleventh EURALEX International Congress* (pp. 1–11). Lorient, France.
- Sinclair, J. (2004b). To complement the dictionary. In Blatná, R. & Petkevič, V. (Eds.), *Jazyka a jazykověda* (pp. 417–444). FFUK, ČNK, Praha.

- Sinclair, J. (2004c). *Trust the Text. Language, corpus and discourse*. London: Routledge.
- Sinclair, J. (1999). A Way with Common Words. In Hasselgård & Oksefjell (Eds.), *Out of Corpus. Studies in Honour of Stig Johansson*. Language and Computers: Studies in Practical Linguistics 26 (pp. 157–179). Amsterdam/Atlanta: Rodopi.
- Sinclair, J. (1998). *The lexical item*. In E. Weigand (Ed.), *Contrastive Lexical Semantics* (pp. 1–24). Amsterdam/Philadelphia: John Benjamins.
- Sinclair, J. (1996a). *The empty lexicon*. International Journal of Corpus Linguistics, 1 (1), 99–119.
- Sinclair, J. (1996b). *The Search for Units of Meaning*. TEXTUS IX (1), 75–106.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (Ed.) (1987). *Looking Up: an Account of the COBUILD Project in Lexical Computing*. London: Collins.
- Stubbs, M. (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Stubbs, M. (1996). *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Oxford/Cambridge, Mass.: Blackwell.
- Šmilauer, V. (1966). *Novočeská skladba*. Státní pedagogické nakladatelství.
- Tesnière, L. (1969). *Éléments de syntaxe structurale. Deuxième édition revue et corrigée*. Paris: Éditions Klincksieck.
- Teubert, W. (2007). Sinclair, pattern grammar and the question of *hatred*. *Words, grammar, text: revisiting the work of John Sinclair*, R. Moon (Ed.), Special issue of *International Journal of Corpus Linguistics*, 12 (2), 223–248.
- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10 (1), 1–13.
- Teubert, W. (1979). *Valenz des Substantivs*. Düsseldorf: Schwann.

Teubert, W. & Čermáková, A. (2004). Directions in corpus linguistics. In Halliday et. al., *Lexicology and Corpus Linguistics*. London: Continuum.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.

Úvod a příručka uživatele (2000). Praha: Český národní korpus.

Zipf, G. K. (1935). *The Psychobiology of Language*. Cambridge, Mass.: MIT Press.

JAZYKOVÉ KORPUSY

(Následující výčet obsahuje pouze jazykové korpusy zmiňované v textu)

Český národní korpus

Podrobné informace o SYN2000, SYN2005, Pražském mluveném korpuse a korpuse Oral 2006 jsou přístupné na: <http://ucnk.ff.cuni.cz/> (listopad 2007).

Vzorek českého jazyk a literatury. (2000). CD obsahující korpus SYNEK a LITERA. Vydal Ústav Českého národního korpusu.

Brown Corpus

Manuál ke korpuse je přístupný na: <http://icame.uib.no/brown/bcm.html> (listopad 2007)

International Corpus of English (ICE)

<http://www.ucl.ac.uk/english-usage/ice/> (listopad 2007)

The Lancaster-Oslo/Bergen Corpus

Manuál ke korpuse je přístupný na:
<http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM> (listopad 2007)

The London-Lund Corpus of Spoken English

Manuál ke korpuse je přístupný na:
<http://khnt.hit.uib.no/icame/manuals/LONDLUND/INDEX.HTM> (listopad 2007)

Survey of English Usage

<http://www.ucl.ac.uk/english-usage/> (listopad 2007).

PŘÍLOHA A. - Substantiva vybraná pro pilotní analýzu valence substantiv v češtině.

Frekvenční pásmo I: 230–240 (195 substantiv, každé osmé):

chorál, stavebnice, podloží, třísko, kadeřnice, ostroh, debl, posilovna, soudnost, chvění, léčka, reggae, bajka, ohledání, zádrhel, kousíček, vlajkonoš, motel, přeřad, nositelka, sterilizace, lampička, dezinfekce, obviňování, škodlivost

Frekvenční pásmo II: 430–450 (154 substantiv, každé šesté):

oddělení, zákoutí, vzpěrač, poskytovatel, čip, opravna, vlaštovka, skifař, sounáležitost, nora, jamka, zpochybnění, vracení, regata, přísada, větrání, exekuce, žokej, maloobchod, mrk, zápor, ražení, saň, happening, měřidlo

Frekvenčního pásmo III: 630–660 (138 substantiv, každé páté):

romantika, zloba, alkoholik, rozběh, rentabilita, cestovatel, výprodej, nadšenec, komplet, medvídek, improvizace, házenkářka, zákonnost, bojkot, dřev, holocaust, police, podplukovník, havel (typ účesu), kozlík, sčítání, výhrůžka, křeace, formalita, dotyk

PŘÍLOHA B. - Sémantické typy genitivů.

Substantiva označená * mají více sémantických typů genitivní valence.

ambasáda	poses. G 4,3 %	<i>dvě ambasády Spojených států</i>
ateliér*	kvalit. G 6,2 %	<i>v ateliéru architektury</i>
ateliér*	poses. G 10,6 %	<i>v ateliéru Bořka Šípka</i>
citát*	kvalit. G 2,3 % zdroj 1,6 % čas.	<i>přímé citáty židovských žalmů citát týdne</i>
citát*	11,3 % gen. auctoris	<i>nějaký citát Aristotela</i>
dlužník*	kvalit. G 1,4 %	<i>dlužníků nájemného</i>
dlužník*	obj. G 4,6 %	<i>největším dlužníkem Fondu národního majetku</i>
dodatek*	kvalit. G 12,3 %	<i>ten dodatek zápisu</i>
dodatek*	2x gen. auctoris	<i>dodatku Poslanecké sněmovny k</i>
dostih*	kvalit. G 4,5 % čas. 4 %	<i>dostih letošního roku nejdražší dostih světa v Dubaji</i>
dostih*	obj. G 1,75 %	<i>dostih klusáků</i>
důležitost	kvalit. G 46,2 %	<i>chápal obrovskou důležitost dokumentu</i>
funkcionář	poses. G 36,4 %	<i>funkcionáři strany</i>
hádká	poses. G 3 %	<i>hádká majitele domu s drsným mladíkem</i>
klávesnice	poses. G 6 %	<i>klávesnice cembala</i>
klika*	poses. G (konkr.) 9,3 % (abstr.) 2,3 %	<i>klika dveří klika Chruščova a Brežněva</i>

kolaps	subj G 34 %	<i>po kolapsu východního mocenského centra</i>
korupce	obj. G 5,2 %	<i>korupce policistů</i>
křivka	poses. G 32 %	<i>křivka těla</i>
letectvo*	poses. G 7,5 %	<i>letectvo Spojenců</i>
mítink*	kvalit. G 0,8 %	<i>mítink letošní sezóny</i>
mítink*	poses. G 11,6 %	<i>mítink občanů s ministrem vnitra</i>
monopol*	explik. G 9 %	<i>monopol moci</i>
monopol*	poses. G 14,4 %	<i>monopolu České pojišťovny</i>
mor*	kvalit. G 1 % čas.	<i>mor minulého měsíce</i>
mor*	(konkr.) obj. G 5,3 %	<i>mor prasat</i>
mor*	(abstr.) subj G 4,2 %	<i>mor komunismu</i>
nález*	kvalit. G 3x	<i>nález miocénu u Hluchova</i>
nález*	obj. G 20,8 %	<i>nález maligního nádoru</i>
nález*	subj G 6,6 %	<i>závazný nález soudu</i>
obliba*	obj. G 11,4 %	<i>obliba značky IBM</i>
obliba*	subj G 7,4 %	<i>ztratila oblibu voličů</i>
odjezd	obj. G 10,6 % neživ. 7,2 % živ.	<i>odjezd vlaku</i> <i>odjezd vojáků</i>
odpočinek	obj. G 4,5 %	<i>je dnem odpočinku těla i duše</i>
operátor*	kvalit. G 12,3 % místo zařiz. předm.	<i>operátor Baltiku</i> <i>operátor Titaniku</i> <i>operátor mobilních telefonů</i>
partnerka	poses. G 7 %	<i>partnerka Brada Pitta</i>

pasáž*	part. G 10,4 %	<i>pasáž americké verze dokumentu</i>
pasáž*	(konkr.) poses. G 1 %	<i>nad vchodem do pasáže divadla Čas</i>
pero	16,6 % gen. auctoris "z pera" poses. G 2,8 %	<i>z pera Hanse Christiana Andersena pera vran</i>
plavec	kvalit. G 4 %	<i>nejlepšího plavce světa</i>
plátno*	kvalit. G 3x vel. 1x čas. 2x obec. 1 % abstr.	<i>plátna velkých rozměrů vlasteneckých pláten minulého století vzácných pláten historické ceny na plátně světové literatury</i>
plátno* (umělec.) (film.)	3,4 % gen. auctoris poses. G 2 %	<i>plátno Paula Cézanna na plátna kin</i>
plemeno	explik. G 17,3 %	<i>plemen psů</i>
pleť	poses. G 3,4 %	<i>pleti obličeje a krku</i>
pocta	obj. G 4,7 % "k poctě"	<i>k poctě J. A. Komenského; pocta nizozemského odboje</i>
podzemí	kvalit. G 21 %	<i>podzemí vichyovské Francie (abstr. i konkr. význam)</i>
pohoda*	kvalit. G 4,2 % (celk. i se subj. gen.)	<i>pohodu vánoc; pohodu bydlení</i>
pohoda*	subj G 4,2 % (celk. i s kval. gen.)	<i>pohodu týmu</i>
pomník	poses. G 22,5 %	<i>pomník T. G. Masaryka</i>
pomocník	poses. G 12 %	<i>pomocník dozorčího</i>
porost*	kvalit. G 13,2 %	<i>nízké i vysoké porosty trávy</i>
porost*	obj. G 1,6 %	<i>porost houby</i>
povstání*	poses. G 1 %	<i>p. Bedri Chánovo ve 40. letech</i>

povstání*	subj G 10,7 %	<i>povstání sedláků</i>
recenze*	obj. G 17,8 %	<i>recenze Vaculíkova a Kunderova románu</i>
recenze*	5,7 % gen. auctoris	<i>recenze pana docenta</i>
rozpaky	poses. G 5,4 %	<i>rozpaky odborné veřejnosti</i>
směrnice*	obj. G 2,5 %	<i>Robertovy směrnice pořádku</i>
směrnice*	23,75 % gen. auctoris	<i>směrnice EU</i>
soustředění*	obj. G 5,6 %	<i>soustředění podnikatelských aktivit</i>
soustředění*	(sport.) poses. G 3,8 %	<i>soustředění národního týmu</i>
soustředění*	subj G (sil: abstr. - 2x, konkr. -4x) (psych.) gen. obj. i subj. (3 %)	<i>soustředění sil poblíž hranic Belgie a Lucemburska</i> <i>soustředění jejich pozornosti na strategii; soustředění motoristy</i>
stanoviště*	kvalit. G 2x	<i>stanoviště klasického provedení</i>
stanoviště*	poses. G 17,6 %	<i>stanoviště OSN v somálském hlavním městě</i>
stáří	subj G 13,8 %	<i>stáří automobilu</i>
sval*	poses. G 5 % (spec.) 1,8 % (obec.)	<i>svaly břišní stěny</i> <i>svaly koně</i>
tolerance*	obj. G 3,5 % (techn.) 3,25 %	<i>s nulovou tolerancí zločinu</i> <i>tolerance jednoho bodu, tolerance vnitřního radiusu</i>
tolerance*	subj G 6,25 %	<i>tolerance veřejnosti k lidským ztrátám</i>
tribuna	poses. G 10,3 %	<i>tribuna amfiteátru</i>
tušení*	obj. G 7,3 %	<i>tušením brzkého konce</i>
tušení*	subj G 1 %	<i>tušení všech</i>

ubytovna	poses. G 19,6 %	<i>ubytovna žadatelů o azyl</i>
vaření	obj. G 10 %	<i>vaření brambor</i>
vila	poses. gen. 7,8 %	<i>kolem vily Havlových</i>
viník	obj. G 27,25 %	<i>hlavními viníky spiknutí</i>
volnost	kvalit. G 17,3 %	<i>volnost pohybu</i>
výdělek	poses. G 6,2 %	<i>výdělky žen zaostávají</i>
vzhled	poses. G 31 %	<i>vzhled této místnosti</i>
vzpoura	subj G 18 %	<i>vzpoura mlouk</i>
zájezd*	obj. G 2,8 %	<i>zájezd příznivců na zápas</i>
zájezd*	subj G 2,8 %	<i>na zájezdech Čedoku</i>
zodpovědnost*	obj. G (okrajově)	<i>zodpovědnost předávání víry</i>
zodpovědnost*	subj G 8,3 %	<i>zodpovědnost architekta za celé dílo</i>
žíla*	poses. G anat. 1 % anat. spec. 1x obec. 2,7 %	<i>žíla Vala Kilmera žíla horní končetiny žíly mramoru</i>

PŘÍLOHA C. - Přehled bezpředložkové genitivní valence podle frekvence

Substantivum	Genitiv	Počet sémant. typů	Substantivum	Genitiv	Počet sémant. typů	Substantivum	Genitiv	Počet sémant. typů
<i>baba</i>	0		<i>ambasáda</i>	4,30%	1	<i>porost</i>	14,50%	1
<i>disident</i>	0		<i>odpočinek</i>	4,50%	1	<i>ateliér</i>	16,80%	2
<i>chudák</i>	0		<i>žila</i>	4,60%	3	<i>plemeno</i>	17,30%	1
<i>moucha</i>	0		<i>korupce</i>	5,20%	1	<i>odjezd</i>	17,80%	1 (2)
<i>obilí</i>	0		<i>rozpaky</i>	5,40%	1	<i>stanoviště</i>	18%	2
<i>smetana</i>	0		<i>zájezd</i>	5,60%	2	<i>vzpouza</i>	18%	1
<i>ves</i>	0		<i>dlužník</i>	6%	2	<i>pero</i>	18,30%	1
<i>žlutek</i>	0		<i>klávesnice</i>	6%	1	<i>oblíba</i>	18,80%	2
<i>housle</i>	<1%	2	<i>výdělek</i>	6,20%	1	<i>podzemí</i>	21%	1
<i>krb</i>	<1%	1	<i>pocta</i>	6,60%	1 (2)	<i>volnost</i>	21,30%	1
<i>kuře</i>	<1%	2	<i>sval</i>	6,80%	2	<i>monopol</i>	23,40%	2
<i>malíčkost</i>	<1%	1	<i>partnerka</i>	7%	1	<i>recenze</i>	23,50%	2
<i>pec</i>	<1%	1	<i>letectvo</i>	7,50%	2	<i>směrnice</i>	23,80%	2
<i>policejt</i>	<1%	1	<i>plátno</i>	7,50%	3	<i>pomník</i>	24,80%	1
<i>sonda</i>	<1%	1	<i>vila</i>	7,80%	1	<i>viník</i>	27,30%	1
<i>šátek</i>	<1%	1	<i>tušení</i>	8,30%	2	<i>nález</i>	28%	5
<i>vdova</i>	<1%	1	<i>zodpovědnost</i>	8,30%	2	<i>vzhled</i>	31%	1
<i>pojistné</i>	1%	1	<i>mitink</i>	9,40%	3	<i>křivka</i>	32%	1
<i>rukavice</i>	1%	1	<i>mor</i>	10%	2	<i>kolaps</i>	34%	1
<i>civilizace</i>	1,60%	1	<i>vaření</i>	10%	1	<i>funkcionář</i>	36,40%	1
<i>bankovníctví</i>	1,70%	1	<i>dostih</i>	10,30%	2	<i>důležitost</i>	46,20%	1
<i>školka</i>	2%	2	<i>tribuna</i>	10,30%	1	<i>příznak</i>	48%	1
<i>koridor</i>	2,30%	3	<i>povstání</i>	10,70%	2	<i>konání</i>	57%	2
<i>zamyšlení</i>	2,30%		<i>ubytovna</i>	11%	1	<i>zástupkyně</i>	57%	1
<i>deprese</i>	2,50%	3	<i>dodatek</i>	12%	2	<i>udržení</i>	71%	1
<i>kompresor</i>	2,60%	1	<i>pomocník</i>	12%	1	<i>respektování</i>	85%	1
<i>samec</i>	2,80%	1	<i>pasáž</i>	12,20%	2	<i>odstranění</i>	88%	1
<i>chata</i>	2,80%	2	<i>operátor</i>	12,30%	3	<i>vyhlášení</i>	88%	2
<i>hádky</i>	3%	1	<i>tolerance</i>	13%	2	<i>překročení</i>	89%	1
<i>recese</i>	3%	2	<i>soustředění</i>	13,60%	3	<i>zavádění</i>	89%	1
<i>pleť</i>	3,40%	1	<i>stáří</i>	13,80%	1	<i>stanovení</i>	94%	1
<i>plavec</i>	4%	1	<i>citát</i>	14%	3	<i>poskytování</i>	96%	1
<i>pohoda</i>	4,20%	3	<i>klíka</i>	14%	2	<i>sklonek</i>	99%	1

PŘÍLOHA D. - Infinitivní valence

Výsledky hledání: [tag="N.*"] [tag="Vf.*"] - celkem 335 061 konkordačních řádků, které byly podrobeny manuální analýze.

ambice	potěšení
cíl (za cíl / s cílem)	potřeba
ctižádost	povel
čas	pověření
čest	povinnost
dar	povolání (rozlišuje 2 lexémy)
důvod	povolení
hrůza	požadavek
chuť	právo
mandát	pravděpodobnost
motivace	pravomoc
možnost	privilegium
myšlenka	problém
nabídka	přání
naděje	předpoklad
nápad	předsevzetí
nárok	představa
nařízení	příkaz
návrh	příležitost
nechuť	přípravenost
nemožnost	rada
neochota	radost
neschopnost	rozhodnutí
nezbytnost	rozkaz
nutkání	rozkoš
nutnost	sen
odhodlání	schopnost
odmítnutí	síla
odpovědnost	sklon
odvaha	slib
ochota	slušnost
omyl	smysl
oprávnění	snaha
plán	souhlas
pocit	strach
pocta	svoboda
pokus	svolení
pokušení	šance

škoda
štěstí
tendence
touha
trend
trpělivost
úkol
umění (jako "schopnost")
úmysl
úsilí

vůle
zájem
zákaz
záměr
záminka
zásada
závazek
způsobilost
zvyk