

Univerzita Karlova v Praze
Filozofická fakulta
Ústav českého jazyka a teorie komunikace

Postavení číslovek v systému slovních druhů

Numerals in the Parts of Speech System

DIPLOMOVÁ PRÁCE

Michaela Paldusová

m.paldusova@seznam.cz

Vedoucí práce: Mgr. Václav Cvrček, Ph.D.

Praha 2012

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně, výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu

V Praze dne

Na tomto místě bych ráda poděkovala Mgr. Václavu Cvrčkovi, Ph.D. za pečlivé, vstřícné a trpělivé vedení práce a podnětné připomínky, svým sestřám za podporu nejen při psaní této práce, všem hudebním múzám za inspiraci, MUDr. Pavlu Kuthanovi za svědomitý a lidský přístup doprovázený lingvistickomedicínskými rozhovory u štěrbínové lampy, všem, kteří nepochybovali, a koneckonců také jedné nemoci za to, že mě naučila dívat se na svět jinýma očima.

Abstrakt

Tato diplomová práce se zabývá problematikou vymezení slovního druhu číslovek a jeho postavením v systému slovních druhů. Nejprve se stručně věnuje historii a současným postojům k obecné slovnědruhové otázce v kontextu zahraniční i české lingvistiky. Následně představuje jednotlivé teoretické koncepce české lingvistiky věnující se otázce slovního druhu číslovek od poloviny 20. století až po současnost. Na základě kontextové analýzy dat z databáze Českého národního korpusu poté ověřuje dosavadní kritéria slovnědruhové klasifikace. Zároveň představuje a testuje na prototypických číslovkových lemmatech nový přístup k otázce vymezení číslovek jako samostatného slovního druhu, jehož základem je kontext zkoumaných jednotek. Na závěr předkládá návrhy, které se opírají o statisticky zpracované výstupy kontextové analýzy.

Klíčová slova: číslovky, vymezení číslovek, kontextová analýza, Český národní korpus, korpusová analýza

Abstract

This thesis deals with the definition of numerals and its position in the system of words classes. First, briefly describes the history and current attitudes to general issue of parts-of-speech classification in the context of foreign and Czech linguistics. Subsequently, a different theoretical concepts Czech linguistics devoted to the issue of speech numerals from the mid-20th century to the present. Based on the contextual analysis of data from the database of the Czech National Corpus then verifies the current classification criteria of parts-of-speech classification. At the same time tests on a prototypical numerals new approach to the definition of numerals as separate species, based on the context units surveyed. Finally submits proposals that are based on statistically processed outputs of contextual analysis.

Key words: numerals, classification of numerals, context analysis, Czech National Corpus, corpus analysis

Seznam použitých zkratek:

A	adjektiva
C	číslovky
CA	číslovky adjektivní
CS	číslovky substantivní
CN	číslovky nesklonné
ČNK	Český národní korpus
ČŘJ	Čeština - řeč a jazyk
D	adverbia
ESČ	Encyklopedický slovník češtiny
H&J	Havránek a Jedlička
I	interjekce
J	konjunkce
L1	bezprostřední levostranný kontext
MČII	Mluvnice češtiny 2
N	substantiva
P	zájmena
P1	bezprostřední pravostranný kontext
PDT	Pražský závislostní korpus
PMČ	Příruční mluvnice češtiny
R	prepozice
T	partikule
V	verba
Z	číslovky numerické

Obsah

1. Úvod.....	7
2. Struktura práce.....	8
3. Jazykový materiál.....	9
4. Přístupy ke slovním druhům.....	11
4.1 Historie slovnědruhové klasifikace.....	11
4.2 Kritéria klasifikace.....	13
4.2.1 Obecně.....	13
4.2.2 Kritérium sémantické.....	14
4.2.3 Kritérium morfologické.....	15
4.2.4 Kritérium syntaktické.....	15
4.3 Česká tradice slovnědruhové klasifikace.....	17
4.3.1 Kritérium sémantické.....	17
4.3.2 Kritérium morfologické.....	18
4.3.3 Kritérium syntaktické.....	18
4.3.4 Současné přístupy ke slovnědruhové klasifikaci.....	19
4.3.4.1 Příruční mluvnice češtiny.....	19
4.3.4.2 Mluvnice češtiny II.....	20
4.3.4.3 Mluvnice současné češtiny.....	20
4.3.4.4 Funkční generativní popis.....	21
5. Vymezení slovního druhu číslovek v mluvnicích, jazykových pracích a monografiích.....	22
5.1 Mluvnice.....	22
5.1.1 Česká mluvnice.....	22
5.1.2 Mluvnice češtiny II.....	23
5.1.3 Příruční mluvnice češtiny.....	26
5.1.4 Čeština - řeč a jazyk.....	27
5.1.5 Mluvnice současné češtiny.....	28
5.2 Slovníky.....	30
5.2.1 Encyklopedický slovník češtiny.....	30
5.3 Monografie.....	31
5.3.1 Příspěvky k české morfologii.....	31
6. Analýza materiálu.....	34
6.1 Metoda - kontextová analýza.....	34
6.2 Materiálová analýza.....	39
6.2.1 Materiálová analýza na obecné slovnědruhové úrovni.....	43
6.2.1.1 Kontext číslovek numerických.....	44
6.2.1.2 Kontext číslovek substantivních.....	46
6.2.1.3 Kontext číslovek adjektivních.....	47
6.2.1.4 Kontext číslovek nesklonných.....	49
6.2.1.5 Zhodnocení podobnosti kontextů slovních druhů statistickými metodami.....	53
6.2.2 Materiálová analýza na úrovni lexémů.....	56
6.2.2.1 Kontext numerických číslovek.....	57
6.2.2.2 Kontext adjektivních číslovek.....	58
6.2.2.3 Kontext nesklonných číslovek.....	59
6.2.2.4 Kontext substantivních číslovek.....	60
7. Závěr.....	62
8. Bibliografie.....	66

1. Úvod

Jak je z názvu patrné, předkládaná práce se bude zabývat slovním druhem číslovek jako celkem, jeho postavením v systému slovních druhů a problematikou jeho vymezení.

Uvědomíme-li si, že slovní druhy jsou charakterizovány na základě tří kritérií, tedy vzhledem k jejich formální stránce (tu zastupuje kritérium morfologické), funkci, kterou plní ve větě (hledisko syntaktické), a vzhledem k jejich významu (kritérium sémantické), přičemž číslovky jsou jako jediný slovní druh vymežovány především na základě kritéria sémantického, je jejich postavení při nejmenším výjimečné a je podle nás legitimním požadavkem ho podrobit dalšímu zkoumání.

Čeština disponuje hned několika prostředky vyjadřování kvantovosti. Od gramatického prostředku čísla (*pes šel, psi šli*) přes prostředky vyjadřující způsoby slovesného děje (*naplakat se, zaplakat si*), stupňování u adjektiv a adverbii (*vysoký, vyšší, nejvyšší, vysoko, výše, nejvýše*), vyjádření násobenosti pomocí reduplikace (*malinkatý, malilinkatý, velikanánský*) a některých prefixů (*převeliký, prastarý*) po některé funkce pádových tvarů (genitiv partitivní - *přivezla jsem jablek*). Slovní druh číslovek je obecně chápán jako základní slovnědruhový prostředek označení kvantovosti.

Hlavním cílem této práce je přiblížit dosavadní teoretické přístupy ke klasifikaci slovních druhů v různých bohemistických i zahraničních lingvistických koncepcích a následně představit především přístupy k vymezení slovního druhu číslovek v české lingvistické tradici a pokusit se na základě analýzy dat Českého národního korpusu tyto přístupy ověřit, případně přinést kritéria nová.

Na rozdíl od dřívějších přístupů má corpus-based analýza (přístup ověřující informace na korpusu), na které budeme tuto práci zakládat, nesporná pozitiva v množství dat, o něž se opírá, a v možnosti jejich statistického zpracování. Jedná se tedy o ryze formální přístup na základě autentického jazykového materiálu.

2. Struktura práce

Tato diplomová práce se skládá ze dvou hlavních částí, teoretické a analytické, v jejichž rámci je rozdělena do pěti kapitol, pokud nepočítáme kapitolu Úvod a tuto kapitolu, která se věnuje struktuře předkládané práce.

Začínáme tedy u kapitoly třetí, v níž si představíme materiál, se kterým budeme pracovat, a pokusíme se poukázat na jeho přednosti ve srovnání s materiály ostatními.

Kapitola čtvrtá nás stručně seznámí s tradicí a vývojem vymezení slovnědruhových tříd, nejen v prostředí české lingvistiky, a přístupy k němu.

V páté kapitole přiblížíme základní koncepce v rámci české lingvistiky věnující se vymezení slovního druhu číslovek a také se pokusíme naznačit problematiku jeho klasifikace.

V rámci šesté kapitoly se seznámíme s metodou, pro kterou jsme se v této práci rozhodli, a podáme vysvětlení tohoto rozhodnutí. Ve druhé části šesté kapitoly se budeme zabývat analýzou získaných dat.

V sedmé kapitole (Závěr) se budeme věnovat hodnocení výsledků analýzy a návrhu možných řešení v rámci tohoto experimentu.

3. Jazykový materiál

Jazykový materiál budeme čerpat ze souboru počítačově uložených nejen psaných, ale i ve formě prepisů zaznamenaných textů mluvených, vytvářeného v rámci akademického projektu, který realizuje Ústav českého národního korpusu, s názvem Český národní korpus (dále ČNK) (dostupný z <http://korpus.cz/>).

K výhodám takového počítačově zpracovaného souboru dat patří např. na rozdíl od internetových vyhledávačů stálost dat a už při jeho vytváření zohledňovaný požadavek vyváženosti, tím pádem co nejvěrnější reprezentativnosti daného jazyka. Reprezentativní ve smyslu psaných textů míníme reálně poměrnou reflexi všech typů textů a promluv. V případě mluvených textů rozumíme reprezentativností vyváženost v základních sociolingvistických kategoriích mluvčích (pohlaví, věková skupina, výše dosaženého vzdělání a oblast pobytu v dětství).

Nalézt tu můžeme jak texty publicistické (korpus SYN2009PUB, SYN2006PUB), odborné (korpus LINK), beletrii (SYN2000, SYN2005), tak i (jak už jsme zmínili) texty mluvené (např. korpus ORAL2008, PMK).

Všechny synchronní korpusy jsou spojené do nereprezentativního korpusu SYN, čítajícího 1,3 miliardy textových slov. Nereprezentativnost je dána zařazením korpusů SYN2006PUB a SYN2009PUB, jehož důsledkem je převaha publicistických textů.

Současně se synchronními korpusy vznikají i korpusy diachronní (DIAKORP, zahrnující texty celkem ze sedmi století vývoje českého jazyka, a DOTKO, diachronní korpus dolní lužické srbštiny) a paralelní, jejichž cílem je pokrytí co největšího množství jazyků (InterCorp).

O složení reprezentativního korpusu SYN2010 si můžeme udělat představu z následujícího grafu.



40 % **beletrie**
27 % **odborná literatura**
33 % **publicistika**

Graf 1: Složení korpusu SYN2010

Korpus prohledáváme prostřednictvím korpusových manažerů a speciálního dotazovacího programu, jež nám umožní vyhledávat samotné slovní tvary, slovní spojení v kontextu, slova charakterizovaná určitými gramatickými kategoriemi, zjišťovat jejich frekvenci v korpusu i původní textový zdroj. Pro naši práci je výhodou právě to, že v něm lze vyhledávat například i podle slovních druhů.

Pro každý slovní tvar je v korpusu zaznamenána tzv. *morfologická značka*, vyjadřující gramatické kategorie kýženého slovního tvaru, a *lemma*, tedy základní tvar slova.

V korpusu samozřejmě nenalezneme všechny vzniklé texty. I přesto nám korpusy zprostředkovávají zatím nejucelenější informaci o podobě současného užití nejen psaného spisovného jazyka a umožňují nám i snadné zpracování a třídění vybraného materiálu, což je oproti např. datům získaným z internetu nesporná výhoda.

V naší práci se budeme opírat o data získaná z korpusu SYN2010 o velikosti 100 mil. textových slov, který oproti publicistickým korpusům (SYN2006PUB a SYN2009PUB) obsahuje žánrově vyvážený a zároveň nejaktuálnější takto dostupný materiál. Pro jeho výběr hovoří i to, že se jedná o korpus referenční, tzn. že jeho podoba je ustálená (na rozdíl od nereferenčního korpusu SYN).

4. Přístupy ke slovním druhům

Potřeba kategorizovat svět kolem sebe, přiřazovat a zařazovat pojmy, které sdílejí určité vlastnosti, do tříd je běžná praxe stará jako lidstvo samo. Řazení slov do tříd tak patří k nejtradičnějším oblastem zkoumání jazyka.

Chceme-li se zabývat číslovkami jako slovním druhem a jejich vymezením, je nejprve nutné si ujasnit, co se pod pojmem slovní druh a číslovka rozumí, a definovat a shrnout různé přístupy, které se touto oblastí zabývají.

Termínem slovní druh se v české lingvistické tradici rozumí „třída slov, z níž se vychází ve výkladech funkčního tvarosloví“ a „jeho charakter je závislý na typu jazyka a na zvolené gramatické teorii“ (ESČ 2002: 17). Anglosaští lingvisté soudí, že slovní druhy jsou „třídy slov, která jsou v jazyce gramaticky odlišena“ (Crystal 1987: 91), nebo skupiny vytvářené jednotkami se společnými vlastnostmi (Schachter 1985: 2).

Na rozdíly mezi českou a anglosaskou tradicí vymezení slovních druhů se ještě podrobněji podíváme, nejprve se stručně zmíníme o historii jejich vymezení obecně.

4.1 Historie slovnědruhov^é klasifikace

Už staroindičtí gramatikové zhruba v 5. stol. př. n. l. rozlišovali čtyři třídy slov - substantiva, slovesa, předložky a částice. Nezávisle na nich přichází ve 4. stol. př. n. l. se svou klasifikací také řecký filozof Platón, který rozlišuje jméno a sloveso. Jeho žák Aristoteles k těmto dvěma třídám, jež odlišoval podle přítomnosti, resp. nepřítomnosti ideje času, přidal další dvě, spojky a členy.

Základy současného třídění pokládá ve své gramatice řečtiny ve 2. stol. př. n. l. Dionýsios Thrax, když vyděluje osm slovních druhů – jméno, sloveso, participium, zájmeno, předložku, spojku, příslovce a člen. Zhruba ve stejné době rozlišoval v gramatice latiny Marcus Terentius Varro slovní druhy čtyři, a to podle schopnosti vyjadřovat kategorie pádu a času. Toto dělení reviduje na základě gramatiky řečtiny v 5.-6. stol. Priscianus a s cílem zachovat původní Thraxem navržený počet druhů místo řeckého členu vyděluje citoslovce. Tady můžeme sledovat sílu tradiční klasifikace slovních druhů. Nejenže se latinské tradice drží z větší části terminologie slovních druhů, ale v některých jazycích, které na rozdíl od latiny členem disponovaly, se autoři gramatik tímto faktem vůbec nezabývali a po latinském vzoru člen jako samostatnou kategorii nerozlišovali

(Černý 1996: 62).

Na základě logických a sémantických definic vymezují ve 13. stol. tzv. modisté, kteří se snažili vysvětlit gramatická pravidla pomocí filozofických postupů, samostatnou třídu adjektiv.

V rámci starověké logiky se také začala rozlišovat slova, která měla význam sama o sobě (autosémantika), a slova, která význam nabyla až ve spojení se skupinou první (syntaktická).

Další pokus o nový pohled na problematiku slovních druhů zaznamenáváme ve 2. pol. 17. stol., kdy se rozvíjí racionalistické pojetí gramatiky, v tzv. gramatice port-royalské. Její autoři se snaží v souladu s tehdejší filozofií gramatické jevy rozumově zdůvodnit. Klasicky vymezují devět slovních druhů, do nichž zahrnují jak člen, tak citoslovce, a to na základě sémantického kritéria na ty, které se vztahují k objektům myšlení, a ty, jež vztahují k formě nebo způsobu jakým přemýšlíme (sloveso, spojka a citoslovce) (Černý 1996: 69).

Ve 20. století se začínají objevovat pochybnosti a návrhy revize slovnědruhové klasifikace, založené na nespokojenosti s tradičními kritérii.

Ve třetím vydání své knihy *Prinzipien der Sprachgeschichte* germanista Hermann Paul kritizuje nejednotnost třídícího principu tradiční klasifikace. Říká, že „obvyklé dělení částí řeči v indogermánských jazycích, tak, jak jsme ho zdělili od antických gramatiků, nespočívá na důsledně aplikovaných logických principech“ (Paul 1920: §244), ale má svévolný, arbitrární charakter.

Co se významového kritéria týče, korespondují podle Paula především substantiva, adjektiva a slovesa se substancí, vlastností a činnostmi bez problémů. Jisté potíže vidí ale např. u deverbativních substantiv, nebo v deadjektivních slovesech, u kterých se význam a forma dostávají do rozporu. To ho přivádí k tvrzení, že pokud máme potřebu každé slovo umístit do určité kategorie, je jakékoliv třídění založené na přísně logických absolutně neproveditelné (Paul 1920, *ibid.*).

O tom, jestli jsou slovní druhy kategorií univerzálního charakteru, a o nepopiratelnosti jejich existence pochybuje v některých částech svého *Kurzu obecné lingvistiky* Ferdinand de Saussure. Podle něho lingvisté neustále pracují „s pojmy, vytvořenými gramatiky, o nichž nevíme, zda opravdu odpovídají konstitutivním faktorům jazykového systému“ (Saussure 2007: 136), a klade si otázky po oprávněnosti vymezení slovních druhů. Jsou-li vymežovány jen na základě vnějazykového principu, nebo zda se jedná o klasifikaci jazykem podmíněnou, která v něm má své místo a je mu vlastní.

Na příkladu francouzského výrazu „bon marché“ Saussure ukazuje, že určovat toto spojení

jako adjektivum má sice smysl (stojí jako typické adjektivum před substantivem), ale formálně se tak nechová. Skládá se ze dvou slov a tato skutečnost se úplně neslučuje s faktem, že „rozdíl mezi částmi řeči má sloužit k jejich klasifikaci“ (ibid.)¹. Z tohoto hlediska upozorňuje na „chybnou, nebo neúplnou klasifikaci“ a dochází k závěru, že „rozlišování slov na slovní druhy není jazyková realita nepopíratelná“ (ibid.).

4.2 Kritéria klasifikace

4.2.1 Obecně

Výše jsme stručně zmínili vývoj klasifikace slovních druhů. Principy, na jejichž základě bylo třídění prováděno, však byly načrtnuty pouze okrajově, proto se nyní na tyto principy podíváme podrobněji.

I základní kritéria klasifikace jsou naznačena už z období starověku a středověku. Při třídění se v různé míře využívalo kritérium (z dnešního pohledu) sémantické, morfologické a syntaktické. Tedy co daný druh označuje, jaké gramatické kategorie vyjadřuje (ve starověku a středověku to byl většinou čas a pád) a jakou pozici ve větě zaujímá.

V průběhu staletí můžeme pozorovat různé tendence se k různým kritériím stavět různě. Od aristotelské sémantické klasifikace, přes Thraxovu snahu zhodnotit všechna tři kritéria a Varronův ryze formální přístup až k port-royalské gramatice, která se opět opírá především o hledisko sémantické.

S ohledem na tehdejší strukturalistické a formalistické tendence se na začátku 20. stol. k třídění slovních druhů staví netradičně Vigo Bröndal, který zakládá jejich klasifikaci výhradně na kritériích sémantických, přičemž vychází ze základních aristotelských kategorií a aristotelské logiky. Na příklad Komárek (2006) však Bröndalovu snahu hodnotí jako příliš mentalistickou a tvrdí, že toto univerzální třídění neumožňuje „adekvátně postihnout specifické rysy slovně druhové stavby konkrétních jazyků“ (Komárek 2006: 17).

Protože obecně je významová báze empiricky velmi obtížně uchopitelná, v současné době se upřednostňuje klasifikace odvoditelná z formálních vlastností lexikálních jednotek.

Jednotlivá kritéria se teď pokusíme konkretizovat a zhodnotit jejich pozitiva, či negativa v případě, že by jich bylo použito jako jediného třídícího principu.

¹ Saussure zde však pod pojmem slovo rozumí slovo grafické, nikoliv lexém jako abstraktní formálně významovou jednotku.

4.2.2 Kritérium sémantické

Jak už jsme zmínili, vymezení slovních druhů pouze na základě hlediska sémantického je problematické a moderní lingvistika s ním příliš neoperuje. Snahy související s užíváním sémantického kritéria hodnotí negativně např. Crystal (1987: 91), když se zamýšlí nad používanou tradiční definicí substantiv a adjektiv. Podle té je substantivum definováno jako „název něčeho“ (v češtině jako názvy osob, zvířat a věcí atd.). Crystal klade otázku, zda je vhodné určovat slovo „krása“ (a podobně další slova s abstraktním významem) jako substantivum, když krásu nemůžeme označit za „věc“. Stejně tak adjektivum „červený“ může být ale chápáno také jako „název barvy“, nejen jako vlastnost. Namísto těchto vágních definic vyzdvihuje snahu soustředit se na rysy formální, které nám mohou ukázat, jakým způsobem se skupiny slov v jazyce chovají.

V českém prostředí kritizuje klasifikaci založenou pouze na sémantice např. Poldauf (1957) ve svém článku pojednávajícím o vyjadřování kvantity, kde se věnuje především k syntaktickému rozboru, protože „význam sám o sobě nemůže rozhodnout o slovním druhu“ (Poldauf 1957: 77).

Třídí-li se slovní druhy výlučně na základě hlediska sémantického, to znamená na základě jejich lexikálního významu, dochází se nejprve k vyčlenění slov, která mají lexikální význam sama o sobě, na jedné straně a na straně druhé k vydělení slov, která význam lexikální postrádají. Jejich význam je označen jako gramatický. První skupina slov se nazývá autosémantika, druhá potom synsémantika. Autosémantika se dále rozlišuje podle toho, zda označují substanci, vlastnost, okolnost, či vyjadřují děj. Synsémantika se konkrétněji podle lexikálního významu nerozlišuje.

Pro jazyky, které disponují invariantními tvary, je slovnědruhovú klasifikace zakládající se na určování izolovaných tvarů slov z pohledu stávající tradice vymezení slovních druhů neefektivní. Pokud izolovaná slova sdílejí ve většině případů jeden a týž tvar a přitom mohou nabývat různých významů (např. anglické *study* - studium, *study* - studovat), pozbývá snaha o klasifikaci² slov bez přítomnosti kontextu na logice. K takovým jazykům patří např. angličtina. Z tohoto důvodu se v anglosaské tradici využívá především kritérium syntaktické, které dokáže vymežit invarianty na základě jejich syntaktické funkce, a hledisko sémantické je opomíjeno. Specifika takového třídění budou uvedena dále.

² Tj. klasifikaci podle stávajících zvyklostí

4.2.3 Kritérium morfologické

Ve flektivních jazycích se jako funkční ukazuje kritérium morfologické. Samotný tvar a způsob flexe, tedy zda se jedná o deklinaci, nebo konjugaci, resp. absence flexe nám signalizuje pravděpodobné zařazení do třídy slov, jež sdílejí podobné charakteristiky. Podle ne/přítomnosti flexe se slova dělí na ohebná, tzn. schopná deklinace, nebo konjugace, a neohebná, vyznačující se neměnností svých tvarů (to ale neplatí stoprocentně, např. česká stupňovatelná adverbia, tvary *rád*, *-a*, *-o* tradičně zařazované mezi adverbia, nebo časovatelná spojka *aby* mohou být podle tohoto hlediska řazena k ohebným slovním druhům, zatímco některé číslovky k neohebným).

Toto kritérium velmi dobře může sloužit k subklasifikaci v rámci ohebných slovních druhů, na rozlišení invariantních tvarů však nestačí.

4.2.4 Kritérium syntaktické

V případě jazyků neflektivních nabývá z výše uvedených důvodů na relevanci hledisko syntaktické, které představuje funkci, jakou může slovo ve větě zastávat.

Čeština jako jazyk flektivní si může dovolit relativně volný slovosled, protože syntaktické funkce, které dané slovo zastává, jsou většinou vyjádřeny jeho formou, souborem gramatických významů afixů připojených ke slovnímu kořenu.

Jinak je tomu na příklad v analytickém jazyce, jakým je už zmíněná angličtina. V důsledku velmi časté homonymie a invariantních tvarů slov mnohdy můžeme slovo zařadit ke konkrétnímu slovnímu druhu, až když víme, na jaké pozici ve větě je umístěno a jakou funkci zastává, tedy jak se v rámci věty „chová“.

Další z důležitých charakteristik je např. podle Schachtera (Schachter 1985: 1) tzv. slovní distribuce. Termínem distribuce, který v 50. letech 20. stol. ve snaze najít čistě formální metodu popisu jazyka zavedl Z. S. Harris (1951), se označuje souhrn všech možných kontextů, do kterých jazyková jednotka, v našem případě slovo, může vstoupit. Právě na základě distribuce je pak daná jednotka blíže hodnocena.

Problematiku určení slovního druhu v angličtině můžeme ilustrovat na příkladu slova *round*, které může z tradičního pohledu vystupovat ve funkci pěti slovních druhů (Crystal 1987: 90). V závislosti na okolí a na funkci, kterou ve větě zaujímá, jej můžeme chápat jako adjektivum v případě věty (1), ve větě (2) jako prepozici, jako sloveso ve větě (3), ve větě (4) bychom ho určili jako adverbium a ve větě (5) se by se jednalo o substantivum.

- (1) *Mary bought a round table.* - Mary koupila *kulatý* stůl.
- (2) *The car went round the building.* - Auto jelo *kolem* budovy.
- (3) *The yacht will round the buoy soon.* - Ta jachta brzy *obepluje* bóji.
- (4) *We walked round to the shop.* - Šli jsme do obchodu *na rohu*.
- (5) *It's your round.* - Jste *na řadě*.

I v češtině najdeme slova, jejichž slovnědruhovou platnost jsme schopni identifikovat až prostřednictvím kontextu. Tradičním příkladem, který znají z testů už žáci základních škol, je slovo *kolem*. Tato forma může ve větě zastupovat tvar instrumentálu nominativu lexému *kolo* (6), nebo plnit funkci předložky (7), případně adverbia (8).

- (6) Po chvíli došel k autu s *kolem* na střeše.
- (7) Marvinovi bylo *kolem* třicítky.
- (8) A *kolem* je božský klid.

Více slovních druhů může být určováno i u nesklonného *fajn*. V závislosti na kontextu je schopno fungovat jako adjektivum (9), adverbium (10), nebo jako citoslovce (11).

- (9) Je *fajn*, že si mám z čeho vybírat.
- (10) Cítím se *fajn*.
- (11) *Fajn*. Najíme se v půl desáté.

Zároveň však existuje také adjektivum *fajnový*, které se svou formou přibližuje k typickým adjektivům. Případy invariantnosti však nejsou v češtině běžné, proto se k třídění využívá zejména hledisko morfologické (viz 4.3.2).

Z předcházejících příkladů vidíme, že to, jak se konkrétně jednotlivé druhy vydělují, závisí nejen na typu daného jazyka, ale i na zvoleném úhlu pohledu. Různé jazyky proto operují s různými počty i typy slovních tříd.

4.3 Česká tradice slovnědruhov³é klasifikace

V následujícím oddíle se pokusíme shrnout dosavadní přístupy k otázce slovnědruhov³é klasifikace v české lingvistice a ukázat potenciální podobu třídění, kdyby byla konkrétní hlediska použita samostatně.

Česká lingvistika pracuje s tříděním tradičním, založeném na hledisku sémantickém, morfologickém a syntaktickém, lépe řečeno na jejich syntéze. Tato kritéria však nejsou u jednotlivých slovních druhů uplatňována se stejnou relevancí.

U autosémantik, slov plnovýznamových, korespondují většinou všechna tři kritéria, zatímco u specifických druhů, jakými jsou číslovky a zájmena, je upřednostňován rys významový. Synsémantika, jejichž význam se označuje za gramatický, se od ostatních slov z velké části odlišují absencí flexe a mezi sebou se liší svým gramatickým významem a syntaktickou funkcí.

Pokusme se nejdříve podívat, jak by mohla vypadat soustava slovních druhů, kdyby se třídila podle jednotného principu. Klasifikačním kritériem by tedy nebyla kombinace všech tří tradičních kritérií, ale vždy kritérium jedno³. Podle Komárka (2006: 15) je takový pohled sice možný, ale „povrchově popisný, inventarizující a v podstatě destruktivní“.

Příklady, na kterých budeme na následujících stránkách v oddílech věnovaných jednotlivým koncepcím demonstrovat, jsou převzaty z prací představujících tyto koncepce. Ostatní příklady a příklady uvedené v rámci analýzy materiálu, pocházejí z databáze ČNK (korpusu SYN2010).

4.3.1 Kritérium sémantické

Jak už bylo řečeno, kritérium sémantické je považováno za nejednoznačné a nepřesné, pro další postup s klasifikací nedostačující.

V případě, že bychom se omezili pouze na to, co daná slova označují, tedy na lexikální význam, pracovali bychom přibližně s následujícími pěti slovními druhy, spadajícími do kategorie autosémantik, třída šestá by byla tvořena synsémantiky.

3 Budeme při tom vycházet ze stávající tradiční slovnědruhov³é klasifikace.

- substantiva - označující nezávislé entity a vlastnosti a děje pojímané jako nezávislé entity (PMČ)
- adjektiva - označující vlastnosti substancí a některé relace (PMČ)
- číslovky - vyjadřující kvantovost
- příslovce - vyjadřující okolnost, nebo vlastnost děje
- slovesa - vyjadřující stav, nebo událost

Zbytek slovní zásoby by při požadavku využití pouze sémantického kritéria zůstal zařazený v široké třídě synsémantik a tvořil by tak značně heterogenní kategorii šestou.

4.3.2 Kritérium morfologické

Klasifikace na základě tvarů jednotlivých slov má pevnou pozici především v indoevropských flektivních jazycích.

Kdybychom celé třídění zakládali pouze na tvarech slov a jejich variantnosti, v češtině by mohly být rozlišovány následující třídy slov, jejichž hlavním znakem by byla schopnost deklinace, resp. konjugace:

- podstatná jména, rozlišující svými tvary gramatické kategorie čísla a pádu
- adjektiva, která rozlišují číslo, jmenný rod, pád, popř. stupeň vlastností⁴
- zájmena, rozlišující osobu, číslo, pád a jmenný rod⁵
- slovesa, vyjadřující kategorie osoby, čísla, způsobu, času a slovesného rodu
- stupňovatelná adverbia, schopná vyjádřit míru okolnosti

Zbylé lexikální jednotky by spadaly do kategorie slov neohebných. Nevýhodou tohoto třídění je kromě nedostatečné specifikace neohebných slovních druhů i problematika neskloňných jmen, která svými tvary patří mezi slovní druhy neohebné (např. substantivum *pyré*) (viz 4.3.4.1).

4.3.3 Kritérium syntaktické

Podle funkce, kterou mohou slova ve větě zastávat, můžeme slova třídit zhruba do sedmi tříd. Mohli bychom vyčlenit druhy, které jsou z tradičního bohemistického hlediska

⁴ Zde ovšem nutno podotknout, že kategorii stupňovatelnosti vyjadřuje jen určité procento adjektiv.

⁵ Podobně jako u kategorie stupňovatelnosti není ani u všech zájmen vyjádřena kategorie jmenného rodu.

schopny plnit funkci větného členu, a druhy, které mají pouze gramatický význam a jako tradiční větný člen nefungují. To bychom ale směřovali hledisko syntaktické se sémantickým, zůstaneme proto u toho, jakou funkci ve větě lexémy plní obecně.

V případě klasifikace pouze na základě kritéria syntaktického by se rozlišovala

- substantiva, jež fungují jako podmět, nebo předmět, případně rozvíjí další substantivum formou přívlastku neshodného
- adjektiva, která rozvíjí substantiva a zájmena jako přívlastek shodný a jako jmenná část přísudku i slovesa
- slovesa zastávající funkci přísudku (tzn. verba finita)
- příslovce, která mohou shodně i neshodně rozvíjet adjektiva a slovesa
- předložky, upřesňující především povahu vztahu jména a slovesa
- spojky souřadící, sloužící ke spojování vět a jejich částí a vyjadřující vztah koordinace těchto jednotek
- spojky podřadící, které také slouží ke spojování vět a jejich částí, mají však význam sémantické závislosti mezi členy spojení

Jak můžeme vidět, uvedené užití jednotlivých kritérií není dle současného přístupu příliš efektivní. Vždy se vydělí relativně velká a zjevně heterogenní část lexikonu, která by zůstala neklasifikovaná, vyžadující v rámci jednotlivých hledisek další specifikaci. Pro tradiční klasifikaci se ustálila kombinace všech tří kritérií v blíže nespecifikované formě.

4.3.4 Současné přístupy ke slovnědruhové klasifikaci

Jednotlivé jazykové příručky a lingvistické teorie preferují různou hierarchii kritérií a utvářejí tak dva hlavní proudy klasifikace.

Zatímco se výklad v Příruční mluvnici češtiny opírá především o morfologickou charakteristiku slovních druhů (a stejně tak i starší Česká mluvnice Havránka a Jedličky), Mluvnice češtiny II (a z ní čerpající příručka Čeština - řeč a jazyk a Encyklopedický slovník češtiny) staví primárně na rysech syntakticko-funkčních, vyjadřujících, jak se slova podílejí na výstavbě věty a textu. Na přístupy k problému slovnědruhové klasifikaci se podíváme na následujících stránkách.

4.3.4.1 Příruční mluvnice češtiny (dále PMČ, 1996)

Věnujme se nejprve prvnímu ze zmíněných přístupů, „většinově morfologickému“, zastoupenému v PMČ. V rámci tohoto popisu jsou na základě tvarosloví nejprve vymezena slova ohebná, dále členěná na časovaná a skloňovaná, která se ohýbají prostřednictvím souborů deklinačních koncovek, a to buď způsobem typickým pro většinu substantiv (deklinační jmenná), adjektiv (skloňování adjektivní, složené), nebo zájmen. K neohebným, jež schopnost deklinační většinou postrádají a jsou dále dělena podle kritérií syntakticko-funkčních, se řadí v této mluvnici adverbia, předložky, spojky, citoslovce a částice.

To, že morfologické kritérium není možné aplikovat bezvýhradně na celou slovní zásobu, už jsme zmínili. Možnost demonstrace tohoto tvrzení nám nabízejí příklady slov tradičně určovaných jako substantiva typu *hajný, krejčí, telecí* apod., u nichž dominuje flexe adjektivní, jména nesklonná jako *atašé, filé*, nebo na příkladu obecně uznávaných adverbií, považovaných za jinak neohebná, *rád a jaktěživ*, jejichž tvary reflektují kategorii čísla i jmenného rodu. Určitým typem flexe je u adverbií také stupňování.

4.3.4.2 Mluvnice češtiny II (dále MČ II, 1986)

Jak už bylo řečeno, MČ II staví především na vlastnostech syntakticko-funkčních, vyjadřujících, jak se slova podílejí na výstavbě věty a textu. Rozlišuje slovní druhy základní (substantiva, adjektiva, verba a adverbia) a nezákladní. V rámci nezákladních druhů vyděluje skupinu návstavbových slovních druhů, napodobujících svými vlastnostmi slovní druhy základní, kam řadí zájmena a číslovky, a nesamostatné, vyznačující se povahou gramatických morfémů.

4.3.4.3 Mluvnice současné češtiny (dále MSČ, 2011)

Nejaktuálnější z mluvnic, Mluvnice současné češtiny, která má mezi existujícími příručkami specifické postavení⁶, se opírá v první řadě o charakteristiky morfologické a činí tak mnohem důsledněji než na první pohled stejně postupující PMČ. Důsledkem toho je např. to, že výše zmiňované (4.3.4.1) adverbium *rád* řadí na základě vyjadřovaných gramatických kategorií mezi adjektiva a stejně nakládá i se slovy typu *možno, lhostejno* etc., v tradičních přístupech hodnocenými jako modální adverbia.

Z hlediska tématu předkládané práce přináší i netradiční pohled na slovní druh číslovek,

⁶ Tato mluvnice na rozdíl od prací předchozích zpracovává jazykový materiál deskriptivně, staví se tedy k jazyku bez nároků na hodnocení v rámci dichotomie spisovný - nespisovný.

jak se budeme moci přesvědčit v kapitole 5.

4.3.4.4 Funkční generativní popis (FGP)

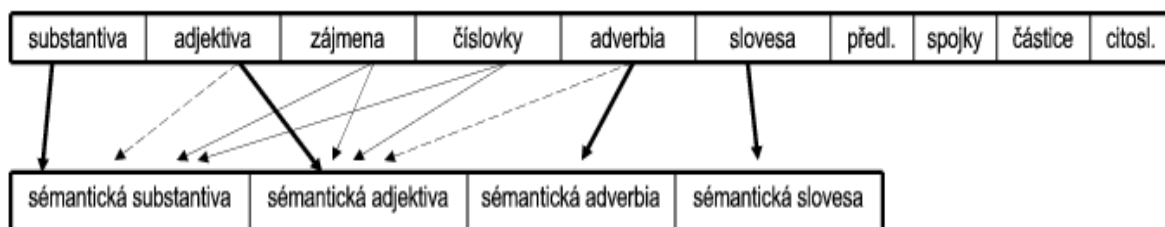
V rámci teorie tzv. Funkčního generativního popisu, který svým názvem odkazuje nejen na tradici Pražského lingvistického kroužku (funkční), ale i na lingvistický přístup N. Chomského (generativní), a anotace, uplatňované při budování Pražského závislostního korpusu (dále PDT), vzniká zcela netradiční slovnědruhovú klasifikace na základě sémanticko-syntaktických charakteristik. Anotace PDT se soustředí na realizaci jazykových jednotek na tzv. tektogramatické rovině, zachycující hloubkovou strukturu.

Podle základních onomaziologických charakteristik (substance, vlastnost, okolnost a událost) specifikovaných vzhledem ke svým syntaktickým funkcím jsou vyčleňovány následující slovní druhy:

- sémantická substantiva - vyjadřující substanci - *otec, kdo, který, tvůj*
- sémantická adjektiva - odpovídají základní onomaziologické kategorii vlastnosti - *příjemnější (pocit), tři (děti), mnoho (lidí)*
- sémantická adverbia - vyjadřující onomaziologickou kategorií okolnosti - *odtud (už to není daleko), (bylo mu) všelijak*
- sémantická slovesa - odpovídající kategorii událost - *přijít*

Zbylé tradiční slovní druhy nejsou na tektogramatické rovině reprezentovány, tudíž nejsou v rámci tohoto třídění rozlišovány.

Vztahy sémantických slovních druhů k druhům tradičním znázorňuje Obrázek 1. Tučnými šipkami jsou vyznačeny „prototypické“ vztahy slovních druhů, tenké šipky ukazují distribuci zájmen a číslovek do tříd sémantických substantiv a adjektiv, přerušované šipky sledují klasifikaci na základě slovtvorných vztahů.



Obr.1: Vztahy sémantických slovních druhů ke slovním druhům tradičním

5. Vymezení slovního druhu číslovek v mluvnicích, jazykových pracích a monografiích

Nejen samotný probíraný slovní druh, ale i stanoviska lingvistů k němu zaujímaná, mají výrazně heterogenní povahu. S obecnými přístupy k otázce slovních druhů jsme se už seznámili, nyní se budeme věnovat tomu, jak vybrané souborné mluvnické práce a některé jazykové příručky, věnující se teoreticky i prakticky třídění slovních druhů, vymezují třídu číslovek, a pokusíme se jednotlivé přístupy shrnout a srovnat. Práce budou řazeny chronologicky, podle data jejich vydání.

5.1 Mluvnice

5.1.1 Česká mluvnice (Havránek a Jedlička 1960, dále H&J)

Nejstarší příručkou, o kterou se v této práci budeme opírat, je Česká mluvnice. Autoři Havránek a Jedlička v ní charakterizují číslovky jako „slova významu číselného“, označující „počet, pořadí apod.“ (H&J 1960: 181). Vymezují je jako druh většinou ohebný a schopný vyjádřit svými tvary kategorii pádu a v některých případech také čísla a rodu. Podle významů mluvnických jsou to „jména podstatná nebo přídavná, nebo někdy i příslovce“, která mají ale společný číselný význam (H&J 1960: 87), jež H&J považují za směrodatný.

Číslovky, které se neshodují ve všech třech klasifikačních kritériích, ale mají s číslovkami společný základ, označují jako číslovky příslovečné (*dvakrát, tolik*), protože „význam číselný je zde rozhodující“ (H&J 1960: 91). Jinak ale ve sporných případech řadí slova „tam, kam náleží podle významů mluvnických“ (ibid.), tedy na základě hlediska morfologického, a potvrzují tak nejednotnost voleného přístupu.

Na základě významu dále rozlišují číslovky **základní**, vyjadřující počet a odpovídající na otázku *kolik (jeden, dvacet, sto)*, **řadové**, označující pořadí a místo v číselné řadě a odpovídající na otázku *kolikátý, který (první, dvacátý)*, **druhové**, které označují množství druhů a na něž se ptáme pomocí otázky *kolikery* a které ve spojení se jmény abstraktními, hromadnými a pomnožnými vyjadřují prostý počet stejně jako číslovky základní (*dvoje dveře, trojí včelstvo*), a nakonec číslovky **násobné**, které sdělují, kolikrát se nějaká věc vyskytuje, kolikrát se znásobila, a reagují na dotaz *kolikanásobný, kolikrát (jednou, dvojnásobný)*.

Všechny výše uvedené druhy vyjadřují počet buď určitý a je možné je zapsat číslicí, v tom případě se jedná o číslovky určité, nebo je pomocí číslic zapsat nelze a pak se jedná o číslovky neurčité (*málo, několikátý, několikanásobný, všechny*).

Zmiňuje se zde také možnost zařazovat mezi číslovky předložkové výrazy, které mají podílný význam. Tyto číslovky jsou v mluvnici nazývány **podílnými** (*po dvou, po několika*).

Za zřejmá substantiva a adjektiva, která mají číselný význam jen zčásti, považují Havránek s Jedličkou výrazy jako *jednotka, dvojka, stovka, jednoduchý, dvojitý* apod.

Jak bylo uvedeno výše, z formálního hlediska popisuje mluvnice číslovky jako slova ohebná, s výjimkou některých číslovek násobných a neurčitých, tedy tzv. číslovky příslovečné (*jednou, dvakrát*).

Následně se Havránek s Jedličkou věnují jejich skloňování a určují jeho pravidla.

5.1.2 Mluvnice češtiny II (tzv. akademická, MČ II, 1986)

Z mluvnických prací se číslovkami nejpodrobněji zabývá druhý díl „akademické“ Mluvnice češtiny. Přístupu uplatňovanému v MČII jsme se obecně věnovali výše, teď se podíváme, jakým způsobem jsou od ostatních slovních druhů odlišovány číslovky.

Připomeňme, že v rámci této mluvnice se vyčleňují slovní druhy základní, tedy ty, které „zobecňují myšlenkové činnosti“, substantiva, adjektiva, verba, adverbia (MČII 1986: 17), nezákladní, to jsou ty, které „svými funkčními vlastnostmi napodobují soustavu slovních druhů základních (neslovesných)“ (MČII 1986: 22) a slovní druhy nesamostatné, tzn. volné morfémové útvary (předložky), slova s významem gramatickým (spojky), komunikační formativy (částice) a specifický druh citoslovcí.

Slova, která jsou charakteristická jednotou významu a funkce, tvoří jádro základních skupin. Tzv. jadernými slovními druhy jsou tak nazývána substantiva ve funkci subjektu nebo objektu, adjektiva v pozici shodného přívlastku, slovesa zastávající funkci slovesného přísudku a adverbia v pozici příslovečného určení. Tyto funkce se nazývají primární.

Rozvinutá soustava základních slovních druhů se určuje na základě významové báze a tzv. funkčního příznaku, který je „reprezentován příslušnou funkcí splňující požadavek jednoty významu a funkce“ (MČII 1986: 17). Pokud funkce tento požadavek nesplňuje, je označována jako sekundární (např. substantivum ve funkci neshodného přívlastku - *hlas kosa*).

Číslovky jsou řazeny mezi druhy nástavbové, jejichž tendence napodobovat soustavu základních slovních druhů je velmi silná (MČII 1986: 22). Stejně jako v České mluvnici jsou zde hodnoceny na sémantickém základě kvantovosti, která nabývá forem substantiv (*sto, milion, pět*), adjektiv (*druhý, dva*) a příslovci (*natřikrát*).

Primární větnou funkcí číslovek může být subjekt/objekt (12), přívlastek shodný (13), nebo příslovečné určení (14).

(12) Závodů se účastní asi *sto/osm* družstev. Koupil *stovku* odznaků.

(13) Víáme *miliontého* návštěvníka.

(14) Nemusíš to říkat *dvakrát*.

Mohou ale taky plnit funkce sekundární, např. funkce neshodného atributu (15), predikativu (16), subjektu/objektu (17), adverbialie (18), tradičního doplňku (19).

(15) Hájíme zájmy *milionů* (lidí).

(16) Jedna a jedna jsou *dvě*.

(17) Každý *sedmý* vyhrává.

(18) Pouštěli tam *po deseti*.

(19) Skončil (jako) *šestý*.

Do slovního druhu číslovek zahrnuje tato mluvnice i výrazy, které mají zároveň rys deiktičnosti a které se posléze formálně substantivizují, adjektivizují, nebo adverbializují (*tolik, kolikátý, několik*).

Podle toho, jakým způsobem vyjadřují počet, dělí číslovky na skupinu těch, které vyjadřují pouze počet a ty následně na určité (počet je vyjádřitelný číslem) a neurčité (počet není vyjádřitelný číslem). Další skupinou jsou číslovky, které vedle počtu vyjadřují i jiné významy, číslovky, jež vyjadřují významy založené na významu počtu a nakonec číslovky od významu počtu abstrahující (pojmenování číslic).

Konkrétněji tedy „akademická“ mluvnice rozlišuje následující druhy číslovek:

- **základní** - označující počet jednotlivin (*jeden, tři, tucet, pět tisíc*)
- **numerické** - vyjadřují počet a pojmenovávají číslo (*nula, jedna, tři sta, jednotky, desítky*)
- **úhrnné** - vyjadřují počet jevů pojatý úhrnem (*dvé, čtvero, patero, několikero*)
- **souborové** - udávající počet souborů (*jedny, dvoje, několikery*)
- **druhov** - vyjadřují počet druhů (*dvoji, tisícery, několikery*)
- **velikostní** - udávají velikost, rozsah výši počítaného jevu (*nulový, tisícový, dvoumilionový, několikamilionový*)
- **násobné** - vyjadřují frekvenci realizace děje jako jeho příznak formou adverbia i adjektiva (*dvakrát, dvojnásobný, dvojitý*)
- **dílové** - pojmenovávají jeden z dílů, na které je celek teoreticky rovnoměrně rozdělen (*půl, čtvrt, třetina, dvousetina*)
- **skupinové** - pojmenovávají skupinu jednotek podle počtu, který ji tvoří (*trojice, pár, tisícovka*)
- **řadové** - vyjadřují pořadí, místo jednotek v řadě (*nultý, první, několikátý, x-tý, za třetí*)

Z předcházejícího přehledu vidíme, že druhy číslovek jsou v MČII nejdůkladněji sémanticky rozlišeny.

Z hlediska stránky morfologické je podle MČII většina číslovek charakterizována vlastnostmi typickými pro substantiva, adjektiva a adverbia, pouze malá skupinka se vyčleňuje charakterem pronominálním. Gramatické kategorie rodu, čísla a pádu nejsou vyjadřovány všemi číslovkami rovnoměrně, některé z nich je dokonce nevyjadřují vůbec. Patří tedy částečně mezi slova ohebná a částečně mezi slova neohebná.

Syntakticky tato třída plní vlastnosti slovního druhu základního, který napodobuje.

V kapitole o číslovkách se také můžeme dočíst o alternativních, neslovnědruhových způsobech vyjadřování kvantity. Patří k nim gramatická kategorie čísla, možnosti stupňování u adjektiv a adverbíí a jiné nečíslovkové slovnědruhové prostředky.

I tato mluvnice, přestože na začátku deklaruje přístup funkčně-syntaktický a tvrdí, že

„obecný význam slovních druhů není možno zjistit zkoumáním vlastností izolovaných tříd slov, ale pouze se zřetelem k jejich funkcím“ (MČII 1986: 16), a věnuje se i primárním syntaktickým funkcím, které číslovky mohou zastávat, vyčleňuje a dále klasifikuje tento druh na základě neurčitých významů sémantických.

5.1.3 Příruční mluvnice češtiny (dále PMČ, 1995)

O Příruční mluvnici češtiny jsme už řekli, že se opírá primárně o kritérium morfologické. Druhy slov rozlišuje podle schopnosti/neschopnosti flexe na ohebné a neohebné, podle lexikálního, resp. gramatického významu dále na autosémantika a synsémantika, zbylé slovní druhy odlišuje jejich gramatická (spojky a předložky), nebo komunikační funkce (částice, citoslovce).

Námi zkoumaný slovní druh charakterizuje jako „slovní druh vyjadřující kvantovost, a to buď počítanou, určitou, vyjádřitelnou čísly, nebo nepočítanou, neurčitou, čísly nevyjádřitelnou“ (PMČ 1995: 301). Zmiňuje také jejich heterogenní slovnědruhovou povahu a označuje je za „substitut, který vyjadřuje kvantum jako substanci (*pět, pětka, pětice*), jako vlastnost (*pátý, pětinasobný, několikátý*) i jako okolnost (*pětkrát*)“ (ibid.).

Výrazy *mnoho, málo, několik, fůra, trochu* klasifikuje jako číslovky neurčité.

PMČ rozlišuje číslovky:

- **základní**, ty označují pouhý počet a pojmenovávají čísla (*jedna, jedenáct*), reflektují povahu jména počítaného předmětu a na jejím základě se člení do následujících skupin, které
 - a) se počítají pomocí **vzestupné řady čísel** (*jeden žák, dva žáci, sto žáků*)
 - b) vyjadřují počet **souborů** (*jedny, dvoje, patero*)
 - c) označují **konečnost počtu** (*oba, dvoje, troje, čtvero*)
 - d) vyjadřují **velikost dílu celku, zlomky** (*polovina, třetina, desetina*)
 - e) tvoří tvary smíšené z **názvů celků a jejich dílů** (*jeden a půl, půldruhého*)
- **řadové**, vyjadřující umístění v pořadí jednotek téhož druhu (*nultý, první, stý, několikátý, poslední*)
- **druhov**é, jež označují počet druhů (*dvojí, trojí, několikery*)

- **násobné**, vyjadřující počet opakování děje (*dvakrát*), násobenost substance (*dvojitý*), srovnání (*dvakrát tak...*), ale i míru (*dvojnásobně*)

5.1.4 Čeština - řeč a jazyk (dále ČŘJ, 1996)

Tato jazyková příručka spíše než obecně teoretické přístupy k vymezení třídy číslovek rozpracovává její tvaroslovnou, morfologickou stránku a na jejím základě jednotlivé podtřídy blíže specifikuje. K vymezení slovnědruhové povahy číslovek říká pouze, že jsou to „slova s číselným významem“ (ČŘJ 1996: 199).

Rozlišuje číslovky určité a neurčité a obojí potom na základě toho, co vyjadřují, podobně jako MČII, a to **základní** (*dvě, pět, několik*), **úhrnné** (*dvé, patero*), **souborové** (*dvoje, paterý, několikeré*), **druhové** (*dvojí, paterý, několikerý*), **násobné** (*jednou, dvakrát, několikrát*), **dílové** (*půl, polovina*), **skupinové** (*dvojice, pětice*) a **řadové** (*první, druhý, stý, několikátý*).

Po tomto relativně stručně zdůvodněném rozlišení se věnuje jejich tvaroslovnému systému, který se opírá o tvarové soustavy základních slovních druhů, tedy o tvary substantiv, adjektiv, zájmen i příslovcí, a nerovnoměrně reflektuje gramatické prostředky jmenného rodu, čísla a pádu.

Zabývá se číslovkami s výhradně **adjektivním skloňováním** (řadové, druhové, souborové a většina násobných)⁷ i s deklinací **kombinovanou** z tvarů krátkých a dlouhých adjektiv (některé souborové - dvoje, čtvero), dále tvarům číslovky *jeden, -a, -o*, které vykazují **deklinaci zájmennou** a v neposlední řadě číslovkám, které jsou „vlastně podstatnými jmény číselného významu“ (ČŘJ 1996: 200), a jejich deklinace je tím pádem **substantivní**. Jednotlivé „substantivní“ číslovky přiřazují k tradičním substantivním vzorům. Podle vzoru *hrad* se skloňují „podstatná jména“ (nikoliv „číslivky“, *ibid.*: 200) *milion, bilion*; ke vzoru *stroj* náleží výraz *tisíc*; tvary *miliarda, polovina, čtvrt* patří ke vzorům ženských substantiv; o číslovce *sto* mluvnické říká, že se skloňuje jednak podle vzoru *město*, jednak se užívá i v nesklonném tvaru.

Číslovky, které se odlišují od zmíněných pravidelných způsobů deklinace (např. *dva, obě, tři, čtyři, číslovky zakončené na -t, několik, mnoho*), se v této práci řadí do skupiny se „zvláštním skloňováním“ a uvádí jejich kodifikované tvary (*„málo lidí, z mála lidí, ale*

⁷ Autorka zde má na mysli zřejmě výrazy typu dvojnásobný, nikoliv výrazy typu dvakrát.

jen k málo penězům“; ibid.: 200). Stručně se věnuje i tvarům kombinovaných číslovkových výrazů.

I tato práce, ačkoliv se v ní objevuje tendence klasifikace číslovek podle formálního, morfologického kritéria, ve výsledku preferuje primárně hledisko sémantické.

5.1.5 Mluvnice současné češtiny (MSČ 2011)

Nejnovější z českých mluvnic, Mluvnice současné češtiny, která vychází z deskriptivního přístupu ke korpusovým datům a pracuje nejen s texty psanými, ale i mluvenými, preferuje při slovnědruhovém klasifikaci především kritérium morfologické. Při charakteristice číslovek reflektuje tradiční přístup, definuje je tedy jako slova „sloužící k vyjádření množství“ (MSČ 2011: 226) a stejně jako předchozí práce připomíná jejich formální heterogenost a specifické vymezování na základě společného významu kvantovosti. Protože se však jejich význam a funkce přibližují přídavným jménům, zájmenům a příslovcím (viz např. 5.1.2), vyslovuje možnost oprávněné pochybnosti o samostatné slovnědruhovém platnosti této skupiny.

Na rozdíl od předchozích prací tato nezachovává tradiční řazení přibližných vyjádření kvantity mezi číslovky neurčité (*mnoho, málo, hodně*). Takové třídění se v této mluvnici považuje za značně vágní, proto tyto výrazy zařazuje podle jejich formy k adjektivům, nebo adverbům, a v rámci těchto slovních druhů je také podrobněji popisuje.

Posléze ale dělí MSČ číslovky víceméně tradičně podle významu a formy na:

- **základní**, k nimž řadí slova vyjadřující prostý počet a bez dalších významů pojmenovávající čísla

- významem i formou se nejvíce přibližující zájmenům (20) a podstatným jménům (21)

(20) *dvě* dcery ~ ty dcery

(21) *pětice* lidí ~ skupina lidí

- jako jejich zvláštní podskupinu jsou vyčleněny číslovky **dílové**, které se používají při pojmenování podílu a popisu zlomků (*třetina*)

- **řadové** - vyjadřují pořadí počítaného předmětu v určité posloupnosti

- významem i formou se přibližují adjektivům (22)

(22) *první* liga ~ nejvyšší liga

- **násobné** - udávají, kolikrát, nebo pokolikáté se daný děj opakuje (Letadlo *dvakrát* přeletělo nad kulometným hnízdem.), poměr mezi vlastnostmi pomocí násobku (Hromádka jahod by na Měsíci vážila *šestkrát* méně než na Zemi.), nebo násobenost počítaného předmětu (Pracuje na *dvójitý* úvazek.)

- významově a formálně se nejvíc přibližují příslovcím (23)

(23) *tříkrát* měř ~ přesně měř

- **druhové** - označují počet souborů, sad nebo druhů počítaného předmětu
- **ostatní** výrazy s kvantifikačním významem - do této skupiny jsou zahrnuta
 - a) „podstatná jména číslovkového původu“ - pojmenování číslic samotných (*jednička, dvojka*) a některých řádů (*půlka, stovka, tisíce*) se substantivními vlastnostmi a pojmenování skupin podle členů (*dvojice, pětice*)
 - b) číslovky, jež se objevují jako součást složenin a jejichž funkce a způsob deklinace jsou dány slovem, s nímž se spojují (*dvoukolový, pětihektarový*)⁸
 - c) adjektiva i slovesa s číselným významem vzniklá slovně druhovým přehodnocením (*milionový, zdvojit*)

Tolik se v MSC dozvíme o teoretickém přístupu k suspektnímu slovnímu druhu a jeho klasifikaci. V mluvnicí dále najdeme informace o frekvenci jednotlivých tvarů v psané i mluvené formě tak, jak se vyskytují v korpusech.

Přestože se MSC staví k číslovkám víceméně tradičně a stejně jako předchozí práce je vyčleňuje a činí tak na základě sémantického hlediska, přináší na ně zcela nový pohled. Ve shodě s její celkovou koncepcí se snaží přiblížit klasifikaci formálním kritériím, resp. kritériu morfologickému, a vyslovuje potenciální možnost řadit číslovky ke slovním druhům, k nimž z tvarového hlediska náleží.

⁸ Pokud tato spojení MSC považuje za číslovky, je podle nás na místě vznést dotaz např. na slovnědruhovou platnost výrazu „tříkolka“.

5.2 Slovníky

5.2.1 Encyklopedický slovník češtiny (dále ESČ, 2002)

Jediná jazyková publikace slovníkového charakteru, se kterou pracujeme, je Encyklopedický slovník češtiny. Podle anotace se snaží shrnout a vyložit jednak otázky struktury, ale i různé teoretické koncepce.

Číslovky definuje jako „ohebný slovní druh vyjadřující kvantitu“ (ESČ 2002: 96), aby dále poukázal na to, že se jedná o druh částečně nesklonný (číslovky násobné). Kvantitu však podle autora hesla mohou vyjadřovat i substantiva (*desetina, stovka*), adjektiva (*poslední*) a adverbia (*moc, trochu*) i zájmena (*tolik, kolik*). Po slovotvorné stránce je zmíněna specifická vlastnost číslovek, kterou je označování vyšší číselné hodnoty slovními spojeními (tzv. kombinované číslovkové výrazy), syntakticky jsou schopny plnit funkce slovních druhů, se kterými se formálně shodují, nebo které zastupují. Ohledně morfologických charakteristik se ESČ krátce věnuje zvláštnostem specifickým číslovek ve vyjadřování kategorií jmenného rodu, čísla a pádu.

Určení druhů číslovek se v ESČ stejně jako v předchozích mluvnických pracích zakládá na sémantické klasifikaci podle toho, jaký počet (počet čeho) označují. V rámci všech takto vymezených druhů číslovek rozlišuje protiklad určitost - neurčitost.

V ESČ jsou rozlišovány číslovky:

- **základní**, které vyjadřují prostý počet substancí (*tři muži, pět stolů*), jejich jádro je tvořeno pojmenováním řady přirozených čísel (*jeden, dva, milion*) a patří mezi ně i starší výrazy pro početní míry (*tucet, kopa*) a tzv. číslovky dílové, vyjadřující, o kolikátou část celku jde (*pětina, desetina*)⁹
- **řadové**, označující umístění v pořadí jevů téhož druhu (*první, druhý, dvacátý první*), řadí mezi ně i tzv. číslovky adverbiální (*Říkal mi to poněkolikáté.*) a výčtové (*za prvé, za x-té*)
- **úhrnné**, vyjadřující počet jevů jako číslovky základní, ale pojatý úhrnem (*tré, čtvero, jedenáctero*)
- **druhové**, které udávají počet druhů jevů označeného substantiva (*dvojí, trojí*)
- **souborové**, jež udávají počet souborů (*patery, dvanáctery*) a ve spojení se jmény

⁹ Výše ale zmiňované jako substantiva číslovkového původu (ESČ 2002: 96).

pomnožnými mají význam číslovek základních (*troje kalhoty*)

- **násobné**, vyjadřující četnost, frekvenci realizací děje, primárně adverbiálně (*dvakrát*), sekundárně adjektivně (*dvojnásobný*)

5.3 Monografie

5.3.1 Příspěvky k české morfologii (Komárek 2006)

Ve své morfologickým otázkám věnované monografii, jejíž pojetí bylo částečně použito při tvorbě MČII, Komárek zařazuje číslovky k nastavbovým slovním druhům a specifikuje je jako lexikální kvantifikátory, prostředky sloužící ke „kvantifikaci propozice“.

Pod pojmem kvantifikace rozumí Komárek „takovou modifikaci propozice, aby propozice mohla být referenčně vztažena k denotátu, jinak řečeno, aby byl ve výpovědi vymezen její význam extenzionální, tj. množství a identita denotátů označovaných v daném případě“ (Komárek 2006: 54). Jedná-li se o kvantifikaci, která se vztahuje k názvům objektů, kvantifikační funkci může plnit prostředek gramatický, tj. kategorie čísla (*pes štěkal - psi štěkali*), lexikální prostředky substantiv a adjektiv (*některý, část, většina*) a speciální kvantifikátory (*jeden, dva, několik*). V případě kvantifikace propozice se kvantifikace může realizovat i determinací predikátu (*většina města byla zničena, často kašlal*). Takovou kvantifikaci podle Komárka vyjadřují různá cirkumstantiva.

Číslovky jako druh lexikálních kvantifikátorů specifikuje jako prostředky, „pro které je kvantifikace funkcí typickou, základní“ (Komárek 2006: 55). Tuto funkci je možno bez potíží přiřadit číslovkám základním, číslovky řadové mají z jeho pohledu vztah ke kvantifikaci pouze nepřímý, zprostředkovaný.

V soustavě českých číslovek rozlišuje Komárek následující druhy:

- **základní**, „vyjadřující prvky specifického systému číselných významů zahrnutých pod pojmem přirozených čísel“ a neadekvátněji zaznamenané v písemných číselných znacích (ibid.: 56)

- různorodost základních číslovek vysvětluje specifickými potřebami desítkové soustavy, kde se vedle nižších, nepočítatelných jednotek (*jeden, pět, dvacet*), objevují i názvy označující větší číselné množství, které počítatelné jsou (*sto, milion, miliarda*) a mají proto většinou substantivní povahu

- základní číslovky dále dělí na

a) substantivní - výrazům jako *dvojice, trojice* přisuzuje v případě, že označují počet, plně substantivní povahu (*dvojice lidí*)

- spolu s výrazy označujícími počet a tvořícími pozůstatek starších způsobů počítání (*pár, tucet, mandel, kopa*) už podle něho netvoří součást číselné řady, ale považuje je za výrazy vyjadřující počet jako skupinu obsahující příslušný počet jednotlivců

b) adjektivní - tato povaha je zřetelná u číslovek 1- 4 (*jeden chlapec, čtyři města*) a pro výrazy jako *četný, mnohý, nesčíslný*, které plynule přecházejí k neurčitým zájmenům a cirkumstantivům

- podle Komárka se nesmíme nechat zmást tvary číslovek 1 - 4, jež jsou zčásti flexe zájmené a zčásti substantivní, neboť rozhodující je v tomto případě právě jejich primární syntaktická funkce shodného přívlastku, tedy funkce typicky zastávaná adjektivy

c) číslovky typu pět - tyto výrazy se vyznačují jednak specifickými rysy syntaktickými (ve funkci nominativu a akuzativu se chovají jako substantiva, v ostatních funkcích je pád počítaného předmětu dán zapojením do věty), jednak tvarovými, které korespondují se zvláštnostmi syntaktickými (zvláštní tvar pro funkci nominativ - akuzativ a jiný tvar pro pády ostatní)

- navzdory tvarovým odlišnostem se v obou tvarech shodují ve funkci přívlastku (přívlastku shodného ve spojení *tři chlapci*, přívlastku neshodného ve spojení *pět vrabců*)

- právě syntakticko-morfologické chování číslovek, tzn. to, že mají adjektivní charakter jak ve funkci nominativ - akuzativ, tak i v pádech ostatních, je podle něj natolik specifické, že ho považuje za konstitutivní jádro tohoto slovního druhu (Komárek 2006: 60)

• **druhové**, které tvoří téměř paralelní řadu (systémovou paralelu postrádají výrazy typu *šest set, deset tisíc* apod.) s číslovkami základními (*dvoje, troji*) a vyjadřují počet druhů

- výrazy typu *dvojaký, obojetný*, vyjadřující kolísavost příznaku, považuje Komárek za plně adjektivní

- **skupinové** typu *čtvero* jsou na rozdíl od stejně tvořených druhových číslovek povahy substantivní a udávají „počet jednočlenných podtříd“

- **násobné**, které vyjadřují „počet opakujících se dějů nebo stavů modelovaných identickou propozicí“ (ibid.: 67), jsou svou formou buď adjektivy (*jednonásobný*), nebo cirkumstantivy (*jednou*) a Komárek zmiňuje možnost chápat je jako obdoby číslovek základních

- **řadové** číslovky, které mohou být také adjektivní (*první*), nebo cirkumstantivní (*poprvé*), a které udávají „místo v homogenní sérii, jejíž členy mohou být počítány“ (ibid.: 68), je možné určovat spíše jako prostředek identifikační a jejich povaze by podle autora lépe odpovídalo označení „řadové identifikátory“ (ibid.)

- na závěr vyslovuje myšlenku, že pro tzv. **podílné** číslovky, které můžeme nalézt v jiných jazycích (např. lat. *bini, terni*), nemá čeština odpovídající lexikální prostředky a podílnost vyjadřuje pomocí předložky *po* a základní číslovkou (*po jednom, po deseti*)

Ve srovnání s bohemistickými přístupy k definicím číslovek z hlediska významového se Komárek pokouší o formalizaci svého přístupu a pojmenovává jejich konstitutivní rys, který je formován specifickým morfologicko-syntaktickým charakterem (viz 5.3.1c). Tento rys je však příznačný jen pro číslovky „typu pět“, další druhy vymezuje tradičně, tedy pouze na základě významu.

Jak jsme se mohli přesvědčit, téměř všechny gramatické koncepce určují zkoumaný slovní druh primárně na základě kritéria sémantického. To, že se jedná o skupinu značně heterogenní, jsme mohli vidět na přehledu její klasifikace (např. výraz *trochu* považuje PMČ za neurčitou číslovku, zatímco ESČ za adverbium). Za nejproblematictější ovšem nepovažujeme rozdíly mezi jednotlivými koncepcemi, ale nejednotnost vymezení některých číslovek, ke kterým dochází v rámci jedné koncepce (například výraz *desetina* je v ESČ jednou charakterizován jako substantivum číslovkového původu, podruhé jako číslovka základní, blíže specifikovaná příslušností mezi tzv. číslovky dílové (ESČ 2002: 96,98)).

6. Analýza materiálu

Předchozí kapitola nám představila různé pohledy na klasifikaci číslovek jako samostatného slovního druhu. Ve většině z nich jsou číslovky neoddiskutovatelnou slovnědruhovou kategorií, přesto mezi nimi i v rámci jich samých existují rozpory ohledně určování konkrétních a podle nás neperiferních jednotek. Máme tu tedy co do činění s převážně subjektivními způsoby klasifikace, což je zřejmě způsobeno absencí jednotného formálního a v rámci konkrétních koncepcí jednotně aplikovaného principu. V této práci se zaměříme jednak na ověření kritérií stávajících, jednak na ověření kritéria, jež by nám umožnilo odhlédnout od subjektivních pohledů na problematiku. Náš postup se pokusí zohlednit inherentní vlastnosti lexikálních jednotek, kontext, do něhož mohou vstupovat.

Jak už bylo výše zmíněno, pracovat budeme s materiálem získaným z korpusu SYN2010, který je korpusem nejaktuálnějším a zároveň nepodléhajícím již žádným změnám. Musíme si však uvědomit, že ačkoliv budeme pracovat s metodou ryze formální založenou na větším množství dat, než které měli lingvisté dosud k dispozici, stále zohledňujeme vzorek (pouze psaného) jazyka, z něhož pro účely práce vybíráme opět jen část, a to omezeného rozsahu. Naše zjištění tedy nemohou být absolutní a neaspírají na vyřešení otázky slovnědruhové platnosti zkoumané třídy slov. S jejich pomocí se pokusíme přispět k této složité problematice a ukázat možný směr dalších výzkumů.

6.1 Metoda - kontextová analýza

Nejprve si pojdme vysvětlit, co přesně rozumíme termínem kontextová analýza, a zdůvodnit, proč jsme se pro její aplikaci rozhodli.

Každá jazyková jednotka se zapojuje do vyšších, nejen textových celků, v rámci kterých funguje a které tvoří její okolí. Taková okolí mohou mít u různých jazykových jednotek různý charakter. Pozornost může být věnována okolí textovému, fonetickému, situačnímu etc. Vzhledem k charakteru námi zkoumaného materiálu a požadovanému výstupu se soustředíme na kontext textový. To znamená, že se budeme zabývat slovním okolím, v němž se zkoumané lexémy objevují.

Na základě tzv. distribuce (distribuce ve významu souhrnu všech možných kontextů, do nichž jazyková jednotka vstupuje) popisoval jazykové jednotky už výše zmíněný (4.2.4) Z. S. Harris (1951). Harris však nezakládal svoje postupy na zjišťování kvantitativních vlastností kontextu, jako se o to bude snažit předkládaná práce.

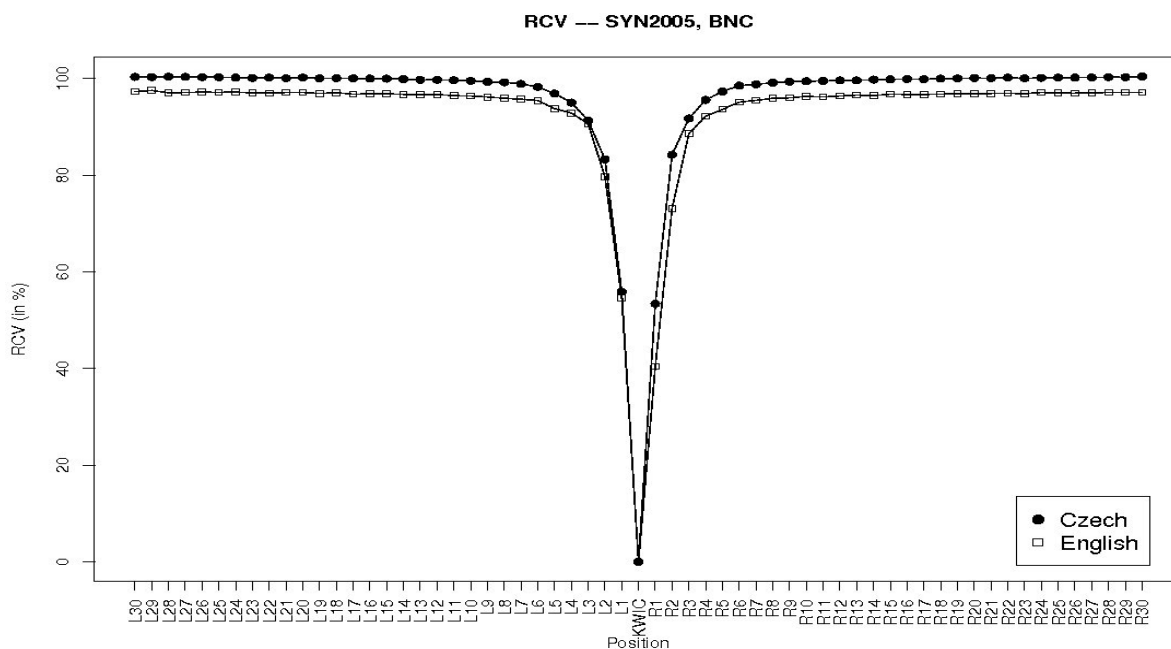
Metoda, která pracuje s textovým okolím jazykových jednotek a kterou zde aplikujeme, je nazývána kontextovou analýzou (Cvrček 2011). Její východiska primárně zakládají na axiomu, vysloveném J.R. Firthem, předchůdcem korpusové lingvistiky, ve znění: „Slovo poznáš podle toho, s čím se spojuje.“ (Palmer 1968: 175 - 176).

Užití této metody tedy předpokládá, že všechna výše zmíněná kritéria slovnědruhé klasifikace se projeví na okolí slova. Podobu kontextu zkoumaného slova tudíž můžeme chápat jako jeho formativní charakteristiku. Rozhodneme-li se pracovat s touto metodou, přijímáme tvrzení, že kontext, do kterého jednotka vstupuje, je utvářen inherentními vlastnostmi jazykové jednotky, jež určitým způsobem reflektuje, zároveň však sám tyto charakteristiky determinuje.

Ve svém článku věnovaném kontextové analýze Cvrček (2011) rozlišuje dva základní typy jazykových kontextů. Všechna okolí na vertikální ose (představíme-li si např. všechny vypsane věty, v nichž se slovo může objevit, pod sebou), hypotetický soubor všech okolí ze všech textů, do nichž jednotka vstupuje (tzn. všechna slova na pravostranných i levostranných pozicích různě vzdálených od slova klíčového ve všech jeho výskytech), nazývá kontextem úhrnným. Na ose horizontální rozlišuje kontext vzdálený, blízký a bezprostřední. Kontextem vzdáleným rozumí všechny jednotky (nejen grafická slova, ale také např. interpunkci) vyskytující se čtyři a více pozic od klíčového slova. Termínem blízký kontext označuje všechny jednotky vzdálené maximálně tři pozice od slova, jemuž je věnována pozornost. V rámci blízkého kontextu vyčleňuje ještě kontext bezprostřední. Tím jsou míněny jednotky, jež slovo bezprostředně předcházejí a následují (tzn. první pozice vlevo a první pozice vpravo).

Předpokládáme, že analogicky jako je konkrétní realizace konsonantů, které mohou podléhat asimilaci znělosti, ovlivněna jejich bezprostředním hláskovým okolím, nikoliv okolím vzdáleným, reflektuje lexikální a gramatické charakteristiky lexému právě také jeho nejbližší okolí. Kontext vzdálený potom poukazuje spíše na charakteristiky stylové a textotvorné (Cvrček, připravuje se). Zároveň se podle výsledků zhodnocení dat z korpusů SYN2005 a British National Corpus (Cvrček, připravuje se) směrem ke klíčovému slovu snižuje variabilita kontextu. To znamená, že čím bližší jednotky budeme zkoumat, tím specifitější informaci o slově získáme. Variabilita kontextu ovšem neroste jenom se vzdáleností zkoumaných okolních jednotek, ale i s frekvencí výskytu klíčového slova. Čím je celková frekvence klíčového slova vyšší, tím vyšší je i variabilita jeho kontextu, tzn. tím vyšší počet kontextových doplnění na různých pozicích můžeme najít. Hodnota variability ovšem neroste do nekonečna (viz. Graf 2). Tuto práci budeme zakládat na analýze co

nejbližšího kontextu zkoumaných jazykových jednotek, aby informace z této analýzy byly co nejkonkrétnější. Na druhou stranu však budeme dbát na to, aby se jejich vypovídací hodnota opírala o co možná největší množství dat, tzn. abychom zjistili a porovnali co největší možné množství takových konkrétních, informací poskytujících kontextů.



Graf 2: Průměrná relativní variabilita slov v češtině a angličtině

Graf 2 znázorňuje relativní variabilitu slov v češtině a angličtině vzhledem ke vzdálenosti okolních jednotek pomocí tzv. *absolutní variability kontextu* (hodnota RCV na ose x). Bližší vysvětlení tohoto ukazatele najdeme ve Cvrčkové připravované práci. Absolutní variabilita kontextu umožňuje mezi sebou v průměru porovnávat různě veliké a různě frekventované skupiny slov. Pro nás je důležité, že z grafu je evidentní klesavá tendence variability kontextu směrem ke klíčovému slovu (Cvrček, připravuje se). Proto budeme náš pokus zakládat na analýze co nejbližšího kontextu zkoumaných jednotek.

To, že pod pojmem bezprostřední kontext slova chápeme ty jednotky, které klíčové slovo jednak předcházejí, tzn. okolí levostranné, jednak následují, tzn. kontext pravostranný, a jeho relevanci v rámci kontextové analýzy slova, jsme si vysvětlili. Nyní se pokusíme naznačit a vyložit bližší určení práce s takto získanými soubory.

Zcela logicky se nám nabízí dvě možnosti výzkumu bezprostředního okolí slova. Jednou z nich je studium vlastností kontextu oboustranného, druhou je analýza každé strany kontextu zvlášť. Podle Cvrčka (připravuje se) se pokusy na specifickém „oboustranném kontextu přibližují testům vzájemné substituovatelnosti dvou jednotek

ve všech jejich kontextech“, což předpokládá existenci stoprocentních substitutů (které jsou v jazyce jevy více než sporadickými) a výzkum jejich shod nepřináší kromě zjištění, že se jedná o stoprocentní substituty, o takových jednotkách další nové poznatky. Dále také ve své práci ukazuje i to, že každá strana kontextu vykazuje jiné preference. To znamená, že analýza jednostranných kontextů zvláště může o dané jednotce přinést ještě konkrétnější informace. Za optimální variantu kontextové analýzy tedy shodně s ním považujeme výzkum jednostranného kontextu s rozlišováním pořadí.

Pod pojmem kontext rozumíme v této práci bezprostřední okolí slova, tedy jednotky, které jej předchází (první pozici vlevo, L1), a jednotky, které jej následují (první pozice vpravo, P1). Vždy bereme v úvahu kontext jednostranný (buď L1, nebo P1, nikdy oba zároveň).

Na rozdíl od analýz slovních druhů, které jsou představeny ve výše zmíněných pracích (Cvrček 2011, Cvrček připravuje se)¹⁰, nebudeme v této analýze zohledňovat pozice, následující po těch, jež jsou obsazeny interpunkčními znaménky (tzn. např. po tečce následující pozice P2).

Vraťme se ale ke zdůvodnění výběru kontextové analýzy a vyslovení hypotézy. Kontextová analýza nám poskytuje komplexnější pohled než samostatná tradiční kritéria tím, že předpokládá, že kontext reflektuje všechna tři kritéria, tedy jak vlastnosti formální, tak i lexikálněsémantické.

Pokud číslovky tvoří samostatnou a svébytnou třídu, předpokládáme, že distribuce jejich kontextů bude odlišná od kontextů slovních druhů, s nimiž mají alespoň v prototypických případech společné formální charakteristiky.

Chceme-li se na základě pokusu s jazykovými jednotkami se společnými vlastnostmi vyjádřit k jejich potenciální podobnosti, resp. nepodobnosti se třídou jinou, naše první kroky musí směřovat k typickým zástupcům těchto tříd, u nichž se koncentruje největší množství znaků, jež je odlišují od tříd ostatních.

Jednotky, které bude tato práce zkoumat, tvoří vzhledem k jejich frekvenci ve vzorku jazyka (tzn. v korpusu SYN2010) tzv. centrum určených slovnědruhových tříd. Předpokládáme totiž spolu s Danešem (Daneš 1965), že „představa jazykového systému jako pravidelně, rovnoměrně, symetricky uspořádané soustavy jednotek je falešná“. Daneš zmiňuje také z tohoto faktu často vyplývající nemožnost jednoznačně určit zařazení

¹⁰ O výsledcích těchto výzkumů se později zmíníme ještě podrobněji.

některých jednotek v systému. Jazykové kategorie nejsou totiž kategoriemi ve smyslu kategorií aristotelských, jež jsou uzavřené a pevně vymezené a u jejichž jednotek se předpokládá jednoznačná přináležitost, resp. nepřináležitost. Kategorie jazykové tedy neoznačujeme jako klasické, ale prototypové. Označení prototypová kategorie Daneš dále vysvětluje následovně: Jednotky přirozeného jazyka se shlukují do skupin se zpravidla koncentrovaným a kompaktním jádrem (centrum, tvořené prototypickými jednotkami), které volně přechází do více či méně rozptýlené periferie. Připodobňuje tuto organizaci k množství „silových polí s gravitačními centry, vzájemně se stýkajícími, prostupujícími, hierarchizovanými“ (Daneš 1965: 4). To však podle něj neznamená, že není žádoucí a možné jednotlivé kategorie v rámci jazyka vymezovat. Naopak kvituje možnost na základě existence centra a periferie tyto kategorie vyčleňovat. Domnívá, že jednotlivé kategorie lze charakterizovat různou mírou jejich podobnosti, resp. rozdílnosti. Poté ale poznamenává, že způsob, jakým byly dosud centrum a periferie určovány a vymezovány, má charakter spíše intuitivní, nikoliv exaktní, a představuje kritéria vlastní, formální. Využití přesné specifikace jevů vzhledem k jejich kvantitativně funkčnímu a frekvenčnímu zatížení však zavrhuje z důvodu neschopnosti určit jednoznačné kritérium, které by rozhodlo, zda jev řadit k centru, či k periférii. V této práci neřadíme zkoumané jednotky blíže centru, resp. periférii na základě „intuitivních soudů“, ale na základě jejich předem daných formálních kritérií (morfologické podobnosti), které většinou vykazují, a jejich frekvence v podle nás pro tuto práci dostatečně velkém¹¹ vzorku jazyka¹².

Náš pokus tedy bude založen na porovnání kontextů prototypických číslovek s kontexty prototypických jednotek výše zmíněných srovnávacích slovních druhů.

V první fázi budeme pracovat na nejobecnější rovině slovnědruhové, to znamená zjišťovat, v sousedství jakých slovních druhů se může určitý druh číslovek vyskytovat, posléze se zaměříme na výzkum kontextu jednotlivých lexémů. Protože zkoumáme obecnou slovnědruhovou platnost nějakým způsobem definované třídy, pokládáme za adekvátní pracovat na obecné slovnědruhové rovině bez další subkategorizace (např. adjektiva posesivní, adjektiva stupňovatelná etc.).

Skupinu číslovek v následující kapitole vymezíme podrobněji a s takto vymezenými jednotkami budeme dále pracovat. Tyto jednotky (resp. jejich kontexty) budeme

11 Vzhledem k předchozím výzkumům.

12 Ovšem stále pouze vzorku.

porovnávat se zástupci předem daných slovních druhů. S těmi budeme pracovat tak, jak jsou vymezeny v Českém národním korpusu. V podstatě tedy zkoumáme vztah námi definované skupiny číslovek ke skupinám ostatních slov právě tak, jak jsou vymezeny v rámci koncepce ČNK.

Jak jsme se mohli v úvodních kapitolách přesvědčit, neexistuje jeden jednotný a absolutně správný přístup, konvence klasifikace slov do slovnědruhových tříd. To, že se značkování slovnědruhové příslušnosti v korpusu v některých bodech zcela neshoduje s koncepcemi v českých lingvistických kruzích většinou přijímanými, nás při našem experimentu nebude determinovat. Důležité je, že řazení do slovnědruhových tříd je provedeno jednotně.

6.2 Materiálová analýza

Z důvodu již zmiňované heterogenity, nejasného a nejednotného vymezení suspektního slovního druhu v odborných pracích považujeme za nutné ujasnit, co přesně pod pojmem číslovky budeme při analýze rozumět, a specifikovat výběr materiálu.

Čistě pro účely předkládané analýzy jsme definovali číslovky jako lexémy, jež je možné zapsat pomocí číslic (*jedna, desetina, první, dvakrát* nikoliv *trojice*). V úvahu nebereme složeniny vzniklé kombinací číslovky a adjektiva (*sedmnáctiletý, prvotní*).

Z korpusu SYN2010 jsme vybrali číslovky s frekvencí výskytu vyšší než 200.¹³ Dále jsme je rozdělili na:

- „numerické“ (dále v tabulkách označované zkratkou Z) - „numerickými“ číslovkami budou nadále označována slova, jejichž primární funkcí je vyjadřovat počet a která zároveň nevykazují formální podobnost s jinými slovními druhy (tzn. jejich paradigma se skládá ze dvou tvarů¹⁴ a nevystupují ve funkci jména počítaného předmětu¹⁵). Druhé kritérium ovšem nespĺňují číslovky pro pojmenování číslic 1- 4. U nich je pro nás klíčová jejich neschopnost vystupovat ve funkci počítaného předmětu.

13 Hranice byla určena zcela arbitrárně. Domníváme se, že relevantní typy číslovkových lexémů jsou u těchto vybraných již zastoupeny. Zároveň by nám takto určená hranice měla poskytnout dostatek dat, o která se budeme opírat, a navíc také jistým způsobem vylučuje možnost zařazení jednotek mezi jevy periferních.

14 Např. tvar pět pro nominativ, akuzativ a pět pro zbylé pády.

15 Např. *dvě pět, *čtyři oba, *hodně sto vs. dvě pětky, čtyři dvojice, hodně stovek etc. Přesto můžeme v jejich okolí číslovky zaznamenat, to je ovšem důsledek pravopisného úzu kombinovaných číslovkových výrazů (např. dvacet pět).

- „substantivní“ (CS) - jako „substantivní“ číslovky chápeme v této části lexémy, které se ve svých morfologických vlastnostech shodují s vlastnostmi typických substantiv (tzn. vyjadřují jeden jmenný rod, dvě gramatická čísla a disponují prostředky pro vyjádření sedmi pádů), stále ale zapsatelné pomocí číslic (*desítka, pětina, milion*).
- „adjektivní“ (CA) - za číslovky „adjektivní“ jsou označeny číslovky shodující se v morfologických vlastnostech s vlastnostmi typických adjektiv (tzn. schopnost vyjádřit tři jmenné rody, obě gramatická čísla a rozlišit sedm pádů) (*první, osmý, sedmdesátý*).
- „nesklonné“ (CN) - číslovky, které disponují pouze jedním invariantním tvarem (tzn. nereflektují jmenný rod, gramatické číslo, ani pády), nazveme nesklonnými (*dvakrát, desetkrát*).

Takto rozdělené skupiny číslovek budou v následujícím pokusu srovnávány s jednotkami, které náleží ke slovním druhům, s nimiž číslovky sdílejí morfologické a syntaktické charakteristiky a u nichž se v předchozích výzkumech už jistá podobnost projevila (Cvrček 2010).

Všechny číslovky s požadovanou frekvencí výskytu jsme vyhledali v korpusu SYN2010 prostřednictvím korpusového manažeru Bonito. Využijeme-li speciální dotazovací jazyk, korpusový manažer nám pomocí jediného příkazu zobrazí všechny tvary daného slova, které se v korpusu vyskytují. Příkaz, který nám zajistí vyhledání všech v korpusu zastoupených tvarů lemmatu „jedna“, má podobu [lemma=“jedna“].

Jak už bylo uvedeno v kapitole týkající se jazykového materiálu, pro každý slovní tvar je v korpusu zanesena tzv. morfologická značka, která uvádí gramatickou charakteristiku zkoumaného slovního tvaru (např. že se jedná o substantivum ve tvaru akuzativu plurálu, femininum etc.). Díky takto označovanému korpusu bylo možné zjistit, jaké slovní druhy se v bezprostředním kontextu číslovek vyskytují. Po zadání příkazu k vyhledání lemmat, která nás zajímala, stačilo využít funkci, jež v korpusovém manažeru Bonito najdeme pod označením „Frekvenční distribuce“. Tato funkce nám poskytne informaci o frekvencích konkrétních lemmat, tvarů slov apod. Při upřesnění, o jakou frekvenční distribuci nám přesně jde (tzn. jestli se zajímáme o frekvenci tvarů samotného klíčového slova, nebo o frekvenci slovních tvarů, či lemmat vyskytujících se v blízkosti klíčového slova a na jaké pozici) jsme určili, že požadujeme seznam frekvencí morfologických značek (tzv. tagů),

kteře se vyskytují na pozici L1, resp. P1 od klíčového slova.

Ze seznamu všech slov s frekvencí výskytů nad 200 jsme tedy vybrali lemmata, která splňovala naši pracovní definici číslovek, a rozdělili je dle formálních charakteristik do čtyř skupin (explanace dělení viz výše v tomto oddíle). Podle frekvencí uvedených u jednotlivých číslovek jsme vybrali lemmata zastoupená se stejnou, nebo nejbližší podobnou četností výskytu v korpusu SYN2010.

U všech těchto lemmat jsme získali zvlášť levostranný (L1) a pravostranný (P1) kontext, který byl tvořen morfologickými značkami. Pro účely tohoto pokusu jsme brali v úvahu pouze první značku, tzn. značku slovního druhu bez další specifikace. Pouze u číslovek jsme zohledňovali, zda nejsou zapsány v podobě číslic (tj. v korpusu odpovídající značkou C=). Tyto jednotky jsme pro jejich jedinečné postavení do analýzy nezahrnovali.

Soupis prvních dvaceti nejfrekventovanějších lemmat, která vstupovala do pokusu, i s frekvencí jejich výskytu v korpusu SYN2010 je uveden v následujících tabulkách (Tab.1 a Tab.2).

	ZÁKLADNÍ		NÁSOBNÉ		ADJEKTIVNÍ		SUBSTANTIVNÍ
210543	jeden	37139	jednou	125952	první	48814	tisíc
155842	dva	20653	poprvé	112181	druhý	46153	milion
83738	tři	10449	dvakrát	27011	třetí	27894	sto
41070	čtyři	4628	tříkrát	9471	čtvrtý	23085	polovina
39352	pět	1695	potřetí	7032	pátý	18045	miliarda
29721	deset	1614	čtyřikrát	4802	šestý	13840	desítka
22944	šest	1078	pětkrát	3690	osmý	12526	stovka
22435	půl	4024	podruhé	3272	desátý	7550	třetina
17551	dvacet	836	desetkrát	3257	dvacátý	6163	dvojice
16064	sedm	761	stokrát	3004	devátý	3556	nula
15573	osm	541	počtvrté	2759	devadesátý	3153	čtvrtina
10923	třicet	535	šestkrát	2705	šedesátý	2924	půle
9669	padesát	490	zaprvé	2128	sedmdesátý	2575	půlka
9498	devět	470	tisíckrát	2080	padesátý	1500	tisícovka
9262	patnáct	378	sedmkrát	2073	nulový	1362	pětina
7664	dvanáct	369	jedinkrát	2049	osmdesátý	1161	trojka
6630	čtyřicet	347	popáté	1820	jedenáctý	1155	desetina
5696	čtvrt	326	dvojnásobně	1707	třicátý	961	pětka
5062	čtrnáct	341	jedenkrát	1656	dvojitý	920	padesátka
4533	jedenáct	270	osmkrát	1419	dvanáctý	710	dvacítko

Tab 1: Vzorek nejfrekventovanějších zástupců zkoumaných číslovkových lemmat

	ADJEKTIVA		ADVERBIA		SUBSTANTIVA
133361	celý	20708	mimo	46925	řada
113354	dobrý	10454	uprostřed	45550	způsob
27030	současný	4633	úspěšně	28263	pocit
9469	severní	1684	paradoxně	23128	prostředí
7024	určený	1616	hladce	18048	energie
4809	zvaný	1078	zuřivě	13828	pravidlo
3693	nápadný	4055	těžce	12523	svoboda
3277	proslulý	836	suše	7544	spisovatel
3255	smrtelný	760	severně	6151	symbol
2998	tajemný	542	chytře	3556	půlnoc
2757	zábavný	534	hbitě	3153	pero
2709	jízdní	490	měkce	2926	zemědělství
2697	komunikační	471	zodpovědně	2576	omáčka
2502	sporný	378	tupě	1500	balvan
2129	kladný	368	prve	1362	montáž
2086	údajný	325	znepokojeně	1161	zaplacení
2073	masivní	346	vzorně	1156	magie
2053	uvolněný	342	bolestivě	1362	montáž
1823	něžný	269	účelně	920	stolička
1703	celosvětový			710	rasismus

Tab.2: Vzorek nejfrekventovanějších zástupců slovních druhů blízkých z formálního hlediska zkoumaným číslovkám

Po těchto krocích jsme pro všechna okolí vytvořili frekvenční distribuci, tzn. zjistili, s jakou frekvencí se číslovky a jednotky s jinou slovnědruhovou platností vyskytují v kontextu těch kterých slovních druhů. Frekvenční distribuci jednotlivých slovních druhů v kontextu číslovek i slovních druhů znázorňují následující grafy. Z nich si také můžeme udělat představu, jaké slovní druhy zkoumané slovnědruhové třídy ve svých kontextech preferují.

Nejdříve ale výše popsany postup pro ujasnění stručně shrňme a uvedme ještě potřebné technické detaily:

1. Z lemmat s frekvencí vyšší než 200 výskytů byly vybrány číslovky a jim formálně podobné slovní druhy (pro číslovky adjektivní adjektiva, pro číslovky nesklonné adverbia etc.) se shodnou, nebo co nejbližší frekvencí.
2. Pro každou takto získanou skupinu bylo zjištěno, jak jsou v jejich bezprostředním levostranném a pravostranném kontextu zastoupeny jednotlivé slovní druhy.
3. Zjištěné kontextové preference skupin, u nichž pozorujeme určité formální analogie, budeme dále mezi sebou porovnávat

Značení slovních druhů se shoduje s konvencí zápisu v rámci ČNK. Seznam zkratk pro slovní druhy vypadá následovně:

A	Adjektiva
C	Číslovky
D	Adverbia
I	Interjekce
J	Konjunkce
N	Substantiva
P	Pronomina
R	Prepozice
T	Partikule
V	Verba

Tab. 3: Zkratky slovních druhů používané v konvenci zápisu ČNK

Při pohledu na grafy, zaznamenávající preference v jednostranných kontextech zkoumaných jednotek, můžeme u určitých tříd zaznamenat jistou podobnost. Vizuální porovnávání rozložení kontextů a následné vyvozování více či méně relevantních závěrů však podle nás nemá požadovanou vypovídací hodnotu. To, zda opravdu existuje nějaká souvislost v kontextových preferencích zkoumaných jednotek, proto v další části ověříme pomocí přesných statistických metod.

6.2.1 Materiálová analýza na obecné slovnědruhové úrovni

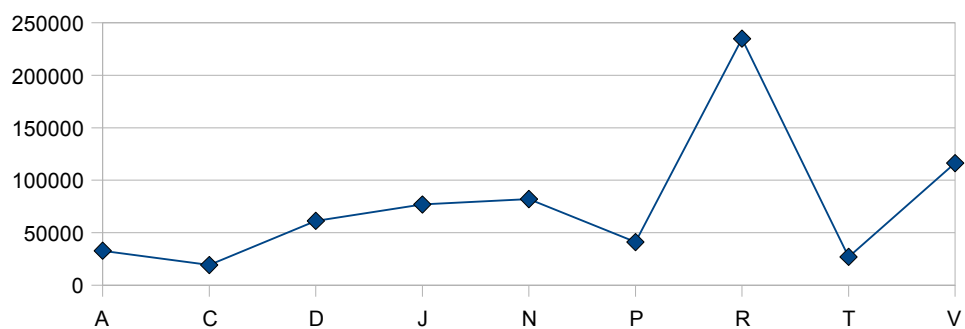
Distribuci kontextů jednotlivých číslovek se pokusíme nejprve zhodnotit vizuálně z grafů, jež toto rozložení znázorňují, poté přistoupíme k vyhodnocení jejich podobnosti, resp. odlišnosti pomocí statistických metod. Osa y v následujících grafech udává frekvenci výskytů jednotlivých slovních druhů, jejichž značky jsou uvedeny na ose x. Hledáme podobnost, či rozdílnost kontextů stejně, resp. přibližně stejně frekventovaných jednotek. Ne všechny srovnávané skupiny si velikostně odpovídají. Zatímco frekvence substantiv může nabývat hodnot desetitisícových, frekvence číslovek „pouze“ tisícových.

V této práci nehledáme absolutní shodu ve frekvencích preferovaných slovních druhů, jež do kontextů vstupují. Zajímají nás tedy četnosti poměrné, tzn. tendence k preferenci určitých slovnědruhových tříd. Nebudeme proto při vizuálním hodnocení grafů analyzovat

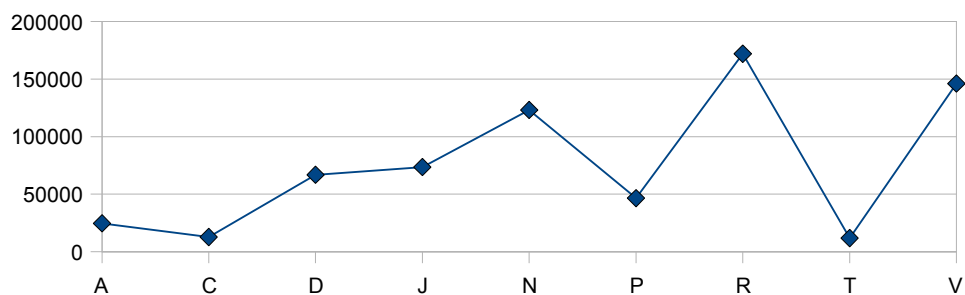
podobnost na ose y, ale podobnost tvarů a průběhů jednotlivých křivek.

6.2.1.1 Kontext číslovek numerických

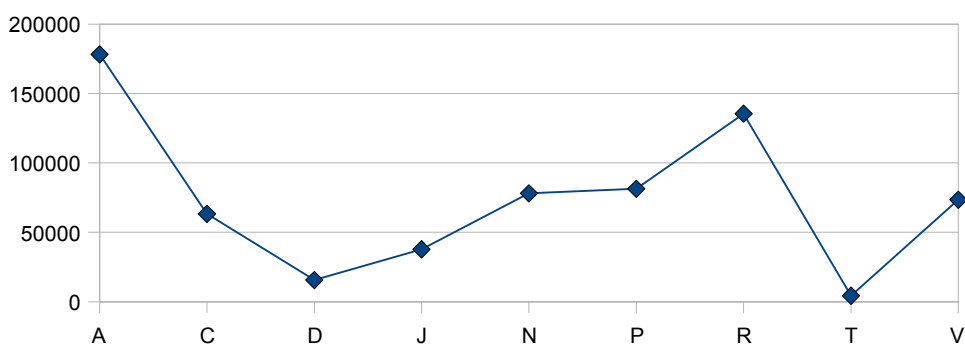
Z grafů zobrazujících kontexty námi zkoumaných numerických číslovek vidíme, že distribuce jejich levostranného kontextu je na pohled velmi podobná rozložení kontextu adjektiv a substantiv. Substantiva se však odlišují velkým podílem adjektiv na pozici L1.



Graf 3: Číslovky numerické - L1



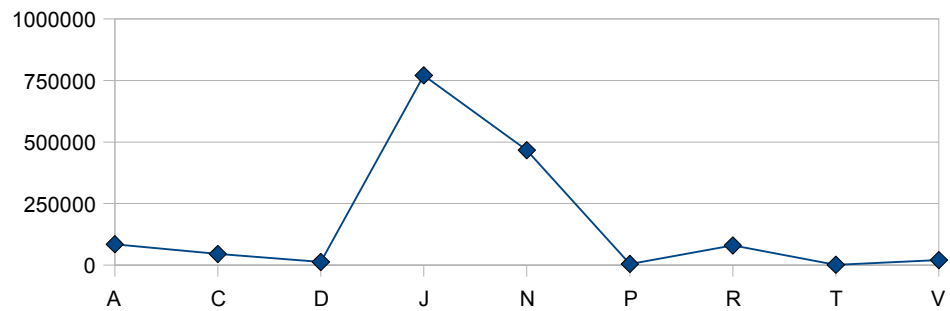
Graf 4: Adjektiva odpovídající frekvenci číslovkám numerickým - L1



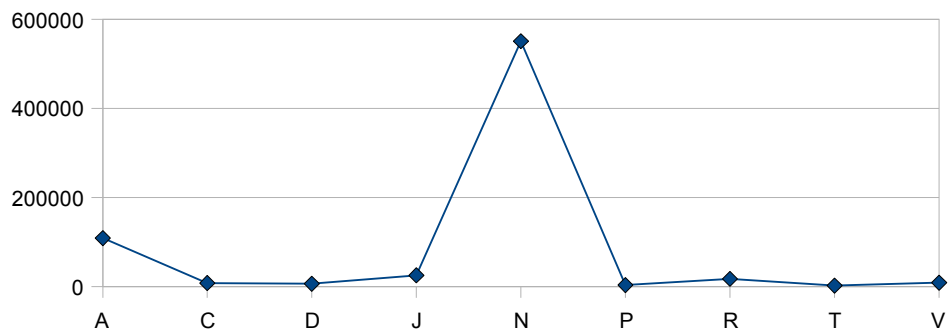
Graf 5: Substantiva odpovídající frekvenci číslovkám numerickým - L1

Poněkud jinak to ale vypadá s rozložením kontextu pravého, který se od rozložení u adjektiv a substantiv viditelně liší. U první zmíněné třídy opět o něco méně než u druhé.

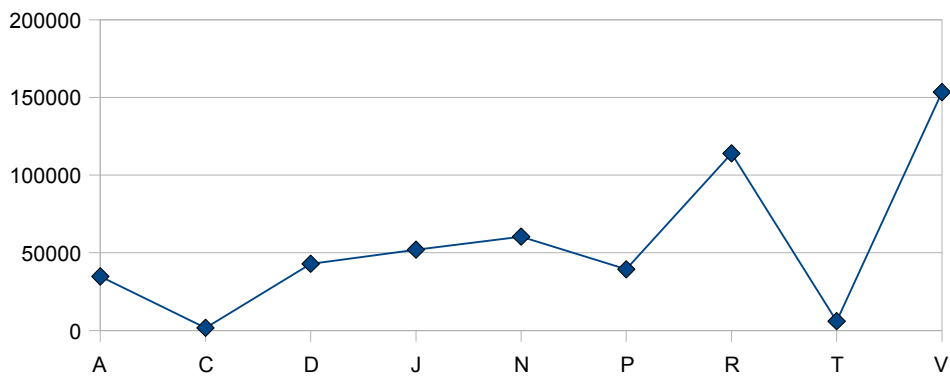
Zdá se, že formální a lexikálněsémantické vlastnosti v korpusu SYN2010 nejčastěji zastoupených numerických číslovek mají potenciál zčásti odlišovat tuto třídu od ostatních.



Graf 6: Číslovky numerické - P1



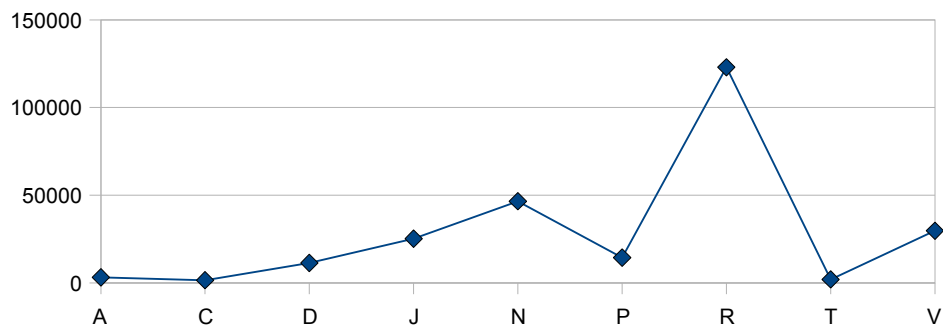
Graf 7: Adjektiva odpovídající frekvenci číslovkám numerickým - P1



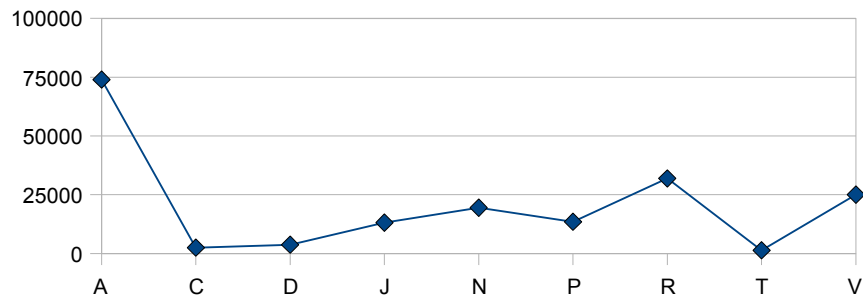
Graf 8: Substantiva odpovídající frekvenci číslovkám numerickým - P1

6.2.1.2 Kontext číslovek substantivních

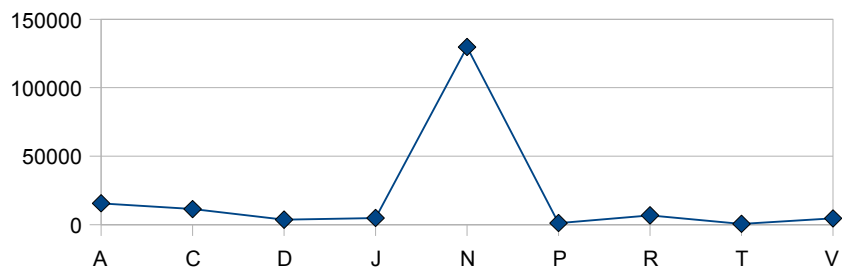
Zatímco číslovky substantivní vstupují především do pravostranného kontextu předložek, substantiva jsme zaznamenali nejčastěji v pravostranném okolí adjektiv (typicky syntakticky shodného přívlastku). Zdá se, že námi zkoumané substantivní číslovky nevyžadují svou sémantikou bližší adjektivní specifikaci tak, jak je tomu u substantiv.



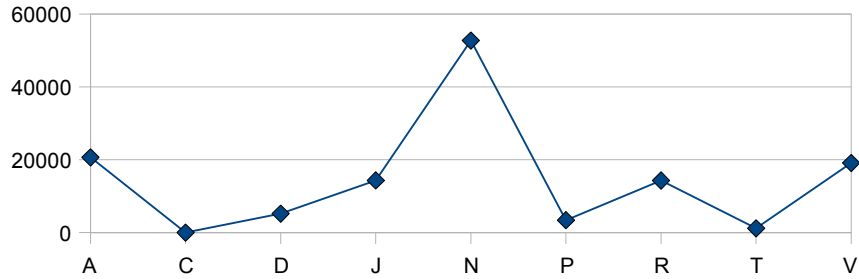
Graf 9: Číslovky substantivní - L1



Graf 10: Substantiva odpovídající frekvenci číslovkám substantivním - L1



Graf 11: Substantivní číslovky - P1

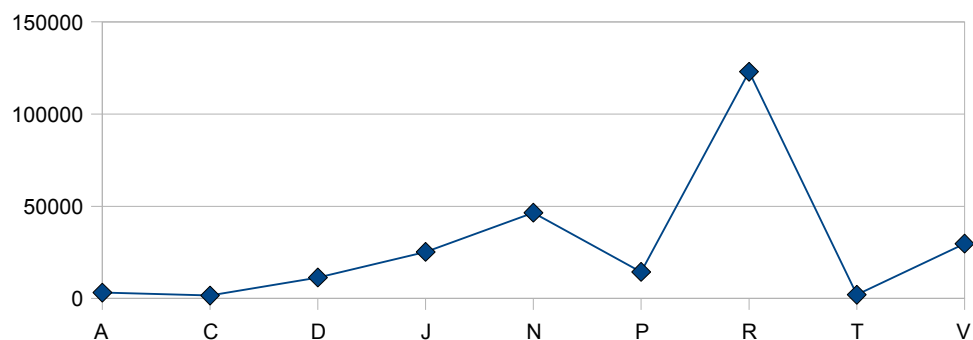


Graf 12: Substantiva odpovídající frekvenci číslovkám numerickým - P1

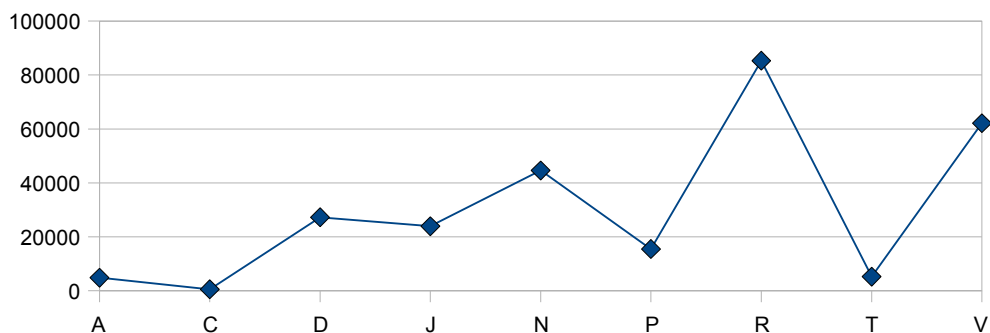
Z grafu 11 je možné vyčíst, že číslovky preferují v pravostranném kontextu substantiva, entity, jež jsou číslovkami blíže specifikovány a které zároveň udávají, čemu příznak, který v sobě číslovky nesou (určení množství), náleží. Nejfrekventovanější kontext této skupiny tvoří zřejmě substantiva v roli jména počítaného předmětu.

6.2.1.3 Kontext číslovek adjektivních

Už z označení, které jsme těmto číslovkám přiřadili, vyplývá, že distribuci jejich kontextu budeme srovnávat s distribucí kontextu slovního druhu adjektiv, s nimiž sdílejí formální charakteristiky.



Graf 13: Číslovky adjektivní - L1

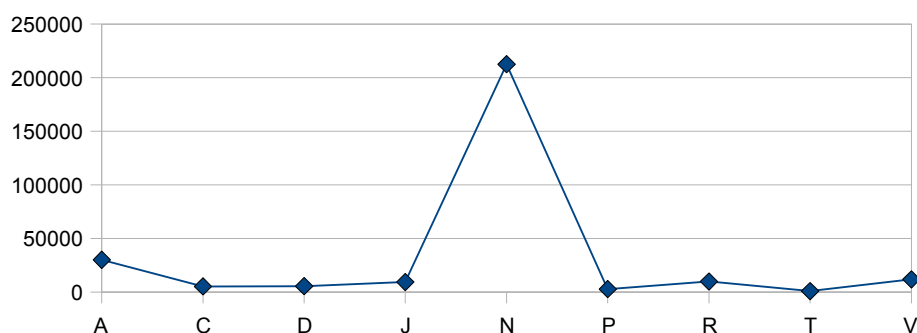


Graf 14: Adjektiva odpovídající frekvenci číslům adjektivním- L1

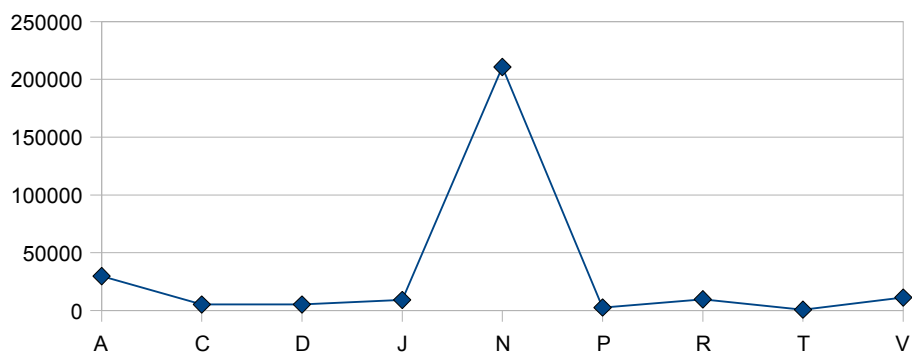
Z grafů 13 a 14 vidíme, že srovnávané třídy mají společné nejen formální vlastnosti, ale i velice podobné rozložení bezprostředního levostranného okolí.

Stejně jako u adjektiv zaznamenáme v kontextu čísel nejčastěji předložky, slovesa se potom významně podílejí na levém kontextu ve větší míře u adjektiv než u adjektivních čísel. Příkládáme to vyšší frekvenci a většímu počtu užití adjektiv ve funkci jmenného přísudku.

Do pravého kontextu shodně vstupují u čísel adjektivních i adjektiv nejčastěji substantiva, což je zřejmě způsobeno funkcí shodného přívlastku, kterou tyto slovní druhy obvykle ve větě plní.



Graf 15: Číselky adjektivní - P1



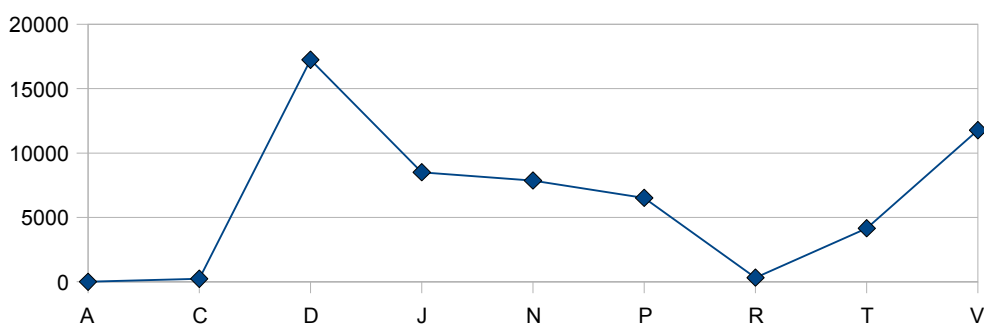
Graf 16: Adjektiva odpovídající frekvenci číslovkám adjektivním- P1

Po zhlédnutí grafů věnovaných číslovkám adjektivním v souvislosti s adjektivy je na místě konstatování, že z hlediska L1 a P1 se jedná o třídy relativně shodné. To znamená, že vzorky adjektivních číslovek z korpusu SYN2010 nevykazují zásadně odlišné formální a lexikálněsémantické charakteristiky v porovnání se stejně, resp. podobně frekventovanými lemmaty adjektivními.

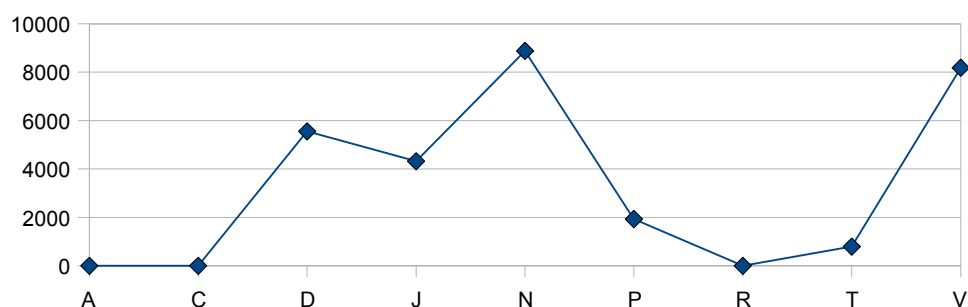
6.2.1.4 Kontext číslovek nesklonných

Jinak je tomu v případě levého kontextu u číslovek nesklonných a adverbíí. Test nesklonných číslovek jsme provedli dvakrát. Nejprve na obecné úrovni, kdy byla testována všechna lemmata, která jsme s určitou mírou benevolence do souboru číslovek nesklonných zahrnuli (viz níže), poté byly tyto číslovky ještě rozděleny na základě formálních kritérií na dvě podskupiny.

V případě testu celého souboru (Graf 17 a Graf 18) vidíme, že v levostranném kontextu obou skupin se shodně jako druhá nejfrekventovanější skupina objevují slovesa. Tím se ale podobnost rozložení jejich kontextů vyčerpává.



Graf 17: Číslovky nesklonné - L1

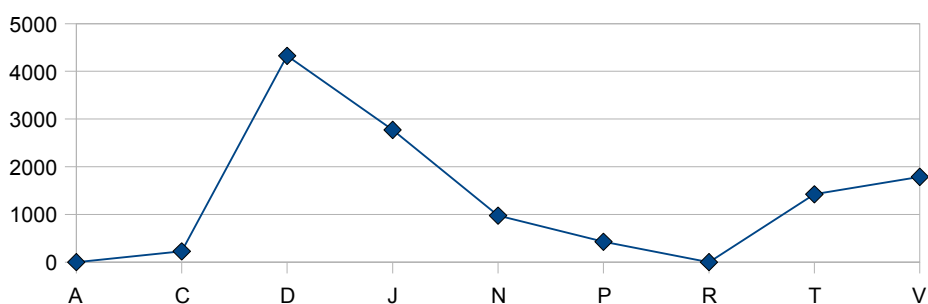


Graf 18: Adverbia odpovídající frekvenci číslovkám nesklonným - LI

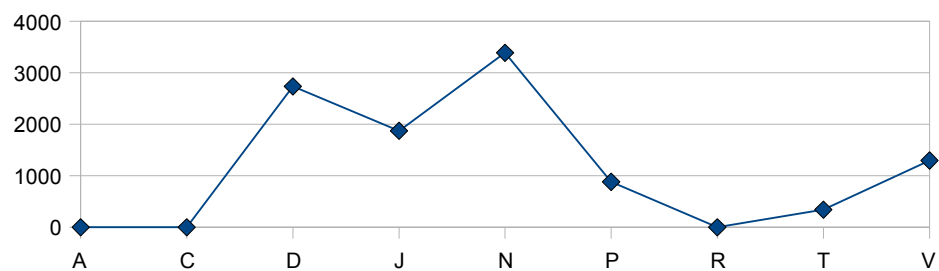
Pokud si pozorně prohlédneme Tabulku 1, zjistíme, že ne všechny zkoumané nesklonné číslovky bychom bez potíží mohli zapsat číslem, tudíž ne všechny odpovídají na začátku uvedenému obecnému kritériu pro vymezení prototypických číslovek.

Číslovky *poprvé*, *potřetí*, *počtvrté* etc. jsme zařadili do pokusu z toho důvodu, že vyjadřují stejně jako adverbia, s nimiž je srovnáváme, nějakou okolnost. Pokud budeme preferovat zápis skládající se ze dvou grafických slov (*po prvé*), druhou část bez obtíží číslicí zapíšeme (ačkoliv to úzu samozřejmě neodpovídá). Obě grafická slova však tvoří jeden lexém, proto jsme zahrnuli i číslovky tohoto typu.

Po vynechání výše zmíněných lexémů (*poprvé*, *potřetí* etc.) a opakování testu, grafy zobrazující distribuci číslovek, tradičně zvaných násobnými, vypadají následovně.

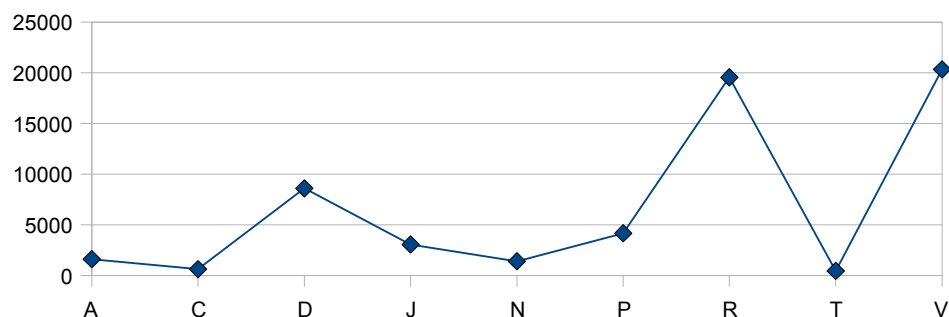


Graf 19: Číslovky nesklonné (násobné) - LI

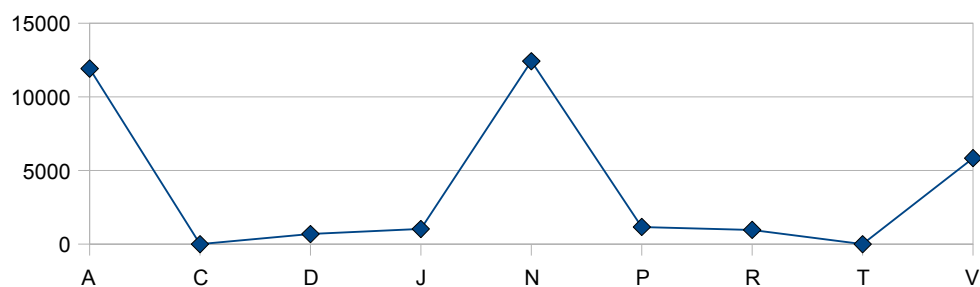


Graf 20: Adverbia odpovídající číslkám nesklonným (násobným) - LI

Teď se podíváme na to, zda se zachová podobnost i v případě kontextu pravostranného.



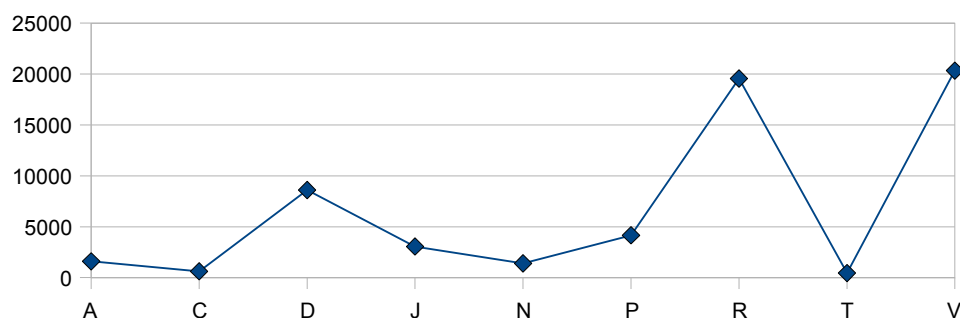
Graf 21: Číslky nesklonné - P1



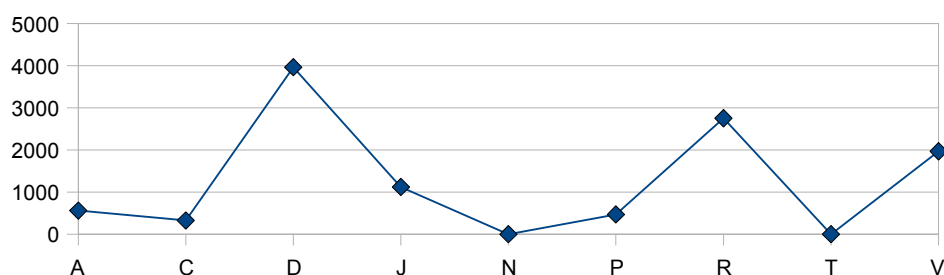
Graf 22: Adverbia odpovídající číslkám nesklonným - P1

Na první pohled je zjevné, že v rozložení pravostranného kontextu tyto dvě zkoumané skupiny přílišnou podobnost nevykazují.

Zkusme znovu z této skupiny vyřadit číslovky typu *poprvé* a zkoumání podrobit pouze číslovky nesklonné v tradičních pracích označované jako číslovky násobné.



Graf 23: Číslovky nesklonné (násobné) - L1



Graf 24: Adverbia odpovídající číslovkám nesklonným (násobným) - P1

Situace se radikálně změnila. Zatímco předcházející distribuce kontextu vykazovala relativní rozdílnost, distribuce kontextu adverbii a číslovek nesklonných (násobných) se zdá být poměrně shodná.

Jestliže lemmata, která jsme zařadili mezi číslovky nesklonné, způsobila takový rozdíl v podobnosti pravostranného kontextu obou skupin, je na místě se ptát, zda přiznat lexikálněsémantickým (neboť syntakticky i morfologicky se shodují se slovním druhem adverbii) vlastnostem číslovek (tradičně nazývaných číslovky násobné) takový význam a zda je řadit v rámci našeho vzorku k jinému slovnímu druhu než k adverbii, či zpochybňovat oprávněnost přiřazení číslovek typu *poprvé* k číslovkám srovnatelným s adverbii (tyto číslovky jsou tradičně označovány jako číslovky řadové).

Jak jsme viděli na grafech, které zobrazovaly frekvenci blíže nespecifikovaných slovních druhů v kontextu čtyř námi vymezených skupin číslovek a jim formálně blízkých slovnědruhových tříd, není na první pohled zjevné, že by číslovky disponovaly natolik specifickými charakteristikami, které by je od určených tříd jasně odlišovaly. Letný pohled ovšem může být matoucí.

Výpovědní hodnotu našeho průzkumu teď ověříme pomocí statistické metody, jejíž závěry by už neměly posuzovat problematiku „od oka“.

6.2.1.5 Zhodnocení podobnosti kontextů slovních druhů statistickými metodami

Jelikož máme co do činění s daty jazykovými, která jsou ve své přirozenosti neparametrická a jejichž rozdělení se nedá označit za normální (normální rozdělení zobrazuje známá Gaussova křivka), a chceme poznat, jaká je vzájemná podobnost mezi dvěma nestejně rozsáhlými skupinami dat, použijeme k vyjádření tohoto vztahu Spearmanův korelační koeficient, který se k posouzení síly takových korelací ve statistice využívá. Čím je souvislost mezi skupinami užší, tím je hodnota koeficientu vyšší. Absolutní kladná korelace (se zvýšením jedné se zvyšuje i druhá proměnná) nabývá hodnoty 1, maximální záporná korelace (se zvýšením jedné proměnné hodnota druhé klesá) -1. Důležité je podotknout, že pod pojmem korelace není míněná kauzalita. Je-li mezi dvěma proměnnými hodnota korelace silná (tzn. nabývá hodnot kolem +/- 0.85), ukazuje to na významnou podobnost tendencí těchto proměnných, nikoliv na to, že by hodnota jedné ovlivňovala hodnotu té druhé. Naopak korelaci pohybující se pod hodnotou 0,4, resp. nad - 0,4, označíme za nevýznamnou.

Pokud se tedy korelace mezi rozložením kontextů zkoumaných skupin bude blížit 1, resp. 0, znamená to, že dané skupiny mají podobné, resp. rozdílné kontextové preference, a připomeneme-li si výchozí axiom, také podobný/ různý význam, funkci i formální vlastnosti.

Výsledky výpočtů Spearmanova korelačního koeficientu uvádíme v následující tabulce.

	L	P
Z - A	0.98	0.93
Z - N	0.3	0.33
CA - A	0.92	0.8
CN - D	0.86	0.26
	0.78	0.87
CS - N	0.32	0.58

Tab.3: Korelační koeficient mezi kontexty zkoumaných slovních druhů, první hodnota u CN je hodnotou koeficientu před rozdělením skupiny (zkratka Z značí číslovky numerické, CA číslovky adjektivní, CN číslovky nesklonné, CS číslovky substantivní)

Zjištěné hodnoty dokládají správnost soudů na základě vizuálního zhodnocení grafů. Tmavě šedé buňky v tabulce vyznačují silnou korelaci mezi kontexty porovnávaných slovních druhů. Zatímco podobnost adjektivních číslovek a adjektiv byla zjevná, o podobnosti číslovek substantivních se substantivy bylo třeba pochybovat, v čemž nás tabulka ujišťuje. Potvrdily se i pochybnosti týkající se číslovek nesklonných.

Největší podobnosti dosahují podle tabulky číslovky numerické s adjektivy. Vysvětlujeme si to podobnou syntaktickou funkcí adjektiv a číslovek, které většinou shodně obsazují pozici před jménem, typickou pro funkci přívlastku shodného (o tom hovoří už Komárek, viz 5.3.1.b)).

Jako relativně autonomní se ukázaly číslovky substantivní (CS), které v mluvnicích najdeme jako zástupce číslovek základních. Máme za to, že i když se v našem vzorku zkoumané jednotky formálně přibližují substantivům, mohly by to právě ony, které mají potenciál argumentu pro existenci slovnědruhové třídy číslovek.

Na našem vzorku se ale také ukázalo, že sémantické kritérium se v případě číslovek v mluvnicích označovaných jako řadové, násobné a základní nezdá být kritériem příliš silným k vymezení číslovek jako samostatného slovního druhu.¹⁶

Otázkou je, jak by dopadla analýza, kdybychom nepracovali s tak úzce vymezeným

¹⁶ Uvědomíme-li si, že formální charakteristiky zkoumaných číslovek víceméně odpovídaly s nimi srovnávaným slovním druhům.

souborem číslovek a tak obecně definovaným termínem, jakým slovní druhy jsou. I slovní druhy, o jejichž samostatné existenci se žádným způsobem nepochybuje, jsou skupinami relativně heterogenními. Reflexe této skutečnosti (např. rozlišování adjektiv přivlastňovacích na jedné straně a měkkých a tvrdých na druhé, nebo odlišení stupňovatelných a nestupňovatelných adverbii) by mohla přinést výsledky přínosnější. Naším cílem bylo podrobit analýze prototypické zástupce číslovek, číslovky určité, na obecné slovnědruhové rovině.

Než se budeme věnovat analýze konkrétních vybraných lemmat, srovnáme výsledky našeho pokusu s výsledky pilotního výzkumu, kterým se v naší práci inspirujeme.

Ve výše zmíněných pokusech Cvrček (Cvrček 2010) podrobil analýze všechna lemmata tehdy půlmiliardového korpusu SYN (složeného z korpusů SYN2000, SYN2005 a SYN2006PUB). Na rozdíl od nás však vyexcerpoval a srovnával kontexty oboustranné, taktéž tvořené značkou slovního druhu (první písmeno v morfologické značce, tzv. tagu). Co je ale v případě jeho analýzy odlišné, je to, že číslovky pro potřeby experimentu nijak zvlášť nevymezoval. Pracoval s touto slovní třídou tak, jak je určena, označena a subklasifikována do jednotlivých druhů v rámci ČNK.

Slovními druhy, s nimiž číslovky porovnával, byla z důvodu zjevné formální a předpokládané kontextové podobnosti substantiva, adjektiva, zájmena a adverbia, vzhledem k případné demonstraci této podobnosti i slovesa, spojky a předložky, u nichž se podobnost s číslovkami nepředpokládá. Adjektiva dále rozdělil na „obyčejná“ a „ostatní“ (jmenné tvary a tvary přivlastňovacích zájmen), adverbia na stupňovatelná a nestupňovatelná.

V tabulce, kterou níže pro srovnání předkládáme, nejsou uvedeny hodnoty Spearmanova korelačního koeficientu, nýbrž koeficientu determinačního. Jeho hodnotu dostaneme, pokud Spearmanův korelační koeficient umocníme. Získané hodnoty jsou v procentech a udávají, kolik procent z variace jednoho souboru je vysvětlitelných variací v souboru druhém.

Číslovky	Subst.	Adjektiva		Pron.	Verba	Adverbia		Prep.	Konj.
		„obyč.“	„ostat.“			stup.	nestup.		
Základní	62,60%	82,80%	67,40%	56,60%	39,90%	47,80%	39,70%	52,50%	40,20%
Řadové	68,30%	87,80%	76,50%	70,80%	46,60%	54,90%	48,10%	56,00%	46,60%
Násobné	27,80%	46,20%	59,00%	56,00%	77,40%	81,10%	85,70%	38,70%	70,50%
Druhové	65,90%	83,20%	85,10%	88,00%	63,20%	70,90%	65,70%	43,80%	61,00%

Tab. 4: Přehled korelačních koeficientů mezi jednotlivými skupinami slov

Na základě vysokých hodnot determinačního koeficientu Cvrček dochází k závěru, že kontextová charakteristika jednotlivých druhů číslovek je velmi podobná charakteristice slovnědruhových tříd, k nimž by z formálního hlediska mohly být číslovky řazeny. Podobnost kontextů podle něj jednoznačně ukazuje na formální, významovou i strukturní blízkost srovnávaných dvojic slovních druhů.

V další fázi pokusu se budeme soustředit na jednotlivá lemmata a jejich kontexty a posuzovat, zda závěry, ke kterým jsme došli na zcela obecné rovině, platí i v případě konkrétních zástupců číslovek.

6.2.2 Materiálová analýza na úrovni lexémů

Pro účely konkrétnější analýzy jsme náhodně vybrali zhruba každý desátý lexém ze seznamu jednotek, které se účastnily výše uvedené analýzy na obecné úrovni. Vynecháme však porovnávání číslovek numerických se substantivy, u nichž se korelace ukázala být nevýznamnou. Naopak ověříme nevýznamnou korelaci kontextů substantivních číslovek se substantivy. A to z důvodu jejich sdílené formální charakteristiky.

Po předchozím vyhodnocení podobnosti kontextů číslovek s jim formálně podobnými slovními druhy bychom mohli k datům přistupovat s určitým očekáváním korelace se zjištěními na obecné úrovni. Nesmíme však zapomínat na zcela jedinečné lexikálněsémantické vlastnosti každé jazykové jednotky. V této části výzkumu bychom

chtěli specifikovat, na jaké úrovni je podobnost kontextů podobná, resp. rozdílná.

Opět musíme mít na paměti, že se jedná o vzorek vzorku, navíc u některých nepříliš frekventovaných jednotek s nepříliš velkými soubory dat.

Porovnávali jsme vždy jeden číslovkový lexém s lexémem jiné slovnědruhové platnosti zhruba stejné frekvence. Snažili jsme se, aby byly zastoupeny jak jednotky s nejvyšší frekvenční charakteristikou, tak jednotky, jejichž frekvence se pohybovala těsně nad hranicí 200 výskytů v korpusu SYN2010, která byla na počátku práce stanovena. U takto zjištěných lemmat jsme opět vyexcerpovali levostranný i pravostranný kontext v podobě značky slovního druhu a hodnoty zastoupení jednotlivých slovních druhů podrobili výpočtu Spearmanova korelačního koeficientu.

6.2.2.1 Kontext numerických číslovek

Ze souboru číslovek numerických byly vybrány číslovky *jeden*, *dvacet*, *čtrnáct*, *devadesát*, *dvaatřicet* a *devětadvacet* a jim frekvenčně odpovídající adjektiva *velký*, *správný*, *uzavřený*, *srovnatelný*, *férový*, *čalouněný*.

Jak dopadlo srovnání v rámci jednotlivých dvojic, si ukážeme nikoliv už na grafech, ale na hodnotách Spearmanova korelačního koeficientu (viz Tab. 5).

	L	P
velký - jeden	0.88	0.77
správný - dvacet	0.47	0.58
uzavřený - čtrnáct	0.67	0.61
srovnatelný - devadesát	0.56	0.04
férový - dvaatřicet	0.67	0.58
čalouněný - devětadvacet	0.87	1

Tab.5: Hodnoty korelačního koeficientu mezi vybranými numerickými číslovkami a adjektivy

Silná souvislost se objevila mezi lexémy *velký* a *jeden*. Číslovka *jeden*, která v některých situacích ztrácí svůj jedinečný, kvantitu vyjadřující význam, a vyjadřuje neurčitost (*jeden* - nějaký), se v tomto vzorku výrazně přibližuje adjektivům. Potvrzuje tedy tendence patrné při obecném srovnání kontextů.

Podobně silnou korelaci vykazují i lexémy *čalouněný* a *devětadvacet*. V levém kontextu obou jednotek se totiž nejčastěji (a téměř výhradně) vyskytla adverbia a prepozice.

V pravém kontextu objevíme výhradně substantiva. To ovšem nemůžeme jednoznačně vysvětlovat jejich podobností. Na hodnotící závěry disponujeme v případě těchto jednotek příliš malým souborem dat.

Opět musíme mít na paměti, že se jedná o vzorek vzorku, navíc u některých nepříliš frekventovaných jednotek s nepříliš velkými soubory dat.

6.2.2.2 Kontext adjektivních číslovek

Korelace u souboru adjektiv a číslovek adjektivních se tendencím zjištěným u srovnávání na obecné úrovni přibližují více, než tomu bylo u korelací mezi číslovkami numerickými a adjektivy. Do tohoto pokusu vstupovaly číslovky *druhý*, *devadesátý*, *devatenáctý*, *jednadvacátý* a *stý*. Pro zajímavost jsme zařadili i číslovky *dvojitý*, *milionový* a *půlmilionový*, které neodpovídají postulovanému kritériu ohledně výběru číslovek a nelze je zapsat číslicí. Frekvenčně odpovídají v korpusu těmto číslovkám adjektivní lemmata *dobrý*, *zábavný*, *indiánský*, *fyzikální*, *infekční*, *předminulý* a *záchranářský*.

Přehled korelačních koeficientů nabízí následující tabulka (Tab. 6).

	L1	P1
druhý - dobrý	0.6	0.93
devadesátý - zábavný	0.15	0.3
devatenáctý - indiánský	0.71	1
dvojitý - fyzikální	0.78	0.87
milionový - infekční	0.6	0.75
stý - předminulý	0.36	1
jednadvacátý - záchranářský	1	1

Tab. 6: Korelační koeficient vybraných adjektivních číslovek a jim frekvenčně odpovídajících adjektiv

Zatímco korelační koeficient nejfrekventovanější dvojice *druhý* - *dobrý*, nabývá relativně vysokých hodnot, korelační koeficient hned druhého nejfrekventovanějšího páru *devadesátý* - *zábavný*, signalizuje podobnost spíše marginálního významu, resp. rozdílnost. Přisuzujeme to na jedné straně jistým způsobem zobecněnému významu číslovky *druhý*, která v některých kontextech nabývá významu lexému *jiný*, *další*, na druhé straně relativně

úzkému významu číslovky *devadesátý*, která vyjadřuje pouze význam číselný, pořadí v nějakém souboru (většina lemmat bezprostředního kontextu je užitá ve významu letopočtu - *polovina devadesátých let*, ale vyskytuje se i spojení *devadesáté narozeniny*). Její kontext je tak specifitější, než je tomu u kontextu číslovky druhý.

Nabízí se tedy otázka, zda číslovka druhý (jakkoliv je vyjádřitelná číslicí a splňuje tedy výše uvedená kritéria pro výběr prototypických číslovek) je opravdu zástupcem prototypu číslovek. Lemmata *jeden* a *druhý* jsou příklady lemmat homonymních. U jejich dalšího výzkumu bychom museli oba významy a způsoby užití oddělit.

V případě lemmat *dvojité* - *fyzikální* jsou korelace obou kontextů velmi silné. Domníváme se, že příčina tkví vedle formální podobnosti i v obecnějším významu číslovky *dvojité*. Stejně tak si vysvětlujeme i silnou korelaci pravého kontextu u dvojice *milionový* - *infekční*.

Velmi zajímavý výsledek zjišťujeme u dvojice *devatenáctý* – *indiánský*. Do jejich pravého kontextu vstupovaly v našem vzorku pouze spojky a substantiva, a to zhruba ve stejných poměrech. Počet ostatních slovních druhů byl na nule, což vysvětluje tak silnou podobnost. Stejná situace nastává i v případě lemmat *stý* – *předminulý* a *jednadvacátý* - *záchranný*, u nichž byl pravostranný kontext tvořen pouze substantivy. Naprosto rozdílný je však kontext levostranný. Dvojice *stý* – *předminulý* koreluje na pozicích L1 velmi slabě. Na rozdíl od toho druhá zmíněná dvojice, *jednadvacátý* - *záchranný*, koreluje stoprocentně.

6.2.2.3 Kontext nesklonných číslovek

Na rozdíl od analýzy na obecné, slovnědruhovú úrovni, kde jsme mohli zaznamenat silnou podobnost, zjišťujeme u analýzy konkrétních lemmat souvislost nevýznamnou. Do této části pokusu vstupovaly dvojice lemmat *dvakrát* - *nejprve*, *desetkrát* - *suše*, *jedinkrát* - *prve*.

	L1	P1
dvakrát - nejprve	0.78	0.5
desetkrát - suše	0.27	-0.27
jedinkrát - řídce	0.7	-0.14

Tab. 7: Korelační koeficient vybraných nesklonných číslovek a jim frekvenčně odpovídajících adverbii

Dvojice, u nichž můžeme zaznamenat větší kontextovou podobnost, jsou dvojice lemmat *dvakrát - nejprve a jedinkrát - řídce*. Ovšem jen v levostranném kontextu.

V podstatě se nám zde potvrzuje platnost kontextové analýzy. Tímto konkrétním pokusem jsme vlastně ověřili sílu kontextové analýzy, tzn. schopnost blízkého kontextu reflektovat gramatické i lexikálně sémantické vlastnosti. Na příklad v bezprostředním levostranném okolí číslovky *desetkrát* se nejčastěji objevují různé druhy částic (částice modální - *asi, snad*, a kvantifikační, neboli měrová adjektiva - *nejméně, aspoň*), v kontextu formálně podobného lemmatu *suše* potom zaznamenáme nejčastěji slovesa, jejichž význam toto adverbium upřesňuje (nejfrekventovanější jsou tzv. verba dicendi - *podotknout, poznamenat*).

Podobně je to i v pravostranném kontextu slov *desetkrát - suše*. *Desetkrát* udává míru něčeho. Proto v tomto kontextu nacházíme nejčastěji slova, která jsou schopná vyjádřit komparaci, nebo slova vyjadřující časový úsek. *Desetkrát* přitom udává, kolikrát je srovnávané slovo jiné (*desetkrát větší, vyšší*), nebo jak často se nějaký jev v rámci časového úseku opakuje (*desetkrát denně*). Pravostranný kontext slova *suše* se mění jen minimálně, ke slovesům mluvení se přidalo i sloveso zasmát se (*suše se zasmál*).

Stejně tak v pravostranném kontextu lemmatu *jedinkrát* nejčastěji objevíme předložky (*jedinkrát v životě*) a slovesa (*jen jedinkrát jsem zažil*), zatímco lemma *řídce* svým významem upřesňuje vlastnost a v jeho pravostranném kontextu převažují adjektiva (*řídce osídlený, řídce chlupatý*).

6.2.2.4 Kontext substantivních číslovek

Přesto, že substantiva a substantivní číslovky mají téměř shodné formální vlastnosti, z hodnot korelací prvních dvou párů v tabulce zjišťujeme, že jejich kontexty žádnou podobnost nevykazují. Jejich syntaktické a sémantické vlastnosti jsou tedy rozdílné.

	L1	P1
polovina - prostředí	0.22	-0.08
tisícovka - balvan	0.15	-0.01
sedmička - sirka	0.74	0.08
dvanáctka - zástěrka	0.75	0.63

Tab. 8: Korelační koeficient vybraných substantivních číslovek a jim frekvenčně odpovídajících substantiv

Zajímavý výsledek poskytuje výsledek srovnání podobnosti kontextu dvojice *sedmička* - *sírka*. Obě slova patří do stejného morfologického paradigmatu, jejich forma je tedy shodná.

Podíváme-li se na levostranný kontext obou slov, zjistíme, že jejich podobnost je relativně silná. Z hlediska pozice L1 je v tomto případě kontext číslovky substantivní opravdu silně podobný substantivu. Pohled na kontext pravostranný ale souvislost nevykazuje téměř žádnou. Vysvětlujeme si to velkou frekvencí skupin vyjadřujících množství nebo pořadí *sedm* (*sedmička vína, sedmá třída, sedmý zub, týdeník Sedmička*), jejichž pojmenování se postupně substantivizuje, což vidíme na L1, zároveň však pokud v pravém kontextu doplňujeme, většinou je třeba tak činíme pomocí jména počítaného předmětu. Tato vlastnost lexému *sírka* zcela chybí.

Ještě větší podobnost se substantivy se projevují u lemmatu *dvanáctka*. Jeho párem je taktéž formálně shodné lemma *zástěrka*. Slovo *dvanáctka* totiž může nabývat různých významů a většinou označuje nějaké entity (dvanáctistupňové pivo, dvanáctá hodina, brokovnice ráže dvanáct, často také skupina dvanácti finalistů). U takových entit již (vyjma poslední) většinou není potřeba dodávat jméno počítaného předmětu.

7. Závěr

Než se budeme věnovat shrnutí této práce a výsledkům, ke kterým jsme v ní došli, je podle nás třeba znovu připomenout, jak úzce vymezená naše data byla a že se jednalo pouze o vzorek vzorku psaného jazyka. Následující soudy a návrhy se tudíž neusilují, ani nemohou usilovat, o všeobecnou validitu. Platné jsou v rozsahu této práce, jež se snaží o rozšíření pohledu na pojednávaný okruh otázek, nikoliv v celkovém nazírání na jazykový systém.

Tématem této diplomové práce je slovní druh číslovek jako celek, problematika jeho postavení v systému slovních druhů a diskuze nad jeho vymezením.

Nejprve jsme se v úvodu práce seznámili s tím, jak pojem slovních druhů chápat, s historií a tradicí jeho vymezení. V souvislosti s tím jsme zjistili, že třídění slov do slovních druhů patří k nejtradičnějším oblastem lingvistiky a již po staletí je jeho podoba téměř neměnná. Zároveň jsme však přiblížili skutečnost, že ne na všechny jazyky je možné aplikovat jeden a tentýž princip klasifikace. Každý jazyk má svou specifickou strukturu a tu je nutné při vyčleňování slovních tříd zohledňovat.

Na přehledu vymezení slovního druhu číslovek v české lingvistické tradici jsme se mohli přesvědčit, že česká lingvistika shodně přiznává číslovkám jejich autonomii v systému slovních druhů, neexistuje však jednotné a v rámci koncepcí jednotně aplikované kritérium. Slovní druh číslovek byl vybrán proto, že se při jeho vyčleňování a určování příslušnosti jednotek často setkáváme s nejasnostmi a nejednotností, plynoucí z vágnosti definic a jeho formální podobnosti s některými slovními druhy.

Rozhodli jsme se proto podrobit prototypické zástupce námi definované třídy lexémů analýze, která využívá prostředky a výhody počítačově zpracovaných souborů textů, jazykových korpusů. Nástroje korpusových manažerů, pomocí nichž z korpusů získáváme data, nám umožňují srovnat nejbližší kontexty na úrovni slovních druhů, do kterých jednotlivé skupiny číslovek vstupují, s kontexty, vyskytujícími se v blízkosti slovních druhů, s nimiž mají číslovky podobné formální charakteristiky. V kapitole 6 jsme se proto pokusili nastínit a ověřit zcela netradiční, korpusovými daty a exaktními metodami podložené kritérium, které se opírá o srovnání nejbližšího kontextu jednotek různých tříd. Výchozí premisou této analýzy byla schopnost bezprostředního kontextu víceméně věrně reflektovat a zároveň determinovat relevantní vlastnosti lexémů (tzn. vlastnosti

morfologické, syntaktické a sémantické). Na základě tohoto předpokladu bylo možné se domnívat, že pokud číslovky tvoří svébytnou a autonomní třídu slov, jejich kontexty, skládající se z obecného značení slovních druhů, budou vykazovat jistou odlišnost ve srovnání s kontexty jim formálně blízkých slovních druhů.

Výhodou kontextové analýzy je, že viditelně nepreferuje žádné z uvedených kritérií slovnědruhé klasifikace. Reflektuje je všechna tak, jak jsou inherentně u jednotlivých lexémů realizována. Neměla by tedy být omezena na jeden konkrétní jazyk.

Číslovky jsme pro účely analýzy rozdělili na třídy, jež vykazovaly formální podobnost se slovními druhy. Při tomto dělení jsme se také opírali o výsledky pilotní analýzy provedené Cvrčkem (2010).

V rámci této analýzy se prokázala velmi silná podobnost levostranného i pravostranného kontextu číslovek numerických, které jsou v odborných publikacích zahrnovány do třídy číslovek základních, s kontextem adjektiv. Ve srovnání s výzkumy založenými na srovnávání oboustranného kontextu (Cvrček 2010), je tato korelace silnější. Další významný výsledek prokázalo testování podobnosti adjektiv s číslovkami se shodnými formálními, tedy adjektivními, vlastnostmi.

Taková podobnost je podle nás podmíněna shodnou formou zkoumaných lemmat, jednak také (především v případě souvislosti číslovek numerických s adjektivy) syntaktickou funkcí, jež je pro tuto formu typická, tedy funkcí přívlastku shodného. Odlišnost kritéria sémantického jsme ve větší míře zaznamenali pouze při komparaci kontextů některých konkrétních lemmat.

Relativně silná souvislost se ukázala i v případě okolí číslovek, které jsme v naší práci označili jako nesklonné a které tradičně patří do třídy číslovek násobných, a adverbii. První pokus, zařadit kvůli adverbiiálnímu významu okolnosti do souboru porovnávaných jednotek i číslovky v mluvnicích označované jako řadové (*poprvé, potřetí*), neukázal na podobnost mezi takto heterogenní skupinou a skupinou adverbii. Zřejmě by tedy bylo třeba v rámci naší práce provést revizi zařazení daného typu číslovek. Po vyřazení číslovek typu *poprvé* a opakování testu, se nepříliš významně snížila korelace levostranného kontextu, zato významnou měrou vzrostla podobnost okolí na pozici pravostranné. Přesto, že se souvislost číslovek násobných a adverbii ukázala o něco slabší, je stále podobnost jejich kontextů a tím pádem i relevantních vlastností, významná. Opět se ale neprojevila na analýze konkrétních lemmat. Tu ovšem pro nedostatečné množství dat považujeme za

vedlejší.

Ve shodě s daty, jež poskytuje předchozí výzkum (ibid.), se neprokázala silnější souvislost mezi číslovkami základními, v této práci nazývanými numerické, a substantivy. Tento výsledek si výše vysvětlujeme především rozdílnými syntaktickými funkcemi, které dané třídy slov zastávají.

Stejně tak nevýznamnou podobnost vykazují kontexty, do nichž vstupují zástupci ze třídy číslovek námi zvaných substantivních a substantiv. Zde se projevil opačný trend než u ostatních porovnávaných skupin. Při analýze konkrétních lemmat se projevil v podobě relativně silné korelace kontextů dvou párů. Takovou souvislost si vysvětlujeme především postupnou substantivizací daných lemmat, která primárně označují, stejně jako substantiva, entity.

Na základě analýzy vzorku poměrně úzce vymezené třídy slov získaného z korpusu SYN2010 a po zhodnocení jejích výsledků ve světle přijatých premis kontextové analýzy, jsme vzhledem ke zmíněné třídě došli k následujícím závěrům. Podobně jako Mluvnice současné češtiny (2010) a Cvrček (2010 a 2011) ve svých výzkumech, v nichž použil postupy, které se od těch našich mírně liší, vyslovujeme pochybnost o jednoznačné autonomii této třídy slov a potřebě ji vymezovat.

V případě významné kontextové podobnosti námi označených číslovek adjektivních s adjektivy a podobnosti okolí číslovek nesklonných (násobných) s adverbii, se v rámci tohoto vzorku přikláníme k označení, které z tradičního pohledu primárně reflektuje charakteristiky formální, sekundárně potom vlastnosti sémantické. Pro termín *číslovky adjektivní* navrhuje inverzi v podobě *adjektiva číslovková*, za vhodnější ale považujeme označení *adjektiva kvantitativní*, pro termín *číslovky nesklonné* potom označení *adverbia kvantitativní*. Zároveň bychom je řadili k těm slovním druhům jež figurují na první pozici názvu.

Do poněkud složitější situace se dostáváme, chceme-li na základě kontextové analýzy navrhnout jednoznačné řešení v případě číslovek, které jsme označovali jako numerické. Kontextová podobnost těchto číslovek s adjektivy je sice velmi silná, zároveň však kromě lemmat označujících číslice 1 - 4 mají specifické paradigma, není tedy mezi nimi a adjektivy formální podobnost. Jazyk však nemůžeme chápat jako soubor diskrétních jednotek, jednotlivých vedle sebe postavených a vzájemně se neovlivňujících paradigmat, ale jako řetězec jednotek interagujících, tvořících syntagmata. K tomuto faktu je třeba

přihlédnout. Jak jsme se u analýzy konkrétních lemmat mohli přesvědčit, okolí slova je schopno reflektovat jak významové, tak formální a syntaktické vlastnosti daného slova. Pokud se tedy kontext, který zohledňuje všechny vlastnosti lexémů a tudíž i všechna kritéria užívaná ke slovnědruhovému klasifikaci, většinou shoduje s kontextem adjektiv, přikláníme se k zařazení této skupiny s označením *adjektiva číselná*.

U skupiny číslovek námi nazvaných substantivní jsme neprokázali významnější souvislost kontextů s jim formálně podobnými substantivy. Z těchto výsledků můžeme vyvodit, že na první pohled zjevná formální podobnost nemusí znamenat bezpodmínečnou příslušnost k téže třídě. Substantiva byla pouze jednou ze skupin, se jejichž kontexty jsme mohli kontexty substantivních číslovek porovnávat. Komparaci je možné (a v případě dalších výzkumů s aplikací uvedené metody i nutné) provést i s adjektivy a zájmeny.

Představili jsme v této práci jeden z dalších možných způsobů zkoumání jazyka a problematiky slovních druhů. Když se vrátíme na začátek této práce do oddílu 4.1, v němž jsme stručně načrtli historii slovnědruhovému klasifikace, narazíme na zmínku o Platónovi, jež byl jedním z prvních, kteří se touto otázkou zabývali. Lexikon rozdělil podle sémantických vlastností na dvě odlišné třídy, jména a slovesa. Protože je však kategorizace¹⁷ na základě významů vágní, nesnadno uchopitelná a umožňující velké množství subjektivních pohledů, postupem času se začala dávat přednost charakteristikám na první pohled patrným, formě a syntaktické funkci. Slovní druhy ale pořád zůstávají kategorií především sémantickou. V případě, že je kontext chápán jako směrodatné kritérium při analýze významu, domníváme se, že by měla kontextová analýza měla být i součástí slovnědruhovému klasifikace.

¹⁷ Tak, jak je používána v rámci klasického vymezování slovních druhů.

8. Bibliografie

Použitá literatura:

- Cvrček, Václav (2010). *Korpusový pohled na postavení číslovek v systému slovních druhů*. In: *Bohemica Olomucensia 2 - Philologica Juvenilia* (red. V. P. Polách). Univerzita Palackého. Olomouc, s. 104–110.
- Cvrček, Václav (et al.) (2010). *Mluvnice současné češtiny*. Karolinum, Praha.
- Cvrček, Václav (2011). *Kontextová analýza a systém slovních druhů*. *Korpus – gramatika – axiologie* 3/2011, (s. 6-32).
- Cvrček, Václav. *Kvantitativní analýza kontextu*, připravuje se
- Crystal, David (ed) (1987). *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge.
- Čechová, Marie (et al.) (1996). *Čeština - řeč a jazyk*. ISV nakladatelství, Praha.
- Čermák, František (2007). *Jazyk a jazykověda*. Karolinum, Praha.
- Černý, Jiří (1996). *Dějiny lingvistiky*. Votobia, Olomouc.
- Daneš, František (1965). *Vztah „centra“ a „perifeire“ jakožto jazykové univerzálie*. In: *Jazykovědné aktuality*, č. 2-3, s. 1-6.
- Daneš., F., Grepl, M., Hlavsa, Z. (eds) (1987). *Mluvnice češtiny 2*. Academia, Praha.
- Encyklopedický slovník češtiny (ESČ)*(2002). Nakladatelství Lidové noviny, Brno.
- Harris, Zellig (1951). *Methods in Structural Linguistics*. University of Chicago, Chicago.
- Hausenblas, Karel (1966). *Spojování číslovek se jménem počítaného předmětu aneb dvacetjedna metrů*. *SlavPrag* 8., s. 301-310.
- Hausenblas, Ondřej (1998). *Poprvé, podruhé, potřetí to jde, proč ne po čtvrté a po sté?* In: *Naše řeč*, roč. 81, s. 53-55.
- Havránek, B., Jedlička, A.(1960). *Česká mluvnice*. SPN, Praha.
- Karlík, Petr a kol. (1995). *Příruční mluvnice češtiny*. Nakladatelství Lidové noviny, Brno.
- Komárek, Miroslav (2006). *Příspěvky k české morfologii*. Periplum, Olomouc.
- Koček, J. - Kopřivová, M. - Kučera, K. (eds.) (2000). *Český národní korpus - úvod a příručka uživatele*. ÚČNK FF UK, Praha.

- Mikulová, Marie (ed.) (2005). *Anotace na tektogramatické rovině Pražského závislostního korpusu*. Anotátorská příručka. ÚFAL MFF UK, Praha
- Palmer, F. R. (ed.) (1968): *Selected papers of J. R. Firth, 1952-1959*. Bloomington, London.
- Paul, Hermann (1920). *Prinzipien der Sprachgeschichte* Max Niemeyer Verlag, Tübingen.
- Poldauf, Ivan (1957). *K vyjadřování kvantity v češtině*. In: Slovo a slovesnost, roč. 18, s. 71-85.
- Saussure, Ferdinand de (2007). *Kurs obecné lingvistiky*. Academia, Praha.
- Schachter, Paul (1985). Parts-of-Speech-System. In: Shopen, Timothy (ed.)(1985). *Language typology and Syntactic Description*. Cambridge University Press, Cambridge

Prameny a internetové zdroje:

- Český národní korpus - SYN2000. Ústav Českého národního korpusu FF UK, Praha 2000. Dostupný z WWW: <<http://www.korpus.cz>>.
- Wessa P. (2012). *Spearman Rank Correlation (v1.0.1)* in Free Statistics Software (v1.1.23-r7). Office for Research Development and Education. Dostupné z WWW: <http://www.wessa.net/rwasp_spearman.wasp/>.