

**UNIVERZITA PALACKÉHO V OLOMOUCI**

**FILOZOFICKÁ FAKULTA**

**KATEDRA BOHEMISTIKY**

Česká filologie



**KVANTITATIVNÍ ANALÝZA FUNKČNÍCH STYLŮ**

**QUANTITATIVE ANALYSIS OF FUNCTIONAL STYLES**

Magisterská diplomová práce

Autor: Lenka Horutová

Vedoucí práce: PhDr. Petr Pořízka, Ph.D.

Olomouc 2016

## **Prohlášení**

Prohlašuji, že jsem magisterskou diplomovou práci vypracovala samostatně za použití literatury uvedené v seznamu na konci práce.

V Olomouci dne 16. 4. 2016

.....

# 1. KVANTITATIVNÍ LINGVISTIKA

Spojení matematiky a lingvistiky není v českém lingvistickém prostředí novou záležitostí. Základy kvantitativní lingvistiky jako „složky matematické lingvistiky, která kvantifikuje (zjišťuje kvantitativní data) jevy různých jazykových rovin a modeluje jejich vztahy realizující se ve větě, v textu, abychom lépe poznali jejich příčinný mechanismus, jejich fungování, jejich stránku formální, ale i sémantickou“ (Těšitelová 1987, s. 8), se u nás formovaly již ve 30. letech 20. století za výrazného přispění lingvistů, jako byli Vilém Mathesius, Bohumil Trnka nebo Josef Vachek. Na počátky mezioborové disciplíny aplikující při studiu jazykových jevů kvantitativní metody měli však kromě lingvistů velký vliv také pedagogové, mj. pedagog Václav Příhoda, který spolu s bohemistou Vladimírem Šmilauerem iniciovali vznik prvního českého frekvenčního slovníku, jenž byl původně zamýšlen pro pedagogické účely (srov. Těšitelová 1987, s. 9). Největším přínosem pro rozvoj české kvantitativní lingvistiky se však staly práce ze 70. a 80. let 20. století,<sup>1</sup> jejichž autorka Marie Těšitelová bývá právem považována za průkopnici kvantitativního přístupu v oblasti české jazykovědy.

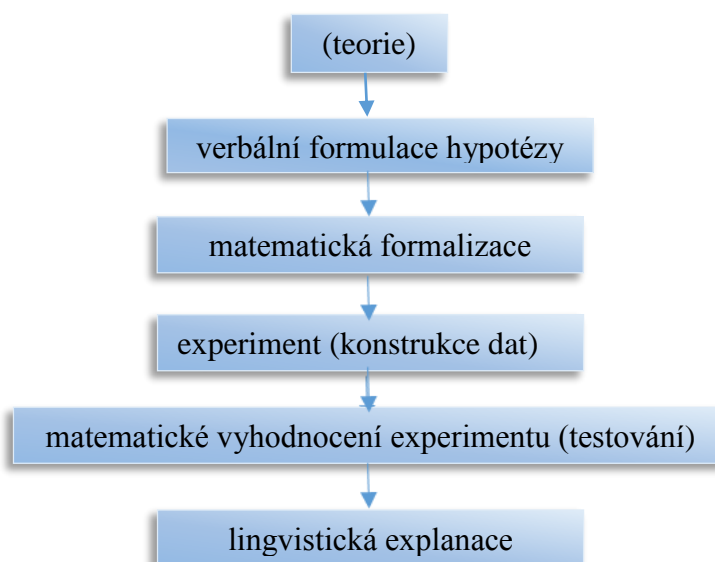
Přestože byly matematické metody implementovány do lingvistického bádání již na počátku minulého století, většina provedených analýz měla často jen deskriptivní charakter. Cílem takovýchto výzkumů se stával pouhý popis jazyka a klasifikace jednotlivých jazykových jevů, doplněné o ilustrativní údaje týkající se kvantifikace (informace o frekvenci, procentuální vyjádření atd.). Radek Čech ovšem upozorňuje, že tento přístup není ničím víc než pouhým popisem s čísly, který neumožňuje vyvodit obecnější závěry ohledně fungování jazyka. „Zjištěné absolutní hodnoty se v praxi většinou normalizují do formy přehledného procentuálního vyjádření. Jenže pouhé konstatování, že pozorované jevy se vyskytují s různou frekvencí (buť normalizovanou), neříká o moc víc než prosté označení jevu číslem.

---

<sup>1</sup> Jedná se zejména o publikace *Otázky lexikální statistiky* (1974), *Psaná a mluvená odborná čeština z kvantitativního hlediska* (1983), *Kvantitativní charakteristiky současné češtiny* (1985), *O češtině v číslech* (1987), *Kvantitativní lingvistika* (1987) aj.

Kvantifikace začíná mít smysl v případě, kdy se snažíme předpokládané (a posléze zjištěné či nezjištěné) rozdíly interpretovat, tj. dáváme do souvislosti změřený rozdíl s vlastností jinou, např. formou jazyka, žánrem, sémantickými vlastnostmi, délkou [...] aj.“ (Čech 2014, s. 177)

Teprve od 70. let 20. století se pomalu začíná uplatňovat jiný pohled na kvantifikaci přirozeného jazyka – pohled, který se snaží popisně-klasifikační povahu dosavadních lingvistických bádání překonat a který si klade za cíl odhalit mechanismy řídící naše jazykové chování pomocí empiricky testovatelných hypotéz. Jako příklad si můžeme vzít dvě slova: „lev“ a „krokodýl“. To, že zjistíme, že se ve stomilionovém korpusu SYN2015 vyskytuje slovo „lev“ ve všech svých tvarech s frekvencí 2 626 a slovo „krokodýl“ pouze s frekvencí 882, nám samo o sobě neumožňuje vyslovit o fungování jazyka žádné závěry. Pokud ovšem počet výskytů vztáhneme k jiné vlastnosti pozorovaných slov, konkrétně k jejich délce, můžeme již objevit jistou zákonitost existující mezi délkou slov a jejich frekvencí a formulovat hypotézu „Čím je slovo frekventovanější, tím je kratší.“, jejíž platnost je snadné, avšak nutné ověřit na rozsáhlejší materiálu. Samotné konstatování, že mezi dvěma či více jazykovými vlastnostmi byla zjištěna vzájemná korelace, ovšem nestačí, ani když tento poznatek doložíme konkrétními číselnými hodnotami. Celé naše kvantitativní šetření dostane význam až na základě relevantní lingvistické interpretace získaných dat, jež je nejdůležitější částí výzkumu. Postup, který se obvykle uplatňuje při kvantitativnělingvistickém výzkumu, Čech, Popescu a Altmann (2014, s. 8) vyjádřili také graficky:



Lingvistické bádání využívající matematické metody vychází obvykle z jisté jazykové teorie, z níž také bývá odvozena hypotéza, jejíž pravdivost je potřeba výzkumem otestovat. Protože však ne každá hypotéza je skutečně na jazykovou teorii vázána, ale její formulace může vzniknout i jiným způsobem, prvotní krok je uveden v závorce. Ověřovanou hypotézu je poté potřeba vyjádřit matematicky, aby mohl být uskutečněn experiment ve formě měření nebo výpočtu. Výsledné hodnoty je následně nutné podrobit statistickému testu, který dokáže s přesně definovanou pravděpodobností chyby (obvykle stanovenou na 5 %) vyhodnotit, zda můžeme testovanou hypotézu přijmout, či zda ji budeme muset zamítnout. Zcela zásadním krokem je však závěrečná lingvistická interpretace poznatků, k nimž jsme na základě experimentu dospěli.

Důležité je si uvědomit, že přestože kvantitativní lingvistika využívá ke svému výzkumu matematické metody, její pozornost je stále soustředěna na studium přirozeného jazyka. To, čím se liší od jiných lingvistických disciplín, tak není předmět zkoumání, ale čistě jen způsob, jakým lingvisté dospívají ke svým závěrům. Metody kvantitativních lingvistů jsou založeny na experimentálním přístupu, který umožňuje mnohé do značné míry intuitivní teze týkající se přirozeného jazyka empiricky testovat. Tím se také kvantitativní lingvistika zařadila po bok exaktních věd, jako jsou fyzika či chemie.

Kvantifikace tedy není cílem kvantitativní lingvistiky, ale je jen prostředkem, který badatelům umožňuje získat intersubjektívni pohled na zkoumanou problematiku a který jim dokáže zaručit, že za dodržení stejných podmínek je kdokoli schopný daný experiment zopakovat vždy se stejnými výsledky.

## 2. FUNKČNÍ STYLY

S ohledem na to, že je práce věnována analýze funkčních stylů, považujeme za nutné nejprve vymezit základní termín „funkční styl“ a objasnit, které z mnoha klasifikačních kritérií, jež se v českém stylistickém prostředí uplatňují, budeme v naší práci respektovat.

Stejně jako jazyk sám, také styly podléhají neustálému vývoji a proměňují se pod vlivem společenských, kulturních a historických událostí. Kolektiv autorů *Současné stylistiky* potvrzuje, že „[j]azyk vždy reagoval na nové jevy v životě společnosti, s novými situacemi plnil nové komunikační funkce a různou měrou sloužil nově se objevujícím dobovým komunikačním potřebám. Docházelo k proměnám jazykové komunikace, což se projevuje ve vzniku a rozvoji různých stylů.“ (Čechová – Krčmová – Minářová 2008, s. 93)

Styl Milan Jelínek v *Příruční mluvnici češtiny* definuje jako „výsledek výběru jazykových prostředků z množin prostředků konkurenčních“ (Karlík – Nekula – Rusínová (eds.) 1995, s. 699). Faktory, jež tento výběr ovlivňují, se zpravidla označují jako slohotvorné činitele, které je možné dále rozlišit v závislosti na spjatosti s osobou autora komunikátu na faktory objektivní (mimopersonální) a subjektivní (personální). Právě na stylových faktorech lze založit jednu z mnoha klasifikací stylů. Milan Jelínek tak např. rozlišuje obecně styly objektivní a subjektivní, v rámci objektivních stylů pak navrhuje třídění na styly projevů psaných a mluvených, připravených a nepřipravených, styly monologické a dialogické, oficiální a familiární aj. (Karlík – Nekula – Rusínová (eds.) 1995, s. 705–706). Nejvýznamnějšími však z množiny objektivních stylů jsou styly funkční.

Současné české stylistické teorie vycházejí z přístupu, který ve 20. a 30. letech 20. století představili ve svých pracích zejména Bohuslav Havránek a Vilém Mathesius. Základním termínem jejich pojetí, které ovlivnilo celé české stylistické poznání, je právě termín funkce, jež stylistika chápe jako záměr autora komunikátu nebo účel, který daný projev v komunikaci plní. Převládající funkce

textu se tak stala spolu s výběrem výrazových prostředků hlavním kritériem při rozlišování jednotlivých funkčních stylů (Čechová – Krčmová – Minářová 2008, s. 28).

Počet a charakteristika stylů nezůstaly v průběhu času neměnné, na Havránka navázali např. K. Hausenblas, M. Jelínek nebo A. Jedlička, kteří v reakci na nově rozpoznané funkce jazyka postupně zformulovali nové názory na stylové diferenciaci (více informací o vývoji představ ohledně funkce jazyka a jeho funkční stylové diferenciaci viz Čechová – Krčmová – Minářová 2008, s. 93–105).

Ani v současné době nejsme schopni podat úplný přehled možných klasifikací funkčních stylů, neboť při jejich vymezení hraje kromě komunikační funkce značnou roli i míra zobecnění – při nižším stupni abstrakce je tak možné namísto stylu publicistického konkrétněji odlišit styl zpravodajský, úvahový, interviewový, přesvědčovací atd. (Karlík – Nekula – Rusínová (eds.) 1995, s. 725). Pojetí funkčních stylů uplatněné v naší práci se přidrží tradičnější české stylové diferenciaci, která respektuje šest funkčních stylů:

- prostěsdělovací,
- odborný,
- umělecký,
- publicistický,
- administrativní,
- řečnický.

Jak bylo nastíněno, přístupů ke stylové diferenciaci češtiny existuje celá řada, žádná klasifikace ovšem není a nemůže být univerzálně platná; stejně jako se postupem času mění a vyvíjí funkce jazyka, mění se a vyvíjí se stále i teorie funkčních stylů.

### 3. VOLBA JEDNOTKY

Dříve než přistoupíme k samotné analýze funkčních stylů, je potřeba si ujasnit, s jakými jednotkami budeme při zpracování kvantitativních dat pracovat. Poměrně problematické se jeví stanovení základní jednotky při aplikaci lexikálněstatistických metod. Za základní jednotku v lexikální statistice bývá zpravidla považováno slovo – to ovšem není v různých jazycích chápáno jednotně. V českém lingvistickém prostředí se tradičně problematika slova vnímá jako opozice slovního tvaru versus lemmatu. „V jazycích s bohatou morfologií, jako jsou jazyky slovanské, se tyto dvě ‚podoby slova‘ v lexikální statistice celkem jasně diferencují, v jazycích s morfologií chudou, jako jsou angličtina, němčina apod., nejsou tyto dvě podoby vždy dostatečně diferencovány.“ (Těšitelová 1974, s. 7) U češtiny jako silně flektivního jazyku tak většina jazykovědců dlouhou dobu preferovala práci s lemmatizovanými texty.

V současné době již ovšem řada lingvistů odmítá volbu lemmatu jako základní jednotky souboru takto jednoznačně přijmout. Např. Čech, Popescu a Altmann (2014, s. 10, 19–21) zdůrazňují, že jazykové jednotky jsou pouze lingvistické nástroje, jejichž prostřednictvím se jen snažíme uvažovat o jazyce; znamená to tedy, že žádná jednotka není jazyku vlastní, přirozená, tudíž je i nemožné některou označit za lepší ve smyslu „lépe odpovídající skutečnosti“ než jinou. Konkrétní volba jednotky by tak měla vždy vycházet z badatelského cíle.

Rovněž Miroslav Kubát (2012) v rámci své magisterské diplomové práce provedl výzkum, v němž se snažil ověřit oprávněnost tradiční volby lemmatu jako základní jednotky lexikostatistiky, a to tak, že veškeré výpočty pro měření bohatství slovníku provedl paralelně s lemmaty i se slovoformami. Závěry, k nimž došel na základě rozsáhlého materiálu ze všech funkčních stylů, naznačují, že ačkoli se výsledky získané při práci s lemmaty liší absolutními hodnotami od výpočtů prováděných se slovními tvary, vzájemné vztahy mezi lexikostatistickými parametry zůstávají vcelku zachovány, ať už kalkulujeme s jakoukoli z těchto jednotek.

Proto se domníváme, že je výhodné při volbě jednotek vzít v úvahu i jistá omezení, která vyplývají z práce s lemmaty: pro výzkum je nutné zvolit pouze texty,



jež prošly procesem lemmatizace, popř. je zapotřebí opatřit si vhodný software, který nám ke slovům jednotlivá lemmata přiřadí. Je však důležité zjistit, jak byla lemmatizace textů provedena, neboť neexistuje jednoznačné pravidlo, jež by nám říkalo, které jednotky bychom měli přiřadit k jednomu lemmatu (jisté problémy způsobují např. komparace adjektiv, přechylování substantiv, negace aj.), tudíž texty z různých zdrojů nemusejí být lemmatizovány stejným způsobem. Další nedostatek potom souvisí s mírou nepřesnosti procesu lemmatizace, protože ať už převádíme slova do jejich základního slovníkového tvaru ručně, nebo strojově, vždy při tomto procesu existuje jisté procento chybovosti. S ohledem na tyto obtíže technického charakteru i s ohledem na srovnatelné výsledky, ke kterým je možné dojít při práci s lemmaty i slovoformami, domníváme se, že je výhodné přihlídnout k mnohem jednodušší práci se slovními tvary a není potřeba za každou cenu zůstat u lemmatu jako tradiční základní lingvistické jednotky.

Přestože nelze žádné jednotky označit výhradně za správné, nebo špatné, existují jednotky, které jsou pro určitý výzkum vhodnější a méně vhodné. Nevyplývá z toho ovšem, že by bylo možné dogmaticky přiřazovat konkrétní jednotky k jednotlivým analýzám jako jediné možné řešení. Každý badatel si může zvolit jednotku, která nejlépe vyhovuje potřebám jeho výzkumu, a jsme přesvědčeni, že je možné diskutovat jen o užitečnosti volby příslušné jednotky, ale ne o jejím absolutním přijetí, či zavržení. V této práci budeme i my v souladu s výše uvedeným v závislosti na statistické metodě používat různé jednotky, které dle našeho přesvědčení nejlépe vyhovují jednotlivým analýzám.

## 4. VÝBĚR ANALYZOVANÉHO MATERIÁLU

Zásadní rozhodnutí, které může značně ovlivnit výsledky každého výzkumu, spočívá i ve výběru vhodného materiálu, jenž bude podroben zkoumání. Protože cílem našeho šetření je analýza funkčních stylů, předmětem našeho bádání by se v ideálním případě měly stát všechny projevy, jež byly realizovány v současné češtině. Protože však tento požadavek není z logických důvodů reálný, musíme se v našem výzkumu omezit na pouhý vzorek jazyka. Výběrový soubor by však měl vždy maximálně vyhovovat cíli bádání, a to jak po stránce kvantitativní, tak po stránce kvalitativní.

### 4.1 Zdroje výběrového souboru

V dnešní době již není žádnou výjimkou, že badatel využívá ke studiu jazyka rozsáhlé jazykové databáze, které shromažďují a zpracovávají texty nejrůznějších druhů. Možnosti, jež nám jazykové korpusy nabízejí, jsou značné, vzhledem ale k tomu, že pro naše zvolené statistické výpočty budeme používat k tomuto účelu vytvořený software, jenž vyžaduje materiál určený k analýze v čistém textovém formátu, naším primárním cílem v souvislosti s jazykovými korpusy byla jen excerpce úplných souvislých reálných jazykových projevů všech funkčních stylů; rozšířené funkce těchto zdrojů jazykových dat jsme tudíž nemohli plně využít.

Výběrový soubor jsme nakonec sestavili na základě textů vyexcerpovaných ze dvou jazykových korpusů – k získání psaných textů a oficiálních mluvených projevů jsme využili největší databázi *Český národní korpus*, neformální mluvené projevy jsme pak vybrali z *Olomouckého korpusu mluvené češtiny*.

#### 4.1.1 ČNK

Jak již bylo zmíněno, nejvýznamnějším zdrojem jazykových dat je v současnosti *Český národní korpus* (dále jen ČNK), jenž laické i odborné veřejnosti zpřístupňuje jazykový materiál o více než dvou miliardách slov. ČNK sestává z více než třiceti různých korpusů, z nichž jsme pro naše účely zvolili obecný korpus SYN, který

obsahuje několik desítek tisíc textů synchronní psané češtiny, tj. textů, které byly publikovány od roku 1990. Z této velké množiny textových projevů jsme postupně vyfiltrovali dostupné komunikáty z jednotlivých funkčních stylů (tj. stylu publicistického, odborného, administrativního a uměleckého), z každé podskupiny jsme pak sami vybrali 20 textů tak, aby byla ve výběrovém souboru zastoupena rozličná témata, v rámci jednotlivých stylů různé žánry a také aby zkoumaný materiál tvořily projevy odlišných produktorů, přičemž jsme zejména v uměleckém stylu preferovali texty známých autorů, které jsou opakovaně vydávány, a současnou češtinu tak stále ovlivňují.

Tímto jsme získali materiál pro analýzu prvních čtyř funkčních stylů, zbylo nám tak ještě zajistit podklady k výzkumu stylu řečnického a prostěsdělovacího. Protože oba dva styly se vyznačují primárně mluveností, bylo zapotřebí k jejich zkoumání využít korpusy mluveného jazyka. Pro kvantitativní analýzu řečnického stylu jsme zvolili specializovaný korpus SPEECHES z databáze ČNK, jenž je sestaven z oficiálních projevů prezidentů České republiky, příp. představitelů předchozích státních útvarů. Jedná se o přepisy připravených prezidentských promluv proslovených u příležitosti významných státních i jiných svátků, přičemž za oficiální podobu projevů, se kterou se v korpusu pracuje, je považována připravená psaná verze, která nereflktuje případná přeřeknutí či jiné drobné změny v realizovaném verbálním projevu (více viz charakteristika korpusu Speeches na stránkách ČNK).

Abychom zachovali jednotné kritérium vzniku analyzovaných textů nejpozději v roce 1990, zařadili jsme do našeho souboru jen projevy posledních tří českých prezidentů, tedy Václava Havla, Václava Klause a Miloše Zemana. Všichni zmínění své proslovy shodně přednesli při výročí vzniku samostatného československého státu, Václav Havel a Václav Klaus poté s projevy vystoupili také na Nový rok, zatímco Miloš Zeman svou další promluvu realizoval v období Vánoc.

Jsme si vědomi toho, že promluvy nejvyšších reprezentantů státu tvoří jen zlomek projevů náležejících do řečnického stylu. Protože jsme však nechtěli tento styl z naší analýzy vyřadit jen kvůli nedostatku různorodého materiálu, rozhodli jsme

se šetření provést i na takto úzce zaměřené skupině textů. Jsme přesvědčeni, že i navzdory nevelké rozmanitosti textového materiálu jsou prezidentské projevy vhodným zástupcem řečnického stylu, proto i závěry, k nimž dojdeme na základě kvantitativních metod, budeme moci považovat za obecně platné.

#### **4.2.1 OMK**

Zatímco podklady pro výzkum psaných komunikátů lze poměrně bez problémů opatřit v dostatečném množství, možnosti týkající se analýzy mluveného jazyka jsou značně omezenější. Soukromá spontánní komunikace, která by nebyla vázána jen na úzce vymezenou teritoriální oblast, ale reflektovala by projevy běžného denního styku na celém území České republiky, je z důvodu technických nároků zachycena jen ve velmi omezeném počtu jazykových databází. Protože korpus ORAL2013 z velké databáze ČNK neumožňuje uživatelům získat přístup k celým souvislým komunikátům, i když jinak výše uvedené požadavky splňuje, pro analýzu prostěsdělovacího stylu jsme zvolili přepisy nahrávek z *Olomouckého korpusu mluvené češtiny* (dále jen OMK), které jsme získali z archivu autora daného korpusu Petra Pořízky.

Velkou výhodou OMK je vždy existence dvou typů transkripcí pořízených nahrávek – badatelé tak mají k dispozici jednak přepisy ortografické, jež jsou vhodnější pro automatické lingvistické nástroje, jednak přepisy fonetické, které reflektují autentickou výslovnost i nespisovně užitá gramatické tvary jednotlivých mluvčích. Vzhledem ke kvantitativnělingvistickým nástrojům užitým v naší práci vyhodnotili jsme jako vhodnější pro náš výzkum verzi ortografickou. Považujeme však za vhodné upozornit na skutečnost, že i ortografické transkripce obsahují nejrůznější strukturní značky a metatextové informace, díky kterým je možné v korpusu vyhledávat jazykové jevy podle zadaných parametrů. Aby však nebyly výsledné hodnoty našeho výzkumu zkresleny, bylo zapotřebí metatextové údaje z vybraného vzorku odstranit.

Pro všechny užitá transkribované projevy platí, že se jedná o neformální spontánní promluvy dialogického charakteru, které v reálném čase trvaly zhruba

20 minut. Zúčastnění mluvčí si během hovoru nebyli vzájemně neznámí a fyzicky se vyskytovali na stejném místě, tzn. že dialogy nebyly vedeny prostřednictvím telefonu nebo jiného komunikačního přístroje. Při výběru jednotlivých komunikátů jsme stejně jako v případě psaných textů dbali na rozmanitost užitého materiálu. Původci mluvených projevů pocházejí z různých míst České republiky (zastoupena je oblast Čech, Moravy i Slezska), dosáhli různých stupňů vzdělání, věkově se pohybují v rozmezí 17–86 let a pracují na rozličných pozicích. Ani tématem hovoru nebyli mluvčí nijak limitováni, vzhledem ale k tomu, že se jedná o rozhovory realizované v každodenních komunikačních situacích, není překvapivé, že se předmětem dialogu často stala rodina, škola, práce, domácnost, současné společenské dění apod.

## 4.2 Kvantitativní hledisko

Jak už bylo uvedeno, při výběru zkoumaného materiálu je nutné vycházet primárně z cíle výzkumu, přičemž v potaz by se měla vzít otázka kvantity i kvality výběrového souboru. Pokud budeme pracovat s příliš malým vzorkem, mohou být naše poznatky do značné míry zkresleny vůči výsledkům, kterých bychom dosáhli při práci s celkovou populací. Ani velký výběrový soubor nám ovšem nezaručuje přesnější výsledky. Radek Čech (2014, s. 176) k této problematice poznamenává: „[...] jakýkoliv obecný korpus je vždy souborem různých textů, jejichž jazyk je ovlivněn autorstvím, žánrem atd. Z toho plyne, že zkoumáme-li nějaký jazykový jev, pravděpodobnost jeho výskytu je pro každý text, autora, dílo atd. jiná, což má následující důsledek: zvětšujeme-li počet pozorování *pomocí* a *uvnitř* korpusu, obecně se nepohybujeme v homogenním prostoru. Extrémní růst počtu sledovaných jevů proto nemusí nutně znamenat extrémní zpřesnění poznatků.“ Čech (2014, s. 177) dále upozorňuje, že vzhledem k charakteru jazyka jako takového je v podstatě nemožné přesně určit, jaký vzorek je dostatečně reprezentativní vůči celkové populaci, již tvoří právě jazyk samotný.

Pokud však nechceme zůstat u pouhé deskripce zkoumaných jazykových jevů, ale budeme usilovat o experimentální analýzu odhalující principy, které ovlivňují uživatele jazyka v jejich jazykovém chování, můžeme vyjít rovněž

z Čechova přesvědčení, že pro testování hypotéz nepředstavuje problematika reprezentativnosti zkoumaných vzorků žádnou velkou překážku. Čech (2014, s. 181) je totiž přesvědčen, že pokud jednotlivé mechanismy, které jsou předmětem ověřování našich hypotéz, skutečně platí, řídí se těmito mechanismy chování jednotlivých jazykových uživatelů, a proto se musejí tyto mechanismy projevit i v jakémkoli jednotlivém mluveném nebo psaném textu.

Vzhledem k tomu, že naši práci pojmáme jako experimentální sondu, budeme během našeho výzkumu pracovat s databází menšího rozsahu. Z každého funkčního stylu podrobíme analýze 20 textů, celkově tak budou vybrané kvantitativní metody aplikovány na 120 jazykových projevech. S ohledem na výše uvedené se přesto domníváme, že i na takto nevelkém počtu analyzovaných textů budeme schopni odhalit jisté tendence, které se v jazyce projevují a které jednotlivé funkční styly od sebe navzájem diferencují. Počítáme však s tím, že do budoucna bude potřeba naše závěry ověřit na rozsáhlejší materiálu, popř. je doplnit dalšími studii.

### **4.3 Kvalitativní hledisko**

Při stanovení výběrového souboru je potřeba vzít v úvahu nejen otázku kvantity, ale je nutné zohlednit také stránku kvalitativní. Jak již bylo naznačeno, někteří lingvisté (mj. Kubát 2015, Čech 2014) důrazně upozorňují, že jednotlivé texty v sobě nesou nejen charakteristiky konkrétních funkčních stylů, ale jsou znatelně ovlivněny také stylem samotného autora, volbou tématu, příslušným žánrem a mnohými dalšími faktory. Pokud tedy do analyzovaného materiálu zařadíme projevy různých autorů, všechny tyto vlivy budou promíchány a my pak nebudeme schopni jednoznačně rozhodnout, zda jsme k výsledným datům dospěli na základě působení konkrétního funkčního stylu, nebo individuálního stylu autorského.

Přestože s uvedeným tvrzením souhlasíme a věříme, že by korpus sestavený z textů jednoho autora vykazoval snáze akceptovatelné výsledky, do našeho analyzovaného souboru jsme zahrnuli komunikáty nejen různých funkčních stylů, ale také různých autorů. Naše rozhodnutí bylo ovlivněno jednak skutečností, že bychom

jen stěží mohli najít českého autora, který by v dostatečném množství produkoval jazykové projevy všech funkčních stylů, jednak by interpretace takto získaných výsledků musela být omezena čistě jen na komunikáty příslušného autora. Navíc jsme přesvědčeni, že i navzdory působení různorodého autorství a různorodých témat budou námi užití statistické metody v textech schopny rozpoznat signifikantní rysy diferencující jednotlivé funkční styly.

Náš výběrový korpus tak tvoří komunikáty šesti funkčních stylů, každý funkční styl pak obsahuje 20 jazykových projevů, vždy od různých autorů a pokud možno také co nejvíce tematicky odlišných. Celkový přehled analyzovaných textů uvádíme v tabulkách 1–6:

**Tabulka 1:** Přehled jazykových projevů zařazených do publicistického stylu

	<b>typ periodika</b>	<b>název periodika</b>	<b>rok vydání</b>
<b>publicistický styl</b>	bulvární noviny	<i>Blesk</i> , 30. 3. 2009	2009
		<i>Blesk magazin</i> , č. 39/2007	2007
		<i>Bulvár</i> , č. 6/2008	2008
		<i>Nedělní Blesk</i> , č. 13/2009	2009
		<i>ŠÍP</i> , 11. 1. 2007	2007
	seriózní noviny	<i>Hospodářské noviny</i> , 5. 3. 2009	2009
		<i>Lidové noviny</i> , 15. 1. 2009	2009
		<i>Metro</i> , 15. 1. 2008	2008
		<i>Mladá fronta DNES</i> , 9. 1. 2004	2004
		<i>Právo</i> , 19. 1. 2004	2004
	regionální noviny	<i>Haló Brno</i> , č. 9/2003	2003
		<i>Karvinský zpravodaj</i> , č. 6/2003	2003
		<i>Kopřivnické noviny</i> , č. 18/2005	2005
		<i>Náchodský zpravodaj</i> , č. 11/2004	2004
		<i>Radniční zpravodaj – Pardubice</i> , č. 9/2001	2001
	časopisy	<i>Nedělní svět</i> , č. 28/2004	2004
		<i>Reflex</i> , č. 48/2007	2007
		<i>Respekt</i> , č. 46/2007	2007
		<i>Týden</i> , č. 21/2004	2004
		<i>Živel</i> , č. 3/2002	2002

**Tabulka 2:** Přehled jazykových projevů zařazených do uměleckého stylu

	žánr	název textu	rok vydání
umělecký styl	romány, novely	Ladislav Fuks: <i>Vévodkyně a kuchařka</i>	2006
		Bohumil Hrabal: <i>Pábení</i>	1993
		Petra Hůlová: <i>Cirkus Les Mémoires</i>	2005
		Josef Škvorecký: <i>Tankový prapor</i>	1990
		Michal Viewegh: <i>Román pro ženy</i>	2001
	povídky	Jan Balabán: <i>Je čas hledat i čas ztrácet</i> (z povídkové knihy <i>Možná že odcházíme</i> )	2007
		Alexandra Berková: <i>Miniromán</i> (z povídkové knihy <i>Knížka s červeným obalem</i> )	2003
		Lev Blatný: <i>Osamělí</i> (z povídkové knihy <i>Servus, Ser-vá-ci</i> )	2003
		Květa Legátová: <i>Zrcadlo ve větru</i> (z povídkové knihy <i>Želary</i> )	2001
		Filip Topol: <i>Zápisky milencovy</i> (z povídkové knihy <i>Tři novely</i> )	2004
	imaginativní texty	Jan Burian: <i>Na shledanou zítra</i>	1995
		Eberhardt Hauptbahnhof: <i>Nedokončený kalendář na tento rok a všechny roky příští</i>	1994
		Jiří Haussmann: <i>Haussmanovy méně známé texty</i>	1999
		Mikuláš Medek: <i>Texty</i>	1995
		Zdeněk Rotrekl: <i>Skryté tváře</i>	2005
	básně	Vladimír Holan: <i>Nokturnál</i>	2003
		František Hrubín: <i>Zpívám</i>	2002
		Josef Kainar: <i>Synkopy</i>	2003
		Bohuslav Reynek: <i>Ostny v závoji</i>	2002
		Jaroslav Seifert: <i>Slavík zpívá špatně</i>	2002



**Tabulka 3:** Přehled jazykových projevů zařazených do odborného stylu

	typ odborných textů	název textu	rok vydání
odborný styl	vědecko-naučná literatura	Lubomír Nátr: <i>Země jako skleník</i>	2006
		Pavel Příbyl – Aleš Janota – Juraj Spalek: <i>Analýza a řízení rizik v dopravě</i>	2008
		Jarmila Riegerová: <i>Rekondiční a sportovní masáže</i>	2007
		Petra Řepová: <i>Dítě a koktavost</i>	2007
		Jaroslav Zlámal – Jana Bellová: <i>Podniková ekonomie a management</i>	2007
	populárně naučná literatura	Jan Kolář: <i>Biologické hodiny rostlin</i>	2006
		<i>Dieta</i> , č. 10/2009	2009
		<i>Ekonom</i> , č. 39/2009	2009
		Benjamin Kuras: <i>Evropa snů a skutečností</i>	2007
		<i>Stavitel</i> , č. 3/2009	2009
	učebnice	Eva Francová: <i>Cestovní ruch</i>	2003
		Jan Ponec – Milič Jiráček: <i>Digitální fotografie</i>	2002
		Zdeněk Puchinger: <i>Daně; Daňová soustava ČR</i>	2003
		Eva Reiterová: <i>Základy statistiky pro studenty psychologie</i>	2003
		Jozef Rosina – Hana Kolářová – Jiří Stanek: <i>Biofyzika pro studenty zdravotnických oborů</i>	2006
	uspořádaná díla	Jan Císař: <i>Světoví dramatici</i>	1997
		Markéta Mikysková – Miriam Prokešová – Marcela Stehlíková: <i>Babiččina kuchařka</i>	2006
		Lydia Petráňová: <i>Domovní znamení staré Prahy</i>	2008
		Jarmila Teplíková: <i>Houby známé a exotické</i>	2004
		Václav Zelený: <i>Rostliny Středozeří</i>	2004

**Tabulka 4:** Přehled jazykových projevů zařazených do administrativního stylu

	<b>název textu</b>	<b>rok vydání</b>
<b>administrativní styl</b>	<i>České centrum čistší produkce – výroční zpráva 1997</i>	1998
	<i>Českoskalický zpravodaj, č. 4/1997</i>	1997
	<i>Efeméra – smlouvy</i>	1996
	<i>Fites – stanovy</i>	1995
	<i>Kolektivní smlouva ČD, s. o., na rok 1999</i>	1999
	<i>Lesní zákon – výňatek</i>	1995
	<i>Město Dačice informuje, ročník 1998</i>	1998
	<i>Návrhy zákonů a vyhlášek Ministerstva práce a sociálních věcí ČR</i>	2002
	<i>Návrhy zákonů a vyhlášek Ministerstva pro místní rozvoj ČR</i>	2002
	<i>Návrhy zákonů a vyhlášek Ministerstva vnitra ČR</i>	2002
	<i>Návrhy zákonů a vyhlášek Ministerstva zemědělství ČR</i>	2002
	<i>Návrhy zákonů a vyhlášek Ministerstva životního prostředí ČR</i>	2002
	<i>Smlouva o vytvoření a užití díla</i>	2003
	<i>Stanovy bytového družstva</i>	1998
	<i>Střední zdravotnická škola Nymburk</i>	1998
	<i>Školní administrativa</i>	1997
	<i>Texty z obecního úřadu Lázně Toušeň I</i>	1995
	<i>Vyhlášky MěÚ Dačice</i>	1995
	<i>Výroční zpráva svazu PRO-BIO za rok 2002</i>	2003
	<i>Zpravodaj SETUZA, č. 1/1997</i>	1997

**Tabulka 5:** Přehled jazykových projevů zařazených do řečnického stylu

	<b>typ projevu</b>	<b>název projevu</b>	<b>rok proslovení</b>
<b>řečnický styl</b>	projev ke dni vzniku samostatného československého státu – Václav Havel	Projev prezidenta republiky Václava Havla ke státnímu svátku České republiky	1999
		Projev prezidenta republiky Václava Havla ke státnímu svátku České republiky	2000
		Projev prezidenta republiky Václava Havla ke státnímu svátku České republiky	2001
		Projev prezidenta republiky Václava Havla ke státnímu svátku České republiky	2002
	novoroční projev – Václav Havel	Novoroční projev prezidenta republiky Václava Havla	2000
		Novoroční projev prezidenta republiky Václava Havla	2001
		Novoroční projev prezidenta republiky Václava Havla	2002
		Novoroční projev prezidenta republiky Václava Havla	2003
	projev ke dni vzniku samostatného československého státu – Václav Klaus	<i>Překonejme minulost přítomností</i> – Projev prezidenta republiky u příležitosti státního svátku 28. 10. 2004	2004
		Projev prezidenta republiky u příležitosti státního svátku dne 28. října 2007	2007
		Projev prezidenta republiky k 28. říjnu 2010 o smyslu našeho státu	2010
		Projev prezidenta republiky k 28. říjnu 2011	2011
	novoroční projev – Václav Klaus	Novoroční projev prezidenta republiky	2006
		Novoroční projev prezidenta republiky	2008
		Novoroční projev prezidenta republiky	2012
		Novoroční projev prezidenta republiky 2013	2013
	projev ke dni vzniku samostatného československého státu – Miloš Zeman	Projev prezidenta republiky při slavnostním ceremoniálu udílení státních vyznamenání	2013
		Projev prezidenta republiky během slavnostního ceremoniálu udělení státního vyznamenání České republiky ve Vladislavském sále	2014
	vánoční poselství – Miloš Zeman	Vánoční poselství prezidenta republiky Miloše Zemana	2013
		Vánoční poselství prezidenta republiky	2014

**Tabulka 6:** Přehled jazykových projevů zařazených do hovorového stylu

	lokality, z níž pocházejí mluvčí	počet mluvčích	věk mluvčích	téma rozhovoru	rok uskutečnění rozhovoru
<b>hovorový styl</b>	Vlkoš, Kyjov	2	51, 25	rodina, domácnost	2010
	Přerov	2	48, 44	počasí, peněžní výdaje, domácí práce	2010
	Ostrava, Dvůr Králové	2	44, 40	škola, aktuální společenské dění, domácnost	2010
	Ostrava, Nová Paka	3	41, 20, 26	nakupování, vaření, pečení, práce	2010
	Brno, Frýdek-Místek	3	22, 24, 20	volný čas, Expo Shanghai, Olympijské hry, škola	2010
	Chodovice, Vrchlabí	2	75, 43	diskuze nad časopisem, studánky	2010
	Krnov	2	28, 17	volejbal, oslava narozenin	2010
	Olomouc	2	36, 23	diskuze při snídani, nemoci, zásnubní prsten	2010
	Vrbno pod Pradědem	3	30, 52, 19	ceny a kvalita zboží, film, facebook, kamarádi	2010
	Nové Město na Moravě	2	47, 20	fotbal, sport, škola	2010
	Zlín	2	22, 19	škola, cestování, oblečení	2010
	Hronov, Jablonec nad Nisou	2	54, 55	práce, počasí, rodina, kouření, zahrada	2010
	Moravské Budějovice, Znojmo	3	19, 57, 86	příbuzní, výchova dětí, vzpomínky z dětství, historka z talk show	2010
	Karviná	2	19, 20	doprava, studium, spolužáci, Olomouc	2010
	Vsetín, Uherské Hradiště	2	46, 22	dovolená, cestování, práce v lese	2010
	Rožnov pod Radhoštěm	2	21, 21	bakalářská práce, pohádky, seriály, partner	2010
	Pardubice, Hradec Králové	2	21, 23	výlet, oslava narozenin, Silvestr, vztahy na facebooku, TV pořad	2010
	Litomyšl	2	20, 21	místní zábava, brigáda, praxe, festivaly	2010
	Havířov	2	20, 22	násilí, pojištění, peníze	2010
	Valašská Bystřice, Hutisko-Solanec	2	21, 21	jména, fotografie, výlet, přátelé, ubytování	2010

## 5. Literatura

ČECH, Radek. Jen popis s čísly? Perspektivy korpusové lingvistiky. *Naše řeč*, 97, 2014, s. 171–184.

ČECH, Radek – DAVID, Jaroslav – DAVIDOVÁ GLOGAROVÁ, Jana. Analýza tematické koncentrace textu: komparace publicistiky Ladislava Jehličky a Karla Čapka. *Slovo a slovesnost*, 74, 2013, s. 41–54.

ČECH, Radek – POPESCU, Ioan Iovitz – ALTMANN, Gabriel. *Metody kvantitativní analýzy (nejen) básnických textů*. Olomouc, 2014.

ČECHOVÁ, Marie – KRČMOVÁ, Marie – MINÁŘOVÁ, Eva. *Současná stylistika*. Praha, 2008.

Český národní korpus – korpusy psané i mluvené češtiny. Dostupné z: <<http://www.korpus.cz>>.

HENDL, Jan. *Přehled statistických metod zpracování dat*. Praha, 2004.

HOFFMANOVÁ, Jana. *Stylistika a... Současná situace stylistiky*. Praha: Trizonia 1997.

KARLÍK, Petr – NEKULA, Marek – PLESKALOVÁ, Jana (eds.). *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, 2002.

KARLÍK, Petr – NEKULA, Marek – RUSÍNOVÁ, Zdenka (eds.). *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny, 1995.

KUBÁT, Miroslav. *Rozsah, rozptýlení a koncentrace lexika v psané a mluvené češtině* [bakalářská diplomová práce]. Olomouc, 2010.

KUBÁT, Miroslav. *Funkční styly z hlediska lexikální statistiky* [magisterská diplomová práce]. Olomouc, 2012.

KUBÁT, Miroslav. *Kvantitativní analýza žánrů* [ disertační práce]. Olomouc, 2015.

KUBÁT, Miroslav – MILIČKA, Jiří. Vocabulary Richness Measure in Genres. *Journal of Quantitative Linguistics*, 10, 2013, s. 339–349. Dostupné z: <[https://www.researchgate.net/publication/258518594\\_Vocabulary\\_Richness\\_Measure\\_in\\_Genres](https://www.researchgate.net/publication/258518594_Vocabulary_Richness_Measure_in_Genres)>.

MATLACH, Vladimír. *Kvantitativně lingvistický software* [magisterská práce]. Olomouc, 2014.

MISTRÍK, Jozef. *Frekvencia slov v slovenčine*. Bratislava, 1969.

POŘÍZKA, Petr a kol. *Olomoucký korpus mluvené češtiny*. Interní materiály KB FF UPOL, archiv autora.

TĚŠITELOVÁ, Marie. *Otázky lexikální statistiky*. Praha, 1974.

TĚŠITELOVÁ, Marie. *Kvantitativní charakteristiky současné české publicistiky*. Praha, 1982.

TĚŠITELOVÁ, Marie a kol. *Psaná a mluvená odborná čeština z kvantitativního hlediska (v rámci věcného stylu)*. Praha, 1983a.

TĚŠITELOVÁ, Marie. *Kvantitativní charakteristiky současné odborné češtiny (v rámci věcného stylu)*. Praha, 1983b.

TĚŠITELOVÁ, Marie. *Současná česká administrativa z hlediska kvantitativního*. Praha, 1985.

TĚŠITELOVÁ, Marie. *Kvantitativní lingvistika*. Praha, 1987.