



FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE



UNIVERZITA KARLOVA V PRAZE – FILOZOFICKÁ FAKULTA
ÚSTAV OBECNÉ LINGVISTIKY

Anna Plasová

Počítání pomnožných jmen v polštině

(How to count pluralia tantum in Polish)

BAKALÁŘSKÁ PRÁCE

Praha, 20. ledna 2014

vedoucí práce:
doc. Mirjam Fried, Ph.D.

Děkuji všem, kteří mi pomáhali během psaní této práce – především vedoucí práce
Mirjam Fried a své rodině.

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

Ve Vratislavi dne 19. 1. 2014

Podpis

Souhlasím se zapůjčením této bakalářské práce ke studijním účelům.

Abstrakt

Pluralia tantum – podstatná jména, která nemají kategorii singuláru – tvoří v rámci polských substantiv oddělenou skupinu, a to především s ohledem na zvláštní způsob kvantifikace, jež pro tato substantiva předepisuje jazyková norma: počítání pomocí specifického typu číslovek, tzv. číslovek souborových.

Protože je však paradigma souborových číslovek málo frekventované a pro rodilé mluvčí obtížné, objevují se v praxi různé kompenzační strategie, pomocí nichž se mluvčí problematikým číslovkám vyhýbají a nahrazují je analytickými konstrukcemi (např. *dwie pary wrót* místo *dwoje wrót*).

V teoretické části práce autorka shrnuje dosavadní stav výzkumu v oblasti počítání polských pomnožných jmen zejména na základě článků polonistů Zygmunta Saloniho a Anny Andrejczuk.

V druhé, praktické části práce autorka na datech získaných z Národního korpusu polského jazyka provádí vlastní výzkum s cílem zjistit, jak mluvčí polštiny skutečně postupují, když chtějí *pluralia tantum* kvantifikovat. Kromě opisných konstrukcí nachází také časté použití základních číslovek, což je v přímém rozporu s jazykovou normou.

Užívání souborových číslovek se i přes to, že je norma vyžaduje, zdá být jevem, jehož frekvence je na poklesu, a to zejména v méně formálních vrstvách jazyka. Úzus se tak dostává do konfliktu s normou, od níž se jazykové povědomí mluvčích evidentně vzdaluje.

Klíčová slova: pomnožná jména, číslovky, korpus, polština.

Abstrakt

Pluralia tantum, czyli rzeczowniki niemające gramatycznej kategorii liczby pojedynczej, stanowią w ramach polskich rzeczowników odrębną grupę, przede wszystkim ze względu na fakt, że norma wzorcowa wymaga stosowania specyficznej odmiany liczebnika – liczebnika zbiorowego – jeżeli chce się ich kwantyfikować.

Paradygmaty liczebników zbiorowych mają jednak małą frekwencję i dla tego stanowią dla rodzimych użytkowników języka kłopot. Z tego powodu pojawiają się różne strategie kompensacyjne mające na celu ominąć kłopotliwą formę liczebnika zbiorowego i zastąpić ją konstrukcją opisową (np. dwie pary wrót zamiast dwoje wrót).

W części teoretycznej pracy autorka przedstawia dotychczasowy stan badań dotyczących problemu liczenia polskich pluraliów tantum, i to przede wszystkim na podstawie artykułów autorstwa polonistów Zygmunta Saloniego i Anny Andrejczuk.

W drugiej, praktycznej części pracy przeprowadza autorka własne badania w oparciu o dane uzyskane z Narodowego Korpusu Języka Polskiego. Ma na celu odkryć, jak rzeczywiście postępują użytkownicy polszczyzny, kiedy chcą wyrazić liczbę jakiegoś plurale tantum. Oprócz konstrukcji analitycznych znajduje również liczne przypadki zastosowania liczebnika głównego, co jest wyraźnie sprzeczne z normą wzorcową.

Stosowanie liczebników zbiorowych – choć ich norma wciąż wymusza – wydaje się zjawiskiem, którego frekwencja się obniża, przede wszystkim, jeżeli chodzi o mniej formalne poziomy języka. W ten sposób uzus wchodzi w konflikt z normą wzorcową, od której świadomość językowa rodzimych użytkowników języka ewidentnie się oddala.

Słowa kluczowe: pluralia tantum, liczebniki, korpus, język polski.

Abstract

Among Polish nouns, *pluralia tantum*—i.e. nouns that lack a singular form—form a distinct group, especially in terms of the specific requirements that the language standard places on their quantifiers: they must be quantified using a special type of numerals, so-called collective numerals. However, since collective numerals are used only rarely, native speakers are not well-acquainted with their inflectional paradigm; this leads to various compensatory strategies which speakers resort to in order to avoid these problematic numerals, replacing them with analytic constructions (e.g. *dwie pary wrót* instead of *dwoje wrót*).

In the first part of the thesis, the author reviews what is currently known about quantification patterns of Polish *pluralia tantum*, drawing mostly on articles by Polish linguists Zygmunt Saloni and Anna Andrejczuk.

In the second, experimental part, the author explores data extracted from the National Corpus of the Polish Language, aiming to uncover the strategies of *pluralia tantum* quantification actually employed by speakers of Polish. Apart from periphrastic constructions, she finds that cardinal numerals are also frequently used, which directly contradicts the language standard.

Though the standard may require it, the usage frequency of collective numerals seems to be on the decline, especially in less formal registers of the language. The speakers' linguistic intuition is thus clearly shifting and creating a discrepancy between language standard and practice.

Keywords: pluralia tantum, numerals, corpus, Polish language.

Obsah

Abstrakt.....	5
Abstrakt.....	6
Abstract.....	7
1 Úvod.....	9
2 Teoretická část.....	11
2.1 Liczebniki zbiorowe v polštině.....	11
2.2 Pluralia tantum v polštině.....	14
2.3 Vztahy mezi normou a územ.....	20
3 Praktická část.....	24
3.1 Zdroj jazykových dat: korpus NKJP.....	24
3.2 Sběr dat.....	26
3.2.1 Obecné otázky.....	26
3.2.2 Seznam PT pro vyhledávání.....	26
3.2.3 Vyhledávání a zpracování výsledků.....	28
3.3 Analýza dat.....	31
3.3.1 Dvě skupiny PT.....	31
3.3.2 Jednotlivé typy kvantifikace.....	34
3.3.3 Vztahy mezi normou a územ II.....	45
4 Závěr.....	49
5 Bibliografie.....	51

1 Úvod

Kategorie *pluralia tantum* (pomnožná podstatná jména) v polštině zahrnuje relativně velký počet různých slov, z nichž některá jsou velice frekventovaná (např. *drzwi* – dveře nebo *spodnie* – kalhoty). Norma spisovné polštiny stanovuje, že se tato slova mohou pojít pouze se specifickým typem číslovek, tzv. číslovkami souborovými, které jsou však pro uživatele jazyka poměrně nezvyklé a obtížné. To způsobuje, že se jim mluvčí polštiny pomocí různých strategií vyhýbají. Jak se to stává v případech, kdy norma trvá na zachování pravidel, která jsou pro mluvčí příliš obtížná, netransparentní či jednoduše málo povědomá, uživatelé jazyka si hledají jiné prostředky, jak vyjádřit zamýšlený význam a daná pravidla při tom obejít, případně je úplně pomíjejí.

Cílem této práce je pomocí Národního korpusu polského jazyka (NKJP) prozkoumat, jak se konkrétní pomnožná jména opravdu počítají ve skutečných komunikačních situacích. Je pravděpodobné, že zjištěné výsledky se budou značně lišit od vstupních předpokladů stanovených normou – preskriptivní příručky, které používání souborových číslovek prosazují, vždy musejí nutně zaostávat za tím, jak lidé jazyk používají v praxi. Výzkum realizovaný ve druhé části této práce tak může pomoci uvést na pravou míru některé mýty a mylné představy, které ohledně této oblasti polské gramatiky přetrvávají a které se odrážejí např. ve vyučovacích osnovách polského jazyka pro rodilé mluvčí i pro cizince.

Poznámka:

Všechny překlady polských a anglických citací jsem pořídila sama. Používaná polská slova či věty jsou vždy označeny *kurzívou*; překládám je při prvním výskytu, při opakovaném pouze v případě, že je zvláště třeba na význam upozornit. Některá polská slova, která jsou co do formy i významu totožná se svými českými protějšky, ponechávám pro jednoduchost bez překladu (např. pol. *jeden* = čes. jeden). Překlady příkladových slovních spojení či vět se vždy snaží co nejpřesněji přiblížit polský originál, a to i za cenu jazykové nesprávnosti v češtině (např. *tři housle, pokud je v polském znění *trzy skrzypce*).

Používané zkratky:

- pády: nom., gen., dat., acc., voc., loc., inst.
- LZ = liczebniki zbiorowe = souborové číslovky (polské)
- PT = pluralia tantum = pomnožná substantiva (polská)
- LG = liczebniki główne = základní číslovky (polské)
- sg. = singulár (gramatické jednotné číslo)
- pl. = plurál (gramatické množné číslo)

2 Teoretická část

2.1 Liczebniki zbiorowe v polštině

Slovnědruhov¹á kategorie číslovek (pol. *liczebniki*) patří v polštině podobně jako v češtině k nejvíce heterogenním v celém systému, což je důsledkem jejich komplikovaného a stále ještě neukončeného diachronního vývoje (Siuciak 2008, s. 14–15). Jednotlivá slova jsou do ní zařazena na základě sémantického klíče – všechna vyjadřují počet, množství, pořadí, počet opakování atp., a to buď přesně (*jeden, dwa, sto, milion*), nebo přibližně (neurčité výrazy množství typu *mało* – málo, *kilka* – několik), *trochę* – trochu). Z morfologického hlediska jsou ovšem nesourodé, rozeznáváme jak číslovky substantivního typu (*jeden, milion*), tak číslovky sdílející morfologii s adjektivy (*trzeci* – třetí, *piąty* – pátý, *kilkukrotny* – několikanásobný) nebo adverbii (*dwukrotnie* – dvakrát). Některé číslovky jsou nesklonné (*óół* – půl), jiné mají vlastní paradigma (*pięcioro* – patero).

Zvláštní skupinou v rámci polských číslovek jsou tzv. *liczebniki zbiorowe* (zkratka LZ). Jak dále uvidíme, tato skupina není zcela totožná s českými číslovkami souborovými, ačkoli s nimi má mnoho společného, avšak jejich název je vlastně ekvivalentní (v zájmu plynulosti textu budu tedy pro tyto polské číslovky používat české označení *souborové*). Mají vlastní paradigma a existují paralelně ke každé základní číslovce od 2 do 99¹ (*dwa* – *dwoje*, *pięc* – *pięcioro*, *dziewiętnaście* – *dziewiętnaścioro*), ale také k některým dalším výrazům množství (*oba* – *oboje*, *kilka* – *kilkoro*).

Jako příklad uvedu paradigma číslovek *dwoje* a *pięcioro* (historicky se skloňování číslovek 1–4 a 5+ vyvinulo z rozdílných základů,² ale dnes jsou pádové koncovky kromě nom./acc./voc. totožné):

¹ Pokud je číslovka vyšší, skládá se ze základní číslovky označující všechny řády kromě desítek a jednotek doplněné souborovou číslovkou, která označuje jeden/dva nejmenší řády.

² Všeobecně panuje shoda, že číslovky 1–4 jsou adverbialního původu (viz např. Siuciak 2008, s. 12), zatímco ty vyšší se vyvinuly ze substantiv.

Nom.	<i>dwoje</i>	<i>pięcioro</i>
Gen.	<i>dwojga</i>	<i>pięcioroga</i>
Dat.	<i>dwojgu</i>	<i>pięciorogu</i>
Acc.	<i>dwoje</i>	<i>pięcioro</i>
Loc.	<i>dwojgu</i>	<i>pięciorogu</i>
Instr.	<i>dwojgiem</i>	<i>pięciorogiem</i>
(Voc.	<i>dwoje</i>	<i>pięcioro</i>) ³

Ve tvarech gen., dat., loc. a instr. se kmen číslovky rozšiřuje o epentetickou souhlásku -g-, která v této pozici na rodilé mluvčí polštiny (podle mých zkušeností) působí neobvykle či „exoticky“. Toto -g- předcházející bezprostředně flektivní koncovku je pro souborové číslovky typické a plní tedy funkci „poznávacího znamení“⁴ (stejně jako koncovka -oro v nom. LZ od 5 výše). Nezvyklost tohoto paradigmatu v kombinaci s relativně nízkou frekvencí výskytu však zároveň zvyšuje obtížnost jeho osvojení pro rodilé mluvčí. O tom, jaké to má důsledky, pohovořím podrobněji v následujících kapitolách.

Užití souborových číslovek je v moderní spisovné polštině omezeno na čtyři typy kontextů:

a) Počítaným objektem jsou děti či jiná mláďata (pol. *nazwy istot młodych*) bez určeného pohlaví (Siuciak 2008, s. 42) zakončené v nom. sg. nosovkou -ę (WSPP, s. 1676). Vznikají tak kolokace jako např. *dwoje dzieci* (dvě děti), *troje źrebiąt* (tři hřbata), *pięcioro cieląt* (pět telat), *dzięsięcioro kociąt* (deset koťat). Podle stávající jazykové normy nelze s touto skupinou substantiv spojovat základní číslovky, tedy např. kolokace **pięc dzieci* je chybná.⁵

b) Počítaný objekt je pomnožné substantivum a je tudíž „defektní“ v oblasti některých gramatických kategorií. V tomto případě jde primárně o chybějící singulár, ale

³ Vokativ je v polštině v podstatě vždy shodný s nominativem (kromě některých vlastních jmen, oslovení, názvů funkcí a hodnot atp.), je tedy otázka, zda ho vůbec ve výčtu uvádět.

⁴ Ve skutečnosti vznikl tento jev kontaminací, protože souborové číslovky 2–4 se v praslovanštině původně skloňovaly podle jiných paradigmat než souborové číslovky 5+. Opět to souvisí s odlišným původem číslovek základních, viz pozn. č. 2 (Siuciak 2008, s. 139–140).

⁵ O rozporu mezi normou a územ pojednává oddíl 2.3.

k tomu se navíc připojuje i defektnost v oblasti rodu – polština v plurálu rozlišuje pouze dva rody: *męskoosobowy*, čili maskulina denotující osoby × *niemęskoosobowy*, čili ostatní maskulina, feminina a neutra (Siuciak 2008, s. 44). Pomnožná jména tedy nepočítáme pomocí základních číslovek (**dwie drzwi*), ale s využitím LZ: *troje skrzypiec* (troje housle), *pięcioro nożyczek* (patery nůžky), *sześcioro grabi* (šestery hrábě).

Vyjadřování počtu u pomnožných jmen je však komplikováno tím, že kromě číslovek se s nimi často pojí i různé nečíslovkové výrazy množství. Přesnější přehled o tomto problému, který leží v centru zájmu této práce, včetně dělení pomnožných jmen do různých skupin, podává následující kapitola.

c) Třetí případ užití souborových číslovek uvedu tímto příkladem⁶:

Na egzaminie było pięciu studentów. (Na zkoušce bylo pět studentů.)

Na egzaminie było pięcioro studentów. (Na zkoušce bylo pět (patero?) studentů.)

Čím se liší tyto dvě věty? Druhá z nich obsahuje informaci navíc, kterou překlad do češtiny sám o sobě nedokáže zachytit: použití souborové číslovky jasně signalizuje, že skupina osob, o níž se hovoří, byla co do pohlaví smíšená – obsahovala nejméně jednoho muže a nejméně jednu ženu. V českém překladu by tedy bylo přesnější uvést například: „Na zkoušce bylo pět studentů a studentek.“ První věta v tomto ohledu takto jednoznačná není. Použitý tvar číslovky (*pięciu*) je sice vyhrazen pouze pro mužský osobní (životný) rod (oproti *pięć* pro mužský neosobní, ženský a střední rod), ten však zároveň slouží i jako generické maskulinum. Informaci, zda se jednalo jen o muže, či zda se jednalo o skupinu složenou z neznámého počtu mužů i žen (mluvčí informaci nezná nebo nechce sdělit), tedy z této věty nelze bez dalšího kontextu jednoznačně vyvodit.

V tomto případě se tedy LZ pojí s podstatnými jmény mužského rodu v plurálu, která fungují jako generická maskulina – např. *studenci* (studenti), *goście* (hosté), *znajomi* (známí), *koledzy* (kolegové/přátelé) apod. Volba souborové číslovky je zde tedy příznaková a záměrná, na rozdíl od generického maskulina explicituje složení popisované skupiny. Jak upozorňuje Saloni (2009, s. 61), tento jev se týká pouze

⁶ Příklad byl převzat z Saloni 2009, s. 61.

maskulin – feminina *osoba* (osoba) či *istota* (bytosť), která také označují člověka obecně bez inherentní specifikace pohlaví, se s LZ tímto způsobem nikdy nepojí.

Ačkoli toto užití může působit zvláště, ve skutečnosti je naopak nejbližší původnímu významu souborových číslovek v praslovanštině: označovaly zde totiž soubory *různorodých* prvků (Siuciak 2008, s. 42).⁷ Nevyjadřovaly tedy pouze kvantitu (to byla dokonce původně spíše druhotná funkce), měly specifický sémantický obsah – a v případě tohoto kontextového užití si jej zachovaly i v moderní polštině.

d) Posledním případem, kde se dnes můžeme s těmito číslovkami setkat, jsou ustrnulá slovní spojení, kde se *liczebniki zbiorowe* objevují z historických důvodů i vedle substantiv, která aktuálně⁸ žádnému z výše uvedených kritérií nevyhovují. Tento případ pro náš další výzkum není příliš podstatný, ale pro úplnost je třeba se o něm zmínit. Poměrně běžně se lze setkat s archaickými kolokacemi *dziesięcioro przykazań* (desatero přikázání), *doktor obojga praw* (doktor práva obojího) nebo *Rzeczpospolita obojga narodów* (Republika obou národů – starý název Polska v období unie s Velkoknížectvím litevským).

Pokusme se nyní o shrnutí: když pomíneme lexikalizované kolokace uvedené pod bodem d), nahlédneme, že souborové číslovky v současné polštině v důsledku historického vývoje obvykle označují denotáty, které nejsou pohlavně vyhraněné či určené (neutrum vlastní nedospělým živočichům, „potlačený“ rod pomnožných jmen, smíšené skupiny lidí obojího pohlaví). Ačkoli to nemusí být na první pohled zřejmé, uvedené skupiny prvků a), b) a c) mají tedy mnoho společných rysů.

2.2 Pluralia tantum v polštině

Nejjednodušší definice *pluralií tantum* (zkratka PT) spočívá vlastně v překladu jejich latinského názvu: jsou to podstatná jména, která v rámci gramatické kategorie čísla disponují pouze plurálem. Klasickými příklady jsou slova jako *drzwi* (dveře), *sanie*

⁷ Významový odstín se pokusím ilustrovat následujícím příkladem. Je z něj patrné, že tuto funkci LZ dnes musíme nahradit např. adjektivem:

„Koupil jsem pět hrnců“ = Koupil jsem pět stejných hrnců.

„Koupil jsem patero hrnců.“ = Koupil jsem pět různých hrnců.

⁸ Otázka, jak se do těchto spojení LZ vlastně dostaly a podle jakých kritérií se volba druhu číslovky řídila ve starších vývojových fázích polštiny, bohužel přesahuje rozměry této práce.

(sáně), *skrzypce* (housle), *spodnie* (kalhoty), *pomyje* (pomyje), *wakacje* (prázdniny).⁹ Hanna Jadacka (2006) upozorňuje na existenci slov typu *lody* (zmrzlina), která by zdánlivě mohla této definici odporovat. K těmto výrazům jako by existoval singulár – v našem případě *lód* – ale ten již má jiný význam. Znamená totiž „led“ a samozřejmě je možné jej pluralizovat ve významu „ledy“ (např. *Lody ruszają w deltę rzeki.* – Ledy plují do ústí řeky.). Slovo *lody* ve významu „zmrzlina“ je tedy pomnožné, stejně jako slova *nauszniki* (klapky na uši), *akty* (právní spisy) apod., protože slova *nausznik*, *akt*, *lód* mají jiný význam.

Obvyklý dojem každého mluvčího, že tato skupina slov je dosti nepočtená, uvádí Jadacka (2006, s. 161) na pravou míru: podle ní představují „podstatná jména, která nemají jednotné číslo, početnou skupinu čítající okolo tisíce lexémů (nepočítaje v to vlastní jména¹⁰)“. Tento soubor je vnitřně heterogenní jak po morfologické (*juwenilia*, *bakalie*, *łaskotki*, *koszary*), tak po sémantické stránce (*dewocjonalia* – devocionálie, *grabie* – hrábě, *targi* – veletrh, *kuluary* – kuloáry, *okulary* – brýle).

Pokusme se nyní rozdělit tato slova do několika skupin tak, aby nám toto dělení bylo užitečné v dalším výzkumu. Mělo by odrážet morfologické, syntaktické či sémantické charakteristiky jednotlivých typů pomnožných jmen. V následujících odstavcích budou představeny dva návrhy polských autorů – jednodušší schéma Zygmunta Saloniho (1976) a více rozvětvené, na Saloniho navazující schéma Anny Andrejczuk (2011).

Prvním a nejjednodušším způsobem, jak tuto slovní zásobu rozdělit, je kritérium životnosti. V polštině je – podobně jako v češtině – kategorie životnosti/neživotnosti vyhrazena pouze substantivům mužského rodu, a to jak v singuláru, tak v plurálu. Jejím ukazatelem je tvar ukazovacího zájmena *ten* (*ci* [život.] vs. *te* [neživot.], např. *ci rodzice* vs. *te drzwi*). Právě toto kritérium bylo zásadní pro Z. Saloniho.

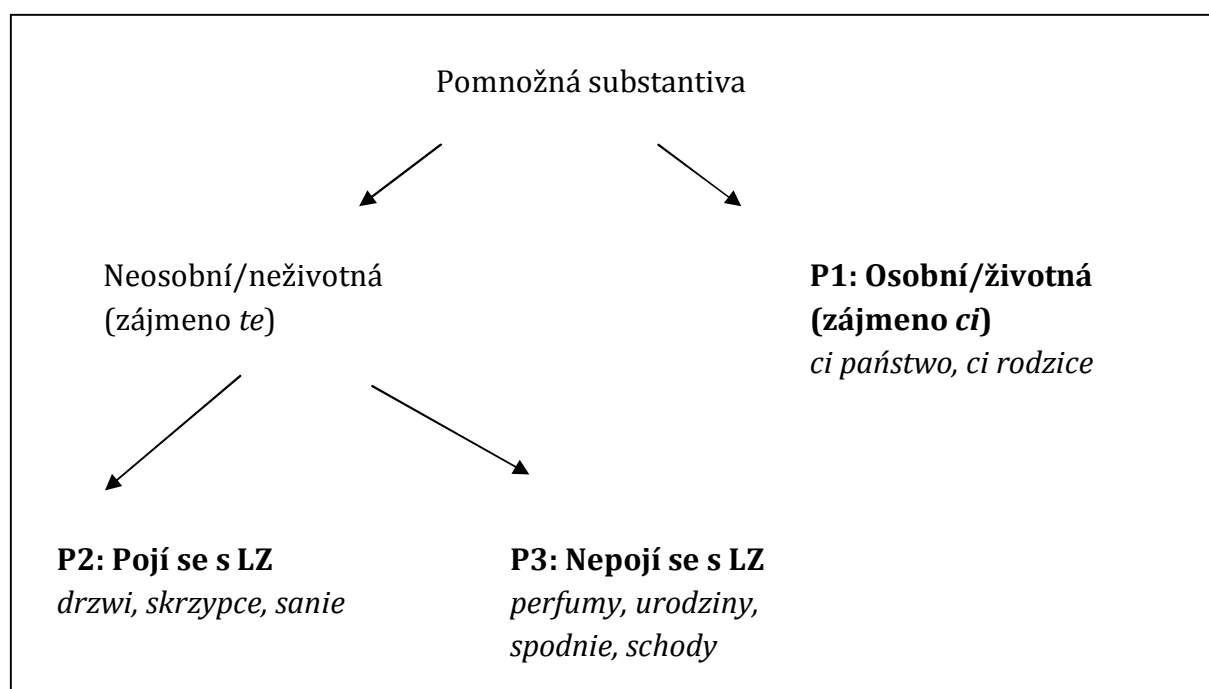
Jak vidíme na schématu 1, Saloni nejprve rozděлил pomnožná jména na životná a neživotná. Životnými se dále nezabýval, zatímco skupinu neživotných rozděлил na dvě podle jediného klíčového kritéria: zda se při vyjadřování jejich množství používá

⁹ Uvedené příklady jsem převzala z Jadacka 2006. Všem odpovídají v češtině rovněž pomnožná slova – většinou se historicky jedná o tytéž lexémy.

¹⁰ Pomnožná vlastní jména nebudou předmětem mého zájmu, věnuji jim proto jen krátkou poznámku. Převážně se jedná o toponyma (*Tychy*, *Katowice* apod.), u nichž otázka vyjadřování množství ztrácí smysl. (Pomiňme sice představitelné, ale vykonstruované formulace typu: „Staré Katowice se mi líbí víc než Nové Katowice – ale oboje Katowice jsou příjemné místo k životu.“) Tato substantiva jsou zajímavá spíše z hlediska problematiky určování rodu, jelikož rod má vliv na výběr vhodné koncovky genitivu (*do Tych* [masc.] vs. *do Tychów* [fem.]).

souborová číslovka (P2) či ne (P3). Upozorňuje přitom, že ve skupině P3 jsou jak výrazy, které se počítají „pomocí výrazů umožňujících kvantifikaci, např. *para*“, tak výrazy, které „lze jen těžko užít v konstrukci s číslovkou“ (Saloni 1967, s. 58–59). Jinými slovy, nezabývá se kategorií počitatelnosti, ale sleduje pouze možnost použití souborové číslovky. Pokud jde o sémantiku, tato skupina výrazů je velmi heterogenní: obsahuje jak názvy předmětů (*spodnie*), tak názvy slavnostních událostí, např. *urodziny* (narozeniny), *chrzciny* (křtiny) nebo substantiva látková (*pomyje, perfumy*).

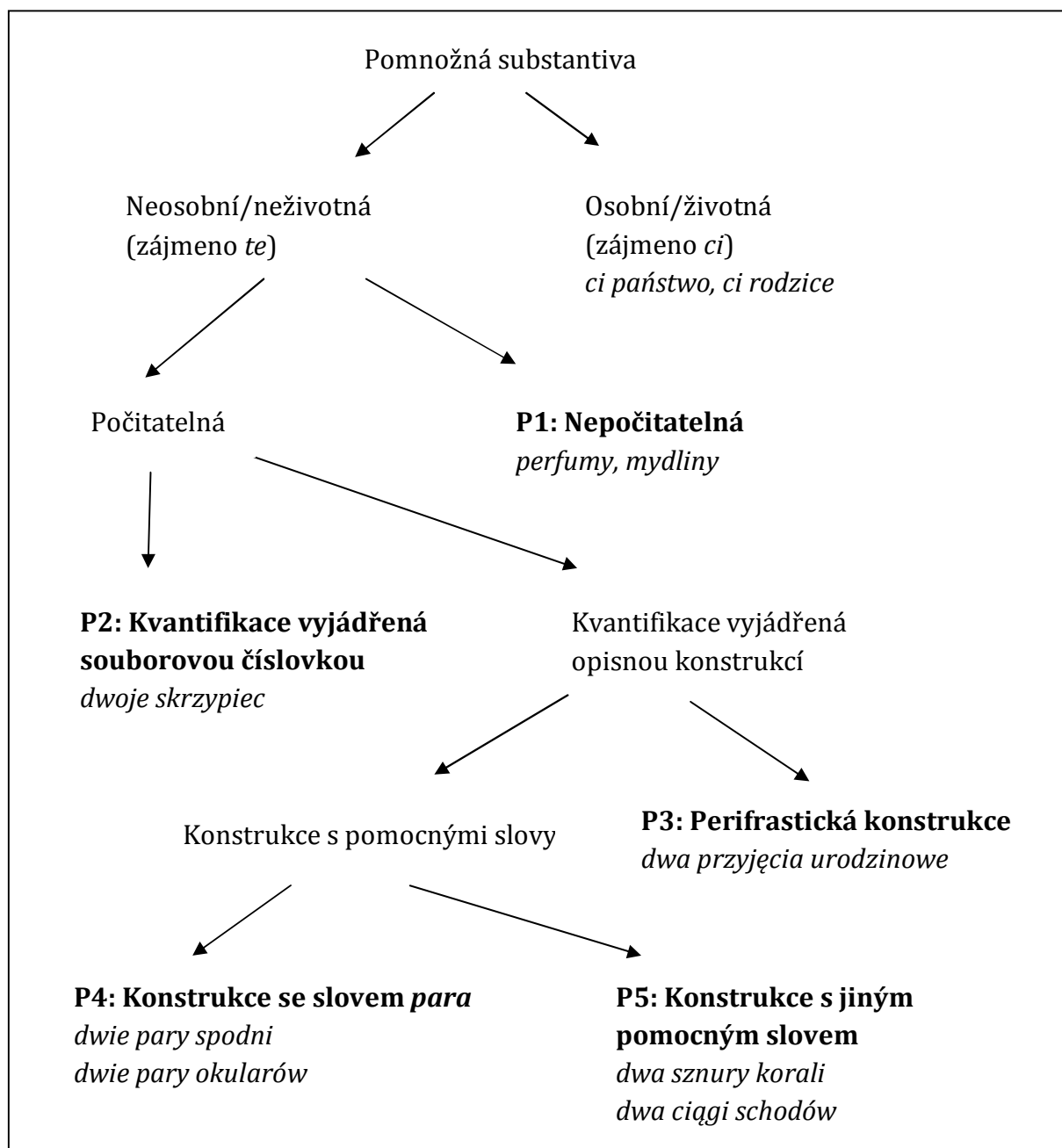
Schéma 1: rozdělení pomnožných jmen do skupin dle Saloni 1976



Je zřejmé, že dělení představené Z. Salonim, které je založeno na dichotomii životnost × neživotnost, neposkytuje příliš mnoho informací o tom, co je předmětem této práce – identifikovat, jak konkrétně se počítají pomnožná jména. Přinejmenším však na tomto místě můžeme říci, že pomnožná jména se budou dělit nejméně do tří zřetelně odlišných skupin a způsoby jejich kvantifikování budou různé, alespoň co se možnosti užití LZ týče.

Druhým kritériem, které se nabízí při snaze pomnožná jména nějak klasifikovat, je jejich počitatelnost a konkrétní způsob počítání. Autorka níže představeného schématu Anna Andrejczuk (2011) vyšla ze Salonioho koncepce a rozšířila ji právě o dělení s ohledem na počitatelnost a konkrétní způsoby vyjadřování počtu.

Schéma 2: rozdělení pomnožných jmen do skupin dle Andrejczuk 2011



Jak je patrné, autorka vychází ze Saloniho: nejprve od sledovaných substantiv oddělí životná jména (*rodzice* ap.). V dalším kroku rozčlení zbývající substantiva (tedy neživotná PT) podle kritéria počítatelnosti. Tak vzniká skupina P1 obsahující nepočítatelné výrazy typu *perfumy, mydliny* (vůně, mydliny).

Pro další dělení se Andrejczuk vrátila k Saloniho schématu a vymezila P2 – skupinu počítatelných neživotných PT, která se počítají pomocí souborové číslovky. Tato skupina

v podstatě odpovídá Saloniho skupině P2 (protože nepočitatelná jména, která Andrejczuk na rozdíl od Saloniho oddělila v předchozím kroku, by sem v každém případě nespádala, jelikož se pomocí LZ v každém případě nepočítají).

Jakmile určíme skupinu „prototypických“ pomnožných jmen počítaných pomocí číslovek souborových, zbývá nám už jen pokusit se vhodně rozdělit zbývající výrazy – tedy pomnožná neživotná PT, která jsou sice počitatelná, ale počítají se určitým „nestandardním“ způsobem. Anna Andrejczuk na základě práce se slovníky i korpusovým materiálem navrhuje tři skupiny.

(1) Skupina P3: Výraz je kvantifikován pomocí perifrastické konstrukce. Daný jev je totiž sice teoreticky počitatelný, ale v praxi to působí poněkud nezvykle, pročež se volí opisná formulace. Ilustruje to uvedený příklad *dwa przyjęcia urodzinowe* (dvě narozeninové oslavy). Andrejczuk to ostatně komentuje i titulem svého článku, který je uveden citací rodilého mluvčího *Dwoje urodzin to brzmi dziwnie* (Dvoje narozeniny, to zní divně.)

(2) Skupina P4: Konstrukce s pomocným slovem *para* (pár). Tento způsob kvantifikace se týká pojmů, které popisují předměty mající zdvojený, symetrický charakter. Nejčastěji se jedná o oblečení či doplňky nošené na párových částech těla (*spodnie, okulary*); jak ostatně uvádí i Zygmunt Saloni (2009, s. 58–59): „This group [...] contains the names of clothes (and similar objects) having a symmetrical structure: *spodnie* ‘trousers’, *majtki* ‘underpants’ [...], *okulary* ‘glasses’ (their dual structure is seen in the *plurale tantum* form, also in English equivalents).“¹¹

(3) Skupina P5: Konstrukce s jinými pomocnými slovy. Týkají se především pomnožných slov označujících předměty, které se skládají z většího množství jednotlivých prvků. Pomocné slovo se pak typicky odvolává na způsob organizace těchto předmětů, který zase souvisí s jejich podobou, viz uvedené příklady *dwa sznury koralí* (dvě šňůry korálí), *dwa ciagi schodów* (dvě řady schodů). Podle toho, jak často a jak systematicky jsou tato pomocná slova užívána, se může velmi lišit status, který jim přisoudíme: škála sahá od formulací vzniklých *ad hoc* podle momentální potřeby mluvčího až po pravidelně se vyskytující kvantifikátory podobného charakteru, jako jsou

¹¹ „Tato skupina [...] obsahuje názvy oblečení (a podobných objektů), které mají symetrickou strukturu: *spodnie* (kalhoty), *majtki* (kalhotky) [...], *okulary* (brýle); jejich duální struktura se odráží v tom, že jsou tato substantiva pomnožná, stejně jako jejich anglické ekvivalenty.“

ty, které známe např. z japonštiny.¹² V praktické části této práce se pokusíme doplnit tuto hypotézu konkrétními příklady z polského korpusu, které by mohly upřesnit naši představu o tom, jak tento způsob počítání PT funguje.

V praktické části této studie budu při zpracovávání jazykových dat vycházet především z druhého uvedeného dělení podle Anny Andrejczuk. Zejména mě budou zajímat skupiny P2, P4 a P5: slova z těchto skupin se skutečně počítají v užším slova smyslu, a to buď pomocí souborových číslovek (P2), nebo pomocných slov se základní číslovkou (P4, P5). Problematický status (podle Andrejczuk dokonce *kontroverzní*) těchto „nestandardních“ způsobů vyjadřování množství u části pomnožných jmen vzbuzuje mnoho otázek a rozhodně není zcela prozkoumán. Budu vycházet z pracovní hypotézy, že všechna zkoumaná slova bude možné zařadit do jedné z těchto skupin. Zároveň bude schéma Anny Andrejczuk otestováno v praxi, přičemž se ukáže, zda je vhodným nástrojem klasifikace PT či nikoli.

Skupinami P1 a P3 se v dalším výzkumu přímo zabývat nebudu. Skupina P1 obsahuje nepočítatelné výrazy, jejichž množství se vyjadřuje zcela jiným způsobem a vždy se týká určitého množství látky (např. „vědro pomyjí“ apod.). Skupina P3 potom obsahuje výrazy, které se nepojí s výrazy množství v užším slova smyslu. Perifrastická vyjádření jsou velmi proměnlivá, až příliš variabilní na to, aby se dala tímto způsobem zkoumat. Anna Andrejczuk (2011, s. 278–279) uvádí, že v nich nelze vysledovat jednoznačný vzorec („nápaditost rodilých mluvčích, pokud jde o tvoření různých, nikoli nutně standardních opisných konstrukcí, je tak veliká, že autorka tohoto článku nebyla schopna odhalit nejtypičtější vzorce“).

Na závěr této kapitoly ve stručnosti představím ještě dělení pomnožných jmen čistě podle způsobu jejich počítání, které navrhuje ve svém článku Hanna Jadacka (2006). Text se kvantifikací PT zabývá spíše okrajově, především se orientuje na polemiku s definicí pomnožných jmen a kritiku konkrétních lexikografických přístupů uplatněných v různých polských slovnících. Autorka však zřejmě jako první badatel vůbec pořídila pomocí retrográdně řazeného slovníku v rámci možností kompletní seznam polských

¹² „Numerativy (pomocná číselná jména) jsou nesamostatná slova, která se připojují k číslovkám. [...] Vhodný numerativ se volí podle typu a vlastností počítaných objektů – odlišné numerativy používáme například při počítání lidí, různých druhů zvířat, strojů, podlouhlých či plochých předmětů atp.“ (Nymburská, Vostrá, Sawatari 2007, str. 141).
Např. *kami san mai* = tři papíry/tři listy papíru; *kami*=papír, *san*=tři, *mai*=numerativ pro „ploché tenké předměty“ (Shimamori, 1994, s. 132).

PT. Když tento korpus prozkoumala, vyvrátila hned několik stereotypních tvrzení o této skupině výrazů, která se stále opakují v gramatických příručkách.

Jadacka precizně prozkoumala, jaká slova patří do skupiny pomnožných substantiv, a představuje tři způsoby počítání pomnožných jmen (s. 166–167), které se v podstatě shodují s výše reprodukováným schématem Anny Andrejczuk, čímž její hypotézu posilují. Tyto tři způsoby jsou:

(1) souborové číslovky,

(2) slovo *para*,

(3) jiný způsob – sem spadají perifrastické konstrukce (přičemž Jadacka podotýká, že se nejedná o počítání v užším slova smyslu) a konstrukce s různými pomocnými slovy.

2.3 Vztahy mezi normou a územ

Jak jsem již předznamenala výše, pro rodilé mluvčí polštiny je paradigma souborových číslovek obtížný jev, který zdaleka ne všichni v úplnosti a správně ovládají. V praxi tedy můžeme často slyšet nesprávné spojení pomnožného jména s číslovkou, a to zejména pokud se jedná o méně frekventovanou kolokaci, kterou mluvčí nemá zažitou jako celek. Sousedství *dwoje drzwi* (dvoje dveře) tedy s vyšší pravděpodobností uslyšíme v souladu s normou, protože se jedná o běžné slovní spojení. Vůbec nejméně se podle mých osobních poznatků z polskojazyčného prostředí chybí, když mluvčí uvádí, kolik má dětí, protože tuto formulaci má sám dobře zažitou, navíc ji slyší i od jiných lidí. Naopak v hovoru o něčem méně obvyklém, například *o piętnaściorgu skrzypiec* (o patnáctero houslích) se mluvčí podle mých praktických zkušeností i tvrzení řady polonistů či učitelů polského jazyka, s nimiž jsem tento problém konzultovala, buďto pokusí souborové číslovce nějakým způsobem vyhnout (viz dále), nebo ji utvoří v rozporu s normou (nesprávný či přímo neexistující tvar).

Ve spojitosti se souborovými číslovkami se často vyskytuje ještě druhý podobně komplikovaný jev. Tím jsou podstatná jména označující mláďata živočichů (včetně lidí), u nichž norma vyžaduje povinně použití LZ (viz kapitola 2.1, bod a)). Tato sémanticky vymezená skupina neuter má zároveň i své vlastní paradigma, v polštině dosti vzácné a pro mluvčí obtížné. Charakterizuje je specifická a typická koncovka nom. sg. –ę (*cielę*,

kocię, dziewczę – tele, kotě, děvče) a rozšiřování kmene o skupinu –ęć– v sg. (gen. *cielęcia, kocięcia, dziewczęcia*) a –ęt–/–ąt– v pl. (nom. *cielęta, kocięta, dziewczęta*, gen. *cieląt, kociąt, dziewcząt*).¹³ Co do typu flexe tato skupina slov tedy odpovídá českým neutřům vzoru kuře: kuř-at-a, hous-at-a, děvč-at-a. Co do absolutního počtu slov je však podstatně menší (sémanticky omezená) a tedy v řeči méně frekventovaná.

Funkční alternativu pro mluvčí představuje původně velkopolská nářeční varianta *cielak, kociak* (místo spisovného *cielę, kocię*), jejíž skloňování probíhá standardně podle vzoru jiných maskulin a kterou norma povoluje užívat ve spojení se základními číslovkami (počítáme tedy *jeden cielak, dwa kociaki, dziesięć kociaków*).

Objektivní oporu tvrzení o komplikovanosti souborových číslovek i výše zmíněné skupiny neuter poskytují vzdělávací osnovy polských základních škol: oba uvedené gramatické jevy jsou během školní docházky vyučovány v rámci hodin mateřského jazyka. Kromě toho, jak jsem se osobně přesvědčila při svém studiu na Vratislavské univerzitě (říjen 2013 – únor 2014), probírají tyto problematické formy i místní studenti druhého ročníku polonistiky a přestože se jedná o výběr spíše nadprůměrně zdatných uživatelů jazyka, ne jeden z nich má potíže utvořit správné tvary souborových číslovek i neuter zakončených na –ę. Dotyční navíc opakovaně uváděli, že tyto formy nepoužívají a nejsou na ně zvyklí.

Ačkoli nemám k dispozici relevantní statistická data (např. z velkého mluveného korpusu, který je v polském prostředí bohužel teprve ve stadiu vývoje), je na první pohled jasné, že ony alternativní (původně nářeční) varianty v mluvené polštině zcela převažují.¹⁴ To, kdy si najde cestu i do psaného jazyka¹⁵, je podle mého názoru jen otázkou pružnosti normotvorných institucí – jakmile bude tato varianta povolena v příručkách, začne se hojně objevovat i v textech oficiálního typu, které procházejí korekturou (tisk, literatura, vědecké práce). Poznámku k tomuto tématu ostatně uvádí ve svém textu z roku 1976 (sic!) i výše citovaný Z. Saloni – píše zde, že původní

¹³ Toto paradigma se z dob společného vývoje slovanských jazyků zachovalo i v češtině, kde ovšem – pokud mohu soudit – není mluvčími považováno za zvláště obtížné.

¹⁴ Kromě zkušeností percepčních (málokdy od někoho takovou formu uslyším) jsem v Polsku získala i hojně zkušenosti s vlastní produkcí: jakožto pečlivý mluvčí, který se polsky naučil až v dospělosti, se při komunikaci s jinými mluvčími snažím tyto formy aktivně tvořit a používat v souladu s normou, přičemž se opakovaně setkávám s překvapenou reakcí typu: „Ty ale divně mluvíš, tohle já bych nikdy neřekl.“

¹⁵ Samozřejmě kromě písemné komunikace na internetu, kde vzhledem k jejímu neformálnímu charakteru převládá výrazně již dnes.

(spisovné) formy se využívají tak málo, že jsou „na hranici toho, co lze označit za pravidelné“ (s. 59).

Rozpor, který zde popisují, je pro tuto práci důležitý proto, že pravděpodobně ovlivňuje jazykové chování mluvčích (na jedné straně chtějí alespoň v některých komunikačních situacích mluvit v souladu s normou, na straně druhé to však neumějí). Lze předpokládat, že se budou složitým formám LZ vyhýbat, vytvoří si nějaké kompenzační strategie (např. pomocná slova a perifráze), nebo budou formy tvořit chybně. Přestože v korpusu, z něhož budu získávat data, jsou pouze texty z oficiálních zdrojů, které s největší pravděpodobností prošly jazykovou korekturou, bude při jejich analýze namísto tento fakt držet v povědomí. Korektoři koneckonců pocházejí mimo jiné z řad absolventů polonistiky, a jak bylo zmíněno již výše, ani oni nemají tyto formy plně zažité.

Anna Andrejczuk v úvodu svého článku říká otevřeně (2011, s. 272): „Ačkoli je propagují vydatelé normotvorných příruček a slovníků, toto pravidlo je pro mnoho uživatelů jazyka divné a nesrozumitelné.“

Wielki Słownik Poprawnej Polszczyzny z produkce vědeckého nakladatelství PWN – předního vydavatele jazykových slovníků a příruček – je právě jedna z takových autorit. Na s. 1672 v hesle *Liczba gramatyczna* (Gramatické číslo) je uvedeno následující:

„Část podstatných jmen nemá obě gramatická čísla. Patří mezi ně [abstraktní pojmy, geografické názvy, substantiva hromadná] a obecná podstatná jména, která mají pouze formu množného čísla, např. *chrzciny, imieniny, rajstopy, sanie, skrzypce, spodnie, urodziny, usta, wrota*; jejich množství je možné vyjádřit pouze pomocí souborové číslovky, např. *troje skrzypiec, czworo sań, pięcioro wrót* (ne **trzy skrzypce, *cztery sanie, *pięć wrót*); spojení se souborovými číslovkami se netýká těch slov, která se mohou pojít s výrazem *para*, např. *trzy pary rajstop, dwie pary spodni*.“

Zajímavá je zejména poslední věta: slovník informuje, že některých PT se spojení s číslovkou souborovou *netýká*. Neinformuje ovšem, která slova to jsou a jak je poznat (uvedené dva příklady oba designují část oblečení určenou na nohy, jmenovitě punčocháče a kalhoty), a navíc není zcela zřejmé, jestli případné spojení s LZ *dwoje spodni* slovník přímo zapovídá jako nesprávné, nebo je jen nedoporučuje, či jeho výskyt prostě nepřipouští.

Na straně 1676 upozorňuje WSPP pod heslem *Liczebnik zbiorowy* na to, že „v současné polštině se souborové číslovky používají čím dál tím méně. Mnoho z nich se vyskytuje v kontextech, v nichž se používá slovo *para*, např. *dwie pary nożyc, cztery pary spodni*, a pak bývají nahrazeny.“ To budí dojem, že polská jazyková norma (tzv. *norma wzorcowa*) doporučuje spojení *czworo spodni*, které „je možné nahradit“ opisem *cztery pary spodni*. To je však v rozporu s předchozí uvedenou citací, kde totiž autoři výslovně uvádějí, že spojení s LZ se *netýká* těch PT, která se *mohou* pojit se slovem *para*. Anna Andrejczuk na tuto nekonzistentnost upozorňuje také a vyjadřuje názor, že „vnitřně nekonzistentní oficiální norma a rozpor s územ mohou vést až k tomu, že uživatelé polštiny tuto normu zcela odmítnou“ (2011, s. 276).

Výchozí situace uživatele polského jazyka tedy není jednoduchá – pokud chce mluvit správně, v souladu s normou (na což jsou v polské společnosti kladeny obdobné nároky jako ve společnosti české), a to i v oblasti těchto pro něj obtížných gramatických jevů, není lehké se zorientovat, jakou formu by si měl vybrat.

Jak se různí mluvčí polštiny v praxi skutečně rozhodují, jak vyjadřují počty různých pomnožných jmen, je předmětem praktického výzkumu v následující kapitole. Autentické texty získané z Národního korpusu polského jazyka ukáží nejlépe, kudy se ubírá úzus.

3 Praktická část

3.1 Zdroj jazykových dat: korpus NKJP

Narodowy korpus języka polskiego (Národní korpus polského jazyka) se nachází na webové adrese <http://www.nkjp.pl>. Jedná se o společný projekt čtyř institucí – dvou ústavů Polské akademie věd (PAN), Katedry počítačové a korpusové lingvistiky Lodžské univerzity a vědeckého nakladatelství PWN.¹⁶ Všechny tyto instituce původně pracovaly na vlastních samostatných korpusech, které byly vytvářeny na míru jejich individuálním potřebám.¹⁷

V případě IPI PAN se jednalo o výzkum polské syntaxe v teoretickém rámci *Head-driven phrase structure grammar*. Práce na korpusu byla zahájena téměř souběžně s tímto výzkumem, protože v této době neexistoval vhodný zdroj polského jazykového materiálu. V letech 2001–2004 tak vznikl nereprezentativní a nevyvážený korpus čítající 250 milionů segmentů. Jako jediný takto rozsáhlý polský korpus byl opatřen morfosyntaktickým značkováním.

V IJP PAN se s budováním korpusu začalo v souvislosti s plánem vypracovat a vydat nový slovník polského jazyka – korpus měl posloužit jako empirický podklad pro lexikografickou práci. Počátkem 90. let byl však projekt slovníku opuštěn. Korpus však existoval dál a byl postupně rozvíjen, sloužil především jako zdroj dat pro výzkum sémantiky a syntaxe polských sloves. IJP neměl v úmyslu jej zpřístupnit veřejnosti, ale vědecká veřejnost k němu přístup měla.

Wydawnictwo Naukowe PWN vydává drtivou většinu slovníků polského jazyka, které v Polsku vycházejí. Již v roce 2000 vyšel první slovník (*Inny słownik języka polskiego*, red. M. Bańko) založený na vlastním korpusu PWN budovaném od roku 1997. Tento korpus měl sloužit především jako zdroj dat pro lexikografické popisy slovníkových hesel a také jako zdroj nejrůznějších příkladů ilustrujících významy těchto

¹⁶ Plné názvy zúčastněných institucí v polštině: Instytut Podstaw Informatyki PAN (koordinátor projektu), Instytut Języka Polskiego PAN, Wydawnictwo Naukowe PWN a Zakład Językoznawstwa Komputerowego i Korpusowego Uniwersytetu Łódzkiego.

¹⁷ Všechny informace o jednotlivých korpusech uvedené v kapitole 3.1 čerpám z publikace Przepiórkowski et al. 2012, s. 1–10.

hesel. Velký důraz byl proto kladen na tematickou a stylistickou pestrost korpusu, a také na polskou literární tradici (např. byly zahrnuty všechny texty povinné středoškolské četby). Rozsah korpusu postupně dosáhl 100 milionů slov.

Lodžský soubor korpusů vznikl ve spolupráci s Katedrou jazykovědy a současné angličtiny Lancasterské univerzity již od roku 1995. Kromě anglickojazyčného korpusu, který se zaměřuje na angličtinu nerodilých mluvčích, zahrnuje především téměř stomilionový korpus PELCRA, který obsahuje proporcionálně vyvážené texty různých typů z mluveného i psaného jazyka. Kromě PELCRA vznikl v Lodži ještě samostatný mluvený korpus o rozsahu 600 tisíc slov.

Jak je vidět z předcházejícího výčtu, polská jazykověda po roce 2000 nedisponovala korpusem, který by byl zároveň rozsáhlý, označkový, vyvážený a reprezentativní. Existující korpusy vždy odpovídaly na potřeby výzkumné skupiny, která iniciovala jejich vznik. Lidské i finanční zdroje se tím poněkud tříštily, což samozřejmě polským jazykovědcům neuniklo. V roce 2006 proto došlo k vytvoření konsorcia *Narodowy Korpus Języka Polskiego*, které brzy získalo finanční prostředky k práci na projektu společného národního korpusu. Koordinátorem projektu se stal IPI PAN v čele s doc. Adamem Przepiórkowským.

Do vznikajícího NKJP byly inkorporovány rozsáhlé části existujících korpusů. Kromě samotné datové báze se pracovalo i na vyhledávacích nástrojích. Dnes je možné prohledávat NKJP pomocí programu Poliqarp (modifikace staršího vyhledávače vyvinutého pro korpus IPI PAN) a programu PELCRA, který původně vznikl v Lodži pro práci se stejnojmenným korpusem. Kromě toho existuje mnoho nástrojů k práci se specifickými subkorpusy, k morfosyntaktickému značkování apod.; podrobnosti o nich je možné (v polštině či angličtině) najít na webu projektu¹⁸.

Národní korpus polského jazyka je všem zájemcům dostupný zdarma na internetu, stejně jako jednotlivé nástroje a vyhledávače. Aktuálně má rozsah cca 1 500 milionů slov a je průběžně doplňován. Kromě polské literatury obsahuje také denní tisk a odborné časopisy, přepsané nahrávky rozhovorů, příležitostné texty (letáky aj.) a texty převzaté z internetu (jak interakce na diskusních fórech apod., tak např. texty z Wikipedie). Jak upozorňují Przepiórkowski et al. (2012), k významu korpusu pro vědeckou činnost

¹⁸ <http://www.nkjp.pl>.

přispívá tematická a žánrová pestrost psaných textů, jakož i sociologická vyváženost (z hlediska věku, pohlaví a oblasti původu) mluvčích zastoupených v orální složce korpusu.

3.2 Sběr dat

3.2.1 Obecné otázky

Během prosince 2013 a ledna 2014 jsem z korpusu NKJP „vytěžila“ jazyková data použitá k praktickému výzkumu v rámci této práce. Korpus NKJP bývá nárazově rozšiřován, ale nepodařilo se mi najít žádné číslování starších verzí. Všechna data jsou proto citována k 1. 1. 2014.

Korpus jsem prohledávala pomocí vyhledávače PELCRA¹⁹, který na rozdíl od druhého dostupného nástroje Poliqarp umožňuje prohledat skutečně celý korpus. Jedno omezení však přeci jen má: není možné zobrazit více než 10 000 výsledků vyhledávání. Vzhledem k tomu, že zkoumaný jev patří spíše k méně frekventovaným, jsem však na tento strop v praxi nenarazila ani jednou.

Vzhledem k tomu, že jsem chtěla získat co nejvyšší počet záznamů, pracovala jsem s kompletním dostupným korpusem NKJP čítajícím 1 500 milionů slov. Tento korpus není vyvážený z hlediska typu ani žánru textů. Dostupný vyvážený subkorpus (300 milionů slov) je však pro naše potřeby příliš malý. Kromě toho se domnívám, že v případě námi sledovaných jevů výhoda získání většího množství výsledků jednoznačně převažuje nad potřebou zachovat vyvážený korpus. Parametr „typ zdrojového textu“ jsem sledovala a uvádím jej u všech záznamů (viz příloha A), takže je samozřejmě možné k němu kdykoli přihlédnout.

3.2.2 Seznam PT pro vyhledávání

Původní záměr sestavit exaktní seznam pomnožných slov, vybrat z něj ta, která spadají do skupin P2, P4 a P5 (podle typických příkladů je můžeme označit *skrzypce*,

¹⁹ PELCRA se nachází na adrese http://www.nkjp.uni.lodz.pl/index_adv.jsp. Čas od času se stává, že je tato webová stránka několik dní až týdnů mimo provoz, zřejmě kvůli nějakým úpravám. Web bohužel neposkytuje informace o odstávkách, což může uživatele poněkud zaskočit. Nakonec se však vždy PELCRA na tuto adresu opět vrátí, trpělivost se tedy vyplatí.

spodnie, schody), a pak v korpusu vyhledat příslušné záznamy, byl poměrně záhy přehodnocen. V žádné dostupné publikaci totiž není uveden seznam pomnožných jmen, a to ani ve zkrácené podobě. Vždy se objevují jen nepočetné příklady (oblíbené jsou zejména *drzwi, skrzypce, spodnie, sanie*).

Zřejmě jediný seriózní pokus o vytvoření seznamu pomnožných slov realizovala Hanna Jadacka. Jak uvádí ve svém článku z roku 2006, posloužil jí k tomu *Indeks a tergo*²⁰, čili slovník, v němž jsou hesla řazena abecedně podle posledního písmene ve slově. Jadacka prošla všechny lexémy zakončené na -a, -e, -i, -y a vytrídila z nich (zřejmě na základě introspekce rodilého mluvčího) ty, které nemají singulár. Získala tak seznam více než 1 000 pomnožných slov. Tento seznam však bohužel není nikde dostupný.

Pokus replikovat postup Jadacké jsem brzy vzdala bez úspěchu. Rozsah slovníku *a tergo* byl příliš velký, navíc mi na rozdíl od rodilé mluvčí Jadacké chybí schopnost posoudit, zda má nějaké méně běžné slovo singulár nebo ne. Většinu slov jsem proto musela kontrolovat ve slovníku, který ovšem často informaci o (ne)existenci singuláru neuvádí. Výsledkem byl seznam mnoha málo frekventovaných slov, která ani nejsou v korpusu dostatečně zastoupena (pokud vůbec). Myšlenka použít k hledání pomnožných jmen přímo korpus narazila na jeho limity: neúplné morfologické značkování, omezený počet vypisovaných záznamů a chybějící softwarové nástroje.

Rozhodla jsem se tedy seznam slov pro vyhledávání sestavit empiricky. Z různých článků týkajících se problematiky pomnožných jmen jsem vypsala příklady, které v nich uvádějí jejich autoři. Tak jsem získala seznam cca 80 slov, z nichž část tvořily velice známé, prototypické příklady pomnožných jmen. Poté následovalo třídění podle významu (omezila jsem analýzu na slova, která označují předměty) a kontrola, zda se vůbec dotyčné lemma²¹ v korpusu vyskytuje v dostatečném počtu. Výsledkem bylo 27 podstatných jmen, která jsem následně podrobila zkoumání. Následující výčet obsahuje všechny i s českým překladem (jen stručným, zabíhat do podrobností je bezpředmětné):

²⁰ DUBISZ, Stanisław. *Uniwersalny słownik języka polskiego*. Tom 7 – indeks a tergo. Warszawa : PWN, 2008.

²¹ Lemma je reprezentativní slovníková podoba slova; každé slovní formě v korpusu je přiřazeno adekvátní lemma, což mimo jiné usnadňuje vyhledávání (vyhledávací nástroj sám vyhledá všechny tvary zadaného lemmatu).

cażki (kleště), *cegi* (kleště), *drzwi* (dveře), *gęśle* (housle), *gogle* (brýle), *grabie* (hrábě), *kalesony* (kamaše), *kombinerki* (kleště), *korale* (korále), *majtki* (kalhotky), *nosze* (nosítka), *nożyce* (nůžky), *nożyczki* (nůžky), *obcegi* (kleště), *obcażki* (kleště), *okulary* (brýle), *rajstopy* (punčocháče), *sanie* (sáně), *schody* (schodiště), *skrzypce* (housle), *slipy* (slipy), *spodenki* (kalhoty), *spodnie* (kalhoty), *szczytce* (kleště), *szorty* (šortky), *widły* (vidle), *wrota* (vrata).

3.2.3 Vyhledávání a zpracování výsledků

Klíčovou částí procesu bylo vyhledávání jednotlivých slov. Ze všech nalezených výskytů bylo potřeba extrahovat ty, kde je daný výraz nějakým způsobem kvantifikován. Toho jsem dosáhla kombinací strojového a ručního třídění – předpokládané způsoby kvantifikace (např. počítání kalhot na páry) bylo možné predikovat a vytržít pomocí vhodně zadaného dotazu. Protože však zároveň od počátku předpokládám, že se mohou objevit také více či méně kreativní způsoby kvantifikace, které predikovat nedokážu, bylo nutné podstoupit poněkud zdlouhavý proces a zbývající záznamy po vytržení predikovaných pozic ručně projít jeden po druhém, zda nevykazují jiné (nepředvídané či strojově těžko dohledatelné) typy kvantifikace.

Výsledek této práce je kompletně reprodukován v příloze A. Pro přehlednost je rozdělen do tabulek podle jednotlivých vyhledávaných výrazů. Každá tabulka má jednoduchou hlavičku, kde je uveden tvar nominativu a genitivu (drtivá většina KWIC²² má právě tento tvar), český význam výrazu a celkový počet výsledků k danému výrazu. Záznamy jsou seskupeny podle jednotlivých typů kvantifikace, jejichž výčet spolu s počtem záznamů je uveden v hlavičce.

Celkem jsem shromáždila 1 653 záznamů (k 22 různým slovům, průměrně tedy 78 záznamů/slovo, ve skutečnosti jsou ale některá slova ve vzorku podstatně hojněji zastoupená než jiná). K datu 1. 1. 2014 by se mělo jednat o kompletní sadu relevantních pozic, které NKJP obsahuje. Opakující se zcela totožné záznamy byly ze vzorku vyřazeny, stejný osud potkal také příklady metajazykového užití těchto výrazů – v rámci polských diskusních fór si uživatelé poměrně často navzájem vysvětlují problematiku počítání

²² Zkratka KWIC – *key word in context*, klíčové slovo v kontextu – odkazuje k podobě, v níž se zobrazují výsledky korpusového dotazu: seznam dokladů (tzv. konkordance) obsahuje vždy výraz, který se shodoval se zadáním dotazu (klíčové slovo), doplněný o levý a pravý kontext. Viz příloha A.

pomnožných jmen. Vyšší četnost slovních spojení, která fungují jako oblíbené příklady, i ukázky chybných formulací by zbytečně zkreslovaly výsledky. Ruční procházení mi umožnilo jich většinu odstranit.

Tabulka v příloze A obsahuje všechna použitá jazyková data z korpusu NKJP. Kromě samotného KWIC je v ní vždy uveden údaj o typu zdroje jazykového materiálu. NKJP obsahuje texty z následujících zdrojů²³, přičemž drtivá většina materiálu, který jsem použila já, pochází z prvních čtyř uvedených – z publicistiky, literatury (prózy) a z internetu (texty i interakce).

Tabulka 1. Typologie zdrojů textů zastoupených v NKJP.

ZNAČKA	POPIS
#typ_publ	publicistika (tisk)
#typ_net_interakt	internet, interakce (např. blogy, fóra s komentáři, diskusní platformy)
#typ_net_nieinterakt	internet, bez interakce (např. domovské stránky, blogy bez komentářů)
#typ_lit_proza	próza
#typ_qmow	quasi-mluvené texty
#typ_fakt	literatura faktu
#typ_urzed	úřední/administrativní texty
#typ_lit	literatura (obecně)
#typ_inf_por	informačně-poradenské texty
#typ_nklas	neklasifikováno (jiné zdroje)
#typ_lit_poezja	poezie
#typ_media	mluvené texty z médií
#typ_nd	populárně-vědecké texty
#typ_konwers	mluvené texty (konverzace)
#typ_lit_dramat	dramata (literárně-dramatické texty)

²³ Tabulka č. 1 s údaji o zdrojových datech je převzata z webu projektu PELCRA (v mém překladu a doplněna o mé vysvětlivky), viz http://www.nkjp.uni.lodz.pl/help.jsp#analiza_rejestru.

Z čistě praktických důvodů jsem do prvního sloupce tabulky v příloze A přidala informaci o typu kvantifikace daného pomnožného jména. Dělení je poměrně hrubé, na podrobnosti se zaměřím při analýze jednotlivých typů. Částečně vychází z hypotéz před provedením výzkumu, ale některé typy jsem oddělila naopak teprve *ex post* při třídění a zpracování dat. Záznamy jsou uspořádány právě podle tohoto parametru, aby bylo hned zjevné, jak je který typ početně (proporčně) v tabulce zastoupen. Následující tabulka č. 2 představuje typy kvantifikace, které jsem při třídění rozlišovala.

Tabulka 2. Kategorie podle typu použité kvantifikace, do nichž byly výskyty cílových lemmat získané z NKJP rozříděny.

ZNAČKA	POPIS	PŘÍKLAD
LG (liczebnik główny)	základní číslovka	Są też cztery skrzypce i bas.
LZ (liczebnik zbiorowy)	souborová číslovka	Przybyło troje sań napełnionych wieśniakami.
LG/LZ	číslovka zapsaná číslicí, proto nelze rozhodnout, zda se jedná o LZ nebo LG	Mieliśmy 18 noszy , a rannych było 20.
sztuka	pomocné slovo <i>sztuka</i> (kus)	sześćdziesiąt sztuk spodni z jeansu [...]
para	pomocné slovo <i>para</i> (pár)	[...] przez dwie pary spodni i bieliznę!
WP _ _ _	jiné pomocné slovo (uvedeno za zkratkou WP)	[...]ustrojona w trzy sznury koralí swojej mamy.
NOM	číslovka substantivního typu	Dziesiątki drzwi otwierały się jedne za drugimi, [...]

3.3 Analýza dat

3.3.1 Dvě skupiny PT

Výzkum prezentovaný v článku Anny Andrejczuk (2011), na nějž touto prací částečně navazuji, je metodologicky poměrně jednoduchý. Autorka vybrala několik nejtypičtějších pomnožných výrazů, rozdělila je na skupiny podle sémantického klíče a v korpusu vyhledávala, jak často se vyskytují společně se souborovou číslovkou, s pomocným slovem *para* a se základní číslovkou. Protože měla výsledků z korpusu velice málo (používala mnohem starší verzi než já), doplnila je prohledáváním internetu pomocí vyhledávače Google. Zjištěné výsledky jsou poněkud hrubé, ale ukazují jisté tendence, podle nichž Andrejczuk vytvořila schéma, které představuji v kapitole 2.2 a které mi slouží jako východisko k vlastnímu výzkumu.

Andrejczuk vyhledávala slova ze tří různých skupin:

- (1) předměty, jejichž tvar je výrazně symetrický (*spodnie, okulary*),
- (2) předměty „jednotitého“ tvaru, takzvaně „typická“ pomnožná jména (*drzwi, skrzypce*),
- (3) názvy svátků, oslav apod. (*urodziny, chrzciny*).²⁴

Slova ze třetí skupiny jsou mimo oblast zájmu této práce, jelikož jejich kvantifikace není počítáním v užším smyslu tohoto slova.²⁵ Dále se tedy budu soustředit pouze na výrazy ze skupin (1) a (2), zcela v duchu toho, jak je Anna Andrejczuk vytvořila. Následující tabulka č. 3 poskytuje přehled všech mnou zpracovávaných slov rozdělených do těchto dvou skupin (některé významově hodně blízké výrazy jsem sloučila do jedné buňky). Jak je vidět ze součtu na konci tabulky, zastoupení obou sémanticky definovaných skupin v kompletu dat je zhruba stejné, což nám dává možnost porovnávat je mezi sebou i co do číselných hodnot.

²⁴ Andrejczuk 2011, s. 276.

²⁵ Obvyklý způsob počítání těchto PT jsou formulace typu *dwa przyjęcia urodzinowe* (dvě narozeninové oslavy). Zde se původní PT *urodziny* transformuje v nominální frázi, v níž je slovo *urodziny* v závislém postavení. Řídícím členem fráze se stává běžné substantivum *przyjęcie* (pl. *przyjęcia*), které se při počítání pojí s běžnou základní číslovkou.

Tabulka 3. Rozdělení zkoumaných lemmat do sémantických kategorií a jejich početní zastoupení.

(1) „symetrické“ předměty (SYM)		(2) „jednolitě“ předměty (JED)	
výraz	počet záznamů	výraz	počet záznamů
<i>okulary, gogle</i>	137, 3	<i>drzwi</i>	543
<i>kalesony, rajstopy</i>	25, 90	<i>wrota</i>	36
<i>spodnie, spodenki</i>	345, 36	<i>grabie</i>	5
<i>nożyce, nożyczki</i>	14, 20	<i>widły</i>	3
<i>obcęgi</i>	1	<i>nosze</i>	30
<i>kombinerki</i>	1	<i>sanie</i>	76
<i>szczypce</i>	13	<i>skrzypce</i>	94
<i>szorty</i>	7	<i>schody</i>	34
<i>majtki, slipy</i>	115, 8	<i>korale</i>	17
<i>cążki, cęgi, obcążki</i>	0	<i>gęśle</i>	0
celkem:	815	celkem:	838

Pokud jde o zastoupení konkrétních typů kvantifikace první skupiny (pro jednoduchost ji označíme zkratkou **SYM**), následující tabulka č. 4 ukazuje na drtivou převahu používání pomocného výrazu „para“. To odpovídá i zjištěním Anny Andrejczuk, která říká, že „způsob vyjadřování počtu pomocí slova *para* v případě pomnožných slov, jejichž denotát má výrazně dvoudílný, symetrický tvar, je v polštině velmi dobře zaveden“. Tento způsob vyjadřování počtu obsluhuje v mém případě 87 % všech případů!

Všechny ostatní způsoby vyjadřování množství jsou oproti němu málo významné. Dvě desítky výskytů překročily ještě výrazy *sztuka* (6 %, pojí se zejména s vyššími čísly zapsanými číslicí, viz dále) a výrazy sdružené pod značku NOM (3 %). Zbývající způsoby kvantifikace se dělí o zbylá 4 % případů.

Souborové číslovky jsou zcela marginalizovány. V celém vzorku, který má přes 800 položek, nalezneme pouhé tři případy užití tohoto druhu číslovek.

Tabulka 4. Četnosti jednotlivých typů kvantifikace (viz tabulka č. 2) u sledovaných lemmat ze skupiny SYM (viz tabulka č. 3).

typ kvant.:	para	LZ	LG	LZ/LG (číslice)	sztuka	NOM	WP
<i>kalesony</i>	21	0	0	0	3	1	0
<i>gogle</i>	3	0	0	0	0	0	0
<i>okulary</i>	113	0	2	3	8	12	0
<i>rajstopy</i>	84	0	1	0	2	2	0
<i>spodnie</i>	325	2	1	0	15	2	0
<i>spodenki</i>	23	0	0	6	6	1	0
<i>nożyce</i>	8	0	1	2	3	0	0
<i>nożyczki</i>	17	0	3	0	0	0	0
<i>obcęgi</i>	1	0	0	0	0	0	0
<i>kombinerki</i>	0	0	0	0	1	0	0
<i>szcypce</i>	9	0	0	0	2	0	0
<i>slipy</i>	7	1	1	1	0	0	0
<i>majtki</i>	95	0	3	2	10	7	0
<i>szorty</i>	5	0	0	0	2	0	0
celkem:	711	3	12	11	52	25	0

Tabulka č. 5 ukazuje zastoupení jednotlivých typů kvantifikace pro jednotlivá slova ze skupiny č. 2 (s jednolitou, ne nutně symetrickou stavbou, pro jednoduchost ji označíme zkratkou **JED**). Teoreticky očekáváme převahu souborových číslovek, protože tato skupina slov podle jazykové normy vyjádřené ve WSPP musí být počítána pomocí LZ, neboť WSPP žádnou jinou možnost nenabízí.²⁶

V praxi skutečně vidíme převahu LZ (68 %), avšak nikoli tak drtivou, jako je tomu v případě první skupiny s pomocným slovem *para*. 15 % případů připadá na LG, která norma vysloveně zapovídá jako chybné. 7 % připadá na *para* a o zbylých 10 % případů se dělí méně obvyklé kvantifikační postupy.

²⁶ „[[]jejich množství je možné vyjádřit pouze pomocí souborové číslovky, např. *troje skrzypiec, czworo sań, pięćcioro wrót [...]*“ (WSPP 2006, s. 1672).

Tabulka 5. Četnosti jednotlivých typů kvantifikace (viz tabulka č. 2) u sledovaných lemmat ze skupiny JED (viz tabulka č. 3).

typ kvant.:	para	LZ	LG	LZ/LG (číslice)	sztuka	NOM	WP
drzwi	8	461	73	0	0	1	0
wrota	7	10	15	0	0	0	4
grabie	2	2	1	0	0	0	0
widły	0	1	0	0	0	2	0
nosze	17	0	1	1	4	3	1
sanie	23	21	21	8	0	5	0
skrzypce	5	71	12	3	0	4	0
schody	0	0	0	0	0	0	34
korale	0	0	0	1	1	0	15
celkem:	62	566	123	13	5	15	54

3.3.2 Jednotlivé typy kvantifikace

Část sledovaných typů kvantifikace jsem předpokládala předem, a to na základě dostupné literatury (Andrejczuk 2011, Saloni 1976 a 2009, Jadacka 2006, WSPP 2006) i vlastních zkušeností z polskojazyčného prostředí. Několik z nich jsem však naopak coby samostatný typ vyčlenila až zpětně po prozkoumání nasbíraného materiálu, protože se v něm vyskytoval nápadně často. Projděme si je popořadě.

Pomocné slovo *para* zvítězilo co do frekvence nad všemi ostatními způsoby vyjadřování množství. Zejména to platí o počítání „symetrických“ předmětů, ale počet jeho výskytů není zanedbatelný ani v případě slov z druhé skupiny. Pokud použijeme slovo *para*, můžeme pak už samotný počet vyjádřit základní číslovkou, což přináší mluvčímu usnadnění. Z jeho hlediska se tak v případě této analytické formulace jedná o velmi ekonomické řešení. Podle mého názoru se do budoucna bude dosah užívání této vazby dále šířit, a to i na slova, s nimiž vlastně sémanticky nesouvisí (nemají „párovou formu“), přičemž lexikální význam tohoto výrazu (pár, tedy dvě části něčeho, které patří

k sobě) bude postupně slábnout. Již dnes to můžeme pozorovat na následujících příkladech²⁷, jež všechny obsahují PT ze skupiny JED.

1. *Langer stał przez chwilę, wpatrując się w **dwadzieścia par drzwi** lypiących na niego czarnymi oczami numerów.*
(Langer chvíli stál a díval se na dvacet párů dveří, které na něj civěly černýma očima číslic.)
2. *Upatrz yli sobie **dwie pary skrzypiec** leżących w szafie.*
(Vyhlédli si dva páry houslí ležící ve skříni.)
3. ***Trzy pary sań** z trzech sąsiednich dworów podjechały pod ganek.*
(Tři páry sání ze tří sousedních dvorců přijely k verandě.)
4. *Na wyposażeniu miała **pięć par noszy** i jeden kuferek wypadkowy.*
(Byla vybavena pěti páry nosítek a jednou lékárníčkou.)
5. *Do tego celu służyła właśnie **druga para wrót**, dziś już nieistniejąca.*
(K tomuto účelu právě sloužil druhý pár vrat, který dnes již neexistuje.)
6. ***Grabi** powinniśmy mieć przynajmniej **dwie pary** - zwykłe, metalowe [...]*
(Hrábí musíme mít alespoň dva páry – klasické, kovové [...])

Dělení denotátů na „zdvojené“ a „jednolité“ má ovšem do jednoznačnosti daleko. Jak je patrné, v první skupině jsou především názvy oblečení párových částí těla (zejména nohou²⁸) a dále názvy náradí a nástrojů, které se jako nůžky nebo kleště skládají ze dvou stejných částí. Ve druhé se nachází poněkud pestřejší směsice předmětů (saně, housle, dveře). I je bychom však mohli vnímat určitou optikou jako zdvojené, zvláště v některých případech. Upozorňuje na to i Zygmunt Saloni (2009, s. 58), který se pozastavuje zejména nad slovem *drzwi*, jež v dnešní polštině figuruje jako prototypický příklad pomnožného jména:

„The most interesting case is the noun *drzwi* ‘door’. A door is usually asymmetrical, however many types of doors are not only symmetrical, but consist of two separate parts, approximately mutually symmetrical. Polish names of such types of door are *plurale tantum*, eg. *wrota*, *wierzeje*, *podwoje*, which co-occur with numeral forms. The border between [those two] sets is not distinct.“²⁹

²⁷ Všechny příklady jsou převzaty z tabulek v příloze A.

²⁸ Chtěla jsem se totiž soustředit pouze na předměty, které tvoří jeden kus a skutečně nemají singulár, takže např. *rękawice* (rukavice) byly z výběru vyloučeny.

²⁹ „Nejzajímavější příklad je substantivum *drzwi* (dveře). Dveře jsou obvykle asymetrické, ačkoli také existuje mnoho typů dveří, které nejenže jsou symetrické, ale skládají se ze dvou oddělených, více méně

V korpusovém materiálu, který jsem shromáždila v příloze A, je jeden příklad, který předchozí Saloniho tvrzení půvabně ilustruje: *dwie pary dwuskrzydłowych wrót* (dva páry dvoukřídých vrat).

V rámci této skupiny slov je rozšířenost počítání pomocí *para* velmi proměnlivá, což není příliš překvapující vzhledem k tomu, že skupina JED je sémanticky více heterogenní než skupina SYM. Pokud jde o *drzwi*, je v podstatě marginální ($8/543 = 2 \%$), stejně jako v případě slova *skrzypce* ($5/95 = 5 \%$). Přičítám to tomu, že obě tato slova jsou velmi ikonickými *pluralii tantum*, mluvčí se s nimi setkávají poměrně často (zejména v případě *drzwi*) a samotná frekvence těchto spojení je udržuje v oběhu. V případě slova *skrzypce* navíc značnou roli hraje velmi ustálený nejběžnější kontext výskytu – jedná se o názvy či charakteristiky hudebních děl určených pro více než jedny housle (např. *Koncert na czworo skrzypiec Antonia Vivaldiego* – Koncert pro čtyři housle A. Vivaldiho, viz příslušná tabulka v příloze A).

Naopak v případě jiných slov je zastoupení slova *para* mnohem vyšší, byť vzhledem k celkově nízkému počtu záznamů můžeme hovořit maximálně o tendenci³⁰: *wrota* ($7/36 = 20 \%$), *grabie* ($2/5 = 40 \%$), *nosze* ($17/26 = 66 \%$), *sanie* ($23/78 = 30 \%$). Za pozornost stojí zejména 30% zastoupení slova *para* u PT *sanie*, které bývá považováno za další prototypické pomnožné jméno typicky se pojící se souborovými číslovkami. Pokud jde o *wrota*, na jejichž „zdvojenou podobu“ Saloni upozorňuje ve výše citovaném úryvku, můj vzorek jeho domněnku spíše nepotvrzuje: jen necelých 20 % výskytů PT *wrota* bylo počítáno pomocí slova *para*, ve většině případů byly použity souborové nebo základní číslovky.

Předměty ze skupiny SYM charakterizuje „párovost“ jejich uspořádání, což se promítá také do užívání slova *para* pro označení jednoho exempláře daného objektu. Z toho zároveň plyne, že mluvčí lexém *para* stále užívají s vědomím jeho lexikálního významu – není to pouze gramatické slovo usnadňující počítání objektů, stále v sobě nese i význam „pár“. Následující příklady ukazují, že kvantifikátor *para* se vyskytuje jak

vzájemně symetrických částí. Polské názvy takového typu dveří jsou *pluralia tantum*, např. *wrota*, *wierzeje*, *podwoje*, a vyskytují se s tvary číslovek. Hranice [mezi těmito dvěma skupinami slov] není zřetelná.“

³⁰ Hodnoty, z nichž vycházejí údaje v %, jsou místy skutečně velmi nízké. Proto je třeba chápat tato čísla především ilustrativně.

ve spojení s číslovkou *jedna*, tak bez ní, nebo dokonce ve spojení s neurčitými číslovkami.

7. *[M]am 2 bluzki różowe i **jedną parę majtek** różowych.*
(Mám dvě růžové halenky a jeden pár růžových kalhotek.)
8. *Siostra natomiast, w tym samym czasie, kupiła sobie **parę majtek**.*
(Sestra si zase ve stejnou dobu koupila pár kalhotek.)
9. *Najlepiej włożyć od razu **kilka par majtek** itd.*
(Nejlepší je obléknout si rovnou několik párů kalhotek atd.)

Pomocné slovo *sztuka* patří k těm, která jsem identifikovala teprve v průběhu analýzy dat. Systematicky se pojí spíše s párovými jmény – ve skupině SYM má 54 zaznamenaných výskytů (což je více, než má kterýkoli další způsob kvantifikace kromě výrazu *para*), zatímco v JED se vyskytl pouze pětkrát (a to převážně u PT *nosze*, které se podle všeho při kvantifikaci chová do značné míry spíše jako slovo ze skupiny SYM).

Zásadní rys tohoto pomocného slova je jeho stylistické zabarvení – vyskytuje se téměř výhradně ve formálnějších kontextech (publicistika či administrativní texty) – obvykle se jedná o informace o krádežích, o výrobních plánech firem nebo např. seznam předmětů tvořících vybavení místnosti. Zároveň se také obvykle pojí s vyššími čísly (stovky, tisíce i miliony), což logicky souvisí s výše popsány kontexty. V následujícím seznamu toto tvrzení ilustrují příklady 10–13, zatímco příklad 14 ukazuje vzácnější opačný případ: slovo *sztuka* se zde objevuje v neformálním kontextu (jedná se o větu z internetového diskusního fóra) a ve spojení se slovem *kilka* (několik, čili malý počet).

10. *Kosmetyki oraz **800 sztuk majtek** i biustonoszy marki „Triumph“ skradli nieznani sprawcy w sklepie.*
(Kosmetiku a 800 kusů kalhotek a podprsenek značky Triumph ukradli v obchodě neznámí pachatelé.)
11. *Fabryka produkuje **tysiąc sztuk spodni** dziennie.*
(Továrna vyrábí tisíc kusů kalhot denně.)
12. *[...] mundury, **5 sztuk kalesonów** i spodenek gimnastycznych, 5 par skarpetek, 2 pary rękawiczek, 2 szaliki.*
([...] uniformy, 5 kusů kamašů a teplákových kalhot, 5 párů ponožek, 2 páry rukavic, 2 šály).
13. *Kiedy jest ładna pogoda, sprzedają około **20 sztuk okularów** w ciągu dnia, kiedy gorsza - 5.*
(Když je krásné počasí, prodám za den kolem 20 kusů brýlí, když je ošklivo, tak jen 5.)

14. [W]yposaz małą w gumowce, wez **kilka sztuk spodni** i rajstop na zmianę.

(Dej malé [myšleno dítěti, holčičce] holiny, vezmi několik kusů kalhot a punčocháčů na převlečení.)

Kdybychom si představu o současné podobě polského jazyka vytvářeli jen za pomoci normotvorných příruček jako je WSPP, museli bychom nutně předpokládat, že **souborové číslovky** (zkratka LZ) budou představovat dominantní způsob vyjadřování počtu přinejmenším u slov ze skupiny JED. Vstupní hypotézy tohoto výzkumu byly však formovány také praktickými zkušenostmi s mluvčími polštiny i odbornými články, které shrnuji v kapitole 2 této práce. Očekávala jsem proto, že frekvence souborových číslovek ve skutečných jazykových projevech bude podstatně nižší, než kolik by vyplývalo z WSPP.

Pokud se podíváme na hodnoty uvedené v tabulce č. 5, mohlo by se na první pohled zdát, že LZ se objevují poměrně frekventovaně. 566 příkladů z celkových 838 (čili přibližně 68 %) představuje souborové číslovky. Pokud si však prohlédneme tabulku celou, zjistíme zde, že z devíti lexémů PT jen u dvou LZ skutečně převažují nad ostatními typy kvantifikace. Jsou to slova *drzwi* (461/543 = 85 %) a *skrzypce* (71/94 = 76 %). Ostatní výrazy z této skupiny už tak přesvědčivé hodnoty nemají: *wrota* (10/36 = 28 %) a *sanie* (21/76 = 28 %) se pohybují pod 1/3 výskytů a *nosze* dokonce nepojí s LZ ani v jednom z 30 zaznamenaných kontextů.³¹ *Grabie* (2/5) a *widły* (1/3) mají tak nízký celkový počet výskytů, že u nich by bylo činit jakékoli závěry unáhlené.

Vysvětlení se nabízí samo – tyto číslovky jsou pro mluvčí obtížné a oni je proto nepreferují. Teoretické a morfologické zdůvodnění tohoto tvrzení je uvedeno v kapitole 2, nyní k němu přidejme praktické, na jazykových datech založené argumenty.

Předně, distribuce souborových číslovek se omezuje na nižší čísla. Zřejmě to souvisí i s mimojazykovou realitou – pokud specifikujeme konkrétní počty různých předmětů, častěji se pohybujeme v malých číslech, protože to odpovídá skutečnosti kolem nás, již popisujeme. Z 569 zaznamenaných případů užití LZ se v 275 případech jedná o číslovku *dwoje*, následuje *troje* (117), *czworo* (68) a *pięcioro* (53). Číslovky 6–10 se vyskytly dohromady v pouhých 21 příkladech.

³¹ Jak jsem již uvedla výše, dělení slov do skupin SYM a JED je dosti problematické. Pokud se přidržíme fyzického tvaru předmětů, patří nosítka spíše do JED podobně jako např. *sáně*. Pokud bychom ovšem naše skupiny určovali na základě dat z korpusu, museli bychom slovo *sanie* umístit do jedné skupiny spolu s ostatními slovy, která se počítají téměř výhradně pomocí pomocného slova *para*.

Souborové číslovky vyšší než deset³² jsou v celém korpusu obecně zastoupeny velmi málo. Po prozkoumání všech výskytů se ukázalo, že se kromě jediného případu (viz příklad 19) nikdy nepojí s názvem předmětu. Jedná se výhradně o užití LZ při počítání osob nebo dětí v duchu bodů a) a c) z oddílu 2.1. Je to nanejvýš logické, protože zejména v užití, které má indikovat skupinu osob obojího pohlaví, vnáší záměrné použití LZ místo generického maskulina informaci navíc, kterou je jinak možné do textu vložit jen pomocí delší konstrukce (například *dwadzieścioro studentów × dwadzieścia studentów i studentek*). Proto jsou podle mého názoru LZ v tomto kontextu podstatně „životaschopnější“ než v případě počítání předmětů vyjádřených pomnožnými jmény.

U čísel 5–10 se navíc opakovaně objevovaly typické kontexty – v případě *czworo a pięcioro drzwi* se jednalo o popisy automobilů (např. inzeráty), u ještě vyšších čísel pak o chodby nebo vagony. Jediný zaznamenaný výskyt souborové číslovky vyšší než 10 ve spojení s pomnožným substantivem pochází z románu *Europejka* Manuely Gretkowské a je zde použit v citelně ironickém kontextu (př. č. 19).

15. *Pieczęć z końca XIX wieku przedstawia **dwoje** skrzyżowanych **grabi**.*
(Pečeť z konce 19. století představuje dvoje zkřížené hrábě.)
16. *Posiada jednak **czworo drzwi**, co oznacza, że jest najszybszym bentleyem w wersji sedan.*
(Má však čtvery dveře, což znamená, že to je nejrychlejší bentley ve verzi sedan.)
17. *[...] **długi, wąski korytarz, w którym stało otworem siedmioro drzwi**.*
([...] dlouhá, úzká chodba, na které byly dokořán sedmery dveře.)
18. *Wchodzę na korytarz – **ośmioro drzwi** z lewej, **ośmioro** z prawej, wszystkie prowadzą do [...]*
(Vyjdu na chodbu – osmery dveře napravo, osmery nalevo, všechny vedou do [...])
19. *Muzyka **dwudziestu czterech skrzypiec** – dworską intrygą.*
(Hudba čtyřiadvaceti houslí je dvorská intrika.)

Základní číslovky (značka LG) by se teoreticky v datech z korpusu naopak měly objevovat jen zcela minimálně. Vzhledem k tomu, že je jazyková norma nepřipouští u žádného z pomnožných substantiv, měly by se vyskytovat jen v textech, které neprošly jazykovou korekturou, což jsou v případě našeho souboru dat především ty se značkou

³² Ačkoli se s nimi příliš často nesetkáme, tvary LZ existují paralelně pro každou základní číslovku až do 99 (*dziewięćdziesięcioro dziewięcioro*), pro číslovku *sto* už LZ neexistuje. U vyšších čísel se používají základní číslovky až do předposledního řádu, poslední dva řády jsou LZ. Např.: *W tej orkiestrze gra **cztery sta pięćdziesięcioro** troje muzyków.* (V tomto orchestru hraje čtyři sta padesát tři hudebníků a hudebnic.)

#typ_net_interakt. Dále se to týká značky #typ_net_nieinterakt (tyto texty někdy korektury mají, např. v článcích na Wikipedii bývají chyby málokdy, ale přeci jen se to liší od situace např. v tisku) a textů z mluveného projevu, jež ovšem nejsou v mém korpusu zastoupeny prakticky vůbec (pouze 6 záznamů).

Praktické zjištění proto poměrně překvapí – ze 135 výskytů základní číslovky v našem vzorku jich 33 pochází z diskusních fór na internetu, 6 z internetových stránek, 12 z literatury a více než polovina – 84 – z tištěných publicistických textů. Je tedy zřejmé, že ačkoli texty procházejí v mediálních institucích korekturou, tendence nahradit souborovou číslovku číslovkou základní je tak silná, že i profesionál orientovaný na dodržování jazykové normy jí někdy podlehne.³³

Porovnáme-li hodnoty LZ a LG z tabulek č. 4 (SYM) a 5 (JED), je patrné, že základní číslovka je častěji alternativou číslovky souborové než výrazu se slovem *para*. (V tabulce č. 4 se LZ objevují třikrát a LG dvanáctkrát, zatímco v tabulce č. 5 najdeme 566 výskytů LZ a 123 výskytů LG.) Neboli – tyto dva typy číslovek si vzájemně konkurují. V případě, že to sémantika výrazu umožňuje (čili v případě slov se skupiny SMY), mluvčí evidentně rovnou volí konstrukci se slovem *para* nebo jinou opisnou formulaci tohoto typu. Váhání mezi užitím souborové a základní číslovky se tak objevuje hlavně v případě slov ze skupiny JED. Proto se také chybné užití základní číslovky (tj. výsledek tohoto váhání) častěji objevuje u slov ze skupiny JED než u slov ze skupiny SYM.

Podíváme-li se na následující příklady, uvidíme, že k záměně druhu číslovek došlo i v tak ustáleném a zavedeném kontextu, jako jsou názvy hudebních skladeb pro vícero houslí.

20. *Na moich **trzech spodniach** usiadł pies?*

(Na mých třech kalhotách seděl pes?)

21. *Okazało się, że **sześć drzwi** było zamkniętych, nikt nie pracował.*

(Ukázalo se, že šest dveří bylo zavřených, nikdo nepracoval.)

22. *Finałem koncertu było wykonanie Koncertu h-moll na **cztery skrzypce**.*

(Vrcholem koncertu bylo předvedení Koncertu h-moll pro čtyři housle.)

23. *Jechaliśmy w **czterech saniach**, które ciągnęło osiem koni.*

(Jeli jsme ve čtyřech saních, které táhlo osm koní.)

³³ Zároveň je však třeba mít na paměti, že soubor dat není z hlediska zdrojů textů vyvážený; publicistické texty v něm tvoří většinu – celých 61 %.

Zvláštní kategorie se při zpracovávání dat dostalo také **číslovkám zapsaným číslicí**. Teoreticky totiž nelze rozhodnout, zda se pod tímto symbolem skrývá číslovka základní nebo souborová (proto jsem pro ně zvolila značku LG/LZ). Autor textu měl pravděpodobně na mysli jen jednu z nich, každý čtenář se ovšem může přiklonit jak k jedné, tak ke druhé možnosti. Celkově jich v našem souboru dat není mnoho, jen necelá 2 % (24/1653), takže celkový „poměr sil“ příliš neovlivňují.

Nelze samozřejmě vyloučit ani možnost, že autor zapsal číslovku číslicí záměrně právě proto, aby se vyhnul problematickému tvaru souborové číslovky. Pravděpodobnější je to zejména u nižších hodnot, u nichž konvence velí spíše zápis slovy. Podívejme se na pár příkladů (první tři jsou z publicistiky, poslední je z diskusního fóra):

24. *Wczoraj 5 **sań** wypełnionych po brzegi szczęśliwymi dziećmi ruszyło spod Miejskiego Teatru.*
(Včera vyjelo od Městského divadla 5 saní po okraj naplněných šťastnými dětmi.)
25. *[...] suita „Z czasów Holberga“ E. Griega i Koncert na 4 **skrzypiec** H-moll A. Vivaldiego.*
([...] suita Z časů Holbergových od E. Griega a Koncert pro 4 housle H-moll A. Vivaldiho.)
26. *Wszystko zaczęło się od zakupu 12 **spodenek** i tyłuż koszulek.*
(Všechno začalo tím, že koupil 12 trenýrek a stejně tolik triček.)
27. *Królowa powiesiła na sznurku 10 **majtek**. Przyszedł złodziej, ukradł 3, ile majtek zostało królowej.*
(Královna pověsila na šňůru 10 kalhotek. Přišel zloděj, ukradl 3, kolik kalhotek královně zůstalo.)

Skupina číslovek označená **NOM** zahrnuje vlastně tři podskupiny výrazů. Všechny je spojuje, že se nacházejí na pomezí číslovek a podstatných jmen, a dohromady se jich v souboru dat vyskytlo tak málo, že by bylo asi zbytečné pro ně vytvářet samostatné kategorie.

Za prvé jsou to číslovky *sto*, *tysiąc*, *milion* (sto, tisíc, milion). Tyto LG nemají svůj paralelní tvar LZ. V souladu s normou tedy např. tisíc dveří = *tysiąc drzwi* (teoreticky lze utvořit tvar **tysięcioro*, ten však v praxi neexistuje). Přestože se tedy jedná o konkrétní číselnou hodnotu, mluví se při použití této číslovky nemusel rozhodovat, jak naložit s problematickým gramatickým pravidlem. Dilematu „souborová či základní?“ byl zcela ušetřen.

28. **Sto sań** skrzypiało po śnieżystej grudzie, sto sań leciało z wichrami w zawody [...] (Sto saní vrzalo na zmrzlém sněhu, sto saní letělo s větrem o závod [...])
29. **Na nabywcę czekać będzie 5 tys. kałesonów** męskich w pakiecie z 1,5 tys. koszul męskich. (Na nového majitele bude čekát 5 tis. pánských kamaší v balíku s 1,5 tis. pánských košil.)
30. **W ciągu roku sprzedaje się w Polsce około 7 milionów okularów**, z czego aż dwa miliony pochodzą z bazarów. (Během roku se v Polsku prodá kolem 7 milionů brýlí, z čeho až dva miliony pocházejí z bazarů.)

Druhou skupinou pod hlavičkou NOM jsou neurčité číslovky vzniklé pluralizací číslovek popsaných v předchozím odstavci: *dziesiątki*, *setki*, *tysiące*, *miliony* (desítky, stovky, tisíce, miliony). Pokud nejsou předcházeny další základní číslovkou (*dwa tysiące* apod.), „přestávají signalizovat konkrétní počet a stávají se neurčitými číslovkami o významu ‚hodně, mnoho‘“ (WSPP, s. 1675). Ani v tomto případě neexistují paralelní LZ a mluvčí se tedy vyhne problematické úvaze, jak si s nimi poradit. Zejména v případě nižších hodnot (*dziesiątki*) považují za možné, že mluvčí použije tuto neurčitou číslovku úmyslně místo toho, aby vyjádřil přesný počet souborovou číslovkou (např. *czterdzieścioro*).

31. **Kostiumolog przynosi dziesiątki spodni** i kurteczek w paski. (Kostymér přináší desítky pruhovaných kalhot a kabátků.)
32. **Dziesiątki drzwi** otwierały się jedne za drugimi, wracała znajomość terażniejszości. (Desítky dveří se otevíraly jedny za druhými, navracelo se vědomí současnosti.)
33. **Tysiące noszy**, które pomyślane były tak sprytnie, że [...] (Tisíce nosítek, která byla tak šikovně vymyšlená, že [...])

Třetí skupinu tvoří substantiva odvozená od základních číslovek (polsky *rzeczowniki odliczebnikowe*) – *dwójka*, *trójka*, *piątka*, *dziesiątka*, *osiemnastka* (dvojka, trojka, pětka, desítká, osmnáctka) atd. Ta představují další možnost, jak se vyhnout použití LZ. WSPP (s. 1677) uvádí tyto formy ve spojení s lexémem *dziecko*, což je i podle zběžného pohledu do korpusu jejich nejčastější užití. Obecně se užívají především při počítání lidí (ať už skupin smíšeného pohlaví nebo dětí), ale jeden příklad (č. 34) se našel i v rámci mého souboru dat s PT.

Jako zvláštnost připojuji ještě ojedinělý příklad z beletrie (č. 35), v němž figuruje dnes již zastaralá nominální číslovka *kopa* (čili 60), která funguje stejným způsobem jako např. *tysiąc*.

34. *Do naszej stałej dyspozycji była **dwójka** zaprzężonych **sań**.*

(Měli jsme k dispozici dvojku zapřažených saní.)

35. *Jedna markietanka statecznego autoramentu, mendel kos, **dwie kopy wideł**, tuzin drągów, dwie baby z kopystkami [...]*

(Jedna markytánka statečného autoramentu, mandel kos, dvě kopy vidlí, tucet drapáků, dvě ženské s vařečkami [...])

Poslední kategorie označená zkratkou **WP** (*wyrazy pomocnicze*, pomocná slova) patří k těm, jejichž existenci jsem předpokládala ve vstupní hypotéze. Odpovídá totiž skupině P5 na schématu Anny Andrejczuk (viz oddíl 2.2), které posloužilo jako východisko při samotném výzkumu. Měla by do něj spadat pomnožná jména typu *schody* nebo *korale*, která označují předměty složené z většího množství malých částí. Počítat by se měla buď pomocí LZ, nebo pomocí pomocného slova, které odkazuje ke způsobu organizace daného předmětu. V praxi se ovšem ukázalo, že obě příkladová slova se počítají výhradně pomocí pomocných výrazů, LZ se ve spojení s nimi neobjevila ani jednou. U dvou lexémů uváděných Annou Andrejczuk jsem začala a nakonec i skončila. Nepodařilo se mi najít žádná další PT obdobného typu, u nichž by se objevovala alespoň nějaká pravidelnost ve volbě pomocného slova. I u obou slov převzatých od Andrejczuk jsem zjistila větší variabilitu, než o jaké hovoří ve svém článku. S nepočtenými příklady nalezenými u pouhých dvou lexémů (*korale*, *schody*) je obtížné cokoli zobecnit. Domnívám se, že spolehlivý přehled o tomto jevu by mohl přinést jen skutečně hodně rozsáhlý soubor dat, k němuž je ještě dlouhá cesta.

Pokud jde o slovo *korale*, byly výsledky poměrně jednoznačné. Dvanáctkrát se objevilo slovo *sznur* (šňůra) a čtyřikrát deminutivum *sznurek* (šňůrka). Tento pomocný výraz se opět často používá i ve spojení s jedním exemplářem předmětu:

36. *Trzy grube **sznury** różowych **korali** od najdawniejszych czasów zakładała co rana.*

(Tři silné šňůry růžových korálí si odjakživa brala každé ráno.)

37. *Boryna wczoraj dał jej **osiem sznurków korali**, wszystkie jakie mu zostały po nieboszczkach...*

(Boryna jí včera dal osm šňůrek korálí, všechny, jež mu zůstaly po nebožkách...)

38. *To znaczy połączyć kilka drewnianych koralików w **jeden sznur koralí**.*

(To znamená spojit několik dřevěných korálků v jednu šňůru korálí.)

V případě slova *schody* je už situace komplikovanější. Anna Andrejczuk uvádí ve svém schématu spojení *dwa ciągi schodów*, já jsem však pomocných slov identifikovala celkem sedm (pro české ekvivalenty viz následující seznam příkladů z korpusu): *ciąg* (14×), *bieg* (11×), *piętro* (3×), *rzęd* (3×), *kondygnacja* (2×), *spirala* (1×), *kolumna* (1×). V tomto případě bylo hledání paralelních a přirozeně znějících českých konstrukcí převážně neúspěšné, pro názornost jsem tedy překládala doslovně:

39. *W wewnętrznym pierścieniu utworzono **trzy biegi** dookolnych **schodów**.*

(Ve vnitřním prstenci byly vytvořeny tři běhy schodišť/schodů kolem dokola.)

40. ***Dwa ciągi** marmurowych **schodów**.*

(Dva tahy mramorových schodišť/schodů.)

41. *[...] na wiadukcie nad ul. Lubicz schodziły do jezdni **dwie kolumny schodów**.*

([...] na viaduktu nad ul. Lubicz vedly k vozovce dva sloupce schodišť/schodů.)

42. *Stąd **dwie kondygnacje schodów** prowadziły w górę i na dół.*

(Odtud vedla dvě podlaží schodišť/schodů nahoru i dolů.)

43. *Żeby do niej dotrzeć, trzeba najpierw pokonać **trzy piętra** kręconych **schodów** w Teatru Wielkim.*

(Aby se k ní dostal, musel nejdříve překonat tři patra točitých schodišť/schodů ve Velkém divadle.)

44. *Pod drzwiami sklepów prowadzi **dwa rzędy** szerokich **schodów**.*

(Ke dveřím obchodů vedou dvě řady širokých schodišť/schodů.)

45. *Zbudowane w jej wnętrzu **dwie spirale schodów** skręcają w tę samą stronę, nie krzyżując się ze sobą.*

(Dvě spirály schodišť/schodů postavené uvnitř [vily] se stácejí na tutéž stranu, vzájemně se nekříží.)

Jak je zřejmé z uvedených příkladů, některá pomocná slova navazují na způsob fyzické organizace schodiště (schodiště jako řada jednotlivých schodů – *ciąg*, *bieg*, *rzęd*, *kolumna* – tah, běh, řada, sloupec), jiná zřetelně odkazují k jeho tvaru (*spirala* – spirála) anebo k prostoru, který pokrývá či spojuje (*piętro*, *kondygnacja* – patro, podlaží). Některá nepochybně vznikla *ad hoc* (č. 45), u jiných lze předpokládat určitou pravidelnost, kterou by bylo záhodno ověřit na větším souboru dat.

Celkově se domnívám, že tato poslední skupina PT nezapadá příliš dobře mezi ostatní. Především je to dáno tím, že k těmto slovům (*schody*, *korale* a další

předpokládání členové této skupiny) vždy existuje singulár. Slovo *koral* nebo *schod* vlastně znamená „jedna část, jeden dílek souboru *schody* či *korale*“. Způsob vnímání těchto předmětů jako „souborů složených z částí“ se tedy dost liší od ostatních PT z mnou zkoumané množiny, které jsou buď zcela jednolitými předměty, nebo jsou vnímána jako předmět složený ze dvou (a *právě ze dvou*) částí.

3.3.3 Vztahy mezi normou a územ II

Když jsme teď získali představu o skutečném jazykové chování mluvčích polštiny, pokud jde o počítání neživotných PT, je načase, abychom se opět podívali na normativní příručku a konfrontovali její doporučení (či nařízení) s realitou.

Jako zástupce normativní autority posloužil *Wielki Słownik Poprawnej Polszczyzny* (WSPP), který je dostatečně rozsáhlý (cca 30 000 výrazů) na to, aby ve formě hesel obsahoval všechna výše zkoumaná PT. Hesla ve WSPP jsou poměrně krátká a orientují se především normativním směrem – kromě informací o pádových koncovkách a významu hesla proto také často obsahují upozornění na jevy, v nichž se často chybuje. Nejlépe to ilustruji, když jedno heslo ocituji v úplnosti:

Spodenki, blp, D. spodenek (*nie*: spodenków): Spodenki dziecięce. Spodenki gimnastyczne, plażowe. Trzy pary spodenek (*nie*: trzy spodenki). Włóżyć (*nie*: założyć, *nie*: ubrać) spodenki. (WSPP 2006, s. 955)³⁴

Jak je zjevné z citovaného hesla, slovník se u pomnožných jmen vyjadřuje právě i ke způsobu kvantifikace. V tomto případě uvádí, že správný způsob počítání slova *spodenki* je pomocí konstrukce *s para*, a naopak výslovně zapovídá použití základní číslovky. K použití souborové číslovky se explicitně nevyjadřuje.

Když jsem shromáždila popisy všech zkoumaných PT, ukázalo se, že slovník je ve svých pokynech poměrně nekonzistentní. Někde uvádí jen správné formy, jinde také explicitně vyjmenovává, které formy jsou nesprávné, a pro některá další PT nedává, pokud jde o kvantifikaci, doporučení vůbec žádná.

Informace z jednotlivých hesel jsem shromáždila do následující tabulky č. 6; druhý a třetí sloupec shrnují, co WSPP doporučuje/nedoporučuje, zatímco ve čtvrtém sloupci je

³⁴ „**Spodenki**, blp [= bez jednotného čísla], gen. spodenek (*ne*: spodenków). Spodenki dětské. Spodenki gymnastické, plážové. Tři páry spodenek (*ne*: tři spodenki). Obléci (*ne*: navléci, *ne*: vzít si) spodenki.“

pro porovnání zopakováno, která forma kvantifikace byla nejběžnější v provedeném výzkumu. Některé řádky nejsou vyplněny, protože dané PT se v souboru dat vyskytlo v příliš malém počtu příkladů. Jako hranici jsem zde stanovila 20 dokladů.

Tabulka 6. Porovnání normou (ne)doporučovaných forem kvantifikace sledovaných PT s nejméně frekventovanější formou kvantifikace stanovenou na základě korpusu.

PT	normativně (WSPP)		empiricky (NKJP)
	správné formy	nesprávné formy	
<i>kalesony</i>	dwie pary kalesonów <i>przest.</i> dwoje kalesonów	dwa/dwie kalesony	dwie pary kalesonów (84 %)
<i>gogle</i>	dwoje gogli <i>pot.</i> dwie pary gogli	x	(málo údajů)
<i>okulary</i>	dwoje okularów dwie pary okularów	x	dwie pary okularów (82 %)
<i>nożyce</i>	dwoje nożyc dwie pary nożyc	x	(málo údajů)
<i>nożyczki</i>	dwoje nożyczek dwie pary nożyczek	x	dwie pary nożyczek (85 %)
<i>spodnie</i>	dwie pary spodni jedna para spodni	jedne spodnie	dwie pary spodni (94 %)
<i>spodenki</i>	dwie pary spodenek	trzy spodenki	dwie pary spodenek (64 %)
<i>obcęgi</i>	dwoje obcęzek/obcęarów	dwa obcęki/obcęgi	(málo údajů)
<i>kombinerki</i>	dwoje kombinerek dwie pary kombinerek	dwie kombinerki	(málo údajů)
<i>slipy</i>	dwie pary slipów	x	(málo údajů)
<i>drzwi</i>	dwoje drzwi	dwa drzwi	dwoje drzwi (85 %)
<i>szczypce</i>	dwoje szczypiec	dwa szczypce	(málo údajů)
<i>grabie, widły</i>	dwoje grabi/ wideł	x	(málo údajů)
<i>sanie</i>	dwoje sań	dwie sanie	dwie pary sań (30 %) dwoje sań (28 %)

			dwie sanie (28 %)
<i>skrzypce</i>	dwoje skrzypiec	dwa skrzypce dwie pary skrzypiec	dwoje skrzypiec (76 %)
<i>korale</i>	sznur korali	x	sznur/sznurek (100 %)

Jak je zřejmé z uvedené tabulky, kombinatorických možností je mnoho. V případě některých PT slovník vydává pouze kladné pokyny (viz např. *slipy*, *grabie*, *widły*), v případě jiných i výslovně uvádí, které (podle autorů slovníku zřejmě často užívané) formy jsou nesprávné (viz např. *sanie*, *kombinerki*, *drzwi*). V některých případech doporučuje pouze jednu formu (většinou LZ, ale nikoli nutně, viz *slipy*) v jiných dává na výběr (pokud ano, tak vždy mezi LZ a *para*).

Kromě jednoho případu slovník nikdy nezakazuje spojovat PT ze skupiny SYM s číslovkou souborovou. Onou zajímavou výjimkou je slovo *spodnie*, u něhož je explicitně uvedeno, že spojení *jedne spodnie* je nesprávné. U výrazu *spodenki*, který má velmi podobný význam, ale je méně frekventovaný, už ovšem v rubrice „zákazů“ vidíme spojení s LG (*trzy spodenki*). Pokyn používat konstrukci s *para* je totožný, LZ v tomto případě nejsou zmíněny.

Porovnání s třetím sloupcem nikde³⁵ nevede k přímému rozporu – to znamená, že se v praxi neprosazuje jako dominantní takový způsob kvantifikace, který by normativní autorita explicitně odmítala. Dokonce se vždy jako nejfrekventovanější objevuje jedna z možností, kterou slovník přímo akceptuje jako správnou. Zásadní však je, že touto nejfrekventovanější formou kvantifikace (kromě slov *skrzypce* a *drzwi*) nikdy není LZ. Ačkoli slovník prosazuje užívání souborových číslovek na prvním místě, mluvčí se jim systematicky vyhýbají.

Poslední poznámka se týká slovníkových hesel jednotlivých PT; u dvou slov ze souboru jsou v informaci o způsobu počítání využity slovníkové kvalifikátory:

- (1) u slova *kalesony* je uvedeno, že forma s LZ (*dwoje kalesonów*) je zastaralá,

³⁵ Jedinou výjimkou z tohoto tvrzení je heslo *sanie*. Slovník jasně doporučuje LZ a odmítá LG, ale v mém souboru dat se všechny tři základní možnosti – LG, LZ a *para* – vyskytují s téměř shodnou frekvencí. Celkově se však jedná jen o 76 dokladů, takže je otázka, jak by se tento poměr změnil při větším objemu dat.

(2) u slova *gogle* je uvedeno, že forma *s para (dwie pary gogli)* je oproti formě z LZ (*dwoje gogli*) hovorová.

Oba tyto údaje směřují k tématu, které nás potenciálně velmi zajímá: v prvním případě diachronní vývoj způsobů počítání PT, ve druhém případě stylistické hodnocení jednotlivých způsobů počítání PT. Bohužel se však tyto kvantifikátory objevují řídce a nahodile – např. u spojení *dwie pary okularów* (významově téměř totožného s *dwie pary gogli*) žádná specifikace tohoto typu uvedena není. Slovník podle všeho není v této otázce příliš konzistentní.

4 Závěr

V teoretické části této práce jsem představila problematiku pomnožných substantiv a souborových číslovek v polštině a doplnila ji o teze a závěry několika polských lingvistů, kteří se této problematice věnovali v uplynulých letech. Zároveň jsem již v této kapitole ve shodě s Annou Andrejczuk naznačila, že úzus se patrně dosti vzdaluje od toho, co normativní příručky (např. WSPP) prezentují jako normu.

V praktické části jsem představila východiska praktického výzkumu, jeho zdroje i postup realizace. Výsledný soubor dat byl poměrně výrazně ovlivněn technickými záležitostmi jako je fyzická dostupnost korpusu NKJP či nástroje k jeho prohledávání, protože jsem i tuto ryze praktickou stránku věci popsala dosti důkladně.

Nejvíce prostoru bylo věnováno analýze nasbíraných dat a interpretaci výsledků. Při tom jsem často využila i vlastní zkušenosti z polského prostředí, které jsem nabyla během studijního pobytu na Vratislavské univerzitě. V jeho rámci jsem měla možnost porovnávat fakta a tendence zjištěné z polského korpusu s chováním mluvčích v každodenních situacích. Kromě toho jsem měla vynikající příležitost zažít v praxi také výuku věnovanou těmto jevům – a to hned ve dvou obměnách: souborovým číslovkám bylo zasvěceno několik hodin na kurzu polštiny pro nerodilé mluvčí a zároveň jim byl věnován i jeden seminář předmětu Kultura jazyka³⁶, který je naopak určen pro rodilé mluvčí – studenty polonistiky.

Ve všech těchto sférách se potvrdila moje výchozí hypotéza, že reálná frekvence užívání souborových číslovek není právě vysoká. Analýza dat získaných z korpusu však ukázala, že místo těchto číslovek se používá poměrně pestrý výběr různých opisů, analytických formulací či různých kompenzačních mechanismů. Celkově mezi těmito postupy převládá užití konstrukce s pomocným slovem *para* (pár), které se postupně gramatikalizuje, tj. ztrácí svůj lexikální význam a jeho užití se stává automatickým.

Když jsem porovнала získané výsledky s normativní příručkou (WSPP – Velkým slovníkem *správné* polštiny), byla jsem ovšem poněkud překvapena. Vzhledem k tomu, kolik času a úsilí se věnuje výuce souborových číslovek, jsem očekávala, že slovník bude jednoznačně prosazovat souborové číslovky jako jediný možný způsob počítání

³⁶ Pod tímto názvem se v Polsku tradičně skrývá normativní mluvnice a pravopis.

pomnožných jmen (tedy těch, která vůbec počítat lze). Opak je ovšem pravdou – souborové číslovky jsou sice téměř vždy uvedeny na prvním místě, normativní autorita však zároveň připouští i alternativní způsoby kvantifikace s výrazem *para* v čele. I vzhledem k tomu se tedy nemohu ubránit myšlence, že se v rámci výuky věnuje tomuto v lepším případě silně lexikálně omezenému, v horším rovnou mizejícímu jevu snad až příliš mnoho prostoru a energie, které by mohly být využity k jiným cílům.

Do budoucna se nabízejí dvě logické cesty rozšíření tohoto výzkumu. První z nich je spíše skromná – replikovat tentýž výzkum na větším objemu dat. Rozšířit seznam různých PT, která jsou do něj zahrnuta, a zároveň podstatně zvýšit počet dokladů k jednotlivým pomnožným jménům. To by však vyžadovalo větší korpus, než je ten, který je k dispozici nyní.

Druhá možnost, jež by byla realizovatelná již se stávajícím korpusem, spočívá v rozšíření výzkumu směrem k jiným skupinám substantiv, které se také počítají pomocí souborových číslovek. Tím se ohnisko zájmu přesouvá od „počítání pomnožných jmen“ k „počítání pomocí souborových číslovek“ a dostávají se do něj jevy popsané v oddílu 2.1, tedy počítání dětí a mláďat a především počítání skupin jedinců obojího pohlaví. Vzhledem k tomu, že se tento jev poměrně dobře slučuje s požadavkem politické korektnosti a v některých případech může elegantně a stručně nahradit jiné formy genderově neutrálního vyjádření (např. tzv. *splitting* nebo dvojí tvary³⁷), je dokonce možné, že v budoucnosti nejen nevymizí, ale jeho frekvence bude opět stoupat.

³⁷ Při *splittingu* se uvádí koncovka *feminina* za lomítkem (např. *student/ka*), ve druhém případě jsou uvedena obě slova (*student* a *studentka*).

5 Bibliografie

Odborná literatura

ANDRZEJCZUK, Anna. Dwoje urodzin to brzmi dziwnie. Norma językowa dotycząca połączeń rzeczowników plurale tantum z liczebnikami a jej realizacja w tekstach. *Język polski*. 2011, č. 4, s. 273–283.

COMRIE, Bernard. Balto-Slavonic. In GVOZDANOVIĆ, J. (ed.). *Indo-European numerals*. Berlin, New York : Mouton de Gruyter, 1991.

DUBISZ, Stanisław. *Uniwersalny słownik języka polskiego*. Tom 7 – indeks a tergo. Warszawa : PWN, 2008.

FIEDLER, Adrian. *Liczebniki w języku polskim i chorwackim (i serbskim) pod względem kategorii żywotności*. 2009. [dostępne z WWW:
[<http://opus.kobv.de/ubp/volltexte/2009/2776/pdf/opole07.pdf>]

GRZEGORCZYKOWA, Renata (ed.). *Gramatyka współczesnego języka polskiego. Morfologia*. 2. wydanie zmienione. Warszawa : PWN, 1998.

JADACKA, Hanna. Pluralia tantum w opisie leksykograficznym. *Lingvaria*. 2006, č. 1, s. 161–169.

MARKOWSKI, Andrzej (red.). *Wielki Słownik Poprawnej Polszczyzny*. Wydawnictwo PWN : Warszawa, 2006.

NYMBURSKÁ, Dita – VOSTRÁ, Denisa – SAWATARI, Mami. *Japonština*. LEDA : Praha, 2007.

PRZEPIÓRKOWSKI, Adam et al. *Narodowy korpus języka polskiego*. Warszawa : PWN, 2012.

SALONI, Zygmunt. Kategoria rodzaju we współczesnym języku polskim. In Laskowski, R. (ed). *Kategorie gramatyczne grup imiennych we współczesnym języku polskim*. Wrocław : Ossolieuum, 1976, s. 43–78.

SALONI, Zygmunt. So-Called Collective Numerals in Polish (in Comparison with Russian). *Studies in Polish Linguistics*. 2009, č. 5, s. 51–60.

SIUCIAK, Mirosława. *Kształtowanie się kategorii gramatycznej liczebnika w języku polskim*. Katowice : Wydawnictwo Uniwersytetu Śląskiego, 2008.

SHIMAMORI, Reiko. *Grammaire japonaise systématique*. Paris : J. Maisonneuve, 1994.

Korpus (zdroj jaz. dat)

Narodowy Korpus Języka Polskiego, Konsorcium NKJP. Cit. 1. 1. 2014. [dostępny z WWW: <http://nkjp.pl>]

PĘZIK, Piotr. *Wyszukiwarka PELCRA dla danych NKJP. Narodowy Korpus Języka Polskiego*. Warszawa : Wydawnictwo PWN, 2012.