

UNIVERZITA PALACKÉHO V OLOMOUCI

Filozofická fakulta

Katedra romanistiky

Lenka Bičová

Frekvenčnost mluvené francouzštiny

**Kvantitativní analýza vybraných částí scénářů francouzských
filmových adaptací ve vztahu k jejich literární předloze**

Magisterská diplomová práce

Vedoucí práce: Doc. PhDr. Jan Holeš, Ph.D.

Olomouc 2014

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně na základě
uvedených pramenů a literatury.

V Olomouci dne 22. dubna 2014

Bc. Lenka Bičová

Poděkování

Mé velké poděkování patří vedoucímu práce Doc. Janu Holešovi za příjemnou spolupráci, ochotu a cenné rady při tvorbě mé Magisterské diplomové práce.

Taktéž děkuji českým filmovým distribučním společnostem AČFK a Aerofilms a také distribučnímu oddělení francouzského studia Pathé, že tak ochotně poskytly filmové scénáře, které byly zapotřebí k provedení textových analýz.

Obsah

Úvod	6
TEORETICKÁ ČÁST	11
Vymezení úkolu teoretické části práce	12
1. Komunikace zájmem lingvistů	12
2. Písmo a řeč jako dva odlišné jazykové kódy (Od řeči k písmu)	13
2.1 Historický pohled	14
2.1.1 Kořeny mluvené a psané francouzštiny	15
2.2 Psaný versus mluvený projev	19
2.2.1 Francouzština slovem a písmem	21
2.3 Boj za ryzí francouzštinu	22
2.4 Pravopisná hrdost versus modernizující prvky současné francouzštiny	25
2.4.1 Jazyk SMS jako odraz mluvené podoby francouzštiny	28
3. Analýza mluveného projevu	31
3.1 Organizace ústního diskurzu (způsoby tvoření)	32
3.2 Slovní zásoba	33
3.3 Slovesný způsob a čas	34
3.4 Syntax	35
3.5 Gramatické a fonetické rysy hovorové francouzštiny	37
3.5.1 Raymond Queneau: příklad hovorové francouzštiny v literatuře	38
3.6 Problematika transkripce textu	40
4. Kvantitativní lingvistika	41
4.1 Počátky a struktura kvantitativní lingvistiky	41
4.2 Lexikální statistika	43
4.3 Gramatická statistika	44
4.4 Ostatní oblasti	44
4.5 Využití kvantitativní lingvistiky v praxi	47
4.5.1 Grafematická statistika a její souvislost s QWERTY	49
4.5.2 Lexikometrická analýza politických projevů	50

4.5.3	Vliv slovní frekvenčnosti na percepci	52
4.5.4	Zkoumání bohatosti slovní zásoby: «la richesse lexicale»	54
4.6	Kritéria pro výběr materiálu k analýze	55
4.6.1	Kvalitativní hlediska	55
4.6.2	Kvantitativní hlediska	56
4.7	Vyhodnocování výsledků analýzy	57
4.7.1	Frekvence, pořadí, rank	58
4.7.2	Zipfovy zákony	60
4.7.3	Průměrná redukovaná frekvence	63
4.7.4	Koeficient disperze	64
4.8	Frekvenční slovníky	64
4.8.1	Základní přehled týkající se lexikální statistiky	65
5.	Kvantitativní lingvistika a mluvený projev ve francouzském jazyce	69
5.1	Korpusy mluveného jazyka	69
5.1.1	Korpusy mluvené francouzštiny	72
5.2	Metoda zpracování jazykového materiálu	74
5.3	Français fondamental a metoda SGAV	77
5.4	Filmové titulky jako korpus pro studium frekvenčnosti	78
	FREKVENČNÍ ANALÝZA	81
	Vymezení úkolu analytické části práce	82
1.	Analyzovaný text	82
1.1	Výběr textu a jeho charakteristika	82
1.2	Rozsah korpusu	85
1.3	Rozdílnost zdrojových médií a adaptační proces	86
2.	Metodologie	88
3.	Statistická analýza vybraných částí literárních děl	94
3.1	Lexikální statistika	94
3.1.1	Frekvenční seznamy: Le Magasin des suicides (scénář a román) a L'Écume des jours (scénář a román)	94
3.1.2	Lexikální komparace scénářů s románovou předlohou	101

3.1.3	Nejfrekventovanější slova	104
3.2	Gramatická statistika	108
3.2.1	Zastoupení slovních druhů v textu	108
3.2.2	Kvantifikace kategorie rodu a čísla u jména	112
3.2.3	Kvantifikace slovesných tvarů	114
	Závěr	118
	Résumé	120
	Bibliografie	122
	Příloha	129

Úvod

Nad souslovím „matematická lingvistika“ leckterý milovník francouzského jazyka pravděpodobně jen nadzvedne obočí, v krajním případě opouští s nelibými pocity rozečtený odstavec. Je pravda, že matematika často nepatří mezi oblíbené disciplíny humanitně orientovaných filologů, mezi které se řadíme i my, přesto lze v této pomezni disciplíně využívající matematických metod na poli lingvistiky nalézt zalíbení. Naše již druhá diplomová práce, kterou čtenáři nyní předkládáme, zabývající se tématem kvantitativní lingvistiky a lexikální a gramatické frekvenčnosti je toho důkazem.

V první práci *Lexikální analýza vybraných částí scénáře filmu Amélie z Montmartru (Úvod do kvantitativní lingvistiky)* z roku 2011 jsme se zabývali výkladem o vzniku pomezni disciplín, vymezením matematické lingvistiky a následnou poměrně detailní charakteristikou kvantitativní lingvistiky, její základní terminologie a podoblastí s důrazem na lexikální a gramatickou statistiku. Nechybí ani zajímavosti týkající se využití statistické analýzy v lingvistické praxi či podkapitola představující nejdůležitější díla lexikální statistiky, a to s důrazem na české a francouzské práce. V praktické části práce jsme na základě vybraných pasáží filmového scénáře k filmu *Le Fabuleux Destin d'Amélie Poulain* (2001; *Amélie z Montmartru*) zhotovili vlastní korpus, který následně podrobujeme kvantitativní analýze za účelem vytvoření frekvenčního seznamu použitých slov. Dílčím úkolem bylo také podívat se na analyzovaný text skrze perspektivu gramatické statistiky a definovat četnost jednotlivých gramatických jevů v textu. Výsledky našeho zkoumání jsme vizualizovali pomocí tabulek a grafů a provedli taktéž slovní vyhodnocení jevů vyplývajících z rozboru. Výstupy z naší analýzy jsme měli možnost porovnat s již existujícím frekvenčním seznamem. Toto srovnání překvapivě odhalilo výraznou shodu u nejfrekventovanějších hesel řazených v obou seznamech na první místa. Ačkoliv jsou korpusy těchto seznamů početně nesrovnatelné, jsou z nich patrné totožné tendence ve frekvenčnosti francouzského lexika. Hlavní ambicí této lingvistické práce bylo přehledně vyložit problematiku dané disciplíny, prokázat platnost teoretických poznatků v praxi a v neposlední řadě také inspirovat čtenáře k dalšímu bádání v oblasti matematické lingvistiky

a probudit zájem o matematický pohled na jazyk, který dokáže přinést zajímavé a překvapivé informace.

Filmový scénář je navíc díky své dialogičnosti velmi zajímavým materiálem ke statistické analýze lexika, protože může do jisté míry suplovat autentický mluvený projev, který je ve své spontánní podobě stále jedním z nejobtížněji zkoumatelných jazykových projevů. To je dáno časovou náročností při sběru dat za účelem vytvoření korpusu či nutností přetvářet zaznamenaný mluvený projev do psané podoby (ovšem bez zásahů do jeho specifické podoby). Filmový scénář či filmové titulky tak představují text, který si zachovává charakter spontánního orálního diskurzu, a tak poskytuje lingvistovi značné usnadnění v počáteční fázi sběru dat.

V následující diplomové práci, volně navazující na práci předchozí, se tedy opětovně věnujeme frekvenčnosti francouzského lexika a gramatických jevů. Studii s názvem *Frekvenčnost mluvené francouzštiny – Kvantitativní analýza vybraných částí scénářů francouzských filmových adaptací ve vztahu k jejich literární předloze* jsme rozdělili do dvou hlavních částí – teoretické a analytické.

V první teoretické části se tentokrát více zabýváme mluveným projevem jako takovým. Poté, co se v první kapitole dotkneme důležitosti komunikace pro člověka a pro psycholingvistická studia, se ve druhé kapitole věnujeme odkrývání historického pozadí vzniku francouzštiny a hodnocení postavení mluvené formy jazyka vůči formě psané. Pozastavíme se i nad jazykem spojeným s novými technologiemi, především nad jazykem SMS. Ve třetí kapitole s názvem *Analýza mluveného projevu* popisujeme lexikální, syntaktická a gramatická specifika spontánních projevů. V další fázi teoretické části připomeneme některé základní pojmy spojené s kvantitativní textovou analýzou tak, aby bylo možné se opět bez problému zorientovat v praktické části naší práce. V kapitolách o kvantitativní lingvistice čerpáme z velké části z naší bakalářské diplomové práce, kde byly již dříve termíny podrobně vysvětleny a popsány. Jednotlivé podkapitoly však obohacujeme o další poznatky. Nově například pojednáváme o projevech frekvenčnosti v praxi, kde se zmíníme

o efektu slovní frekvenčnosti na vizuální percepci či o grafematické statistice a její souvislosti s klávesnicí QWERTY. V poslední kapitole teoretické části, zasvěcené kvantitativní lingvistice ve spojení s mluveným projevem, vytváříme přehled nejvýznamnějších korpusů mluveného jazyka, a to jak zahraničních, tak francouzských a českých. Přibližujeme metodu zpracování jazykového materiálu určeného pro tvorbu korpusu a vracíme se též do poloviny 20. století, kde představujeme významnost kvantitativního výzkumu mluveného jazyka pro didaktiku výuky francouzštiny. Teoretickou část uzavíráme představením článku o filmových titulcích coby korpusu pro studium frekvenčnosti.

Ve druhé analytické části pracujeme s mluveným projevem na bázi kvantitativní analýzy diskurzu. Výchozím textovým zdrojem, sloužícím následně k vytvoření korpusu, je pro nás i tentokrát francouzský filmový scénář. Naše bádání ve sféře lexikální a gramatické frekvenčnosti jsme tentokrát pojali více komparativně. Záměrně volíme dva scénáře nedávno vzniklých filmových adaptací francouzského románu, jejichž slovník srovnáváme s jejich literární předlohou. Tato metoda má za cíl zkoumat, jak se mění kvantitativní charakter použitého slovníku v závislosti na formě narativního sdělení (tedy psaný román vs. „mluvený“ film), pracuje-li se v obou ztvárněních syžetu se stejnou fabulí. Prostřednictvím takového rozboru se máme možnost dozvědět více o lingvistické stránce adaptačního procesu, popřípadě se zamyslet nad tím, je-li filmová adaptace pouze vykrádáním literárního díla, za které byla dlouho považována.

V analytické části práce zpracováváme vybrané pasáže z těchto děl:

- **Scénář filmové adaptace v původním francouzském znění**

1. *L'Écume des jours (Pěna dní)*, režie: Michel Gondry, scénář: Luc Bossi, Francie/Belgie, 2013, 125 min., distribuce v ČR: Aerofilms.

2. *Le Magasin des suicides (Obchod pro sebevrahy)*, režie a scénář: Patrice Leconte, Francie/Kanada/Belgie, 2012, 85 min., distribuce v ČR: AČFK.

- **Románové dílo v původním francouzském znění**

1. Vian, Boris: *L'Écume des jours*, Paris : Jean-Jacques Pauvert, 1998. 315s.
2. Teulé, Jean: *Le Magasin des Suicides*, Paris : Julliard, 2008. 157s.

Zatímco romány v originálním francouzském znění nebylo v knihovnách či na internetových obchodech typu Amazon.fr tolik obtížné sehnat, filmové scénáře jsme získali jediné díky ochotě českých distribučních společností AČFK a Aerofilms, které dodaly francouzské originály, a poděkování patří také distribučnímu oddělení studia Pathé ve Francii.

Kvantitativní lingvistika je oborem dosti specifickým. V českém prostředí se touto problematikou velmi aktivně v minulosti zabývala především Marie Těšitelová, jejíž publikace nám byly, obzvláště při sestavování teoretické části, velkou oporou. K našim účelům byla nejvhodnější její práce s názvem *Kvantitativní lingvistika* (1977). Autorka se zde zabývá jak jednotlivými oblastmi oboru, tak kvantitativní lingvistikou obecně, pozastavuje se u nejdůležitějších pojmů a zákonitostí při matematické práci s textem a veškerá tvrzení dokládá příklady ze své mnohaleté praxe. *Využití statistických metod v gramatice* (1980) a *Otázky lexikální statistiky* (1974) jsou poté pracemi, ve kterých se již cíleně zaměřuje na speciální odvětví kvantitativní lingvistiky. V metodologii nám byl velkým rádcem spis kolektivu autorů Jelínek, Bečka, Těšitelová (*Frekvence slov, slovních druhů a tvarů v českém jazyce*, 1961), v jehož úvodní části je možné dozvědět se mnoho o výzkumných metodách, postupech a zásadách při tvorbě tohoto slovníku. Velmi poučný a zajímavý zdroj informací představují *Dějiny lingvistiky* (1996) Jiřího Černého, ve kterých najdeme samostatnou kapitolu o kvantitativní lingvistice a teorii informace. Korpusovou lingvistikou se v současnosti zabývá však především Ústav Českého národního korpusu FF UK, který má na svém kontě mnoho publikací. Jeho činnost se soustředí na vytváření synchronních i diachronních korpusů psaných a mluvených. Velkou inspirací nám byla publikace Jitky Šonkové *Morfologie mluvené češtiny – Frekvenční analýza* (2008). Autorka zde pojednává o historii korpusové lingvistiky a také se soustředí na praktický výklad o možnostech a postupech při tvorbě korpusu mluveného jazyka. V kapitolách o historii francouzštiny jsme se opírali o knihu Michèle Perret *Introduction à l'histoire de la langue française*

(2008). Při kompletování kapitol zabývajících se mluvenou formou jazyka jsme ocenili práce Claire Blanche-Benveniste, která velmi přehledným způsobem vysvětluje mechanismus mluveného jazyka. Co se týká gramatické analýzy, jedním z hlavních zdrojů se stala metodická gramatika francouzštiny od Martina Riegela a kolektivu. *Grammaire méthodique du français* (2009) je ucelenou prací, která na gramatiku nahlíží jak z teoretického, tak z praktického hlediska. Její stránky jsou bohaté na konkrétní příklady doplňujících výklad, a proto je tato gramatika velmi dobrým rádcem při studiu francouzského jazyka. V práci jsme využili také poměrně velké množství internetových článků a studií týkajících se problematiky kvantitativní lingvistiky, mluvené formy jazyka a slovní frekvenčnosti.

Kvantitativní analýza přinesla v minulosti mnoho nových poznatků o jazycích. Výsledkem jsou vědecké studie a celá řada frekvenčních slovníků v mnoha jazycích. Přestože se naše práce rozsahem svého korpusu stále nemůže s výše zmíněnými slovníky měřit, doufáme, že její četba bude obohacujícím a zajímavým zdrojem informací o mluvené podobě francouzštiny a jejím lexiku a že se případně stane také podnětem pro další studium matematické lingvistiky, slovní frekvenčnosti či alespoň motivuje ke zhlédnutí a přečtení uměleckých děl použitých k analýze.

TEORETICKÁ ČÁST

Vymezení úkolu teoretické části práce

V první části naší práce se budeme soustředit na charakteristiku písma a řeči jako dvou odlišných jazykových kódů, přehledně definujeme gramatické znaky mluveného projevu a rekapitulujeme pojmy svázané s kvantitativní lingvistikou, a to jak obecně, tak konkrétně ve spojitosti s mluveným jazykem. Cílem teoretické části práce je představit čtenáři důležitost mluveného projevu, jeho přirozenost a jistou dominanci nad jazykem psaným vyplývající již z jazykové historie či z úvahy, že od nepaměti lidé většinou neměli problém s komunikací ústní formou, ale ne všichni uměli číst a psát.

1. Komunikace zájmem lingvistů

Komunikace je již od počátku lidstva jednou ze základních potřeb člověka. Slouží ke každodennímu mezilidskému kontaktu a má mnoho podob. Kontaktovat lidi kolem nás je nám zkrátka vlastní a usilujeme o to všemi možnými prostředky – písemně, ústně, symboly, znaky, znakovou řečí, řečí těla, přímo či nepřímo. Hlavní roli zde samozřejmě hrají jazyky a především náš jazyk mateřský.

Také současná lingvistika se v rámci studia jazykového systému stále více zaměřuje na zkoumání komunikativního procesu. Lingvisté zkoumají regionální dialekty, slangy, argot a zaznamenávají i obohacování jazyka přicházející z mluvy mládeže – například nové expresivní výrazy, sémantické posuny, novotvary, slovní zkracování apod. Vznikají také synchronní a diachronní korpusy, jak mluveného, tak psaného projevu, mapující stav jazyka (Šonková 2008, s. 13).

V souvislosti s pragmatickým obratem¹ se navíc zformovala také paralingvistika zabývající se jevy mimo jazykový systém, které ale přesto jsou

¹ V 70. letech 20. století vzniklo v souvislosti s tzv. *pragmatickým obratem* několik dalších lingvistických disciplín. Jazykovědci se doposud zabývali pouze praktickými otázkami jazykového systému jako celku (*langue*) a o promluvy (*paroles*), realizované v určitý čas na určitém místě, nejevili příliš zájem.

nedílnou součástí komunikačního procesu, neboť ho doprovázejí a doplňují. Mezi paralingvistické prostředky patří např. neverbální signály, jako jsou mimika či gesta často souhrnně nazývány jako *řeč těla*. Reakce na neverbální signály jsou někdy dokonce silnější než na slova. Argyle, Alkema a Gilmour² ve své studii prokazují na to, že pokud se setkáme s neverbálním sdělením, které je v nesouladu se sdělením verbálním, je pravděpodobnost, že uvěříme neverbálnímu signálu, pětkrát větší! Neverbální signály jsou nevědomé, a snad proto je lidé často považují za závažnější indikátory „skutečných“ postojů mluvčího než to, co právě říká. Zdá se tedy, že neverbální aspekty komunikace mají poměrně velký význam. „Jednou z nejnápadnějších forem nonverbální komunikace používané v rozhovoru nebo v diskurzu je metakomunikace. Metakomunikace je obecný název pro různé dimenze, jimiž lze obohatit řeč.“ (Hayesová 2011, s. 29). Sem patří *intonace*, která má moc změnit význam sdělení (např. při použití sarkastického tónu), dále *řečový registr* (deklamační, formální, informativní, familiární, intimní), odmlky nebo zvuky jako „hm“. Výhodou řeči těla může být jistá univerzálnost vůči jazykovým dorozumívacím prostředkům. Na druhé straně je potřeba si uvědomit, že takový způsob komunikace je do značné míry ovlivněn kulturou, ze které mluvčí pochází.

2. Písmo a řeč jako dva odlišné jazykové kódy (Od řeči k písmu)

Kromě jiných specifických oblastí výzkumu se v každém jazyce nabízí zkoumání dvou paralelně existujících a vzájemně se prolínajících jazykových kódů – tedy spisovné psané podoby jazyka a jeho běžné mluvené podoby. Tyto dva jazykové kódy mají ve francouzštině své historické kořeny, o kterých pojednáme v jedné z následujících podkapitol.

² Argyle, Michael, Alkema, Florisse, Gilmour, Robin. 1971. „The communication of friendly and hostile attitudes by verbal and non-verbal signals“. *European Journal of Social Psychology*, 1971, č.1, s. 395–402.

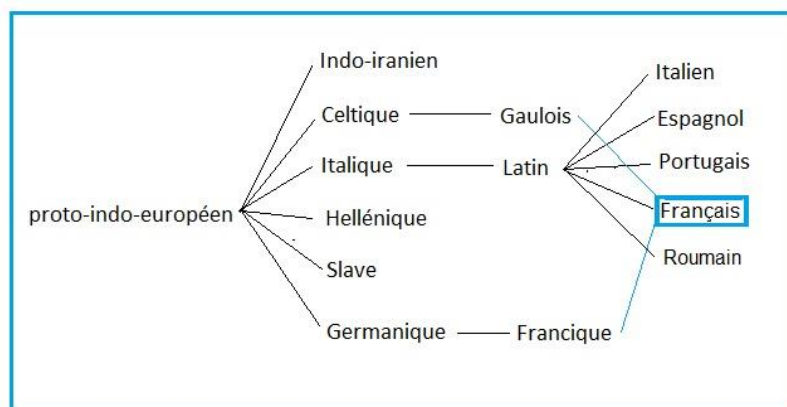
2.1 Historický pohled

Bylo by poměrně odvážné snažit se určit, kdy se zrodila mluvená řeč, nicméně počátky písemnictví, které se objevilo ve druhé fázi po mluveném projevu, lze stanovit poměrně přesně. Jeho nástup byl pozvolný. Kolem roku 3300 před naším letopočtem se objevilo v Mezopotámii sumerské písmo, poté kolem roku 3000 př. n. l. bylo písemnictví doloženo v Egyptě a následně v 17. století př. n. l. v Číně a ve Střední Americe. Dnes převládá názor, že různá písemnictví se sice zrodila nezávisle na sobě, bylo k tomu však zapotřebí stejných podmínek – společnosti disponující průmyslem, aktivním obchodem a organizovaným státem. Vznik písemnictví jde tedy ruku v ruce s vyvíjející se společností a naopak písemnictví se podílí na rozvoji společnosti, která jím vládne. Dochází k prohlubování kognitivních procesů, ke stabilizaci ekonomiky, práva, vědy ale i náboženství. Nyní je možné uchovávat spisy, smlouvy a důležité listiny (Riegel a kol. 2009, s. 51).

Z historického hlediska býval psaný jazyk obecně považován za pravou a skutečnou podobu jazyka, zatímco jeho mluvená varianta byla označována za nestabilní a degenerovanou formu nehodnou vědeckého zkoumání. „Na druhé straně se už v 19. století začala tato situace měnit a lingvisté začali studovat mluvený jazyk jako takový. Rozvoj fonetiky ve Velké Británii studium mluveného jazyka dále podpořil a mluvený jazyk byl už počátkem dvacátého století lingvisty považován za primární a psaný jazyk za sekundární, jenž byl odvozen od mluveného.“ (Šonková 2008, s. 15). Přesto v rámci strukturální lingvistiky nebyl výzkum mluveného jazyka vždy důsledný a mluvená podoba jazyka byla nadále považována za nesystematickou a nereprezentující složitou strukturu jazyka.

2.1.1 Kořeny mluvené a psané francouzštiny

Podívejme se nyní na historické kořeny těchto dvou základních způsobů realizace v případě francouzského jazyka. Jak vlastně vznikl francouzský jazyk?



Zdroj: Perret 2008, s. 20.

Prehistorie francouzštiny je poměrně neznámá a svědectví o jazyce sahají pouze do šestého století před naším letopočtem. Kmeny tohoto období, jako jsou Ligurové³ a Iberové⁴, neměli však na vývoj francouzštiny velký vliv. Prostřednictvím provensálštiny se do francouzštiny dostalo jen několik slov od Řeků usazených v koloniích na březích Středozemního moře.

Poté, v pátém století př. n. l., přišli na území budoucí Francie Galové hovořící jazykem keltského původu, který měl v té době mnoho dialektů. Galové však působili jednotně a to i v oblasti zvyků a náboženství, jehož kněžími byli druidové. Henriette Walterová ve své knize *Francouzština známá i neznámá* připomíná známý názor, že „galský jazyk je praotcem francouzštiny“. Podle ní je však potřeba věnovat se tomuto tvrzení důkladněji (Walterová 1993, s. 12).

Walterová souhlasí, že Galie po obsazení Římany přijala jazyk svých dobyvatelů⁵. Co se týká galštiny, ta je nám jako jazyk nedostatečně známá,

³ Okupující jihovýchod Francie, Švýcarsko a hornatý italský sever.

⁴ Kteří přišli ze Španělska až k Loře.

⁵ Přičemž je potřeba si uvědomit, že galsky se mluvilo ještě dlouho po dobytí Římany. Latina pocházející od Římanů se stala oficiálním jazykem na galsko-římském území, zatímco galština přetrvávala jako mateřský jazyk kolonizovaného obyvatelstva. Galsko-římská civilizace potrvá až do 5. století (Perret 2008, s. 26).

o čemž svědčí pouze několik desítek slov galského původu, která se ve francouzštině zachovala (např. místní jména, jména božstev či osob). Z galštiny nám zbylo pouze několik nesrozumitelných nápisů, jejichž datace lze stanovit s přesností plus minus dvě stě let (Perret 2008, s. 26). Jak poznáme, že slovo pochází z galštiny? Michèle Perret uvádí ve své knize domněnku, že galské slovo je takové, které v sobě splňuje tři podmínky:

- Existence toho termínu v jazyce je doložena velmi brzy (možná už na úrovni latiny)
- Tento termín neexistuje v klasické latině
- Méně nebo více příbuzné formy tohoto slova existují v jiných keltských jazycích (*irlandais, cymrique, gaélique* – irský, kymerský, galský)

Současná podoba francouzského jazyka svědčí o pravdivosti tradiční klasifikace francouzštiny coby románského jazyka pocházejícího z latiny. Dobyť Galie Římany, přinášejícími na území latinu, je skutečně nejstěžejnějším momentem v historii francouzského jazyka. Walterová ale zdůrazňuje, že „útvarem, který posloužil jako zdroj vzniku románských jazyků, nebyla takzvaná *klasická latina*, jazyk psaný, jazyk děl velkých autorů prvního století př. Kr. (Cicerona, Caesara, Tita Livia, Vergilia), ale latina pojmenovaná jako *vulgární*, tedy ta, kterou Římané mluvili v každodenním životě. Proto je třeba francouzštinu charakterizovat jako řeč vzešlou z vulgární latiny importované do Galie římskými dobyvateli.“ (Walterová 1993, s. 12). Latina, kterou se hovořilo v Galii, byla tedy nespisovná podoba každodenní mluvy obsahující zjednodušené, obrazné, hovorové až argotické formy.

Podoba této vulgární latiny, kterou se Galové učili prostřednictvím komunikace s římskými legionáři a obchodníky, byla poznamenána také řečí germánského kmene sídlícího v oblasti dnešní Paříže. Roku 486 vtrhli do galsko-římského království Frankové sídlící dosud na území dnešní Belgie. Pod vedením Chlodvíka I. porazili Vizigóty a v roce 534 připojili Burgundské království k území Franků. Roku 496 konvertoval Chlodvík ke křesťanství⁶, což

⁶ Křesťanství je již od 3. století oficiálním náboženstvím Říše římské.

mohlo částečně přispět k tomu, že po několika stoletích bilingvismu Frankové téměř opustili⁷ svůj vlastní jazyk (*le francique* – franštinu) a přijali latinu – oficiální jazyk dominující na kolonizovaném území, který akceptovali také jako jazyk náboženství. Frankové nicméně ovlivnili vývoj latiny na galském území tak, že jí vtiskli svůj germánský „akcent“, a tak přispěli k utvoření základu mluveného projevu v oblasti Ile-de-France, nářečí, které posloužilo jako zdroj současné francouzštiny (Walterová 1993, s. 12, 13). Ve francouzském slovníku nalezneme ještě nyní více než 400 slov franckého původu jako například slova spojená s venkovským životem (*blé, jardin, bois, troupeau*) či výbrojí (*hache*) (Perret 2008, s. 29, 30).

Kdy se vlastně začalo mluvit namísto latiny francouzsky? V čem se starofrancouzština lišila od latiny?

Za zrození francouzštiny lze považovat rok 813, datum biskupského koncilu v Tours, který projednával fakt, že lid již neumí latinsky, tudíž je zapotřebí se k němu nadále obracet pouze v jeho jazyce. Nemyslí se tím však v jeho mateřském jazyce, nýbrž v jazyce germánského původu (tzv. jazyce *tudesque*) nebo v *lingua romana rustica*, které byly v tu dobu jedinými, jimž věřící rozuměli. Kněží začali tímto hovorovým románským jazykem psát, a tak upevňovali jeho postavení. Jednoho dne pak tento jazyk nahradil latinu. Tento zlom lze považovat za počátek románských jazyků a konec vulgární latiny, respektive latiny vůbec. Od tohoto období se mluví o latině jako o mrtvém jazyku, přestože v tehdejší době ji kněží nadále používali pro všechny listiny, které považovali za „hodnotné“ – jako byly historické, teologické, filozofické a didaktické spisy. Latina byla tedy jazykem kultury, administrativy a vzdělávání a byla považována za oficiální jazyk (*langue officielle*), který byl však vyloučen z mluvené formy komunikace (Perret 2008, s. 36).

Dalším klíčovým momentem je rok 842 a *Štrasburské přísahy* (*Serments de Strasbourg*), diplomatický dokument, ve kterém se dva z vnuků Karla

⁷I později se totiž udržel bilingvní charakter komunikace, především u politicky dominantní menšiny germánského původu. Potomci Chlodvíka (Clovis) – Merovejci – mluvili germánským mateřským jazykem stejně jako Karolinci (Carolingiens), potomci Karla Velikého (Charlemagne).

Velikého – Ludvík Němec a Karel Holý⁸ – zavazují vzájemně si pomáhat v boji proti vnukovi třetímu – Lotharovi. Z lingvistického pohledu spočívá hodnota Štrasburských přísah hlavně v jazycích, kterými byly sepsány. Od dob Karla Velikého bylo území dnešní Francie pouhou součástí jeho germánské říše, což se nezměnilo ani během vlády jeho syna Ludvíka I. Pobožného⁹. Mezi jeho třemi syny (dědici) však vypukly války, protože si každý z nich přál mít stejně bohaté království. Ve snaze ukončit spory vymysleli jejich vyjednavací, že se Francká říše rozdělí na 3 části. Západ, budoucí Francie, byla připsána Karlu Holému, východ, na kterém později vzniklo Německo, patřil Ludvíkovi (později řečenému Němec) a střední území přisouzené Lotharovi dostalo jméno Lotharingie. Toto rozdělení území na Západofranskou, Východofranskou a Středofranskou říši bylo schváleno *Verdunskou smlouvou* roku 843. O rok dříve se Ludvík a Karel spojili proti Lotharovi a každý sepsal již zmíněnou přísahu v jazyce toho druhého, aby zaručil vzájemné spojení. Ludvík tedy sepsal přísahu v románském jazyce, Karel v germánském *tudesque*. Jejich vojska se následně také podrobila slavnostní přísaze (Perret 2008, s. 36, 37). Tento text je považován za první psaný pramen, jenž byl vyhotoven ve starofrancouzském a staroněmeckém jazyce.

V této podkapitole jsme stručně shrnuli počátek zrodu francouzského jazyka, a to jak jeho mluvené, tak psané podoby. Následovalo štěpení na nářečí v souvislosti s geografickou oblastí užívání jazyka a také mnoho vývojových změn. Pro podrobnější seznámení s historií francouzštiny doporučujeme publikaci od Michèle Perret s názvem *Introduction à l'histoire de la langue française* (2008). Naším cílem při seznámení se s historickými kořeny bylo ověřit si významný vliv komunikace a mluveného projevu na podobu francouzského jazyka.

⁸ Francouzsky Louis le Germanique a Charles le Chauve.

⁹ Louis le Pieux.

2.2 Psaný versus mluvený projev

Mluvený a psaný jazyk můžeme v podstatě chápat jako dva odlišné systémy, které je zajímavé porovnat z lingvistického hlediska. Zatímco mluvená forma se neustále vyvíjí, psaný jazyk má tendenci se ustalovat. Rozdíl mezi psaným a mluveným projevem se tak víc a víc prohlubuje, psaný projev se stává autonomním a přestává reprezentovat projev mluvený. I když lidé užívali řeč dlouho předtím, než začali užívat písmo, je nyní mluvený projev často chápán jako sekundární. Je však potřeba si uvědomit, že vzhledem k tomu, že oba jazykové kódy, jak psaný, tak ústní, mají své přednosti i slabiny, nelze jednoznačně označit jeden z nich za primární a druhý sekundární. V rámci jazykové struktury je potřeba je vnímat jako doplňující se součásti tvořící jeden celek.

Jedním ze způsobů klasifikace je pohled z hlediska produkčních a percepčních mechanismů. Obecně lze říci, že zatímco mluvený projev představuje zvukovou realizaci vytvořenou řečovými orgány, pro jejíž percepci je podstatná aktivizace sluchového smyslu, psaná informace je oproti tomu grafickou realizací myšlenkového záměru subjektu, která vyžaduje specifickou materiální podporu a zapojuje do percepce zrak (Riegel a kol. 2009, s. 52). Tato rozdílnost v produkci projevů nám však sama o sobě nevysvětlí, s jakým množstvím odlišností se lze setkat v mluvené formě jazyka oproti jeho formě psané. Debata na toto téma se stala znovu aktuální zejména v posledních letech, kdy se přistoupilo k analýzám komunikačních technik spojených s novodobými technologiemi (emaily, chaty, SMS atd.), z nichž některé jsou považovány v podstatě za převedení mluveného projevu do psané formy, i když je jasné, že zprávy tohoto druhu řadíme do písemných realizací jazyka.

Uvědomíme-li si, že veškerá mezilidská komunikace je kódována způsobem buďto fonickým, nebo grafickým, dojdeme k závěru, že tato dvě média vytváří v rámci své fyzické realizace jasnou dichotomii: vysílaná či přijímaná zpráva je nutně formulována ústně či písemně, aby mohla být informace v ní obsažená srozumitelně sdělena. Ústní a písemná komunikace je odstupňována na základě svých vlastností. Přesnější stupnici definoval lingvista

Peter Koch spolu s Wulfem Oesterreicherem¹⁰, když rozčlenili sdělení do dvou opozičních pólů charakterizujících zprávy podle jejich *blízkosti* (*proximité*) či *vzdálenosti* (*distance*). Slovo „blízkost“ používají ve smyslu slova „přímost“.¹¹ Koch a Oesterreicher rozlišují například soukromou komunikaci od veřejné, důvěrně známého mluvčího od neznámého, silnou emocionalitu projevu od slabé emotionality, dialog od monologu či tematickou svobodu od pevně stanoveného tématu. Mluvená forma je reprezentantem komunikační blízkosti a písemně zpracovaný text je nejlepším příkladem komunikační vzdálenosti. Ve spontánní ústní konverzaci uběhne pouze minimální čas dělící emisi a recepci. Mluví se navíc nachází v časoprostorově determinované situaci. Komunikace na dálku je oproti tomu odlišná. Vysílání a příjem informace zde neprobíhají v dané situaci současně, ale existuje mezi nimi kratší či delší zpoždění. Mnoho sdělení však tvoří přechod mezi těmito dvěma protipóly. Například hlasitá četba psaného textu je charakterizována jako komunikačně vzdálená, zatímco osobní dopis mezi dvěma přáteli užívající grafické kódování představuje komunikační blízkost (Riegel a kol. 2009, s. 53).

Formy ústních a písemných sdělení jsou určovány komunikačními podmínkami. Písemná komunikace poskytuje dostatečný čas pro vytvoření sdělení. Pisatel se může k již složenému textu vrátit a dále jej editovat, kontrolovat a doplňovat. Během spontánního mluveného projevu mluvčí produkuje a vysílá svou zprávu téměř simultánně. Všechny chyby a nepodařené formulace zůstanou součástí sdělení a mluvčí je již nemůže vzít zpět, aniž by narušil plynulost toku řeči prostřednictvím opravného zopakování, hezitačních zvuků, řečových pauz či přeformulování. To však zanechá nerasmazatelné stopy ve sdělení samotném. Tento základní rozdíl ve formulaci objasňuje určitou preciznost a dokonalost psaného projevu oproti projevu mluvenému, který může působit místy zmateně a nést známky autokorekce mluvčího či přerušování ze strany posluchače. Jak jsme již zmínili výše, mluvený projev je doprovázen

¹⁰ Koch, Peter, Oesterreicher, Wulf. 2001. Langage parlé et langage écrit. In *Lexikon der romanistischen Linguistik (LRL)*, vol. 1. ed. G. Holtus, M. Metzeltin, and C. Schmitt, 584–627. Tübingen: Max Niemeyer.

¹¹ Výše zmíněné rozdělení funguje jako bližší charakteristika kanálů, kterými proudí informace k recipientovi. Jejich blízkost či vzdálenost určuje do jisté míry i rychlost recepce a percepce. Tedy to, jak pohotově a přesně je recipient ovlivněn mnoha okolnostmi schopem přijmout vysílanou zprávu a pochopit ji.

množstvím informací pocházejících z tónu hlasu, intonace a síly hlasového projevu, z gest a z mimiky, které dále podmiňují správnou percepci sdělení (Riegel a kol. 2009, s. 52–55).

2.2.1 Francouzština slovem a písmem

Od 17. století hrála psaná podoba francouzského jazyka ve Francii velmi prestižní roli vycházející především z významného postavení klasické literatury a filozofie, která nastolila normu jak v literatuře podléhající v té době přísným pravidlům (*les règles*), tak v gramatice. V roce 1635 zakládá kardinál Richelieu Francouzskou akademii (*Académie française*), která má za cíl kodifikovat normu francouzského jazyka. Již od počátku se Akademie snažila ustálit podobu francouzského jazyka tím, že mu dá jasná pravidla, a tak z něj udělá čistou řeč srozumitelnou pro všechny. Ústní projev je terčem kritiky preskriptivní jazykové normy. Jeho nedokonalosti vyvěrající z přirozené povahy ústní komunikace jsou z normativního pohledu degradací či ochuzováním francouzštiny oproti pevně stanovené písemné normě.

Výraz *français parlé* má dvojitý význam: používá se pro obecné pojmenování francouzštiny používané při řeči (zvukové komunikaci), má ale také negativní konotace jako synonymum slov *populaire* či *familier* (v překladu *lidový*), která označují uvolněný styl francouzštiny, její degradaci, zkrátka francouzštinu diskreditující jazykovou normu (Riegel a kol. 2009, s. 55, 56). Jak píše Françoise Gadet, vedle *français standard*, která je objektem gramatických analýz, existuje ještě *français populaire* vyznačující se vlastními lingvistickými rysy, která je jazykem znevýhodněných sociálních skupin, jazykem lidu. Kde takový jazyk najdeme? Jako příklad uvádí Gadet film o pařížském výrostkovi s Jeanem Gabinem *Nesahejte na prachy* (*Touchez pas au grisbi*, R: Jacques Becker, 1954). Vysvětluje, že od lidového jazyka (*langue populaire*) musíme ještě odlišit argot, který byl dlouho jazykem zločinců, kteří jej používali ke komunikaci v rámci svých skupin (Gadet 1992, s. 3,4). Deskripce jazyka je tedy postavena na opozici mluvené podoby francouzštiny, považované za cosi chybného a vadného, a „správného“ psaného jazyka odpovídajícího normě.

Nemůžeme přitom omezovat *français parlé* na *français familier* nebo *populaire*, protože jak jazyk psaný, tak mluvený se vyznačují růzností registrů a i když je normativní cenzura u psaného jazyka daleko silnější, obě jazykové varianty zahrnují obměny v oblasti sociální, geografické či stylistické, podle nichž je třeba je rozlišovat.

2.3 Boj za ryzí francouzštinu

Jak píše Henriette Walterová, francouzská lingvistka a emeritní profesorka na Univerzitě v Rennes, „cizince dosud udivuje postoj Francouzů k jejich vlastnímu jazyku.“ (Walterová 1993, s. 15). Podle Walterové a její knihy z roku 1988 (rok vydání ve Francii) Francouzi nepovažují svůj jazyk za nástroj, který je možno si přizpůsobit vlastním komunikačním potřebám, ale dívají se na něj jako na platnou konvencemi a tradicemi svázanou instituci, která je poměrně rigidní a takřka posvátná. Francouzi jsou podle ní „velmi tvrdě vedeni k tomu, aby uznali jako správná jenom ta slova, která jsou uváděna ve slovníku. Pokud nějaké slovo ve slovníku nenajdou, velmi přesvědčivě, ale v rozporu s jakoukoli logikou, prohlašují, že toto slovo není francouzské a tedy jednoduše neexistuje. A to i přesto, že je použili a jeho významu rozumějí.“ (Walterová 1993, s. 16). Za vytvoření nového slova byl kritizován již ne jeden Francouz. Walterová uvádí případ slova *cohabitateur*, které můžeme přeložit jako „spolubydlící“, namísto existujícího *colocataire*. Slovo vzniklo v roce 1986 a jeho autorem byl Coluche, oblíbený francouzský herec a komik, kterého můžeme u nás v Česku znát například z komedie Clauda Zidiho *Křídýlko nebo stehýnko* (1976), kde zazářil po boku Louise de Funèse v roli jeho syna tíhnoucího k cirkusu.

Je pravdou, že Francouzi jsou i dnes známí pro svou snahu udržet si čistou podobu jazyka, který bytostně střeží, váží si ho a chrání ho například od anglicismů. Přestože ve francouzštině existuje mnoho výpůjček z angličtiny týkajících se například oblastí hudby¹², kinematografie¹³ či společnosti

¹² Např. *best of*, *CD*, *live*, *mix*, *old school*, *rocker*, *freestyle* a jiné.

¹³ Např. *making-of*, *blockbuster*, *road-movie*, *storyboard*, *zapping*, *vidéo* a jiné.

obecně¹⁴, existují poměrně výrazné tendence této expanzi výpůjček zamezovat. V roce 1964 sepsal francouzský puristický lingvista René Étiemble své slavné dílo *Parlez-vous franglais?*, ve kterém hodnotil využívání anglických slov ve francouzštině jako chybu a snobismus. Automaticky předpokládal, že Francouzi od výpůjček sami upustí, když budou informováni o směšnosti a neužitečnosti této módy.

Během následujících let se ve Francii zformovala řada institucí, které boj proti anglo-americké lingvistické invazi vzaly za svůj. Uvedme Paula Camuse a jeho společnost z roku 1958 s názvem *Défense de la langue française*, která měla podobně jako Francouzská akademie za cíl ochránit kvality francouzského jazyka – jeho jasnost, harmonii a preciznost – a která si také vzala za úkol šířit jazyk dále do světa. Od roku 1970 dohlíží na snižování počtu anglicismů komise s názvem *Commissions ministérielles de Terminologie*. Následovalo několik iniciativ v podobě komisařství *Commissariat Général de la Langue Française* (1984) či rady jménem *Conseil supérieur de la langue française* (1989). I v současné době existuje více institucí podobného rázu. Jmenujme především *La délégation générale à la langue française et aux langues de France* (DGLFLF)¹⁵ – oddělení zřízené v roce 2002 při francouzském Ministerstvu kultury a komunikace – nebo *Commission Générale de Terminologie et de Néologie*, sdružení zaštitěné předsedou vlády, v rámci něhož existují ještě specializované komise soustředící se na terminologii informatiky, průmyslu, obchodu, služeb či na finanční terminologii, názvosloví atd. Jejich práce spočívá v obohacování francouzského jazyka prostřednictvím specializovaných termínů, v propagaci používání francouzských výrazů na úkor anglicismů zejména v ekonomické, soudnické, vědecké a technické oblasti a v neposlední řadě v rozvoji frankofonie.

Ve Francii existují také zákony vztahující se k obraně jazyka. Roku 1975, konkrétně 31. prosince vznikl zákon řečený *Bas-Lauriol*, který přikazoval užívání francouzštiny například v následujících případech: v pojmenováních, nabídkách, prezentacích, psaných či mluvených reklamách atd. Tento zákon

¹⁴ Např. *squatter*, *hippie*, *check-point*, *punk*, *real life*, *underground* a jiné.

¹⁵ Více informací zde: http://www.dglf.culture.gouv.fr/dglf_presentation_anglais.htm.

dopomohl k potrestání firem, které ve Francii nepoužívají francouzštinu. Britská společnost British Airways tak byla odsouzena za to, že šířila letenky vyhotovené jen v angličtině. Další zákon pocházející z nedávné minulosti vstoupil v platnost 4. srpna v roce 1994 a byl pojmenován jako *Toubon*. Tento zákon například nařizoval, aby se všichni francouzští účastníci konferencí a kongresů pořádaných ve Francii vyjadřovali při svých vystoupeních pouze ve francouzském jazyce (Engstrøm 2005, s. 85–87).

Hlavním iniciátorem slovníkových změn a zavádění neologismů (často ve snaze potlačit vliv angličtiny) však stále zůstává Francouzská akademie. Příkladem jejich působení na jazyk může být vysvětlivka ke slovu *digital* na jejích webových stránkách¹⁶, která zní následovně:

L'adjectif *digital* en français signifie « qui appartient aux doigts, se rapporte aux doigts ». Il vient du latin *digitalis*, « qui a l'épaisseur d'un doigt », lui-même dérivé de *digitus*, « doigt ». C'est parce que l'on comptait sur ses doigts que de ce nom latin a aussi été tiré, en anglais, *digit*, « chiffre », et *digital*, « qui utilise des nombres ». On se gardera bien de confondre ces deux adjectifs *digital*, qui appartiennent à des langues différentes et dont les sens ne se recouvrent pas : on se souviendra que le français a à sa disposition l'adjectif *numérique*.

On dit: *Une montre à affichage numérique.*
 Un appareil photo numérique.

On ne dit pas: *Une montre à affichage digital.*
 Un appareil photo digital.

Ve vysvětlivce se dozvíme, proč není vhodné používat ve francouzštině slovo anglického původu *digital*, a jsme upozorněni na to, že korektní podoba francouzštiny užívá pro tyto účely přídavného jména *numérique*, což dokládají demonstrativní příklady. *Numérique* patří, stejně tak jako například *ordinateur* (neboli *computer*) či *logiciel* (neboli *software*), mezi úspěšně implementované výrazy, které byly národem přijaty za vlastní. Současným příkladem zlepšujících návrhů Francouzské akademie je podnět Érika Orsseny, jehož záměrem je nahradit v současnosti hojně užívané anglické akronymum vulgární povahy

¹⁶ Viz: <http://www.academie-francaise.fr/dire-ne-pas-dire/neologismes-anglicismes>.

WTF (What the Fuck) francouzštějším *QPJA (Quelle Partie de Jambes en l'Air !)*. Výraz *WTF* vyjadřuje silné udivení podbarvené nevěřícností. Nové Orssenyho akronymum spoléhá na to, že „není nezbytné utíkat se k vulgaritám, abychom vyjádřili udivení.“ „Výrazem *QPJA* se se obracíme přímo na mládež,“ jak tvrdí Orssena. Tato zkratka, stejně jako zkratka Xaviera Darcose *OMEAS (Oh Mon Etre Abstrait Supérieur !)* nahrazující příslovečné *OMG (Oh, My God)*, se stane součástí příštího vydání akademického slovníku francouzštiny coby respektované a oficiální obdoby slavných *WTF* a *OMG*. Právě tento direktivní a poněkud násilný přístup puristů k jazyku je často kritizován jak ze strany lingvistů, tak ze strany běžných uživatelů, kteří se s takovými změnami, které mnohdy příliš nekorrespondují s realitou, nemohou ztotožnit (Le Courrier des Echos, 2013).

2.4 Pravopisná hrdost versus modernizující prvky současné francouzštiny

Čas od času můžeme zaslechnout tvrzení, že Francouzi už neznají pravopis. Výjimkou nejsou stížnosti posluchačů a diváků na moderátory v rádiích a televizích či školní diktáty prošpikované četnými chybami. Tato situace ovšem nastává snad v každé zemi a ani česká média se mnohdy nevyhnuou výrokům, o jejichž gramatické či sémantické správnosti či vhodnosti by se dalo s úspěchem polemizovat, o zaměňování předložek „kvůli“ a „díky“ jehož jsme pravidelně svědky v českých médiích či filmech (ať už zahraničních dabovaných či českých) nemluvě.

Je pravda, že Francouzi jsou si složitosti svého jazyka dobře vědomi a není tak náhodou, že se z pravopisu a jeho správné interpretace stává takřka národní sport. Orientovat se bezchybně ve všech jazykových nuancích je totiž úkol hodný pravopisného přeborníka. Denně můžeme slyšet například na rozhlasových stanicích *Radio France* pořady či diskuze o lingvistice, pravopisné soutěže typu „vyhláskujte do telefonu správně slovo *ennuyeux*“ a tak dále.

Článek zveřejněný v listopadu 2013 na webových stránkách radia *France Bleu*, jehož autorkou je novinářka Marina Cabiten, představil výsledky ankety publikované deníkem *Le Parisien – Aujourd'hui en France* a organizované skupinou s názvem *Les Timbrés de l'Orthographe* ve spolupráci s institutem Mediaprism. Osmdesát procent Francouzů zde hodnotí sebe sama jako dobré znalce pravopisu. V závěsu za touto informací stojí však fakt, že ve skutečnosti ubývá počet těch, jejichž pravopis se pohybuje na slušné úrovni. Osm respondentů z deseti se tedy považuje za dobré znalce pravopisu a 26 % dotazovaných dokonce tvrdilo, že prakticky nedělají chyby. Ale, jak shrnuje Marina Cabiten, anketa se neptala pouze na názory, ale testovala také jazykovou úroveň respondentů. Z 8 000 testovaných studentů dosáhla pouze polovina dobré úrovně bez hrubých chyb. Průzkum se zabýval také vztahem Francouzů k jejich jazyku. Většina zúčastněných považuje tento jazyk za těžký (62 %), ale i tak jsou proti jakémukoliv zjednodušení, 65 % rovněž považuje diktáty za neodmyslitelnou součást výuky. A v čem dělají Francouzi nejčastěji chyby? Tak například v převádění složených slov (jako *chasse-neige*) do plurálu (až z 56 %), v rozdílném časování budoucího času a kondicionálu, poté v množném čísle u číslovek, ve dvojhláskách a v příčestí minulém (Cabiten, 2013).

Zdá se, že pravopis a s ním spojená otázka cti či národní hrdosti přitahuje stále více lidí. „V roce 1985 uspořádal časopis *Lire* z podnětu Bernada Pivota první pravopisné přebory.“ (Walterová 1993, s. 240). Přeboru se tehdy zúčastnilo 50 000 osob a podle údajů z peoplemetrů sledovalo klání u obrazovek více než 5 miliónů diváku. I v současné době se pravopis těší velké pozornosti znalců a milovníků jazyka. Každoroční turnaj v pravopisu organizuje již několik let zmíněná skupina *Timbrés de l'Orthographe*, která vznikla z iniciativy nakladatelství s názvem *Éditions de l'Opportun*. Za podpory poštovní společnosti *La Poste* a Ministerstva národního vzdělávání (*Le ministère de l'Éducation nationale*) proběhne v dubnu 2014 již 4. ročník soutěže, jehož regionálního kola se zúčastní 10 000 finalistů ve 23 městech

Francie. V červnu 2014 se pak v Paříži uskuteční velké národní kolo, kde se mezi sebou utká nejlepších 500 finalistů.¹⁷

I přes snahy zachovat pravopisnou korektnost jazyka jsou i ve francouzštině, tak jako například v angličtině, výrazné tendence přizpůsobovat jazyk moderním technologiím, což vyžaduje například zkracování a stručnější formulace v zájmu urychlení komunikace či flexibilního přizpůsobení se formátu. Nejedná se však pouze o zkracování, ale v nemalé míře také o vznik novotvarů expresivního typu, které se již vymykají pravidlům tvoření běžným pro francouzský jazyk. Jedná se často o nový slovník vycházející z mluvy mládeže. Prvními, kdo si uvědomil „naléhavost rozdílu mezi jazykem mladých a jejich rodičů, byli novináři: v *20ans*, *Lire*, *Le Nouvel Observateur*, *Express* nebo *Quotidien de Paris*, ti vydávali od počátku 80. let seznamy nových slov s komentáři a výkladem.“ (Walterová 1993, s. 275). Během četby se pak „staří“ (tedy ti, kterým bylo více než 25–30 let) měli možnost seznámit s některými slovy často slýchanými u mladých lidí, jejichž významu až doposud zcela nerozuměli. Takové seznamy vznikají dodnes. Deník *Le Figaro* či televizní kanál *TF1* například uveřejnili seznam oblíbených neologismů i s jejich výkladem. Jedná se o průzkum společnosti *OpinionWay* pro festival XYZ uskutečněný v říjnu 2012. Podle něj si celkem 94 % mladých Francouzů ve věku 18–24 myslí, že zavádění nových slov do jazyka je v pořádku, pokud nové výrazy nebudou pro jazyk degradující a najde se pro ně využití. V opačném případě to 74 % respondentů považuje za špatné. Ve vybraném seznamu nových slov vidíme, že mnoho z nich je zasvěceno digitálním technologiím (Quillet, 2012).

- ATTACHIANT = jedná se o slovo, které je oblíbené u 33 % mládeže. Osobu, která je *attachiante* je těžké snášet, ale přesto nás něčím přitahuje.
- ORDINOSAURE = jedná se o počítač hodící se do starého železa. Česky bychom pravděpodobně použili něco jako *stará šunka* či *vykopávka*.

¹⁷ Více informací zde: <http://timbresdelorthographe.com/>.

- AIMEUSE = tento výraz označuje ženu, která má ráda věci a lidi.
- SE FAIRE ELECTRONIQUER = být oklamán elektronikou – situace, kdy se elektronika spikne proti svému uživateli.
- BONJOIR = univerzální mix mezi *bonsoir* a *bonjour*.
- PHONARD = slovo kritizující označení osoby, která špatně zachází se svým mobilním telefonem.
- PHOTOPHONER = znamená fotit se telefonem a posílat fotku svým starým rodičům, abychom dali najevo, že je vše v pořádku. Jedná se o kompromis mezi SMS a povinnou návštěvou rodičů.

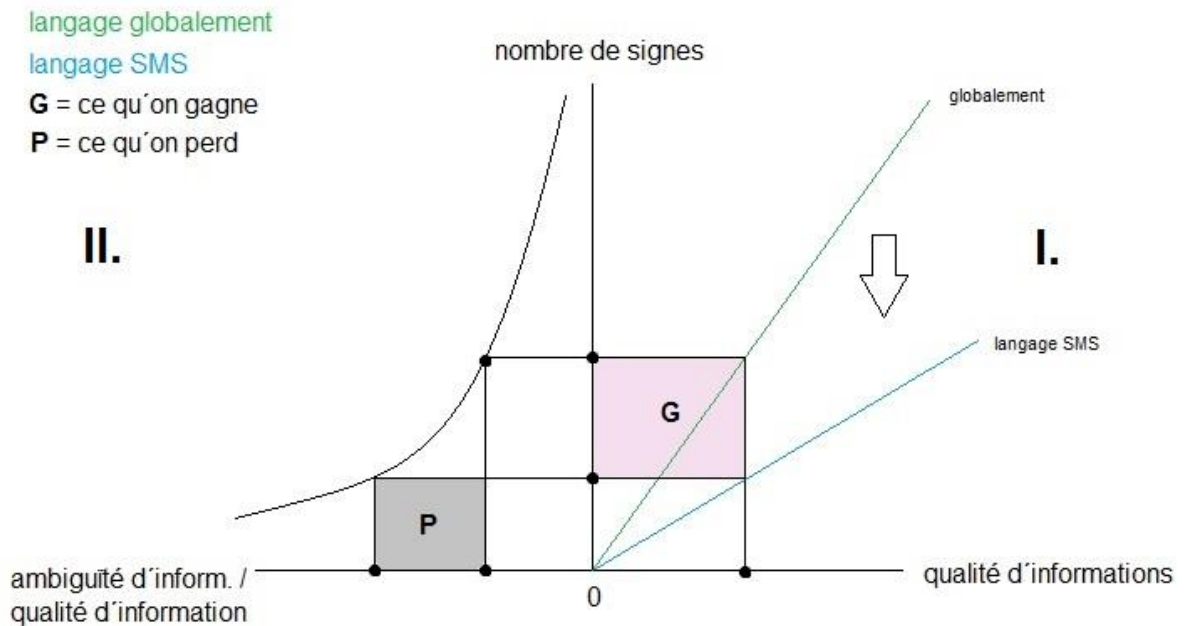
Jelikož se naše diplomová práce zabývá převážně mluveným projevem ve francouzštině, nesmíme opomenout ani již dříve zmíněný jazyk chatů a SMS, který můžeme v převážné většině případů považovat za transpozici mluveného projevu do psané formy, i když tyto zprávy stále řadíme do písemných realizací jazyka.

2.4.1 Jazyk SMS jako odraz mluvené podoby francouzštiny

Velmi inspirativní analýzou jazyka SMS je prezentace belgického ekonoma, umělce a filozofa působícího mimo jiné i v oblasti lingvistiky. Laurent D'Ursel ve své videopřednášce *Le langage SMS ou L'impossible révolution démocratique* s podtitulem *Qui est bien, tchatte bien*, tvrdí, že přesto, že toto užití jazyka může být považováno za primitivní, pro zvládnutí jazyka SMS je potřeba dokonale ovládat jazyk jako takový a umět ho „emancipovat“. Podle D'Ursela jsou to právě ti nejlepší z nejlepších, kteří umějí zvládnout a využívat jazyk SMS. Svou tezi dokládá tak, že vysvětluje, jak tato jazyková modifikace funguje a poukazuje na jistou složitost jejího tvoření (D'Ursel, 2013).

Hlavním principem jazyka SMS je ekonomie znaků a písmen, jde o snahu „jít přímo k jádru věci“ a o tendenci minimalizovat počet znaků vedoucích k vyjádření zprávy. Podle D'Ursela jde svým způsobem o ztělesnění

motta „čas jsou peníze“. D'Ursel přibližuje fungování jazyka a následně jazyka SMS na následujícím grafu.



Graf Lurenta D'Ursela mapující obecné principy jazyka ve vztahu k principům jazyka SMS.

Na uvedeném grafu vidíme, co používáním jazyka SMS získáváme (G) a co naopak ztrácíme (P). Vysvětleme si toto schéma podrobněji.

Součástí grafu je vertikální osa zaznamenávající počet znaků (*nombre de signes*) a osa horizontální, jejíž ubíhání do prvního kvadrantu (I.) znázorňuje kvalitu informace (*qualité d'information*). Naopak její část zasahující do druhého kvadrantu (II.) mapuje hodnotu mnohoznačnosti (či nejednoznačnosti) informace (*ambiguïté d'information*) a s ní opět svázanou kvalitu (*qualité*).

Přímka načrtnutá v prvním kvadrantu, znázorňující přímou úměru (s názvem *globalement*), popisuje obecně známé tvrzení, že čím je ve sdělení použito více znaků, tím více stoupá kvalita informace. Naproti tomu hyperbola ve druhém kvadrantu zaznamenává proměnu mnohoznačnosti informace v závislosti na počtu znaků. Čím více znaků obsahuje sdělení, tím menší je

hodnota mnohoznačnosti a naopak, s klesajícím množstvím znaků roste úroveň mnohoznačnosti¹⁸.

Schéma jazyku SMS je znázorněno v I. kvadrantu jako přímka s názvem *langage SMS*. Jedná se o zakreslení faktu, že jazyk SMS se v rámci svých snah o jazykovou ekonomii definuje jako komunikace zachovávající stejné množství informace¹⁹ při současném snížení počtu znaků. Rozdíl počtu znaků²⁰ při zachování hodnoty kvality informace je v grafu označen jako G a vyjadřuje hodnotu, která je naší výhodou při používání jazyka SMS.

Problematika jazyka SMS má však i svou negativní stránku, která je zaznačena ve II. kvadrantu. Zestručněním zprávy dochází, jak víme, ke zvýšení nejednoznačnosti sdělení. Logickým důsledkem tedy je, že i když jazyk SMS dokázal zachovat hodnotu kvality informace při redukci počtu znaků, projeví se to negativně na hodnotě mnohoznačnosti, která úbytkem znaků vzroste. Právě rozdíl hodnot mnohoznačnosti produkované při obecném jazykovém sdělení a při sdělení jazykem SMS je v grafu označen jako P, což je údaj popisující naše ztráty při aplikaci zestručňujícího jazyka SMS.

Další část své prezentace věnuje Laurent D'Ursel technické stránce jazyka SMS. Vysvětluje logiku a způsob jeho tvoření, které se ve většině případů odkazuje na fonické ztvárnění jazyka. Zmiňuje několik postupů, mezi nimi například používání znaků, jako jsou plus, minus, rovnítko a tak dále – např. „pa=“ (*pas égal*), „pa+“ (*pas plus*), „@+“ (*à plus*), „parX“ (*parfois*); eliminaci samohlásek – např. „bcp“ (*beaucoup*), „slt“ (*salut*), „yx“ (*yeux*); dále fonetizaci na základě výslovnosti písmen – např. „LCKC“ (*elle s'est cassée*), „NRJ“ (*energie*), „FaC“ (*effacer*), „ouvR“ (*ouvert*) atd. Při této příležitosti vysvětluje pravidla psaní velkých a malých písmen. Velká písmena nabádají ke čtení stejným způsobem jako v abecedě. „R“ se tedy bude číst jako [εr], kdežto „r“ přečteme jako [r]. Odtud rozdíl například mezi „ouvR“ použité v SMS čtené

¹⁸ Jedná se především o mnohoznačnost významovou, neboť francouzština je stejně jako mnoho dalších jazyků řečí bohatou na jazykové nejednoznačnosti.

¹⁹ Rozumějme stejné množství informací při totožném sdělení, na které jsou však uplatňovány obecné jazykové zákonitosti a nikoli zákonitosti svázané s jazykem SMS.

²⁰ Během obecného používání jazyka a oproti němu také jazyka SMS při stejném sdělení.

jako [u:vr] a „ouvR“ čtené jako [uve:R]. Dalším způsobem tvoření je užívání čísel a aglutinace nebo odloučení. Příkladem je „2m1“ (*demain*) lišící se od „2 m1“ (*deux mains*). Jako příklad dalšího nápaditého užívání čísel uveďme „parf1“ (*parfum*), „po6tif“ (*positif*), „re6clé“ (*recyclé*) či „6Tmatik“ (*systématique*).

Není potřeba příliš zdůrazňovat, že pro jazyk SMS existují také speciální slovníky shrnující užívané výrazy. Uveďme internetový slovník *DictionnaireSMS.com*²¹, který umožňuje abecední vyhledávání potřebných slov. Tam najdeme výrazy jako „ab1to“ (*à bientôt*), „bi1“ (*bien*), „avanC“ (*avancer*), „6garett“ (*cigarette*) a jiné.

Vidíme tedy, že sepsané a mluvené podoby francouzštiny se obzvláště u současného jazyka SMS užívaného mladými lidmi projevuje velmi výrazně. Patrná je především fonetizace a ekonomizující tendence tak, jak je známe z mluveného projevu.

3. Analýza mluveného projevu

Gramatická tradice upřednostňuje od nepaměti jazyk psaný před mluveným. Není však pochyb o tom, že psaný i mluvený projev využívá stejné jazykové struktury a bylo by tudíž nesmyslné stavět do opozice morfosyntax psanou a mluvenou. Abychom mohli gramaticky popsat mluvený jazyk, je nutné se s jeho morfosyntaxí seznámit blíže, nicméně to neznamena izolovat analýzu mluvy od zbytku gramatiky. V následujících podkapitolách se blíže seznámíme s mechanismem organizace diskurzu, se slovní zásobou a s charakteristikou slovesných kategorií a syntaxe spjatých s mluveným projevem. Dále si shrneme některé gramatické a fonetické rysy hovorové francouzštiny a kapitolu uzavřeme stručným výkladem problematiky transkripce mluveného slova do psané podoby, což je základní postup při analýzách mluveného projevu.

²¹ Viz web: <http://www.dictionnaire-sms.com/>.

3.1 Organizace ústního diskurzu (způsoby tvoření)

Charakter ústního diskurzu se do velké míry odvíjí od podmínek jeho tvoření. Jak píše Claire Blanche-Benveniste ve své knize *Le français parlé, Études grammaticales*, studovat *français parlé* (mluvenou francouzštinu) znamená studovat obecně řečeno diskurzy, které nejsou předem připravené. Protože „když vytváříme předem nepřipravené promluvy, skládáme je postupně během jejich vytváření a toto vytváření na nich zanechává stopy.“ (Blanche-Benveniste 1990, s. 17). Když tyto „stopy“ příznačné pro orální formu jazyka převedeme do písemné podoby, jejich následná četba působí nepřirozeným dojmem. Nejsme zvyklí číst například nedořečená slova či váhavé zvuky a vidět je napsané. Nicméně studium těchto kompozičních „stop“ může být velmi zajímavým předmětem pozorování. Chceme-li srovnávat „nepředpřipravený“ mluvený text s textem psaným, bylo by správnější přirovnat ho k textu napsanému nanečisto, tedy k jakémusi návrhu či konceptu psaného textu. Výhodou při pozorování je, že procesy, které při tvorbě ústního projevu používáme, nejsou nijak individuální a můžeme v nich rozpoznat pravidla, která se týkají každého mluvčího (Blanche-Benveniste 1990, s. 17).

Jak jsme již zmínili dříve, orální diskurz obsahuje opakování, zaváhání, falešné začátky, přerušení, narušení koncepce, nedokončené promluvy, autokorekce atd. Posluchač je přímo přítomen procesu produkce diskurzu a mluvčí má možnost komentovat a upravovat význam toho, co právě říká (například volbu termínu s určitým zabarvením, intonací, gesty, mimikou a tak podobně). Během orálního projevu dochází také k dialogizaci textu (tedy v případě, že se nejedná výhradně o monolog). Dialog je založen na vzájemném konsensu, který upravuje „předání slova“ mezi účastníky. Znamená to, že každý z členů dialogu poskytne mluvčímu určitý prostor k promluvě, kterou může také sám doplňovat a reagovat na ni. Za takových podmínek probíhá dialog harmonicky a přehledně. V chaotických interakcích, jakými mohou být například televizní debaty, hádky či rozhovory početnějších skupin, není „předávání slova“²² respektováno a mluvčí si navzájem „skáčou do řeči“

²² Tedy pořadí, ve kterém se členové účastní dialogu.

nebo se přerušují. Klíčovými prvky pro fungující dialog jsou především okolnosti promluv a také referenční rámec, v jakém se daná situace realizuje. (Riegel a kol. 2009, s. 60–61).

3. 2 Slovní zásoba

Stavět do protikladu slovní zásobu mluveného a psaného jazyka není věcí příliš užitečnou. Výběr slov je vždy ovlivněn jazykovým registrem, tematickou oblastí, o které se hovoří, jazykovými a mentálními schopnostmi mluvčího či pisatele, jeho společenským statutem nebo příslušností k určité sociální skupině, ale také jeho momentální náladou, zdravotním a psychickým rozpoložením, nemluvě o různých indispozicích a dalších okolnostech, ať už rušivých či příznivých.

Jak zmiňuje Claire Blanche-Benveniste, důležitý diskurzivní rozdíl oproti psané podobě spočívá ve způsobu vytváření pojmenování (přesněji *travail de dénomination*). Během ústního projevu přemýšlí mluvčí nad co nejpřesnějším slovem vhodným k formulaci myšlenky tak, aby její transformace do jazyka odpovídala kvalitám jeho myšlenkovému konstrukt. Toto přemýšlení má za následek užívání určitého typu „vycpávkových“ slov či přesněji řečeno explicitních metajazykových komentářů (např. *comme on dit, comment dire, non pas..., mais...* atd.) či vágních termínů (*truc, machin, chose, faire* atd.). Jak uvádí Blanche-Benveniste, jedná se například o ujišťování se, zda jsme použili to „správné slovo“. V knize je uveden následující příklad:

« c'était une maison très huppée²³ – huppée je sais pas si c'est bien le mot qu'on doit dire mais enfin... » (Blanche-Benveniste 1990, s. 26).

Jindy mohou účastníci debaty při vymýšlení správného slova spolupracovat. Jako v příkladu s *krčním límcem*, ve francouzštině *la minerve* (používaném na krční páteř po úrazu), kdy mluvčí naznačuje gesty, směřujícími na krk, význam slova, které má na mysli (Blanche-Benveniste 1990, s. 26–29):

²³ Překlad: zazobaný, hogo fogo, luxusní (restaurace atp).

- „n°1 – Ça fait un... comment on dit ? Je sais plus... une chose, là
n°2 – Une écharpe ? Un col roulé ? Un...
n°1 – Mais non ! Le truc blanc qu'ils ont, ceux qui se sont cassés la, le,
 euh...
n°2 – Ah oui !
n°1 – la la la la cheva–
n°2 – La minerve !
n°1 – Voilà ! La minerve (Revelli 116, 15)“ (Blanche-Benveniste 1990, s. 29).

Existují vypracované seznamy slovní zásoby mluveného projevu vystavěné na hovorové rovině jazyka. Zajímavým, i když očekávaným, zjištěním je fakt, že frekvence užití slov a slovních druhů se při písemném a mluveném projevu liší. Na počátku 50. let 20. století byl představen výzkum s názvem *Français fondamental*, který definoval jádro 800 až 1000 nejfrekventovanějších slov užívaných mluvčími během ústního projevu. Na prvních místech tohoto seznamu se objevily především determinanty, zájmena, spojky a předložky. Tento gramatický základ se objevuje i v psaném projevu. O *Français fondamental* a o dalších důležitých studiích věnovaných tématu mluvené francouzštiny pojednáme v jedné z následujících kapitol.

3.3 Slovesný způsob a čas

Užití slovesných tvarů se u obou forem jazyka výrazně liší.

První osoba množného čísla (*nous allons*) je v mluvené formě, obzvláště v hovorové rovině, často nahrazována zájmenem *on* (*on va au théâtre?*), což s sebou přináší změnu slovesné koncovky (či tvaru) odpovídající třetí osobě jednotného čísla.

Podmínky pro užití způsobů a časů použitelných jak v mluvené, tak v psané formě nejsou identické. Například indikativ přítomnosti (*présent de l'indicatif*) je rozšířenější v ústním podání, kde splňuje téměř omnitemporální funkci, když vyjadřuje přítomnost a stejně tak minulost či budoucnost. Co se týká konjunktivu, tak ho podle skupiny *GARS* (*Groupe aixois de recherches en*

syntaxe) používají v ústním projevu všichni mluvčí. Skupina dokonce kvantifikovala jednotlivé případy jeho užití. Nejčastěji se vyskytuje po neosobních slovesech²⁴ (*faire, y avoir, falloir*) – ve 35 % užití, dále po slovesech s osobním podmětem (*je veux, attend que*, atd.) ve 28 %, po spojovacích výrazech *pour que, pour pas que a avant que* (11,5 %). Méně frekventovaný je ve vedlejších větách vztažných (4 %).

Také slovesné časy se přizpůsobují mluvenému projevu. Minulý čas prostý (*passé simple: je partis*) a imperfektum konjunktivu (*imparfait du subjonctif: qu'il partit*) jsou časy určené takřka výhradně pro psaný projev. Budoucí čas prostý (*futur simple: je partira*) bývá často nahrazován opisným tvarem budoucího času, tj. vazbou pomocného slovesa *aller* s infinitivem neboli blízkou budoucností (*future proche: nous allons partir*), i když významová nuance při užití těchto dvou vyjádření budoucího času není ekvivalentní.

3.4 Syntax

Model kanonické věty²⁵ se pro spontánní mluvený projev téměř neužívá, neboť v mluveném projevu býváme často svědky nebo tvůrci anakolutu neboli vyšinití z větné stavby. Jedná se o takový druh větné stavby, kdy například začneme větu, poté přeskochíme na jinou myšlenku, kterou do té věty vložíme, a následně pokračujeme ve větě původní. Příkladem anakolutu ve francouzském jazyce je věta: *Madame Dubois, comme Pierre, ne mangent pas leurs escalopes*. Zde se jedná o smíchání dvou myšlenek, které gramaticky nedává smysl. Musely by se formulovat následovně: *Ni Madame Dubois ni Pierre ne mangent pas leurs escalopes* nebo *Madame Dubois, comme Pierre, ne mange pas son escalope*. Věty mohou být také nekompletní či nedokončené. Například: *J'ai acheté une nouvelle voiture, qui – Regarde ce gars !* a tak dále. Stejně tak odpovědi na otázky jsou často eliptického

²⁴ To jsou ta, která se skloňují pouze se 3. osobou jednotného čísla.

²⁵ Kanonická věta (*phrase canonique*) je v zásadě vzorová věta. Jedná se o větu jednoduchou (tedy ne složenou), oznamovací (deklarativní – ne např. tázací), neutrální ve všech rovinách (ne negativní, bez vytýkání větného členu – bez *mise en relief*, ...) a s pořadím slov, které je co nejjednodušší a ve své funkci nejcharakterističtější. Pořadí větných členů je následující: *podmět – přísudek-přísllovečné určení / předmět*.

charakteru a vynechávají tak méně podstatné části vět: *Vous partez quand? – À 10 heures.*

Ústní promluvy mají obecnou tendenci zjednodušovat. Příkladem je vynechání první části negace, tedy vypuštění *ne*: *Tu veux ou tu veux pas?* V souvislosti s rizikem nejednoznačnosti dáváme při otázce přednost intonaci (*Tu viens?*) či použití *est-ce que* (*Est-ce que tu viens?*) před inverzí. Tyto zjednodušující postupy ovšem úzce souvisejí s rovinou jazyka (například vynechání *ne* je považováno za hovorovou záležitost). V zájmu expresivity a ekonomie se také poměrně často užívá jednočlenných vět neslovesných jako: *Excellent, ce vin ! ; Ce tableau. quel chef-d'œuvre!* V mluveném projevu se mnohdy setkáme i s indikátory, které slouží k lepší orientaci ve větné struktuře, jako jsou *alors* či *voilà* nebo *quoi*, které uvozují konec výpovědi. Mluvčí budí zvědavost či přitahuje pozornost posluchače prostřednictvím slov s fatickou (neboli kontaktovou) funkcí, když používá formulace typu *hein, n'est-ce pas?, bon, vous voyez, vous savez, vous me suivez?* a jiné.

Už jsme se také několikrát zmiňovali o opakování slov v rámci spontánního ústního projevu. Nejčastějším uspořádáním repetitivních prvků je jejich postavení v antepozici a postpozici současně: *ça c'est les monuments dijonnais ça*. Opakování slov se děje z různých důvodů – zaváhání, chyba, nutnost zopakování pro správné pochopení, zdůrazňování důležitosti, stylistické procesy.

Důležitou součástí jsou tzv. *deiktická slova*, neboli *déictiques*²⁶, poukazující na kontext výpovědi. Setkáme se kupříkladu s častějším užitím *ça* nesoucím deiktickou funkci (*Ecoute un peu ça*). Tradičně považujeme *ça* za hovorovou podobu *cela*, ale zdá se, že jeho využití přesahuje hranice jazykové roviny. Používá se zejména při formulacích neosobních vět: *Ça marche très bien, merci*. Dalším jevem často se vyskytujícím v mluveném projevu jsou tzv.

²⁶ Slovem *déictique* se ve francouzské lingvistice označují prvky, které mají funkci ukazatele. Vztahují se k okolnostem výpovědi, identitě mluvčího, času a prostředí promluvy: *moi, toi, ceci, cela, maintenant* jsou výrazy označované jako *déictique*.

présentatifs, neboli prvky zpřítomňující výpověď. Jedná se zpravidla o spojení *il y a* a *c'est*, která uvozují promluvu.

Nesmíme opomenout ani expresivní termíny. Interjekce jako *oh*, *euh*, *bah*, *aïe* jsou vícevýznamové a jejich funkce se ustálí až při situaci samotné. Setkáme se také se zvukomalebnými slovy: *crac*, *boum*, *hi han*, *miaou*, *cataclap*, která vyvolávají dojem reálného zvuku nebo na něj upomínají.

Závěrem lze říci, že s mluvenými projevy přichází do morfosyntaxe jazyka specifické struktury a obměny, ale stále se objevují i pravidla psané formy, která se zde projevují s různou frekvenčností (Riegel a kol. 2009, s. 60–67).

3.5 Gramatické a fonetické rysy mluvené francouzštiny

Shrňme si nyní některé gramatické a výslovnostní rysy, se kterými se setkáme při kontaktu s mluvenou francouzštinou:

1. Jak už bylo zmíněno, setkáme se s absencí *ne* v záporu.

Je (ne) vois pas ma voiture.

2. Ve výslovnosti vymizí souhláska „l“ u osobního zájmena *il*.

l(l) part quand ?

3. Dochází také k nahrazení zájmena *nous* zájmenem *on* či k absenci osobních zájmen vůbec.

Nous savons pas quoi faire. – On sait pas quoi faire.

(Il) Faut y aller !

4. Ve výslovnosti se vytratí samohláska *u* u osobního zájmena *tu*.

T(u) as mangé ?

5. Často také dochází k utlumení osobního zájmena *je*.

(Je >> J' >>) Ch'te connais !

6. Vymizení koncové souhlásky *l* ve výslovnosti.

C'est incroyab(le), ça !

7. Vynechání inverze v tázacích větách – dává se přednost užívání *est-ce que*.

Est-ce qu'on a pu l'interroger ?

8. Souřadná spojka *que* nahrazuje jiné spojky.

Y a personne d'autre que (à qui) tu pourrais parler?

*La fille **que** (dont) j'te parle, **elle** s'appelle Gisèle.*

9. Opakování podmětu prostřednictvím osobního zájmena.

Jean, i(l) va me téléphoner quand il a fini.

Les enfants, ça dérange.

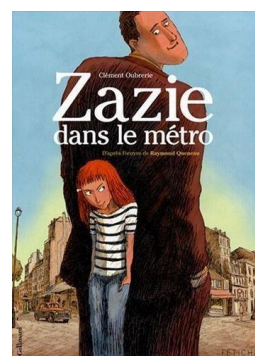
10. Nicméně jazyk se v mluvené formě pouze nezjednodušuje. Objevují se i komplikující tendence jako je několikanásobná negace.

*J'ai **pas** vu **personne**.*

Na výslovnosti hovorové francouzštiny se podílí také zvýšené tempo řeči, které s sebou přináší další zjednodušování či vynechávání výslovnostních elementů (Mougeon 1995, s. 96– 97; Français facile 2014).

3.5.1 Raymond Queneau: příklad hovorové francouzštiny v literatuře

V souvislosti s hovorovým jazykem nemůžeme nezmínit jeden z nejznámějších případů přepisu hovorového jazyka do literatury. Jedná se o román z roku 1959 *Zazie dans le métro*, dílo francouzského spisovatele a jednoho ze zakladatelů skupiny OuLiPo Raymonda Queneaua. Jde o román vynikající nejen vtipem a bizarností svých postav, ale také novátorským stylem textu, jehož dialogy jsou napsány nespisovným jazykem, a využívají i fonetického přepisu slov. Roku 1960 adaptoval režisér Louis Malle román do filmové podoby a v hlavních rolích se představili Catherine Demongeot a Philippe Noiret. V roce 2008, u příležitosti 50. výročí 1. knižního vydání, zpracoval Clément Oubrerie příběh do komiksu.



Queneau se podle svých slov již od dětství zajímal o cizí jazyky, ale také o hovorový jazyk a argot. Mluvenou francouzštinu považoval za velmi odlišnou

od francouzštiny psané a zašel až tak daleko, že mluvenou a psanou podobu francouzštiny prohlásil za dva jiné jazyky.²⁷ Ve svém článku s názvem *Écrit en 1955* vysvětloval, že francouzština se skládá ze dvou jazyků, z nichž jeden je jazyk, který se učíme ve škole (jazyk spisovný, psaný) a který je chráněn oficiálními institucemi jako je Francouzská akademie, a dále je to jazyk hovorový či lidový, který je autonomní a mnohem modernější. Queneau ho nazývá termínem *néo-français* (tedy „neofrancouzštinou“). Spisovná francouzština je matkou, neofrancouzština jejím dítětem.²⁸

Queneau uvedl v *Zazie* v praxi svou teorii o jazykové reformě a pomohl zrodu „nového jazyka“. Na první pohled vidíme pravopisnou obměnu u mnoha slov, stejně jako pro neofrancouzštinu charakteristickou transkripci mluveného jazyka. Příkladem těchto změn je například *égzagère*, *quèque chose*, *escuses* a *gzactement*. Tyto pravopisné změny však nejsou konstantní a v knize se setkáme s dalšími grafickými podobami téhož slova. Obecně se změny týkají slovní zásoby a pravopisu. Ve slovní zásobě se objevují hovorové výrazy, zkratky, argot, sprostá a vulgární slova, cizí jazyk a neologismy. V pravopise zaznamenáme fonetickou transkripci, zkratky, vázání a vliv cizích jazyků (Roskam 2010, s. 13). Uveďme si nyní úryvek z novátorského románu, v němž má hovorový jazyk tak významnou roli:

Gabriel fit une tentative pour jeter un coup d'oeil sur les profondeurs.

– Merde, qu'il dit en se reculant, ça me fout le vertige.

Il s'épongea le front et embauma.

– Moi, qu'il ajoute, je redescends. Si vous en avez pas assez, je vous attends au rez-de-chaussée.

Il est parti avant que Zazie et Charles aient pu le retenir.

– Ça faisait bien vingt ans que j'y étais pas monté, dit Charles. J'en y ai pourtant conduit des gens.

Zazie s'en fout.

– Vous riez pas souvent, qu'elle lui dit. Quel âge que vous avez ?

– Quel âge que tu me donnes ?

– Bin, vzêtes pas jeune : trente ans.

– Et quinze de mieux.

²⁷ Queneau, Raymond. 'Écrit en 1937' *Bâtons, chiffres et lettres*. Paris : Gallimard, 1965. s. 12

²⁸ Queneau, Raymond. 'Écrit en 1955' *Bâtons, chiffres et lettres*. Paris : Gallimard, 1965. s. 66

- Bin alors vzavez pas l'air trop vieux. Et tonton Gabriel ?
- Trente-deux.
- Bin, lui, il paraît plus.
- Lui dis pas surtout, ça le ferait pleurer.

(Raymond Queneau: *Zazie dans le métro*. Paris : Gallimard, 1959. s. 113–114)

Román byl několikrát přeložen do češtiny a je pravděpodobně jedním z nezábavnějších a nejpřístupnějších textů tohoto významného autora francouzské moderní literatury.

3.6 Problematika transkripce textu

Jak zdůrazňuje Blanche-Benveniste, mluvený projev není možné studovat jen z ústního podání a spoléhat se čistě na paměť. Jedině vizuální reprezentace toho, co bylo řečeno, nám zaručuje plné pochopení smyslu a také možnost srovnávat různé úryvky mluveného textu. Dokud nebyl vynalezen diktafon, zaznamenávali pozorovatelé krátké útržky rozhovorů pouze písemně za použití zpřesňujících poznámek a fonetických symbolů. Přenosné diktafony znamenají téměř nekonečné možnosti. Specifické studie zabývající se fonetikou (výslovnost a prozódie) nutně využívají fonetických symbolů. Pro studium gramatických či diskurzivních fenoménů však není tento systém příliš pohodlný, protože existuje jen málo osob, které jsou schopny soustředěně číst kupříkladu dvacet stran textu v podobě fonetické transkripce.

Většina velkých korpusů mluveného jazyka je zaznamenána písemně a je doplněna dalšími údaji. Setkáváme se tak s množstvím potíží spojených s transpozicí mluveného slova do psané podoby. Jedním z problémů je zachycení některých prvků charakteristických pro fonickou stránku jazyka: hlasová modulace, rozdílné zabarvení a přízvuk, melodie a intenzita. Některé studie zaměřené na mluvený jazyk se soustředí na komunikační situace, konverzaci, na kontext interakcí, na předávání slova mezi mluvčími, doprovodnou gestikulaci a obecněji na samotný akt promluvy, který pro tyto účely bývá často doprovázen videonahrávkou. Ani grafická podoba prozaického

psaného textu příliš nepřeje realistickému ztvárnění mluveného slova. Váhání, úpravy, „přeřeky“, stejně jako překrývání slov (kdy mluví více lidí naráz) se v lineární podobě znázorňují poměrně obtížně. Stejně tak interpunkční znaménka jako je tečka, čárka či vykřičník nemohou dostatečně věrně tlumočit expresivitu spojenou například s ironií, otálením či nepřímou otázkou.

To jsou pouze některé z komplikací vznikajících při přepisu mluveného slova do písemné podoby tak, jak je uvádí Claire Blanche-Benveniste. V naší práci se však problematikou transkripce nebudeme zabývat podrobněji, neboť se s ní při analýzách filmových scénářů přímo nepotýkáme (Blanche-Benveniste 2010, s. 33–35).

4. Kvantitativní lingvistika

Následující čtvrtá kapitola bude z velké části vycházet z naší předchozí práce zaměřené na úvod do kvantitativní lingvistiky. Prostřednictvím kapitol publikovaných již v naší bakalářské diplomové práci si připomeneme počátky a strukturu této pomezí disciplíny, shrneme specifické termíny a pojmy a představíme si základní principy kvantitativní práce s jazykem. Zaměříme se také na využití kvantitativní lingvistiky (někdy také nazývané statistické lingvistiky) v praxi, kde doplníme původní text o nové zajímavosti. Tento vhled do problematiky nám později, v praktické části práce, pomůže lépe pochopit a interpretovat výsledky naší kvantitativní analýzy.

4.1 Počátky a struktura kvantitativní lingvistiky

S prvopočátky kvantitativních metod v jazykovědě bylo možné se setkat již ve starověku. O opravdu seriózních pokusech zavedení statistického hlediska do jazykovědného výzkumu ovšem hovoříme teprve od přelomu 19. a 20. století. Podívejme se nyní na to, jak se statistické výzkumy staly součástí lingvistických analýz a v jakých oblastech našly své uplatnění.

Jazyk slouží svým uživatelům, členům příslušného jazykového společenství, především k dorozumívání. Chceme-li dospět k plnému poznání zákonitostí jak v gramatické stavbě jazyka, tak v jeho slovní zásobě, je velmi užitečné seznámit se také s kvantitativními poměry v jazyce (Jelínek; Bečka; Těšitelová 1961, s. 9).

Při práci s texty si odborníci záhy uvědomili, že jazykové jednotky se vyznačují různou četností výskytu, tj. že jejich frekvence se od sebe vzájemně liší. Některé jevy tak v textu figurují v mizivém množství, zatímco jiné v hojném či přímo ve velmi vysokém počtu. Tento poznatek se netýká pouze slov, ale lze jej aplikovat na všechny roviny jazykových jednotek, tedy například také na slabiky, fonémy, hlásky, písmena, ale také na gramatické jevy jako časování, rody, pády, slovesné časy atd. V rámci kvantitativní lingvistiky rozlišujeme aplikaci kvantitativních metod, zejména statistiky, „podle jazykových plánů, lingvistických oblastí a pojmů, popř. oblastí aplikace vůbec.“ (Těšitelová 1977, s. 8). Podle toho, která část jazykové struktury je metodě podrobena, hovoříme podle Marie Těšitelové o 2 základních oblastech kvantitativní lingvistiky:

- lexikální statistika
- gramatická statistika (zde řadíme statistiku morfologickou a syntaktickou)

Do „ostatních oblastí“ kvantitativní lingvistiky patří například statistika:

- fonologická
- grafematická
- stylistická
- typologická
- glottochronologie a statistika týkající se vývoje jazyka

4.2 Lexikální statistika

Lexikální statistika (dále LS), kterou se zabývá i naše práce, je jednou z nejpropracovanějších a nejstarších disciplín kvantitativní lingvistiky. Cílem statistiky bývá „širší kvantifikace slovní zásoby jako celku, dále pak jejich složek, popř. lexikální stránky jednotlivých stylů, textů, autorů.“ (Těšitelová 1977, s. 52).

Vývoj LS probíhal ve dvou etapách. V první etapě, která trvala od začátku 20. století asi do 60. let, se o rozvoj disciplíny nejvíce zasloužili pedagogové, psychologové a těsnopisci, úloha lingvistů byla zatím jen druhotná. Pozornost se soustředila především na zjišťování zákonitostí ve slovníku jednotlivých jazyků. V tomto období vznikaly první práce LS. Druhá etapa začíná zhruba v 60. letech 20. století. Od první etapy se liší především tím, že o problematiku LS se nyní více zajímají lingvisté ve spojení s matematikou. Rozvíjí se také zájem o frekvenci jednotek menších než slovo. Velké změny nastávají i v oblasti zpracovávání dat pomocí moderních technických prostředků.

Hlavním úkolem lexikální statistiky je vytváření frekvenčních slovníků či frekvenčních seznamů slov. Na základě těchto číselných podkladů lze provádět další kvalitativní rozbor. Frekvenční seznamy a slovníky je tedy potřeba vnímat jako materiál, který má dopomoci k řešení problému, má být výchozím bodem pro určitou jazykovou hypotézu či jejím potvrzením.

Lexikální statistika pracuje se slovem, coby jednotkou souboru. Jako „slovo“ se tu jednak chápe tvar slova, jednak lexém. „V jazycích s bohatou morfologií, jako jsou jazyky slovanské, se tyto dvě „podoby slova“ v lexikální statistice celkem jasně diferencují.“ (Těšitelová 1974, s. 7). V jazycích s chudší morfologií nejsou naopak tyto dvě podoby mnohdy dostatečně rozdílné. Gustav Herdan v této souvislosti zavedl termíny *type* a *token*.²⁹ Marie Těšitelová vymezuje tuto problematiku třemi termíny: *slovo* (či *tvar slova*) pro jakoukoliv

²⁹ Výrazu *token* se užívá pro označení výskytu tvaru slova, pojem *type* poukazuje na přítomnost lexému, lexikální jednotky (Těšitelová 1974, s. 7).

jednotku textu, *různý tvar slova* (např. *svetr, svetrem* atd.) a *různé slovo* (odlišné lexémy). Při vymezování jednotky analýzy záleží vždy na typu jazyka a na účelu dané analýzy.

4.3 Gramatická statistika

Vedle statistiky lexikální, které je věnována relativně větší pozornost, řadíme do základní oblasti také statistiku gramatickou. S gramatickou problematikou se střetneme tehdy, máme-li vymezit jednotku pro lexikální statistiku. To se týká obzvláště roviny morfologické. U jazyků, jejichž morfologie není tak bohatá (např. u francouzštiny, angličtiny, němčiny), nepředstavuje stanovení jednotky souboru takový problém jako například u jazyků slovanských, které se vyznačují morfologií poměrně bohatou. Při analýze textu, tzn. lineárně uspořádaných jednotek, se setkáváme s problematikou syntaktickou. Na rozdíl od statistické práce s lexikem máme v gramatické analýze kvantitativního typu jiný cíl. Netřídíme již jednotky na různé kategorie a určení jejich frekvence, distribuce a uplatnění v textu pro nás není jediným výstupem. V gramatice jednotky (či kategorie) především hierarchizujeme a kvantifikujeme jejich vztahy, což nám umožní zjistit frekvenci a distribuci gramatických jevů. Tyto výsledky jsou klíčem k vytvoření modelů vysvětlujících vzájemnou závislost a podmíněnost gramatických jevů při jejich fungování v textu (Těšitelová 1980, s. 8).

Pomocí gramatické statistiky je možno matematicky postihnout frekvenci užívání různých morfologických či syntaktických jevů v textu, jako jsou rod, číslo, čas, pád u substantiv; poměr vět jednoduchých a složených, frekvence větných členů apod.

4.4 Ostatní oblasti

Ostatním oblastem kvantitativní lingvistiky jsme se podrobněji věnovali v naší předchozí práci. V této kapitole si tedy připomeneme pouze základní

informace týkající se jednotlivých odvětví, protože jak již bylo řečeno, statistická metoda není aplikována pouze na slovní zásobu či gramatiku, ale zasahuje také do jiných oblastí jazyka. Tak je možné setkat se se *statistikou fonologickou, grafematickou, stylistickou, typologickou* či s *glottochronologií*.

Fonologická statistika kvantifikuje v textu segmentální jevy, tzn. *fonémy*³⁰ i vyšší jednotky, vznikající jejich kombinací (slabika), a také jevy suprasegmentální, které můžeme charakterizovat jako zvukové prvky založené na prozodických prostředcích řeči, tzn. jedná se o slovní přízvuk a intonaci. Pomocí statistických metod zkoumáme rozložení fonémů v textech, kvantitativní poměry uvnitř fonologického systému, distribuční poměry uvnitř fonémů a tak dále (Těšitelová 1977, s. 102). Marie Těšitelová uvádí ve své příručce nejdůležitější práce z této oblasti. Zde zmíníme pouze S. Gsella, který pro francouzštinu zkoumal frekvenci fonémů v próze a poezii autorů 19. a 20. století.

V polovině 20. století se objevila grafematická statistika, která se stala středem zájmu především pro odborníky z oblasti teorie informace. Šlo o stanovení hodnot *entropie*, tzn. míry neurčitosti pokusu, který má více pravděpodobných možných výsledků, a *redundance*, která udává procento nadbytečných jednotek, pro ten který jazyk. „Entropii lze zjišťovat na úrovni grafémů, ale též fonémů, morfémů i slov.“ (Těšitelová 1977, s. 49). Hlavní význam redundance je v tom, že zabezpečuje spolehlivost sdělení. *Grafémem* rozumíme nejmenší jednotku psaného jazyka, která je již dále nedělitelná. Grafémy zahrnují jak písmena, znaky a piktogramy znakového písma, tak číslice a interpunkční a jiná znaménka. Výsledky grafematické statistiky se stejně jako u fonologické analýzy hodnotí současně běžnými koeficienty (hodnota frekvence) a koeficienty teorie informace (entropie).

Ke konci 20. století si našly statistické metody cestu i do stylistiky a poetiky. Pod pojem stylistické lingvistiky se řadí mimo jiné i problematika

³⁰ Fonémem rozumíme nejmenší součást zvukové stránky jazyka, která není nositelem významu, ale je schopna ho rozlišit; ve francouzštině například dvojice *père/mère*: [p] a [m] jsou dva fonémy, které jsou samy o sobě schopny rozlišit význam jinak stejných slov

studia verše. Jazykový styl je obecně chápán jako charakteristický způsob organizace jazykového projevu, ať už psaného či mluveného, který vychází od autora směrem k posluchači či čtenáři. Osobitý styl každého autora se pak zakládá na výběru slov a jazykových prostředků. Aplikace statistických metod nadále charakterizuje daný styl z hlediska kvantitativního. Tak je možné lépe odlišit jednotlivé funkční styly a určit, co je pro tvorbu toho kterého autora typické. Základem pro analytickou práci je volba charakteristiky a její následné hodnocení. Nejběžnější charakteristikou je slovo a jeho frekvence v textu. Tak vznikají slovníky či frekvenční seznamy slov zpracovávající díla světových autorů, které nám na základě zaznamenaného a frekvenčně zpracovaného lexika přibližují výjimečnost autorova literárního stylu. Další stylistickou charakteristikou může být například délka věty, počet vět v souvětí, poměr vět jednoduchých a složených, studuje se ale také slabika, grafémy atd. V poezii je dobrým základem pro analýzu rytmické uspořádání verše. Početně se zde analyzují versologické jevy jako rytmus, metrum, rým, asonance atd. Stylometrická metoda se využívá i při řešení tzv. sporného autorství. Principem je stanovení potřebných statistických charakteristik v textu, jehož autorství je sporné. Stylistické prvky textu jsou následně porovnány s jinými díly, jejichž autoři jsou známí a zároveň přicházejí v úvahu jako možní autoři textu sporného. Text je pak přisouzen tomu autorovi, jehož stylistika vykazuje nejbližší podobnost se sporným textem. Například již více než 150 let se neustále otevírá otázka, kdo je skutečným autorem her, které jsou připisovány Williamu Shakespearovi.

Použití statistických metod se ukázalo být užitečným také v typologických studiích. „Vedle uvádění základních statistických dat (absolutní a relativní frekvence jazykových jevů) setkáváme se ve statistické typologii např. s různými měrami typologické blízkosti a podobnosti jednotlivých jazyků.“ (Těšitelová 1977, s. 118). V analýzách se setkáme například s indexem syntézy, aglutinace, derivace a inflexe tak, jak je uvádí Joseph Greenberg (Greenberg 1960, 178–194), americký lingvista známý především svým bádáním v oblasti lingvistické typologie a genetické klasifikace jazyků.

Glottochronologie, neboli též lexikostatistika, je vyhraněná metoda historické lingvistiky, která se za pomoci statistických aplikací zabývá měřením příbuzenských vztahů mezi jazyky, a to zejména na základě změn ve slovní zásobě a studia genetické a vývojové charakteristiky jazyka. Glottochronologická metoda pracuje s tzv. indexem rychlosti mizení slov ze slovníkového jádra, procentem shodných dvojic slov ve slovníkovém jádru zkoumaných jazyků a s tzv. časovou hloubkou, „která stanoví, kdy nastala mezi jazyky taková diferenciacce, že lze mluvit o různých jazycích, i když příbuzných.“ (Těšitelová 1977, s. 121). Glottochronologií se od 50. let 20. století zabývali především dva američtí badatelé, Morris Swadesh a Robert B. Lees, kteří společně zavedli pojem *slovníkové jádro*. Za slovníkové jádro se obvykle považuje 100 slov, která mají častý výskyt a dostatečnou stabilitu. Problémem je pak výběr těchto slov, neboť samo jádro má tendenci se obměňovat, a to rychlostí, která není konstantní. K mutacím tedy dochází nepravidelně, což glottochronologický výzkum značně omezuje (Houser, 2004).

4.5 Využití kvantitativní lingvistiky v praxi

Po krátkém seznámení s konkrétními obory spadajícími do kvantitativní lingvistiky se vraťme k charakteristice kvantitativní lingvistiky obecně. Na základě již řečených informací o disciplíně si dovedeme představit, že využití výsledků z kvantitativních výzkumů je poměrně rozmanité.

Například zpracování slovní frekvenčnosti se stává východiskem pro mnoho návazných pozorování, úvah a vědeckých prací v jazykovědě i mimo ni. Podobné výzkumy mají velký pedagogický význam. Umožňují jasně stanovit jádro slovní zásoby toho kterého jazyka, a tak nasměrovat studenta studujícího cizí jazyk (ale samozřejmě také jazyk mateřský) k jazykovým základům, které by si měl přednostně osvojit.

Stejně cenným přínosem je i statistika gramatická. Její výsledky, týkající se frekvence a distribuce určitých gramatických jevů v jazyce, nám umožňují přesnější stanovení učební hierarchie. Např. z výsledků frekvence slovesných

časů pro český jazyk, které máme k dispozici, vyplývá, že nejvyšší frekvenci mají slovesa v přítomném a dokonavém tvaru, v didaktickém systému tvoří tedy tvary základní. Potom následuje imperativ, kondicionál přítomný a pasivní tvary. Přechodník, zvláště pak minulý, má frekvenci zanedbatelnou. Ve výuce začínáme nejběžnějšími, nejfrekventovanějšími, a tedy nejpotřebnějšími jevy. Postupně se přesouváme k obtížnější gramatice, na jejíž pochopení je žákům vyhrazen větší časový prostor než u jevů základních. Nemohou se totiž opřít o své jazykové zkušenosti a aktivní používání složitějších (či méně známých) slovesných časů vyžaduje dlouhodobější přípravu. V případě, že cílem výuky má být pouze pasivní znalost daného slovesného jevu (ve francouzštině např. *indicatif passé simple*), jsou časové dimenze, v rámci kterých lze látku zvládnout, jiné (Jelínek, Bečka, Těšitelová 1961, s. 97).

Hodnotné je i zjištění poměru slovních druhů v textu. Nehledě na to, že je jednoduše zajímavé zjistit jejich procentuální rozložení v textu (tak, jak jej předkládáme v praktické části naší práce), může být takový výzkum podnětným „zejména pro slohové vyučování, neboť zjištěné výsledky mohou být východiskem ke stanovení jazykových prostředků typických pro rozmanité druhy slohové.“ (Jelínek, Bečka, Těšitelová 1961, s. 97).

Co se týká grafematické statistiky, Jiří Černý ve svých *Dějínách lingvistiky* zmiňuje zajímavý fakt vztahující se k využití frekvenčnosti v praxi, když připomíná, že „např. na frekvenci písmen museli brát ohled tiskaři, kteří při své práci potřebují mnohem více exemplářů takových písmen, která mají v daném jazyce vysokou frekvenci, než písmen s frekvencí nízkou.“ (Černý 1996, s. 249). Dále se Černý zmiňuje o okolnostech vzniku Morseovy abecedy, kdy Samuel Morse přiřadil nejsložitější symboly písmenům s nízkou frekvencí a naopak nejjednodušší znak (tečku) přiřadil písmenu e, které má v angličtině (a i ve francouzštině) nejvyšší frekvenci.

4.5.1 Grafematická statistika a její souvislost s QWERTY

S grafematickou statistikou nepřímo souvisí i rozložení písmen na počítačové klávesnici. V článku publikovaném v roce 2006 na serveru Tchnet.cz, jehož autorem je Pavel Kasík, se dozvíme, že přemýšlení o tom, kam umístit které písmeno, započalo už v roce 1860, kdy se objevil první komerčně využívaný psací stroj. Nejednalo se tehdy o dnešní typ psacího stroje, ale o stroj, který díky svému kulovitému tvaru dostal název *Writing Ball*. „Člověk by přirozeně předpokládal, že na takto masově využívané periférii se projeví nějaké plánování. Že někdo napřed pečlivě rozváží, jaké klávesy budou používány často, jak bude psaní probíhat a jak optimalizovat počet přehmatů. Přesně taková analýza skutečně proběhla. Bohužel, její autor měl na mysli pravý opak toho, co bychom očekávali.“ (Kasík, 2006).

Tehdejší psací stroje se však potýkaly s problémem vycházejícím z jejich mechanické podstaty. Jednotlivá písmena se při „datlování“ nesměla potkat, jinak hrozilo zaseknutí mechanismu, s čímž jsme se běžně setkávali i u psacích strojů z 20. století. Jak píše Pavel Kasík, „v roce 1860 se editor Milwaukeekých novin Charles Sholes rozhodl sestavit takové rozložení kláves, které by zásekům bránilo. Jinými slovy, jeho cílem bylo uspořádání klávesnice, které by písaře při vyťukávání textu co nejvíce zpomalovalo.“ (Kasík, 2006). Takového efektu šlo dosáhnout pouze přeskupením písmen, která se původně nacházela na klávesnici v abecedním pořadí. A tak si Sholes zjistil, která písmena se v anglickém jazyce nejčastěji používají, a umístil je na klávesnici co nejdál od sebe. Tak vznikla klávesnice QWERTY pojmenovaná podle prvních písmen horní řady klávesnice. V roce 1873 prodal Sholes svůj nápad společnosti Remington a tato klávesnice se stala uznávaným standardem.

Následovaly adaptace spojené s jednotlivými jazyky. Francouzi vyměnili Q a W s písmeny A a Z. Také písmeno M se přesunulo o řádek výše. Tato klávesnice, známá pod pojmem AZERTY, působí komplikace například cizincům ve francouzských nebo belgických internetových kavárnách. Přesto, že moderní počítače nemají s psacími stroji téměř nic společného (minimálně odpadají ony mechanické problémy), podoba klávesnice zůstává nezměněna.

V roce 1931 proběhl pokus o vytvoření nové podoby rozložení klávesnice, kdy dr. August Dvorak navrhl klávesnici tak, „aby v základní řadě, na které spočívají prsty, byla nejčastěji používaná písmena. Bez přehmatávání lze napsat čtyři sta používaných slov, což je čtyřikrát více, než u QWERTY rozložení.“ (Kasík, 2006). Přesto, že je tato varianta jednoznačně efektivnějším řešením, náklady na přeškolení písařů se údajně ukázaly být natolik vysoké, že byly vyšší než potenciální úspora. Porovnejme drobné změny na francouzské klávesnici oproti klávesnici anglické.



Francouzská klávesnice



Anglická klávesnice

4.5.2 Lexikometrická analýza politických projevů

V naší bakalářské práci jsme uvedli výsledky lexikometrické analýzy zveřejněné na začátku roku 2011 na serveru Linternaute.com. Korpus byl sestaven z projevů, rozhovorů a prohlášení, které proběhly mezi 1. lednem 2000 a 30. červnem 2010, přičemž se výzkum týkal dvaceti politických

osobností. My jsme soustředili svou pozornost na osobnost Ségolène Royalové, členky Socialistické strany, která v roce 2007 kandidovala na funkci prezidenta republiky, a na Nicolase Sarkozyho, tehdejšího francouzského prezidenta. Nicolas Sarkozy byl rekordmanem v užívání první osoby – zájmena *je* a *moi*. Ségolène Royalová se stala šampionkou ve volbě výrazů z lexikálního pole rodiny jako je *mère*, *père*, *enfant* a *parents* (Abou El Khair; Deshayes, 2011).

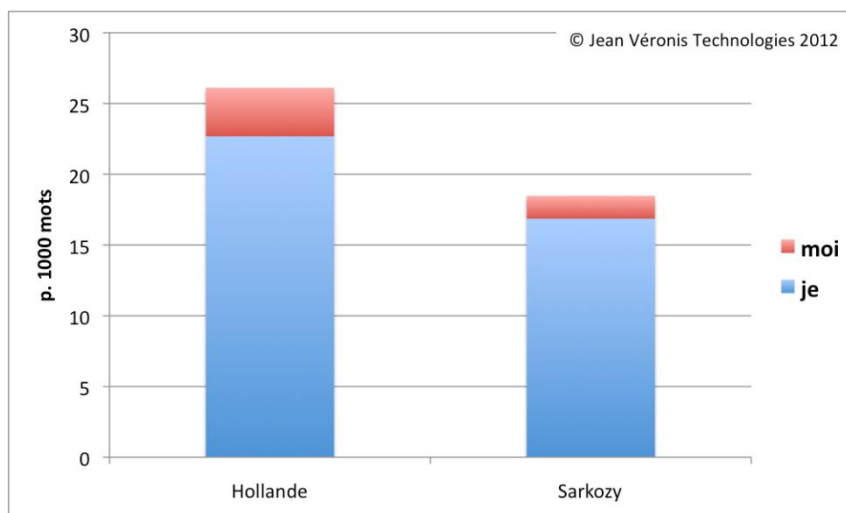
V rámci téže studie byly analyzovány také výroky současného prezidenta François Hollande. Z korpusu byla eliminována slova, která jsou považována za „neužitečná“. V zásadě to znamená, že byly vyřazeny dva typy slov. Za funkční či pouze pomocná slova byly označeny některé determinanty, předložky, příslovce či spojky jako *le*, *la*, *de*, *des*, *avant*, *après*, *pour*, *car*, *ensuite*, *très*, *trop* atd. Druhou vyřazenou skupinou jsou dvojsmyslná slova, jejichž smysl je obtížné přesně interpretovat (výrazy *politique*, *conclusion*, *forme*, *exercice* a další). Účelem bylo poskytnout údaje o slovním rejstříku a způsobu vyjadřování jednotlivých politiků. V případě François Hollande se jako výrazná projevila tendence častého užívání zájmena *nous*.



Poměrně často užívá také slov *gouvernement*, *gauche* a *ministre*. Jako stoupenec určitého ekonomického realismu používá významnou měrou také slova *crise* a *croissance*. V porovnání se svými politickými kolegy vyniká 7 krát častějším užitím slova *socialiste*, ale také *crédit* a *impôt* (5,7 a 5,3 krát častější), což potvrzuje, že daňový a peněžní systém je jeho oblíbeným tématem. Během analyzovaného období použil François Hollande také 5 krát častěji slovo *élections* (Abou El Khair, Deshayes, 2011).

V květnu 2012 uveřejnil Jean Véronis, profesor lingvistiky a informatiky na univerzitě v Aix-Marseille, na svém blogu studii s názvem *Débat: Moi, François Hollande*. Zde se soustředí na předvolební televizní duel mezi

Françoisem Hollandem a Nicolasem Sarkozym a všimá si četnosti používání zájmen u obou protagonistů. Viděli jsme, že Nicolas Sarkozy je králem *je* a *moi*. 2. května 2012, během televizní debaty, to však byl François Hollande, kdo převzal pomyslné žezlo. Zatímco Sarkozy a Royalová užívali toto zájmeno 17 a 18 krát za 1000 slov (v roce 2007), Hollande zvýšil laťku na 22,6. Tím však neporazil svůj vzor (jazykový a gestický), François Mitterranda, který se svými 24 *je* za 1000 slov obhájí prvenství dodnes (Veronis, 2012).



4.5.3 Vliv slovní frekvenčnosti na percepci

Problematika frekvenčnosti se dotýká také kognitivní psychologie zkoumající jazyk či psycholingvistiky neboli psychologie jazyka, která se soustředí na psychologické mechanismy, které stojí v pozadí užívání jazyka. Slovní frekvenčnost má totiž jasnou souvislost s percepcí slov v jazyce. V případě vizuální percepce, která je zkoumána pomocí *tachistoskopu*³¹, se práh percepce nachází v opačném poměru vůči frekvenčnosti každého slova. To znamená, že čím je frekvence slov vyšší, tím je percepční práh nižší, což v důsledku značí, že nejfrekventovanější slova budou rozpoznána rychleji než slova málo používaná. To dokládá, že naše lexikální znalosti v sobě zahrnují

³¹ Přístroj pro zkoumání vizuálního vnímání, pozornosti a reakčního času, který po velmi krátkou dobu umožňuje promítnout stimul v podobě např. obrázku, fotky, psaného slova atd. Pro měření prahu percepce je možné postupně navyšovat dobu prezentace stimulu do chvíle, kdy bude správně rozeznán subjektem. Požadovaný údaj přichází v řádu setin vteřiny, protože expoziční čas dostačující pro rozpoznání například slova je velmi krátký (jedná se o 40–100 milisekund). Měří se, jaké je nejmenší množství informace (vyjádřené v čase), které dovoluje subjektu správně identifikovat slovo.

také údaje o možnostech (a četnosti) užívání jednotlivých slov, což představuje způsob jisté redukce nejistoty.

V roce 1969 vytvořil John Morton, profesor na Institutu kognitivní neurovědy v Londýně, tzv. *logogenový model* zabývající se percepcí řeči. Jeho jednotky se nazývají *logogeny* a jsou popisovány jako jednotky vysvětlující mechanismus porozumění psané (vizuální logogen neboli psaná/zraková podoba informace) a mluvené podobě jazyka (auditivní logogen – zvuková podoba). Každá lexikální jednotka – logogen – má své fonologické, syntaktické a sémantické specifikace, které se určitým způsobem „zapisují“ do paměti subjektu. Velmi zjednodušeně se dá říct, že každé slovo má svou obdobu v logogenu, který si můžeme představit jako určité shromaždiště informací. Ve chvíli, kdy je subjektu představen stimul například v podobě slova a logogen je dostatečně nasycen informacemi, dochází k rozpoznání slova. Každý logogen má svůj aktivační práh. Aby mohlo dojít k percepci a dekódování stimulu (slova), musí subjekt získat o slově určité množství informací. V případě, že je slovo frekventované, logogen potřebuje k nalezení slova doplnit méně informací, protože aktivační práh je u známého slova nižší než u neznámého a vědomí informací o slově se tak udržuje v neustálé polo-aktivované permanenci. Tento proces se uplatňuje například u učení se novým slovům v cizím jazyce. Čím častěji slovo opakujeme, tím lépe si ho zapamatujeme – sníží se aktivační práh slova a my dříve dospějeme například k jeho sémantickému dekódování (Andrieux 1993, s. 14–17).

Běžný uživatel jazyka využívá tohoto mechanismu intuitivně, například když se snáze naučí to, čemu je každodenně vystavován nebo když v četbě u méně známých slov zpomaluje, zatímco u slov frekventovanějších čte rychleji, popřípadě i plynuleji. Bylo také vypořádáno, že ve frekventovanějších slovech se dělá méně chyb (při čtení či psaní).

4.5.4 Zkoumání bohatosti slovní zásoby: «la richesse lexicale»

Bohatost slovní zásoby je jedním ze společenských ukazatelů jazykových hodnot. Není tedy divu, když například literární umělecké dílo hodnotíme více či méně objektivně podle bohatosti slovní zásoby. Umělec, který oplývá květnatým jazykem a aktivně užívá mnoho slov, která mu dovolují barevněji či výstižněji popsat děj, vyprávět příběhy či vyjádřit pocity, je jistě zajímavějším autorem nežli spisovatel, který může být kvůli chudé slovní zásobě považován za člověka bez fantazie či talentu. Z pohledu *psychometrie*³² je rozvoj slovní zásoby svázán s rozvojem inteligence. V didaktice je rozvoj slovní zásoby také velmi důležitým prvkem. V jazyce si lze nejprve osvojit základy, které představují nejfrekventovanější slovní zásobu³³ a mohou posloužit k výuce druhého jazyka. Bohatost slovní zásoby představuje jeden z nejčastějších problémů lexikální statistiky. Jak se dá změřit bohatost slovní zásoby? Statisticy jako Yule nebo Herdan či lingvisté jako Guiraud, Muller, Dugast či Brunet vyřešili problém po svém. Jedni hledají index, který by odrážel rozmanitost slovní zásoby v textu, druzí navrhují spíše metodu srovnání, která by umožnila klasifikovat jedny texty ve vztahu k druhým a to na základě jejich lexikální různorodosti. Jedni vychází z teoretických modelů, druzí se opírají o empirické poznatky. Je to však právě statistické zkoumání jazyka, které za pomoci matematických postupů a vzorců umožňuje určovat bohatost slovní zásoby (Cossette 1994, s. 16).

Vidíme, že kvantitativní výzkumy přinášejí do lingvistiky, a nejen do ní, mnoho nových údajů, které neustále aktualizují poznatky o slovní zásobě i morfologii. Kromě toho vytváří materiál pro další bádání, obzvláště pro tzv. *aplikovanou lingvistiku*³⁴, pro počítačový překlad a tak dále. Na druhé straně je zřejmé, že výzkum kvantitativní stránky jazyka sám o sobě neobsáhne všechny informace o jazyce. Systém každého přirozeného jazyka je velmi složitý

³² Jedná se o vědu zabývající se měřením psychických jevů, jako jsou například osobnostní charakteristiky, znalosti či schopnosti a posuzování proměnlivost těchto jevů a jejich vzájemnými vztahy.

³³ Tak jako v případě *français fondamental*.

³⁴ Aplikovaná lingvistika využívá poznatků deskriptivní a teoretické lingvistiky. Mezi oblasti aplikované lingvistiky patří např. výuka cizích jazyků, jazyková terapie (logopedie), forenzní lingvistika, či manuální a strojový překlad.

a vyžaduje množství dalších charakteristik k tomu, aby mohl být popsán. Vezměme v úvahu jeho složku polysémickou, citovou a volitivní, hierarchii jednotlivých prvků, obsahovou a formální strukturu sdělení, funkci poetickou a podobně. Tyto jevy je nemožné popsat prostřednictvím statistických metod. Statistika má pouze pomocný charakter. Nicméně, jak víme z předchozích kapitol, kvantitativní lingvistika se neomezuje pouze na sestavování jednoduchých frekvenčních seznamů. Aplikuje také složitější metody, kde využívá například teorie informace ke zjišťování entropie a redundance jazykových jednotek (Černý 2008, s. 200–201).

4.6 Kritéria pro výběr materiálu k analýze

Výběr jazykového materiálu pro statistickou analýzu provádíme z hlediska *kvalitativního* a *kvantitativního*.

4.6.1 Kvalitativní hlediska

Výběr materiálu z kvalitativního hlediska se řídí především cílem, který při dané statistické analýze sledujeme. Frekvenční slovník pro potřeby školní výuky se bude svým výběrem materiálu lišit například od potřeb vědeckých. S ohledem na to můžeme vysledovat několik kritérií, podle nichž se tento výběr provádí. Marie Těšitelová uvádí v zásadě 5 kritérií:

- lingvistické
- psychologické, sociologické
- tematické
- sémiotické

Například lingvistické měřítko se uplatňuje tehdy, je-li cílem analýzy zkoumání jazyka vůbec, jazyka jednotlivých funkčních stylů (styl umělecký, publicistický, běžně dorozumivací) či jazyka jednotlivých autorů nebo děl. Při lexikální analýze jakéhokoliv jazyka jako celku nastává problém volby vhodného materiálu. Tradičním postupem je výběr textů z několika funkčních stylů, aby se

předešlo výskytu ojedinělých slov. Důležitý je poměr mezi styly, tedy to, jakou váhu budeme každému z nich připisovat. K dosažení objektivity při zkoumání jazyka je vhodné zaměřit pozornost také na formu jazykových projevů, nejen psaných, ale také mluvených, což je cílem této diplomové práce.

Zajímavou disciplínou je sledování distribuce zájmen v textu z hlediska sociologického. Roku 1969 vyšel v Paříži slovník televizních politických projevů bývalého francouzského prezidenta Charlese de Gaulla. V jeho tzv. provoláních (*les discours-appels*) bylo nejfrekventovanějším slovem zájmeno *je*, dále *vous* a *Français*, čemuž lze ze sociologického hlediska přiřkládat určitý význam. Je ovšem třeba si uvědomit, že vysoká frekvence zájmen ve francouzštině je dána specifickým vztahem osobních zájmen a slovesných tvarů.

Pro výběr materiálu z hlediska kvalitativního nejsou stanovená žádná přesná pravidla či zásady. Záležití především na názoru těch, kdo analýzu provádějí.

4.6.2 Kvantitativní hlediska

Dalším důležitým předpokladem pro jakoukoli statistickou analýzu je výběr materiálu z hlediska kvantitativního. Spolu s výběrem z hlediska kvalitativního rozhoduje o výsledku statistických analýz. Kvantitativním hlediskem rozumíme výběr určitého množství textu. Jedná se většinou o vzorky z různých jazykových či literárních stylů, děl jednotlivých autorů atd., které vytváří materiál o určitém rozsahu (*korpus*), o němž se při analýze opíráme. Jak jsme již zmínili, jedná se o „vzorek“, tzn., že při kvantitativní analýze nepracujeme s celými texty, ale pouze s jejich částmi.

Rozsahem korpusu rozumíme celkový počet jednotek, s nimiž pracujeme. Neexistuje číslo, které by vymezovalo, kolik jednotek je třeba analýze podrobit. Například v lexikální analýze není stanoveno, kolik tisíc slov musíme analyzovat, abychom dospěli k objektivnímu konstatování, že zjištěná frekvence slov platí pro zkoumaný jazyk obecně nebo pro některou ze

speciálních oblastí jazyka. Otázkami rozsahu výběru korpusu se zabývala např. R. M. Frumkinová. Došla k závěru, že aby mohla být frekvence slov stanovena s co největší přesností, je třeba stanovit „relativní chybu výsledků měření (frekvence)“. K výpočtu této „relativní chyby“, označované jako delta, slouží speciální matematická formule (Těšitelová 1977, s. 27).

Co se týká způsobu výběru materiálu, existuje více postupů. Uvádí se výběr náhodným a mechanickým způsobem, dále výběr oblastní atd. Při náhodném výběru je třeba zajistit, aby každá jednotka měla stejnou pravděpodobnost dostat se do výběru. Četné výzkumy prokázaly, že relativně stejně dlouhým textům odpovídá různě dlouhý slovník (počet různých slov v textu) a „že jsou výrazné rozdíly jak mezi texty jednotlivých stylů, tak i mezi jednotlivými texty.“ (Těšitelová 1977, s. 30).

Obecně je pro provedení kvantitativního rozboru vhodný co největší korpus. Čím je materiální základna rozsáhlejší a všestrannější, tím je větší pravděpodobnost, že výsledky budou odpovídat objektivní skutečnosti.

4.7 Vyhodnocování výsledků analýzy

Nyní si představíme některé pojmy, které ke kvantitativní lingvistice neodmyslitelně patří, neboť právě díky nim jsme schopni výsledky vyhodnocovat, popisovat a třídit.

Výstupy z kvantitativní analýzy jsou představovány prostřednictvím kvantitativních ukazatelů, charakteristik, koeficientů, indexů, matematických vzorců, atd. Jejich charakter je většinou relativní a je třeba podrobit výsledky dalšímu ověřování a srovnávání, což může vést k navrhování nových analytických postupů ve snaze minimalizovat chyby.

4.7.1 Frekvence, pořadí, rank

Jedním ze základních výsledků je zjištění *frekvence*, tzn. četnosti jednotek, které bývají uspořádány ve formě frekvenčního seznamu a seřazeny zpravidla podle klesající frekvence. Jako příklad uveďme seznam francouzských slov, který byl sestaven lexikologem Étienne Brunetem a zpracovává celkem 1500 slov, jež se ve francouzštině vyskytují nejčastěji. Začátek frekvenčního seznamu slov uspořádaných podle klesající tendence vypadá takto:

1.	<i>le</i>	1050561
2.	<i>de</i>	862100
3.	<i>un</i>	419564
4.	<i>être</i>	351960
5.	<i>et</i>	362093
6.	<i>à</i>	293083
7.	<i>il</i>	270395
8.	<i>avoir</i>	248488
9.	<i>ne</i>	186755
10.	<i>je</i>	184186

Každou jednotku (zde *lexém*) ve frekvenčním seznamu provází údaj o *absolutní frekvenci*, tj. hodnota popisující počet výskytů všech tvarů daného slova v korpusu. Ve *Frekvenčním slovníku češtiny* z roku 2004 můžeme tuto hodnotu najít pod zkratkovým označením *FRQ*. Každý lexém je také označen hodnotou udávající pořadí dané jednotky. Jedná se v zásadě o průběžné očíslování za sebou následujících jednotek seřazených podle klesající frekvence. S klesající frekvencí přibývá jednotek (např. slov, grafémů atd.), které mají stejnou frekvenci. Jde o empirický fakt, se kterým je třeba pracovat. K jednotkám, jejichž frekvence se neliší, přiřazujeme tzv. *rank*. Jedná se o číslo, které odpovídá jen jednomu údaji o frekvenci. Marie Těšitelová (1977, s. 38) vysvětluje tento jev na příkladu z Čapkovy knihy *Život a dílo skladatele Foltýna*. Uvádí úryvek z frekvenčního seznamu, který nám objasňuje rozdíl mezi rankem a pořadím:

pořadí	slovo	rank	frekv.
50.	<i>umělec</i>	50	58
51.	<i>chtíti</i>	51	57
52.	<i>hráti</i>	52	56
53.	<i>u</i>	53	55
54.	<i>až</i>	54	54
55.	<i>nějaký</i>		54
56.	<i>od(e)</i>		54
57.	<i>všechn</i>		54
58.	<i>říkati</i>	55	53
59.	<i>vel(i)ký</i>		53
atd.			

Vidíme, že do ranku 53 jsou hodnoty pro rank a pořadí totožné, protože slova mají různou frekvenci. Do ranku 54 patří již 4 slova se stejnou frekvencí, která jsou řazena abecedně. Jak připomíná Jiří Černý „rozdíl mezi rankem a pořadím se bude stále zvětšovat a bude velmi výrazný na konci seznamu, kde bývá v jednom ranku velké množství slov se stejnou (nízkou) frekvencí.“ (Černý 1996, s. 255).

Mezi slovem a jeho frekvencí existuje mnoho vazeb. Pierre Guiraud ve své práci *Les Caractères statistiques du vocabulaire* uvádí, že existuje například vztah mezi rankem a přirozeností fonémů. Francouzština je totiž oproti ostatním jazykům charakteristická lehkou predominancí labiálních souhlásek³⁵, jejich relativní frekvence je tedy neobyčejně zvýšená v nejužívanějších slovech a poté progresivně klesá. Další vazbu nalezneme mezi rankem a etymologií. Nejčastěji užívaná slova jsou zároveň slovy, která se v jazyce vyskytují nejdéle (*les mots anciens*) a slova lidová. Guiraud tedy shrnuje problematiku tak, že slovní frekvence má souvislost s fonickým, morfologickým, sémantickým a etymologickým charakterem slov. Je odrazem

³⁵ Jedná se o souhlásky, jejichž výslovnost je realizována za pomoci rtů. Odlišujeme souhlásky bilabiální (např. „m“, „p“, „b“) a labio-dentální („f“, „v“).

jazyka a jeho proměn. Když například se nějaké slovo stane módním či se ve společnosti objeví nový problém či fenomén, okamžitě se to odrazí také na frekvenčnosti daného slova či celého lexikálního pole zabývajícího se onou problematikou (Guiraud 1954, str. 4).

4.7.2 Zipfovy zákony

Problematiky vztahu mezi absolutní frekvencí slov a jejich rankem se týkají tzv. *Zipfovy zákony*. Jejich formulaci vytvořil George Kingley Zipf (čtete /cipf/), americký lingvista německého původu a profesor harvardské univerzity, který se zabýval studiem statistických náhod v různých jazycích. Na přelomu 20. a 30. let se věnoval zkoumání relativní frekvence hlásek, přičemž dospěl k několika pozoruhodným zjištěním.³⁶ Zajímal se také o psychologické a fyziologické vlivy, které působí na mluvčího, a ovlivňují jeho řečový projev. Hlavní zásadou je podle Zipfa tendence k co nejmenší námaze a tudíž k jazykové ekonomii v projevu. Svou teorii nazval *psychobiologií* a navrhl vyčlenit v lingvistice speciální disciplínu – tzv. *biolingvistiku*, která by zkoumala jazykové jevy z hlediska „sdělovacího chování člověka“. Tyto své názory formuloval především v díle *Psychologie jazyka (The Psychology of Language. An Introduction to Human Ecology, 1949)*. (Černý 1996, s. 254).

Největší zásluha se však Zipfovi přisuzuje za jeho bádání v padesátých letech 20. století, kdy se snažil odhalit a matematicky zformulovat zákonitosti slovní zásoby přirozených jazyků. Výsledky jeho práce v této oblasti jsou známy pod názvem Zipfovy zákony. „Zipf vychází z hypotézy, že v jazyce existují tendence udržet rovnováhu mezi frekvencí slova a počtem slov, která tuto frekvenci mají.“ (Těšitelová 1977, s. 39). Jedná se v zásadě o dvě síly, *sjednocující s rozlišující*.³⁷ Jedna má tendenci produkovat slova s co největší frekvencí a snížit tak jejich množství. Proti ní působí druhá síla, která se snaží

³⁶ Tato zjištění přehledně shrnuje Jiří Černý takto: „a) hlásky a jejich třídy vystupují v různých textech téhož jazyka se stejnou frekvencí; b) ve všech jazycích vystupují neznělé hlásky asi dvakrát častěji než znělé; c) čím obtížnější je hláska z hlediska artikulace, tím nižší má frekvenci atd.“ (Černý 1996, s. 253).

³⁷ Marie Těšitelová takto překládá termín *forces of unification and diversification* pocházející ze Zipfova díla *Human Behaviour and the Principle of Least Effort* (Cambridge, 1949).

ovlivnit tvorbu jazykových promluv směrem k co největší rozmanitosti ve slovní zásobě. Vzniká tedy velké množství slov, která mají poměrně nízkou frekvenci. „Příčinou první síly je tzv. ekonomie mluvčího, příčinou druhé tzv. ekonomie posluchače, protože snahou obou je, aby svou činnost (mluvení a slyšení) prováděli při co nejmenší námaze.“ (Těšitelová 1977, s. 39).

První Zipfův zákon se týká vztahu mezi rankem r (uspořádání slov od nejvyšší frekvence k nejnižší) a frekvencí f . Součinem obou veličin získáme konstantu k . První Zipfův zákon tedy zní následovně:

$$r \cdot f = k$$

r znamená rank slova, který je zde používán bez přihlídnutí k počtu slov s touž frekvencí. Dle naší výše zmíněné terminologie se tedy jedná spíše o pořadí

f jedná se o absolutní frekvenci slova

k je konstanta

Jinými slovy: vzorec popisuje vztah slova mezi frekvencí slova a jeho rankem, který se vyznačuje nepřímo úměrnou vzájemnou závislostí obou veličin. Čím je rank slova nižší, tím má slovo vyšší frekvenci a naopak. Zipf s úspěchem ověřoval svůj zákon na několika textech, zejména na románu Jamese Joyce *Odysseus*, a také na několika různých jazycích. Při výzkumu některých děl české literatury však přišla Marie Těšitelová na to, že Zipfovy zákony jsou víceméně spolehlivé pouze ve střední části frekvenčního slovníku. Tento zákon byl všeobecně podroben velké kritice řady lingvistů a ukázalo se, že pravdivost těchto vzorců je třeba i nadále ověřovat. Velkým kritikem oněch vzorců byl francouzský matematik polského původu Benoît B. Mandelbrot. Připustil, že Zipfovy zákony mají svůj smysl, avšak zákonitosti mezi rankem a frekvencí slov jsou schopny popisovat pouze v hrubých rysech bez schopnosti postihnout detaily. Sám pak zformuloval tzv. *harmonický* a tzv. *kanonický* zákon, které mají uvedený vztah charakterizovat výstižněji (Těšitelová 1977, s. 39).

Předchůdcem tohoto prvního zákona byl francouzský matematik a stenograf Jean-Baptiste Estoup, který také formuloval zákonitosti při uspořádání slov podle klesající frekvence, a to v podstatě stejným způsobem ($f \cdot r = k$), avšak mnohem dříve než Zipf. Z toho důvodu se někdy hovoří o *zákonu Estoupově – Zipfově*.

Druhý Zipfův zákon se týká poměru mezi frekvencí slova a počtem různých slov, která tuto frekvenci mají. Zákon vyjadřuje fakt, že čím je frekvence nižší, tím více slov tuto frekvenci má:

$$a \cdot b^2 = k$$

- a počet slov s jistou frekvencí
- b frekvence
- k konstanta

Zipf tomuto druhému zákonu rovněž nepřipisuje platnost obecnou, ale omezuje ho pouze na slova s nepřilíživou velkou frekvencí, která pokrývají značnou část slovníku. I z tohoto zákona slova s vysokou frekvencí vylučuje. Na základě výzkumů lze tomuto zákonu vytknout jeho opomíjení rozsahu textu. Zákon není funkční pro text o libovolné délce

V posledním, třetím zákonu, se Zipf pokusil matematicky vyjádřit vztah mezi frekvencí slova a počtem jeho významů:

$$\frac{m}{\sqrt{f}} = k$$

- m počet významů daného slova
- f frekvence daného slova
- k konstanta

Podle tohoto zákona je počet různých významů (tzv. *polysémie*) větší u slov s vyšší frekvencí, což dle výzkumů Marie Těšitelové, která se Zipfovými zákony podrobně zabývala, platí především pro formální slova (tj. předložky), nikoli například pro česká substantiva, kde počet významů není na frekvenci slova nijak závislý a s klesající frekvencí počet jejich významů jen zvolna klesá.

Také Zipf se při formulování tohoto zákona soustředil především na slova formální. Vycházel z předpokladu, že délka slov je nepřímo úměrná jejich frekvenci a tedy že nejfrekventovanějšími výrazy jsou zároveň slova nejkratší, což bylo ostatně v mnoha výzkumech potvrzeno. Zmíněnými nejkratšími slovy bývají právě slova formální jako předložky, spojky a tak dále. Tato zjištění mají viditelnou souvislost s principem ekonomie, o kterém jsme se zmínili výše. Ekonomický princip se projevuje také v jazykovém vývoji, a to prostřednictvím například „apokopického“ krácení slov (ve francouzštině např. *métro(politain)*, *apéro (apéritif)*, *mécano (mécanicien)*, *fac (faculté)* atd.), kdy slova kratší mají jednoznačně častější frekvenci než slova dlouhá (Černý 1996, s. 256).

4.7.3 Průměrná redukovaná frekvence

Vedle základních údajů, jako jsou frekvence, rank a pořadí slova, mohou jednotku určovat i další kvantitativní charakteristiky, jako jsou údaje o průměrné redukované frekvenci, údaje o počtu stylových skupin či počtu textů tak, jak byly využity například v českém frekvenčním slovníku z roku 1961³⁸ či ve *Frekvenčním slovníku češtiny* z roku 2004.

Průměrná redukovaná frekvence, na rozdíl od klasické absolutní frekvence, bere v úvahu rozložení výskytu slova v celém korpusu, tzn., že zohledňuje počet pramenů, ve kterých se dané slovo vyskytuje. Tak zamezuje tomu, že nějaké slovo bude označeno za obecně velmi frekventované bez ohledu na to, že se často vyskytuje například pouze v jednom z textů nebo žánrů tvořících součást korpusu.

³⁸ Jelínek, Jaroslav, Bečka, Josef Václav, Těšitelová, Marie. *Frekvence slov, slovních druhů a tvarů v českém jazyce*, Praha : Státní pedagogické nakladatelství, 1961. 585 s.

4.7.4 Koeficient disperze

Koeficient *disperze* (rozptýlení), stejně jako průměrná redukovaná frekvence, charakterizuje rozptýlenost frekvence jednotky (slova, grafému, atd.) v jednotlivých textech či žánrech korpusu. Jedná se o jeden z koeficientů, který ve svých frekvenčních slovnících zavedl Alphonse Juilland, švýcarský lingvista. Je zřejmé, že údaje o frekvenci nelze chápat pouze mechanicky. „Hodnota frekvence slova je tedy dána nejen absolutní frekvencí, nýbrž i počtem knih a skupin, v nichž je slovo doloženo.“ (Jelínek, Bečka, Těšitelová 1961, s. 30).

Uvedené pojmy a vzorce, spolu s pojmy entropie a redundance, které jsme zmínili výše, považujeme za základní terminologii vztahující se k hodnocení výsledků kvantitativní analýzy. Je však třeba si uvědomit, že jednotlivé oblasti kvantitativní lingvistiky mohou pro svou charakteristiku vyžadovat další specifické koeficienty a výpočty.

4.8 Frekvenční slovníky

Jak již bylo řečeno v našem výkladu, jedním z hlavních cílů kvantitativní lingvistiky je tvorba lexikografická, tzn. vytváření slovníků a frekvenčních seznamů. Statistické výzkumy, jak víme, se uplatňují v různých oblastech jazyka: od jeho jednotek grafematických, fonologických a lexikálních po zkoumání stylistiky, typologie a měření příbuznosti mezi jazyky. Dosud nejčastěji používaná je statistika lexikální, zabývající se kvantifikací slovní zásoby. Soustředí se buďto na daný jazyk jako celek, nebo na určitý text, dílo či autora. Vznikají jednak jednoúčelové seznamy zpracované na základě poměrně málo rozsáhlého korpusu, jednak frekvenční slovníky (dále jen FS), které využívají ke statistikám rozsáhlého a reprezentativního materiálu. Takové slovníky často obsahují kromě frekvenčního seznamu i seznam abecední a také mnohé další údaje potřebné pro pokud možno co nejpřesnější stanovení frekvenčnosti. FS představují významný a zajímavý zdroj informací nejen pro lingvisty, ale také pro sociology, psychology, matematiky, pedagogy atd.

Jednotlivé slovníky se liší:

- rozsahem analyzovaného materiálu
- stanovením slovní jednotky, s níž se bude v textu pracovat³⁹
- výběrem textů k rozboru (v otázce žánrů či stylů)
- technikou zpracování (ruční či strojové)

Pro některé jazyky existuje už v dnešní době několik různých FS, většina jazyků však stále nemá ani jeden. Představme si nyní některá nejdůležitější díla z této disciplíny, a to s důrazem na práce české a na slovníky pocházející z frankofonních oblastí. Není pochopitelně možné opomenout ani práce z jiných zemí, které sehrály ve vývoji lexikální statistiky významnou roli (Černý 1996, s. 257).

4.8.1 Základní přehled týkající se lexikální statistiky

Jedním z prvních, kteří se statistice ve slovní zásobě věnovali, byl německý stenograf F. W. Käding. Ten v letech 1897 – 98 publikoval první FS vůbec. Jednalo se o *Slovník četnosti výskytu německého jazyka (Häufigkeitswörterbuch der deutschen Sprache, 1897)*. Na základě rozsáhlého materiálu (cca 11 000 000 slov), získaného zejména z právnických a obchodních spisů, byl schopen formulovat zajímavá tvrzení týkající se především slov s vysokou frekvencí. Podle něj totiž „prvních 15 nejfrekventovanějších výrazů představuje asi čtvrtinu celého slovního materiálu běžného německého textu, prvních 66 slov s největší četností výskytu pokrývá již celou polovinu textu a prvních 320 výrazů již téměř tři čtvrtiny textu.“ (Černý 1996, s. 249). Ke zvládnutí tří čtvrtin běžného německého textu by nám tedy mělo postačit naučit se prvních 320 nejfrekventovanějších slov. Kromě jednotek vyšších uvádí také frekvenci jednotek nižších, tedy předpon, přípon, kmenů,

³⁹ Např. autoři *Frekvenčního slovníku češtiny* „považují složený tvar slovesný typu „byl bych šel“ za jeden slovní tvar a uvádějí jej spolu s jinými tvary pod heslem „jít“. Jejich slovenští kolegové ovšem chápou slovo důsledně jako jednotku grafickou a proto by tento tvar považovali za 3 různé jednotky.“ (Černý 1996, s. 260)

slabik, hlásek a písmen. Kädingův slovník se stal východiskem a inspirací pro většinu následně vznikajících německých prací.

Hodnotná díla vznikala také v Anglii. K nejstarším FS patří dílo L.P. Ayrese. Slovník z roku 1915, založený na 368 000 slov sestavených z dopisů obchodních a soukromých, je zpracován monograficky, což je v dnešní době již ojedinělé s ohledem na velké množství analyzovaného textu. Nejznámějšími FS určenými pro vyučování v angličtině jsou slovníky Edwarda Lee Thorndika, které postupně vyšly ve třech verzích. Jádro zde tvoří 10 000 nejfrekventovanějších slov, která byla získána z korpusu o rozsahu 4 565 000 slov. Korpus byl sestaven z textů literatury pro mládež, z učebnic, příruček atd. Uvedený materiál byl později rozšířen o dalších pět miliónu slov, na jejichž základě pak vznikla druhá verze slovníku. V obou verzích byla frekvence slov ohodnocena tzv. kreditovými čísly.⁴⁰ Ve třetí verzi od tohoto systému upustil. Také Thorndikovy slovníky se staly zdrojem pro další lingvistické práce.

Co se týká lexikálně statistických prací ve francouzštině, prvním dílem, které se zabývalo frekvencí slov, byl *Le Vocabulaire d'un journal* sestavený autorem jménem Bony v roce 1920–1921. Jednalo se o statistické zpracování jednoho čísla novin *Le Temps*. Pokus o sestavení frekvenčního slovníku učinil V.A.C. Henmon. Vyhotovil frekvenční seznam o 4 000 slovech na základě 400 000 slov. Tento seznam byl roku 1930 rozšířen a přepracován G. E. Vander Bekem. Tento slovník s názvem *French Word Book* byl určen pro didaktické účely, zejména pro vyučování francouzštiny jako cizího jazyka. Na slovníku Vander Bekově založil např. J. P. Tharp svůj *The Basic French Vocabulary* (New York, 1939), obsahující 3 340 hesel. Věnuje zde pozornost nejen frekvenci, ale také druhům slov apod. Roku 1970 vydal Alphonse Juilland frekvenční slovník pro současnou francouzštinu. Po španělském (1964) a rumunském (1965) slovníku se jednalo o další dílo v sérii frekvenčních slovníků zpracovávajících románské jazyky v jejich současné podobě. V přípravě byl také slovník portugalský a italský. Všechny jeho slovníky jsou

⁴⁰ Jedná se o způsob hodnocení statistických údajů užívaný E.L. Thorndikem, ve kterém přihlíží k distribuci frekvence slov napříč celým materiálem i jeho součástmi (skupinami). Kreditové číslo uvádí zároveň frekvenci i distribuci slova. Bližší informace je možno získat v knize Marie Těšitelové *Kvantitativní lingvistika* (1977 s. 46-47).

sestaveny podle stejných parametrů. Korpus tvoří 500 000 slov vybraných z pěti žánrů (dramata, umělecká próza, eseje, technická literatura a periodika), přičemž texty jsou zpracovány strojově. Ve slovníku nalezneme jak seznam abecední, který je složen z 5 082 slov s koeficientem užití větším nebo rovným 3, tak seznam frekvenční, v němž jsou slova seřazena podle klesajícího koeficientu užití, frekvence a disperze. Juillandovy slovníky jsou užitečné také z hlediska teoretického pohledu na lexikální statistiku, který přináší, a zavedení koeficientů disperze a užití slova. O studium frekvence se opírá i tzv. *Français élémentaire*, která má cizincům usnadnit učení základů francouzštiny. Práce se zaměřuje na frekvenčnost slov v mluveném projevu, čímž se odlišuje od dosavadních slovníků a seznamů slov založených většinou na analýzách projevů psaných. Na mluvené projevy se orientuje také *Dictionnaire de fréquence des mots du français parlé au Québec : fréquence, dispersion, usage, écart réduit*. (New York, 1992), jehož autory jsou N. Beauchemin, P. Margel a M. Théoret. Za užitečnou považujeme poměrně nedávno vzniklou bilingvní práci s názvem *A Frequency Dictionary of French: Core Vocabulary for Learners* (Kanada, USA, 2009). Autory jsou Deryle Lonsdale a Yvon Le Bras. Uvádějí 5 000 nejfrekventovanějších francouzských slov s anglickým překladem a vzorovou větou. Kromě toho obsahuje slovník ještě 27 tematicky zaměřených tabulek shrnujících ta nejfrekventovanější slova k danému tématu.

Snad největších současným vydavatelem frekvenčních slovníků je britské nakladatelství Routledge, které mezi lety 2005 a 2013 vydalo sérii celkem 11 frekvenčních slovníků. Mezi publikacemi najdeme frekvenční slovníky mapující španělštinu (2005), němčinu (2005), portugalštinu (2007), mandarínskou čínštinu (2009), francouzštinu (2009), současnou americkou angličtinu (2010), češtinu (2011), arabštinu (2011), japonštinu (2013), ruštinu (2013) a dánštinu (2013). Slovníky jsou vystavěny na korpusech kombinujících mluvený jazyk a fikční a non-fikční literaturu, přičemž rozsah zkoumaného materiálu se pohybuje od 4,2 miliónů (němčina), přes 20 miliónů (francouzština) po 385 miliónů slov (současná americká angličtina). Podrobnější informace

o jednotlivých slovnících a jména jejich autorů lze vyhledat na oficiálních stránkách nakladatelství.⁴¹

Z prací zabývajících se lexikální statistikou českého jazyka se zastavme u již zmíněného frekvenčního slovníku češtiny, jehož autory jsou J. Jelínek, J. V. Bečka a M. Těšitelová.⁴² Tento slovník, určený původně k didaktickým účelům, se opírá o 1 623 527 slov získaných ze 75 textů. Je uvedena nejen frekvence slov, ale kvantifikována je také četnost výskytu některých gramatických kategorií (slovních druhů, pádů a rodů substantiv atd.). *Frekvenční slovník češtiny* z roku 2004 vychází z korpusu, který při svém rozsahu 100 milionů slov zajišťuje vysokou spolehlivost předkládaných dat týkajících se psané češtiny. Slovní materiál byl zpracován automatickými metodami, po kterých však následovaly rozsáhlé manuální úpravy. Korpus je založen na proporcčně vyváženém výběru textů ze všech žánrů psané češtiny. Jako nedostatek uvádí nepřítomnost autentických mluvených textů. Mluvený jazyk zde byl zastoupen redigovanými projevy politiků. Současná podoba beletrie se však také často blíží mluvené češtině. Vůbec prvním slovníkem svého druhu představujícím autentickou mluvenou češtinu je však *Frekvenční slovník mluvené češtiny* z roku 2007, jehož autorem je František Čermák. Slovník vychází z Pražského mluveného korpusu, založeného na sociolingvisticky reprezentativních nahrávkách rozhovorů.

Obecně lze říci, že první významnější frekvenční slovníky začaly vycházet až před druhou světovou válkou a těsně po ní. Jednalo se o několik desítek různě pojatých slovníků pro němčinu, angličtinu, francouzštinu a další jazyky. Tyto texty se vyznačovaly pracným manuálním zpracováním psaných pramenů a rozsahem korpusu, který se pohyboval průměrně okolo jednoho milionu slov. Proto je Kädingův slovník z roku 1897 tolik ceněn. Se svou materiálovou základnou okolo 10 milionů slov byl a dodnes je ojedinělý. Jak uvádí František Čermák v úvodu k *Frekvenčnímu slovníku češtiny* (2004, s. 7), „výsledky a poznatky takto dosažené se ukázaly jako užitečné mj. už za 2. světové války, při tvorbě Basic English, koncepce zjednodušené a rychle

⁴¹ Viz web: http://www.routledge.com/books/series/RFD/page_1/.

⁴² *Frekvence slov, slovních druhů a tvarů v českém jazyce*, Praha 1961 (zkr. FSC)

naučitelné angličtiny pro zahraniční letce, která se později stala modelem i pro jiné jazyky.“ Stejně jako ostatní slovníky, měly by být i FS průběžně aktualizovány a obohacovány. V dnešní době je tento proces značně usnadněn prostřednictvím počítačového zpracování dat, který odstraňuje úmornou manuální dřinu a zajišťuje větší spolehlivost. Většina moderních slovníků je dnes také opatřena CD, která umožňují pohodlné prohlížení adresáře v elektronické podobě či jeho prohledávání podle uživatelem zadaných kritérií.

5. Kvantitativní lingvistika a mluvený projev ve francouzském jazyce

Jelikož naším záměrem je aplikace výše uvedené teorie na mluvenou francouzštinu, představme si nyní korpusy mluveného jazyka, metodu jeho vypracování a práce a výstupy vztahující se ke statistické analýze mluvené formy jazyka, neboť ta kvůli své specifické povaze vyžaduje jiný přístup kvantitativních metod, než jazyk psaný.

5.1 Korpusy mluveného jazyka

Vedle práce s psaným jazykem se pozornost korpusové lingvistiky soustředí také na práci s mluvenou formou jazyka. Tyto korpusy mnohdy obsahují autentické nahrávky skutečného mluveného jazyka bez normalizace a kodifikace, a tak představují vlastnosti a rysy každodenní mluvy bez příkras a filtru spisovnosti. Navíc takový korpus či následně vzniklý slovník mluveného jazyka naznačuje leccos o tom, kde se dá daný jazyk zjednodušit (protože ve své mluvené podobě už se dávno zjednodušil), aniž by utrpěla jeho funkčnost. Na základě takového korpusu můžeme zkoumat například morfologii jazyka tak, jak to učinila práce Jitky Šonkové s názvem *Morfologie mluvené češtiny: Frekvenční analýza* vydaná roku 2008. Jak píše František Čermák v předloze k této publikaci, „práce vyčerpávajícím způsobem pokrývá v uceleném pohledu tvarosloví mluvené češtiny, které Pražský mluvený korpus odráží, a popisuje ho v množství grafů a tabulek, v nichž jsou vidět proporce jevů uvnitř širší třídy (což je sám o sobě jev u nás v gramatické tradici nevídaný). Uživatel tak mj. snadno

dospívá k poznání, kde je centrum a těžiště dané kategorie, kde periférie atd. Z obecného hlediska je tuto práci třeba chápat také jako první ucelený, na korpusu založený a vyčerpávající popis celé jedné jazykové roviny vůbec.“ (Šonková 2008, s. 5). Využití korpusů je téměř neomezené a je využitelné v teoretické a počítačové lingvistice, ale také v morfologii, syntaxi, lexikologii a lexikografii, sémantice, ve studiu jazykové variace a v dalších lingvistických odvětvích.

Řady jazyků, které disponují korpusy jak psaného, tak mluveného jazyka, jejichž analýza se provádí kvantitativně v počítačových databázích, se neustále rozšiřují. Podíváme se nyní na mluvené korpusy několika jazyků a na závěr se zaměříme na korpusy mluvené francouzštiny a češtiny.

Anglický jazyk, který je po čínštině druhým nejrozšířenějším jazykem ve světě, již delší dobu disponuje oběma druhy korpusů. Prvním počítačovým korpusem vytvořeným v letech 1961–1964 v USA byl *Brown Corpus of Written American English*, který se skládá z 500 textů po 2 000 slovech z 15 literárních žánrů. Starším korpusem je *Survey of English Usage Corpus*, který byl převeden do databázového prostředí počítačů poměrně nedávno a obsahuje texty založené na jazyce psaném a mluveném ve stejném poměru. Jeho část obsahující mluvené projevy se nazývá *London-Lund Corpus of Spoken English*, je kolekcí 87 textů po 5 000 slovech a čítá necelých 450 000 slov ze šesti různých mluvních situací. Pokud se zaměříme výhradně na mluvené korpusy, nemůžeme nezmínit 10 miliónový *The Spoken Corpus* vytvořený Longmanem jako součást *BNC (British National Corpus)*. Jedná se o korpus založený na přirozené, spontánní konverzaci nahrané z přednášek, obchodních schůzek a neformálních rozhovorů po oficiálních příležitostech. Nahrávky mapují projevy 2 000 mluvčích.

Další korpus mluvené americké angličtiny o 249 000 slovech byl sestaven na University of California Santa Barbara, je založen na nahrávkách promluv osob s celých Spojených států a reprezentuje mluvčí nejrůznějšího etnického a sociálního původu, věku, povolání a pohlaví. Kromě přímé konverzace zachycuje korpus telefonní hovory, promluvy v zaměstnání, na

schůzích, kázání a tak dále. *Santa Barbara Corpus of Spoken American English* je součástí *The International Corpus of English (ICE)*, který byl založen v roce 1990 za účelem shromažďování národních a regionálních variant angličtiny patnácti výzkumnými týmy v celosvětovém měřítku.

Dalšími příklady mluveného korpusu je švédský *Spoken Language Corpus of Swedish* na Göteborg University, který je založen na z 50% audio a 50% video nahrávkách přirozených konverzací pořizovaných od roku 1980. *The Spoken Dutch Corpus* je zase databází mluvené holandštiny používané v Nizozemí a ve Vlámku (Šonková 2008, s. 16–17).

Koncem 80. let byla práce na sestavení korpusu zahájena i u nás. Z počátku šlo mluvený korpus, po roce 1990 však tento projekt pod vedením Františka Čermáka přerostl v rozsáhlý *Český národní korpus*. Jeho součástí je také *Pražský mluvený korpus (PMK)*, v jehož případě bylo potřeba pořídit nahrávky mluvené češtiny, přepsat tyto záznamy do elektronické podoby a současně vytvořit vlastní počítačové programy, které by zvládly získaný materiál zpracovat. „Na základě více než 100 hodin nahrávek 504 mluvčích z Prahy a okolí, kteří podle věku, pohlaví, vzdělání a různé mluvní situace vytvořili 16 homogenních skupin, byl autory PMK v letech 1988–1996 pořízen mluvený korpus o velikosti 780 822 pozic (z toho 548 091 slovních tvarů, 232 231 interpunkčních znamének a závislých částí víceslovných jednotek).“ (Šonková 2008, s. 17). Český národní korpus, projekt Ústavu Českého národního korpusu na Filozofické fakultě Univerzity Karlovy v Praze (ÚČNK), od roku 1994, kdy začal fungovat, sestavil mnoho korpusů psaných a mluvených, synchronních i diachronních. V případě mluvených (synchronních) korpusů máme k dispozici již zmíněný Pražský mluvený korpus zveřejněný v roce 2001, Brněnský mluvený korpus o 490 000 slovech (2002), což je přepis nahrávek brněnské mluvy z 90. let 20. století, dále korpus vyučovacích hodin s názvem SCHOLA2010, po kterém následovaly korpusy ORAL2006, ORAL2008 a ORAL2013. Posledně zmiňovaný korpus je dosud v českém prostředí nejrozsáhlejším reprezentativním korpusem neformální mluvené češtiny a čítá 2,79 miliónů slov. K dispozici jsou také cizojazyčné webové korpusy zveřejněné

v roce 2013, mezi nimi také frWaC, což je lemmatizovaný⁴³ korpus francouzštiny o 1 milionu a 350 miliónech slov opatřený značkami. Informace o všech dostupných korpusech jsou k dispozici na oficiálních webových stránkách Českého národního korpusu.⁴⁴

Na základě korpusů často vznikají frekvenční slovníky či frekvenční seznamy, přičemž současným trendem je vytváření frekvenčních seznamů vycházejících z filmových titulků či z televizních a filmových scénářů, jak je tomu v případě frekvenčního seznamu mapujícího epizody seriálu Simpsonovi. Seznam byl zhotoven na základě seriálových titulků ze serveru OpenSubtitles.org.⁴⁵

5.1.1 Korpusy mluvené francouzštiny

Frankofonních korpusů nejen psaných, ale i mluvených, existuje celá řada. Výzkumy týkající se mluvené francouzštiny na území Francie, kterými se budeme v této kapitole zabývat⁴⁶, probíhaly již od 70. let. V roce 1977 vznikla revue *Recherches Sur le Français Parlé*, kterou založil kolektiv vědců z katedry Francouzské lingvistiky na Université de Provence vedený Jeanem Stéfanim. Členové této katedry se spojili do výzkumné skupiny nazvané *Groupe aixois de recherches en syntaxe* neboli GARS, která se částečně zajímala o syntax mluveného jazyka, což bylo dosud velmi málo prozkoumané odvětví jazyka. První čísla revue se zabývala sociolingvistikou W. Labova. Revue vycházela jednou ročně⁴⁷ až do roku 2003 a každé číslo obsahovalo přibližně deset článků.⁴⁸

Roku 1998 byl zahájen projekt *Le Corpus de Référence du Français Parlé (CRFP)* – jednalo se o vytvoření korpusu mluveného jazyka, kterého se

⁴³ Lemmatizace je proces, kdy je slovo převedeno do základního tvaru - tzv. *lemma*. Například slovo "rodinách" je převedeno na slovo "rodina".

⁴⁴ Viz web: <https://korpus.cz/struktura.php>.

⁴⁵ Frekvenční seznam dostupný zde: <http://pastebin.com/anKcMdvk>.

⁴⁶ Dále existuje například *Corpus de français parlé au Québec (CFPQ)* či korpusy sestavené ve Švýcarsku či Belgii.

⁴⁷ Přehled jednotlivých čísel a jejich obsah možno dohledat na <http://sites.univ-provence.fr/delic/rsfp/>.

⁴⁸ Viz web: <http://sites.univ-provence.fr/delic/rsfp/>.

ujal tým z Université de Provence pod vedením Claire Blanche-Benveniste. Tento projekt vznikl na zakázku delegace s názvem *Délégation à la langue française*, spadající pod Ministerstvo kultury, které výzkum kompletně financovalo. V roce 2000 přešel projekt pod vedení Jeana Veronise⁴⁹ a tým *DELIC (Description Linguistique Informatisée sur Corpus)*. Tento korpus čítá 440 000 slov a zahrnuje 134 nahrávek (36 hodin mluveného slova). Data byla posbírána ve 37 větších či menších městech severní a jižní Francie a také v pařížském regionu. Druhy mluvních situací byly celkem tři. Pro zhotovení korpusu se uplatnily nahrávky soukromých promluv (každodenní reakce, zážitky z dětství atd.), dále promluv profesních (mluvčí hovořili o své profesi či práci) a veřejných (radiové vysílání, konference, politické schůze atd). Mezi důležité parametry patřil také věk, pohlaví a vzdělání mluvčích⁵⁰.

Roku 2005 vznikl pod vedením Paula Cappeaua a Magali Sejjidové *Inventaire des corpus oraux*, což je projekt z iniciativy *Délégation générale à la langue française et aux langues de France (DGLFLF)* při Ministerstvu kultury a komunikace. Tento projekt má za cíl systematizovat aktuální informace o existujících korpusech a rozřadit je na základě předem daných kritérií. Jedná se o přehledný seznam existujících korpusových projektů doplněný o jejich charakteristiku, který má umožnit jednotlivým výzkumným týmům lepší vzájemnou komunikaci a výměnu informací (Sejjido, Cappeau 2006).

Zajímavým projektem je program, který byl spuštěn roku 2007. Jedná se o *Corpus de la parole* Ministerstva kultury a komunikace. Internetový projekt umožňuje on-line přístup k transkribovaným a digitalizovaným textům. Jedná se o zpracování různých jazyků vyskytujících se na území Francie – tedy jak v hexagonu, tak v zámořských oblastech. Korpus umožňuje lépe porozumět bohatosti tohoto jazykového dědictví.⁵¹

V roce 2012 byl na pařížské Université Sorbonne Nouvelle Paris 3 pod vedením Soniy Branca-Rosoffové dokončen projekt *Discours sur la ville* –

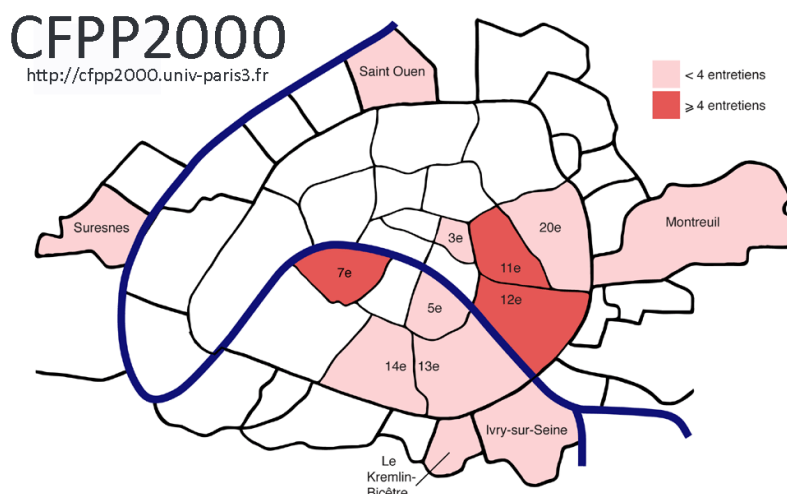
⁴⁹ O Jeanu Veronisovi jsme se zmínili již dříve v souvislosti s jeho frekvenční analýzou politické debaty.

⁵⁰ Více informací v PDF zde: <http://sites.univ-provence.fr/~veronis/pdf/2004-presentation-crfp.pdf>.

⁵¹ Více informací zde: <http://corpusdelap parole.tge-adonis.fr/>.

*Corpus de Français Parlé Parisien des années 2000 (CFPP 2000)*⁵². Korpus byl sestaven na základě nahrávek dlouhých rozhovorů s obyvateli Paříže a její periferie. Byl sestaven dotazník, na jehož podkladu se hovořilo o životě obyvatel, o jejich vztahu ke čtvrtím, ve kterých žijí, o urbanismu, o globalizaci a jejím dopadu na gastronomii, na módu, či o situaci ostatních jazyků. Účelem bylo vyvinout korpus určený ke studiu lingvistické praxe a současného francouzského jazyka, neboť dnešní doba přímo vyzývá k podobným výzkumům, neboť rozvoj internetu a počítačových programů velmi usnadňují práci a transkripci, třídění a archivaci dat.

Vizualizace pařížských čtvrtí, ve kterých proběhl sběr dat pro vytvoření korpusu.



5.2. Metoda zpracování jazykového materiálu

V této kapitole se budeme zabývat tím, jakým způsobem lze pracovat s jazykovým materiálem. Metodologii vytvoření korpusu a jeho následné analýzy, kterou užíváme v naší práci, představíme následně na začátku analytické části.

Jak již bylo řečeno, výzkum mluveného jazyka si lze částečně zjednodušit použitím zástupných textů, jako jsou scénáře a dialogové listiny. Důkladný výzkum, který se má stát reprezentativním zmapováním situace

⁵² Branca-Rosoff, Sonia, Fleury, Serge, Lefeuvre, Florence, Pires, Matthew. 2012. „Discours sur la ville : Corpus du Français Parlé Parisien des années 2000 (CFPP2000)“. Paris : Université Sorbonne Nouvelle Paris 3, 2012 [citováno 14. 1. 2011]. Dostupné z WWW: <<http://cfpp2000.univ-paris3.fr/CFPP2000.pdf>>.

mluveného jazyka na daném území, musí být však pojat komplexněji. V takových případech je obvyklé vycházet z nahrávek projevů mluvčích z dané lokality, které jsou následně statisticky zpracovávány. Ve výzkumu je pak kladen důraz především na výběr a záznam mluvčích, převedení jejich nahrávek do textové podoby, okódování textů a jejich elektronické zpracování.

Při výběru respondentů je potřeba si vymezit sociální parametry. Při vyhotovování pražského či pařížského mluveného korpusu stanovíme například, zda respondent musí být narozen v dané lokalitě nebo je dostačující fakt, že je jejím dlouhodobým obyvatelem. Toto kritérium nám předem eliminuje vlivy dialektů a zaručí kompaktnost jazyka. Dále se pracuje se sociolingvistickými skupinami. V případě Brněnského korpusu, zkoumajícího jazykovou situaci na Moravě, byli mluvčí vybíráni tak, že reprezentovali všechny společenské skupiny uživatelů českého jazyka. „Podle rodu, věku, vzdělání a mluvní situace tvořili šestnáct sociolingvistických skupin, v jejichž rámci byly promluvy podrobeny frekvenční analýze.“ (Šonková 2008, s. 20). Mluvčí byli nahráni v řízeném/formálním i ve spontánním/neformálním rozhovoru. V případě řízeného rozhovoru je zvykem pokládat mluvčím předem stanovené otázky obecného charakteru – například z oblasti vzdělání a povolání, života ve městě, rodinného života a tak dále. Jitka Šonková vysvětluje, že v případě Pražského mluveného korpusu „otázky byly formulovány způsobem, který měl mluvčí stimulovat k pokud možno uvolněné a přirozené odpovědi, ale ne nutně spisovné. Otázky v sobě záměrně obsahovaly prvky spisovné i nespisovné – neformální prvky, aby se mluvčí necítili být vázáni.“ (Šonková 2008, s. 20). Nahrávání často komplikuje fakt, že samo vědomí mluvčího, že je nahráván, vede k jazykové sebestylizaci. Po chvíli si však mluvčí zvyknou a jejich jazyk je opět přirozený. Dotazník PMK se skládal například z těchto otázek (povšimněme si, jak diskurz mluveného jazyka převeden do psané podoby působí nezvykle, až rušivě):

- V poslední době proběhly a probíhají různé změny ve školské soustavě. Co si myslíte o současné škole, když ji srovnáte s dřívější? Jak budeme vychovávat děti, taková bude jedinou společností.
- Otcové našich otců bývali často přísní a brali na děti pásek. Myslíte si, že dnešní rodiče trestají děti málo nebo hodně?

- Ve větších městech jsou lidi k sobě dost nesnášenliví a bezohlední, každé si všímá jenom sám sebe. Lidi z venkova zase říkají, že mají k sobě blíž. Máte stejný nebo jiný zkušenosti.
(Šonková 2008, s. 21)

Jazykový materiál vzniká následně přepisem nahraných záznamů. Před uskutečněním nahrávek je potřeba mluvčí alespoň obecně informovat (například: jedná se o výzkum, který si má za úkol zjistit názory na určité společenské otázky). Skutečný cíl – výzkum jazykového úzu mluvčích – nemusí být předem zmíněn, mluvčí však musí souhlasit s pořizováním nahrávek.

Pravidla přepisu nahraného materiálu se liší korpus od korpusu. Neexistuje žádná přesně daná metodologie. Je zapotřebí si určit, do jaké míry a zdali vůbec promluvy redigovat (například oprava přeřeků), stanovit si systém značek atd. „Přepis nahrávek byl založen na zásadě, že všechny tvary slov, které nemají varianty, byly přepisovány spisovně, i když byly vysloveny jinak,“ vysvětluje Šonková (Šonková 2008, s. 23). V případě pařížského mluveného korpusu byl k transkripci nahrávek ve formátech WAVE a MP3 použit program *Transcriber*⁵³. Ortografický přepis jedné hodiny monologického mluveného projevu trvá obvykle kolem 40 hodin. Tato doba se navyšuje na 70 hodin v případě přepisu dialogů a mluvních situací, kdy si respondenti skákají do řeči, mluví jich více naráz a tak dále. Softwarový program pracující na bázi transkribování audio signálu výrazně zjednodušil jak samotný průběh transkripce, tak následnou synchronizovanou četbu transkripce a audio signálu (Branca-Rosoff, Fleury, Lefevre, Pires 2012, s. 10).

Přepisem nahrávek vznikne lineární text – jazykový korpus, který může být následně podroben frekvenční analýze, která má stejné či podobné rysy jako analýza psaného projevu. Frekvenční zpracování bývá provedeno za pomoci specializovaných korpusových programů (např. v českém prostředí korpusový manažer Bonito).

⁵³ Viz web: <http://trans.sourceforge.net/en/presentation.php>.

5.3 Français fondamental a metoda SGAV

Frekvenčnost mluveného jazyka sehrála roli i v didaktice výuky francouzštiny jako cizího jazyka (*FLE – Français langue étrangère*).

Během 2. světové války posílila angličtina svou pozici a stávala se více a více jazykem mezinárodní komunikace. Poměrně silná pozice francouzského jazyka ve světě byla ohrožena. Proto bylo ve snaze zachovat šíření francouzštiny do zahraničí potřeba vymyslet novou koncepci výuky. Proto se francouzské ministerstvo starající se o národní vzdělávání rozhodlo pověřit komisi, která vypracuje tzv. *français fondamental*, omezenou pouze na slovník považovaný za základní. V roce 1954 byly publikovány výsledky lexikálních studií, které vypracoval CREDIF neboli Centre de Recherche et d'Étude pour la Diffusion du Français vedený Georgem Gougenheimem. Jednalo se o frekvenční analýzu mluveného projevu, jejímž výstupem byly dva seznamy. Français fondamental prvního stupně obsahovala 1475 slov. Druhý stupeň čítal 1609 slov. Français fondamental je považována za základ první etapy výuky francouzštiny jako cizího jazyka, avšak byla často kritizována lingvisty. Častými argumenty byla nutnost aktualizace seznamu, protože prefabrikované dialogy nepředstavují jazyk pravdivě.

V polovině 50. let tedy Petar Guberina z Institutu fonetiky na Univerzitě v Záhřebu formuloval teoretické základy metody SGAV (structuro-globale audiovisuelle). S Paulem Rivencem z École normale supérieure v Saint-Cloud ve Francii definovali principy aplikace metody na pedagogiku. Od roku 1960 spolupracovali také s Raymondem Renardem (Belgie).

Prvním materiálem vypracovaným podle metody SGAV je *Voix et Images de France* (VIF) z roku 1962 a je určený pro děti od 8 do 11 let.⁵⁴ Výuka sestávala z videomateriálu obsahujícího obrázky, které ilustrovali mluvené slovo – jednalo se tedy o metodu spojující obraz se zvukem a postavenou na komunikační situaci, dialogu a obrazu. Účelem této výuky bylo naučit studenty

⁵⁴ Ukázka lekce z této metody je k vidění na webové adrese: <http://ticsenfle.blogspot.cz/2012/11/voix-et-images-de-la-france-1962.html>.

především každodennímu hovorovému projevu (*langue parlée familière a français quotidien parlé*). Důraz byl kladen na porozumění, které má vždy předcházet produkci. Přednost je dána mluvenému slovu před psaným. Na jazyk totiž metoda nahlíží jako na způsob vyjádření a ústní komunikace a psaná forma jazyka je považována pouze za odvozeninu od slova mluveného. Jazyk je akusticko-vizuálním souborem. Proto je v této metodě kladen důraz i na gesta a nonverbální vyjadřování obecně a není zapomenuto ani na afektivní faktor jazyka, tedy na vyjádření emocí a pocitů (Germain 1993, s. 153–157).

5.4 Filmové titulky jako korpus pro studium frekvenčnosti

O slovní frekvenčnosti zkoumané na základě filmových titulků již v naší práci byla řeč. Tým badatelů složený z Borise Newa, Marca Brysbaerta, Jeana Veronise a Christoha Palliera popisuje svůj kvantitativní výzkum vycházející z titulků ve studii s názvem *The use of film subtitles to estimate word frequencies* (2007). Jak je vidět, je internet v dnešní době velmi snadnou cestou, jak získat podklady pro měření slovní frekvenčnosti i v jiném jazykovém registru, než je psaný text.

Boris New tvrdí, že je obtížné nalézt seznamy mapující četnost výskytu mluveného jazyka. Je to podle něj dáno složitostí zpracování materiálu a metodou přepisování záznamů. Softwarové programy je sice možné použít, ale jejich spolehlivost není dokonalá. Limitováno je tak i množství jazyků, pro které je mluvený korpus vytvářen. Zdůrazňuje užitečnost vytváření takových korpusů, neboť se zdá, že frekvenční seznamy psaných textů neodrážejí správně četnost slov každodenního života, jako jsou slova spojená s jídlem, odíváním, sociálními interakcemi a tak dále. Ideálním „mluveným“ korpusem by podle něj byl takový materiál, který by obsahoval vše, co lidé poslouchají a říkají během dne. Vytváření takového korpusu by ale bylo velmi drahou záležitostí.

Proto se ve svém výzkumu soustředí na titulky filmů a televizních programů dostupné na internetu, které jsou pro něj zdrojem transkribovaného mluveného textu. Tento typ textu se vyznačuje dvěma zajímavými znaky. Za

prvé zprostředkovává mezilidskou komunikaci na pozadí prostředí, které je díky obrazu zcela jasně definováno, za druhé pro mnoho lidí představují filmy a televizní programy značnou část přísunu informací o jazyce, uvážíme-li, že současný odhady týkající se sledování televize hovoří o průměrné době 3–4 hodin denně strávených před televizní obrazovkou.

Výzkumníci se opřeli o tisíce souborů s francouzskými titulky, které byly volně k dispozici na internetu a celkem stáhli titulky k 9 474 filmů a televizních seriálů, které pocházeli ze 4 kategorií:

- otitulkované francouzské filmy čítající 1,9 miliónů slov
- otitulkované anglické a americké filmy o 26,5 miliónech slov
- otitulkované anglické a americké televizní seriály (*Přátelé*, *Ally McBeal* atd.) o počtu 14,9 miliónů slov
- otitulkované neanglické evropské filmy, jako *Goodbye, Lenin!*, čítající 2,5 miliónů slov

Většina filmového materiálu tedy pocházela z angličtiny, což proporčně odpovídá převaze anglosaské kinematografie ve světě. Následovala editace textu, odstranění časových údajů atd., což byla jediná část procesu, která se musela dělat ručně. Časová náročnost této fáze zpracování byla přibližně 2minuty/film. Tímto způsobem byl za pomoci dalších postupů vytvořen korpus, který obsahoval 52 miliónů francouzských slov, což je výrazně větší množství než v případě jakéhokoliv jiného korpusu mluvené francouzštiny.

Existovalo však znepokojení ohledně platnosti výstupů ze statistického měření titulků. Koneckonců, titulky jsou oproti textu, který je pronášen ve filmu, téměř vždy zkrácené a upravené. Nezahrnují zaváhání nebo výslovnostní chyby, které jsou součástí mluvy. Navíc námět filmu či seriálu staví na určitém tématu. Například se daleko častěji setkáme s cizoložstvím a s policií než je běžné v reálném životě průměrného účastníka psycholingvistického experimentu. Aby se přišlo na to, jestli se obavy zakládají na pravdě, byl výstup otestován dvěma způsoby. Jedním z nich je porovnáním vzniklého korpusu s již existujícími zdroji. Došlo k porovnání s korpusem *Corpus de Référence du*

Français Parlé (CRFP; Equipe DELIC, 2004) a byla objevena shoda v 5 206 záznamech (cca 73 %), což je obstojný výsledek.

Ve snaze vypátrat kořeny nesouladů mezi těmito dvěma seznamy se dále tým zabýval zkoumáním hesel, která se vyznačovala výrazně vysokou či výrazně nízkou frekvenčností na jednom ze dvou seznamů. Hesla titulkového korpusu vykazující zvýšenou frekvenčnost oproti korpusu CRFP se týkala většinou slov spojených s policií a vyšetřováním, což koresponduje se současným hojným výskytem kriminálních seriálů a krimi filmů. Dále to byly typické hovorové výrazy (*dieu, salut, désolé, papa, con* atd.), které byly častější u titulkového korpusu, který je na rozdíl od monologického charakteru CRFP založen na interakcích osob. Opačná situace – tedy vyšší výskyt slova v CRFP, než v titulkovém korpusu – nastala především u slov, která se používají pouze v některých francouzských regionech (*pétanque, lyonnaise* atd.). Další skupina slov byla svázána s francouzskou administrativou (*municipalité, collectivités* atd.), což pravděpodobně odráží druh otázek, který byl v rámci CRFP účastníkům kladen (Jaké je vaše zaměstnání?). Třetí kategorií jsou zvukomalebná slova (*euh, bé, hum* atd.). Na závěr to byly typicky francouzské výrazy, které v mnoha filmech nefigurovaly (*viticole, charcutier, viticulture* atd.) (New, Brysbaert, Veronis, Pallier 2007).

Boris New a jeho tým představili novou metodu, jak velmi dostupnou cestou vytvořit korpus mluveného jazyka. Jejich materiál je navíc možné neustále aktualizovat za pomoci titulků z nově vznikajících filmů a seriálů. V porovnání s jiným korpusem mluveného jazyka se výsledky obou prací z velké části shodovaly, což lze považovat za úspěch svědčící o relevantnosti tohoto výzkumu.

FREKVENČNÍ ANALÝZA

Vymezení úkolu analytické části práce

Poslední kapitolou teoretické části práce jsme se již dostali k tomu, čemu se bude věnovat následující analytická část. Až dosud jsme akcentovali mluvenou francouzštinu a její morfosyntaktické a lexikální vlastnosti. Nyní ji podrobíme kvantitativní analýze, jejíž součástí bude lexikální a gramatická statistika. Cílem je vytvořit frekvenční seznam slov užitých ve vybraném textu, dále stanovit procentuální rozložení jednotlivých slovních druhů v textu a procentuální či číselné rozložení jednotlivých gramatických jevů v daném textu.

V naší analýze nebudeme vycházet z filmových titulků, jak tomu bylo v případě Borise Newa, ale rozhodli jsme se analyzovat filmové scénáře. Ty vykazují podobný charakter, ale ve srovnání s titulky nejsou kráceny a plně odpovídají původnímu rozsahu replik pronesených ve filmu. Filmové scénáře budeme porovnávat s jejich literární předlohou, tzn. s románem, na jehož základě vznikla analyzovaná filmová adaptace. Frekvenční seznamy obou druhů diskurzu podrobíme komparaci a na základě získaných údajů se pokusíme vysledovat zákonitosti či souvislosti mezi textovým stylem, jeho lexikální vybaveností a využitím vybraných gramatických jevů.

1. Analyzovaný text

1.1 Výběr textu a jeho charakteristika

Při výběru textu k analýze jsme si stanovili několik kritérií, přičemž určujícím bodem byla dostupnost daného filmu, filmového scénáře a románové předlohy v originálním francouzském znění. Významnou roli ve výběru sehrála také naše osobní preference. Korpus určený k analýze tvoří vybrané pasáže z těchto děl:

- **Scénář filmové adaptace v původním francouzském znění**

1. *L'Écume des jours (Pěna dní)*, režie: Michel Gondry, scénář: Luc Bossi, Francie/Belgie, 2013, 125 min., distribuce v ČR: Aerofilms.
2. *Le Magasin des suicides (Obchod pro sebevrahy)*, režie a scénář: Patrice Leconte, Francie/Kanada/Belgie, 2012, 85 min., distribuce v ČR: AČFK.

- **Románové dílo v původním francouzském znění**

1. Vian, Boris: *L'Écume des jours*, Paris: Jean-Jacques Pauvert, 1998. 315s.
2. Teulé, Jean: *Le Magasin des suicides*, Paris: Julliard, 2008. 157s.

V obou případech se jedná o poměrně nedávné filmové adaptace románových předloh. *L'Écume des jours* (česky *Pěna dní*) je neznámějším románem Borise Viana, který jej poprvé vydal v roce 1947. Jedná se o dílo plné fantazie a symboliky, kde nalezneme prvky absurdity a existencialismu. Bezstarostné vyprávění o spokojené lásce se snoubí s pomíjivostí štěstí a se strachem z nemoci a z bolestivé smrti. V díle se vyskytuje také postava Jeana-Paula Sartra (v knize Jean-Sol Partre), který je intelektuální modlou jedné z románových postav – Chicka. Chick je vášnivým sběratelem všeho, co s Partrem souvisí, a stává se tak karikaturou obsesivního a oddaného fanouškovství. Filozofie Jeana-Paula Sartra se v *Pěně dní* neobjevuje, pouze názvy jeho knih jsou parodovány prostřednictvím slovních hříček typu *La Lettre et le Néon* místo *L'Être et le Néant (Bytí a nicota)* a podobně. Román byl adaptován na filmové plátno již v roce 1968, kdy se režie ujal Charles Belmont a do hlavních rolí obsadil Jacquesa Perrina a Marie-France Pisierovou. Na konci dubna loňského roku se román dočkal nového filmového zpracování. Režisér Michel Gondry⁵⁵, který se spolu s Lucem Bossim⁵⁶ podílel i na tvorbě scénáře, vybral pro roli Colina Romaina Durise, Chloé ztvárnila Audrey Tautou, Chicka Gad Elmaleh a v roli Nicolase se představil Omar Sy, kterého můžeme znát jako představitele hlavní role v oceňovaném filmu z roku 2011 *Intouchables* (režie: Olivier Nakache, Eric Toledano). Filmová *Pěna dní* získala v roce 2014 cenu César za Nejlepší výpravu (*Meilleurs décors*) a byla nominována v kategoriích Nejlepší hudba a Nejlepší kostýmy.

⁵⁵ Známé jsou jeho filmy *Věčný svit neposkvřené mysli* (2004) či *Nauka o snech* (2006).

⁵⁶ Luc Bossi je současně producentem filmu.

Autorem novely *Le Magasin des Suicides (Obchod pro sebevrahy)* je francouzský spisovatel, kreslíř a scénárista Jean Teulé. Jedná se o příběh vyznačující se bizarním a morbidním humorem. Námět zpracoval režisér Patrice Leconte a vytvořil tak poměrně zdařilý animovaný muzikál. Zatímco román *Pěna dní* je známým dílem francouzské literatury a není třeba ho více komentovat, *Obchod pro sebevrahy* je současnou literární tvorbou, proto si jeho námět představíme blíže.

Rodina jménem Tuvache vlastní již několik generací krámk splňující bezchybně veškerá očekávání svých zákazníků. Ta nejsou nikterak malá, neboť se jedná o bizarní město, které se zmítá v depresi a skepsi. Jedinou touhou jeho obyvatel se dřív nebo později stává úspěšně dokonaná sebevražda. Tuto poptávku pokrývá právě krámk Mishimy Tuvache a jeho ženy Lucrécie. Spolu se svými dětmi Marilyn a Vincentem se starají o „blaho“ svých klientů tak, že jim nabízejí kvalitní smrt. Jejich firemní filozofie se drží hesla „jestliže se vám nepovedl váš život, tak ať se alespoň povede vaše smrt“. V regálech svého obchodu nabízejí lana a provazy, smrtící pavouky a hady, flakóny s jedy všeho druhu a všech cenových kategorií, betonové kvádry, které vás spolehlivě přidrží na dně, či nabroušený samurajský meč pro pravé „tvrďáky“, kteří si svou smrt chtějí opravdu vychutnat. Přehlídku sortimentu nám přezpívají v první písni, jejíž refrén shrnuje vše podstatné: Vive le suicide! (Ať žije sebevražda!)⁵⁷.

A do tohoto prosperujícího podnikání přichází další potomek jménem Alan. Alan je od pohledu úplně jiný než zbytek rodiny a jeho optimismus a radostná povaha dělá uměřeným a zasmušilým rodičům starosti. Nejen, že je zde riziko, že svým smíchem demotivuje kupující, kteří už tak občas potřebují k sebevraždě trochu popostrčit, ale ještě ohrožuje podnik svými destruktivními nápady ve jménu spokojeného života a lásky. Sestru s nadváhou, která se považuje za ošklivou, přesvědčí, že je přitažlivou orientální tanečnicí, a dá jí tak nový pohled na sebe a na svět. Z pubertální gothičky se na konci filmu stává zamilovaná romantička a z obchodníků se smrtí obchodníci s palačinkami.

⁵⁷ Ke zhlédnutí zde: <https://www.youtube.com/watch?v=Syw8q4vgc3A>.

Jak je vidět z předchozích řádků, obě tato literární díla se vyznačují poměrně specifickými lexikem. V případě *Pěny dní* se setkáme se slovními hříčkami a novotvory jako je *Jean-Sol Partre, bigle moi*, tanec v rytmu swingu, či *piano cocktail* – piano seřízeno tak, že je schopno míchat koktejly podle partitury, která je na něm přehrána. V *Obchodě pro sebevrahy* se pak často setkáváme se slovy náležejícími do lexikálního pole spojeného se smrtí, sebevraždou a způsoby, jak umřít – jedy, meči, palnými zbraněmi, kulkami, provazy a tak dále. Právě zmíněná slovní zásoba dělá z obou knih lingvisticky zajímavý materiál k analýze, který ovšem poměrně málo odráží slovní zásobu non-fikčního světa. Proto je účelem analýzy těchto děl vyzkoumat přednostně kvantitativní poměry vyskytující se v těchto syžetech. Teprve poté bude možno zamyslet se, například prostřednictvím porovnání s jiným existujícím frekvenčním seznamem, nad obecnější platností našich tvrzení.

1.2 Rozsah korpusu

V předchozí diplomové práci jsme se věnovali kvantitativnímu rozboru korpusu o 1962 slovech vytvořeného ze scénáře k filmu *Amélie z Montmartru*. V této práci máme především snahu o zhotovení rozsáhlejší materiálové základny sloužící k analýze, protože je známo, že „pravděpodobnost obecné platnosti zjištěného pořadí slova vzrůstá s rozsahem materiálu užitého k tabulaci.“ (Jelínek, Bečka, Těšitelová 1961, s. 19). I přes naši snahu objektivizovat výsledky pomocí analýzy většího korpusu jsme si však vědomi rezerv, kvůli kterým naše analýza zůstává stále spíše představením jazyka z matematického pohledu a není možné ji prezentovat jako jevy obecně platné pro francouzský jazyk jako celek⁵⁸. Přesto však lze vysledovat tendence, které se přibližují objektivním hodnotám jazyka v návaznosti na typ textu (psaný vs. mluvený). Počet analyzovaných slov zařazených do frekvenčního seznamu můžeme vidět v tabulce:

⁵⁸ A to právě i kvůli již zmíněným specifickým slovní zásoby vybraných děl. V oblasti gramatické statistiky, kde se charakter lexika tolik neprojevuje, jsou však poznatky aplikovatelné i obecněji.

	Filmový scénář	Románová předloha
<i>L'Écume des jours</i>	2 007	4 362
<i>Le Magasin des suicides</i>	1 205	3 323

Celkem jsme tedy analyzovali 10 897 slov, což je 5 a ½ krát více než v naší předchozí práci. Tato materiálová základna vznikla výběrem konkrétních scén z románu a paralelně k tomu odpovídajících scén z filmu. Na bázi totožného lexikálního pole užitého jak v románu, tak ve scénáři je pak možné vyzkoumat proměnlivost četnosti konkrétních slovních tvarů v závislosti na médiu a uvědomit si syžetový charakter textů související s množstvím slov, která byla použita při stejném sdělení v románu a protikladně k tomu ve scénáři. Komparace literární předlohy a jejího zpracování do scénáře nám také dává možnost nahlédnout do dramaturgické problematiky adaptace. Při čtení a analytické práci samotné jsme se pak zamýšleli nad tím, jakým způsobem adaptátor pracuje s jednotlivými scénami⁵⁹ a v jaké míře se inspiruje původním textem⁶⁰, jak pozměňuje rytmus děje či vyškrtává filmově méně atraktivní pasáže.

Při pohledu do tabulky výše si na první pohled povšimneme rozdílnosti v počtu analyzovaných slov. Jak již bylo řečeno, prostředkem k analýze se stalo porovnání totožných scén z románu a scénáře popisujících stejnou dějovou situaci. Jak je patrné z čísel, filmový scénář využívá ke sdělení stejných informací méně jazykových prostředků. Jazykové prostředky jsou kompenzovány audiovizuálním jazykem filmu a mluvené slovo je tak doplněno filmovými znaky. Z tabulky můžeme vyčíst, že text filmového scénáře odpovídá přibližně 36–46 % románového textu.

1. 3. Rozdílnost zdrojových médií a adaptační proces

Jak jsme se již přesvědčili, ze samotného charakteru filmu vyplývá, že jeho scénář je oproti textu původní adaptované předlohy značně zestručněn.

⁵⁹ Například pozměňuje jejich pořadí.

⁶⁰ Adaptace věrná či volná.

Zatímco jedním z úkolů románu je vést čtenáře k představivosti podporované četným líčením prostředí, osob, věcí, dějů, časových souvislostí a niterných představ, povaha filmu, jakožto jiného média využívajícího jiné znakové soustavy, se vyznačuje svou vícevrstevností, která kromě jazykového projevu zahrnuje i obraz, hudbu, zvuky, ruchy či střih schopné emotivně působit na diváka. Prostřednictvím práce s filmovou syntaxí, střihovými postupy či proměnami mizanscény může režisér nechat zaznít věty a slova obsažené v románovém protějšku, aniž by je ve filmu explicitně vyslovil.

Filmové adaptace byly dlouho považovány za pouhé kopie hodnotnějších literárních předloh. Adaptační myšlení se začalo rozvíjet až v souvislosti s etablováním studia filmu na amerických a britských univerzitách (60. a 70. léta) (Bubeníček 2010, s. 7). V té době panoval v intelektuální společnosti názor, že dosavadní audiovizuální podoba kanonických děl, která byla vybrána k adaptaci, pouze potvrzuje tvrzení o nezfilmovatelnosti románu a nemožnosti adekvátní filmové reprezentace fikčního světa. K tomuto postoji přispěli samotní filmoví badatelé, kteří se až na výjimky (například André Bazin) fenoménu příliš nevěnovali a to ani přes velký počet natočených adaptací. Situace se proměnila až v 80. a 90. letech 20. století, kdy nastal určitý interdisciplinární obrat, a adaptace se stala tolerovanou součástí badatelského diskurzu humanitních disciplín.

Jeden z prvních teoretiků filmové adaptace George Bluestone se ve své knize *Novels into Film* (1957) zabývá právě rozdílností obou médií, která zapřičiňuje, že filmová adaptace se vlastně nikdy nemůže podobat své předloze. Tuto odlišnost spatřoval Bluestone především v jazykové povaze literatury a ve vizuálně-prezentačním charakteru filmu. Každý filmař se podle Bluestona stává novým autorem a literární předloha představuje pouze základ jeho vlastní tvorby (Bluestone 2003, s. 61–62). Podle Bluestona je jedním z důvodů negativního nahlížení na adaptace zdůrazňování rozdílů mezi recepcí a percepcí jako dvou „odlišných způsobů porozumění světu“ (Bluestone 2003, s. 20). „Intelektuálně náročná“ recepce literatury je poměřována se způsobem zakoušení hraného filmu. Zatímco četba vzbuzuje dojem intelektuálního úsilí, sledování filmu je především zábavou, odpočinkem a potlačováním vlastní

imaginace. Recepční estetik Wolfgang Iser označil film za „bezmezerovité médium“, které dokáže fikční postavu zcela zpřítomnit, zatímco při čtení je čtenáři ponechán prostor k tomu, aby si hrdinu vykreslil na základě své vlastní imaginace (Bubeníček 2010, s. 9). S tímto označením filmu jakožto „bezmezerovitého média“ si dovoluujeme nesouhlasit, neboť je to právě stříhová skladba a filmové postupy, které umožňují divákovi rekonstrukci viděného v jeho mysli. Například užití elipsovitého stylu vyprávění je základem pro syntézu narativní struktury v mysli diváka, který, ač je svědkem explicitně prezentovaného obrazu, si nemůže být jistý například jeho věrohodností, pravostí či uceleností a zapojuje imaginaci ve snaze vyplnit narativní mezery.

2. Metodologie

Metodologie této práce se téměř shoduje s metodologií naší předchozí bakalářské práce. Proto v kapitole popisující mechanismus zacházení s texty vycházíme z velké části z již dříve formulovaného postupu, který však doplňujeme o nová kritéria, vyplývající z nových zjištění a potřeb, ve snaze vytvořit relevantní kvantitativní analýzy.

V naší analytické části se nesoustředíme pouze na lexikální frekvenčnost slov v textu, ale také na zjištění frekvence slovesných tvarů, početnost rodů a čísel u substantiv a kvantifikaci slovních druhů v textu. Jelikož při sestavování frekvenčního seznamu ani při následných dílčích analýzách jsme neměli k dispozici žádné specializované počítačové programy, veškerá data jsou zpracována ručně a jejich správnost je několikrát ověřována. Při statistické práci s textem se nám opět osvědčil textový editor Microsoft Word a také Excel, tentokrát verze 2010.

V první fázi bylo potřeba vybrat z románů a scénářů pasáže určené k tabulaci.⁶¹ Z textu byly vyříděny dějově odpovídající úryvky ze dvou románů a dvou dialogových listin (scénářů). Scény byly zvoleny čistě

⁶¹ Materiál vybraný k analýze je částečně součástí přílohy. Text bohužel není možné uvést v plném rozsahu s ohledem na přání distribučních společností, které nám texty poskytly.

na základě naší osobní preference. V případě scénářů je nutné zdůraznit, že do tabulace byly zařazeny pouze dialogové promluvy aktérů příběhu. Záměrně jsme vyřadili scénické poznámky postihující okolnosti promluv, intonaci a rozpoložení postav, uspořádání mizanscény a další režijní a dramaturgické pokyny.

Druhým krokem bylo určování slovních druhů. Jako nejspolehlivější variantu ručního zpracovávání jsme zvolili vizuální, barevné odlišování jednotlivých slovních druhů v textu. Během této fáze jsme se místy setkávali s pochybnostmi ohledně správného přiřazení slova k odpovídající kategorii. Ke správné klasifikaci nám významně posloužila francouzská gramatika *Le Bon usage*, jejímž autorem je Maurice Grévisse a také výkladový slovník *Le Robert & Clé International*, jehož autorem je Josette Rey-Debove. Za užitečný považujeme v tomto směru také frekvenční seznam uveřejněný na webových stránkách francouzského Ministerstva školství, jež sestavil lexikolog Étienne Brunet⁶².

Nezbytnou součástí postupu bylo následné rozdělení textů do deseti diferencovaných seznamů (dle počtu slovních druhů). Byly tak vytvořeny čtyři samostatné textové okruhy, které odpovídaly dvěma scénářům a dvěma románům. Práce s textem pak probíhala v rámci každé textové kategorie zvlášť a bylo tak vytvořeno celkem 4 x 10 seznamů s jednotlivými slovními druhy. Sečtením slov v rámci každého slovního druhu jsme zjistili celkový počet analyzovaných slov v rámci jednotlivých textových kategorií (scénáře vs. romány). Tyto seznamy nám umožnily přehlednější práci a výsledkem byla charakteristika materiálu z hlediska procentuálního a početního rozložení slov v jednotlivých textech. Získaná data jsme zanesli do tabulky a částečně i do koláčových grafů.

Při zjišťování lexikální frekvenčnosti jsme využívali textový editor Word a Excel. S jejich pomocí jsme byli schopni určit četnost výskytu slov a na principu práce s textem upraveným do sloupce (na každém řádku jedno slovo)

⁶² Frekvenční seznam je k dispozici v příloze.

a jeho abecedního a číselného seřazování tvořit frekvenční seznamy zaznamenávající vzestupně kvantitativní hodnoty slov v textu. Abecední seřazení slov posloužilo ke sdružení stejných slov ve sloupci pod sebe. Následně byla před slova vpisována čísla zaznamenávající četnost výskytu. Sestupné seřazení číslic v kombinaci se slovem vytvořilo seznam slov seřazený od slov s nejčastějším výskytem po slova s frekvencí 1. Na tomto principu byly zpracovány všechny seznamy s jednotlivými slovními druhy. Poté, co byl takovému postupu podroben každý slovní druh, spojili jsme všechny dílčí seznamy do jednoho. Slovní druhy se nám tak promíchaly a za pomoci drobných úprav mohl vzniknout finální frekvenční seznam řazený od nejčastěji se opakujících jednotek až po jednotky jednou zmíněné.

Při rozepisování textu na jednotky určené k lexikální statistice se vyskytovaly komplikace spojené s otázkou, co můžeme považovat za jedno slovo, a co naopak již hodnotíme jako soubor více slov. Až na výjimky chápeme slovo jako grafickou jednotku. Více slov oddělených spojovníkem chápeme však například v našem frekvenčním seznamu jako jedno heslo. Klíčem k zacházení se získanými daty nám byla následující kritéria:

Podstatná jména (substantiva)

- Vyskytuje-li se v textu samostatné slovo, které je uvedeno jak v singuláru, tak v plurálu (*un compte; des comptes*), do frekvenčního seznamu je slovo zařazeno pouze ve své singulární podobě a jeho tvar v množném čísle je samozřejmě započítán do číslice výskytu.
- Pokud se substantivum nachází v textu pouze v plurálu, ve FS⁶³ je zaneseno jako plurálový tvar (tedy nikoli jako singulár).
- Vlastní jména a slova, jejichž části jsou odděleny spojovníkem, chápeme jako jedno heslo (*Jean-Sol, hara-kiri, arrière-cuisine, boogie-woogie, nouveau-né, poisson-chat, Haut-le-Cœur, grand-chose* a jiná).

⁶³ FS = frekvenční seznam.

- U vlastního jména *Jean-Sol Partre* postupujeme následovně: pokud se jméno i příjmení vyskytují společně, počítáme je jako jedno heslo. Pokud se v textu vyskytne pouze *Jean-Sol* či *Partre*⁶⁴ bez příjmení či křestního jména v okolí, započítáváme každou část jména jako samostatné heslo. Stejně postupujeme také u jména *Duke Ellington* či u pojmenování psí rasy označené jako *Jack Russel*. Toto zacházení se jmény jsme si osvojili z toho důvodu, že podle našeho názoru nejsou dostatečným svědectvím o frekvenčnosti ve francouzském jazyce, neboť například *Jean-Sol Partre* je přezdívka vynalezená Borisem Vianem a *Duke Ellington* není jméno francouzského původu. Přesto jsme však tato slova nechtěli ze seznamu vyškrtnout, neboť jsou významnou součástí diskurzu, který analyzujeme. Četnost vlastních jmen však není naším primárním zájmem.
- Cizojazyčná slovní spojení (např. *Black and Tan Fantasy* – název písně) započítáváme jako jedno heslo, neboť nám připadalo málo efektivní dělit toto anglické spojení dále na slovní druhy.
- S novotvory Borise Viana jako *biglemoi* či *pianoctail* zacházíme jako s ostatními existujícími slovy a z frekvenčního seznamu je nevynecháváme, neboť jsou svědectvím o Vianově lingvistické osobitosti.
- Co se týká následného kvantitativního zpracování kategorie rodu a čísla u substantiv, na rozdíl od pravidel, která jsme si stanovili při tvorbě frekvenčního seznamu, respektujeme u gramatické statistiky kategorii čísla v takové podobě, v jaké se vyskytla v textu. Do statistiky tedy řadíme slovo *compte* jak v singuláru, tak v plurálu, pokud jsou obě varianty v textu přítomny.

Přídavná jména (adjektiva)

- Vyskytuje-li se v textu samostatné heslo, které je uvedeno jak v maskulinu, tak v femininu (*mort, morte; bon, bonne*), do

⁶⁴ To znamená, že „Partre“ se v textu vyskytuje například třikrát, kdežto „Jean-Sol Partre“ pouze jednou.

frekvenčního seznamu je slovo zařazeno pouze ve tvaru maskulina.

- Vyskytuje-li se v textu samostatné heslo, které je uvedeno jak v singuláru, tak v plurálu (*toute, toutes*), do frekvenčního seznamu je slovo zařazeno pouze ve své singulární podobě a jeho tvar v množném čísle je samozřejmě započítán do číslice výskytu.
- Pokud se adjektivum nachází v textu pouze ve femininu či pouze v maskulinu, ve FS je zaneseno jako tvar feminina (či masulina).
- Číslovky jsou na základě normativní gramatiky Maurice Grévisse řazeny do adjektiv.
- Do frekvenčního slovníku zařazujeme pouze číslovky vypsané slovy. Numerická označení typu 638 byla vyřazena.
- Tvary adjektiv, jejichž slova jsou oddělena spojovníkem (např. složená číslovka *trente-huit*), jsou počítány jako jedno heslo.

Determinanty

- U determinantů respektujeme čísla i rody, u hesel je tedy nijak neupravujeme a zaznamenáváme je tak, jak se vyskytují v textu.
- k determinantům řadíme i tvary vzniklé amalgamem – například *au, aux* (vzniklé spojením předložky *à* a determinantu *le, les*) či *du, des* (*de + le, de + les*).

Zájmena (pronomina)

- U zájmen respektujeme čísla i rody a hesla uvádíme ve tvaru, ve kterém se vyskytují v textu.
- Slovo *l'une* či *l'autre* započítáváme jako jedno heslo.
- Slova oddělená spojovníkem započítáváme jako jedno heslo (*celle-ci, celle-là*).

Slovesa (verba)

- Slovesa uvádíme v infinitivu.
- Slovesné tvary typu *il a écrit* chápeme v naší lexikální statistice jako tři slova a dělíme tento tvar tedy na zájmeno *il* a dvě slovesa (*avoir a écrire*).

- Zvratnou podobu sloves typu *se lier* chápeme jako 2 slova.
- Vyskytuje-li se v textu současně zvratná i nezvratná podoba slovesa (*faire; se faire*), pak uvádíme toto sloveso v jeho infinitivním tvaru se zvratným zájmenem v závorce: *faire (se)*. Klíčová je pro nás přítomnost slovesa *faire*, proto se na určování frekvenčnosti jeho zvratné podoby záměrně nesoustředíme a číslo uvedené u slovesa v sobě zahrnuje jak výskyt zvratných, tak nezvratných tvarů slovesa.
- Přísudek jmenný se sponou jako například *(la maison) est faite (de)* vnímáme jako spojení slovesa a přídavného jména.
- Co se týká následného kvantitativního zpracovávání slovesných tvarů, při určování slovesného času vnímáme tvar *(il) a trouvé* jako jednu jednotku a *a trouvé* řadíme tedy celé do *passé composé* (minulý čas složený).

Předložky (prepozice)

- Předložky *par-dessus, pendant que, pour que, avant de, autour de, à travers* a jim podobné jsou počítány jako jedno heslo.

Spojky (konjunkce)

- Spojky *tandis que, dès que* či *quand même* jsou počítány jako jedno heslo.

Příslovce (adverbia)

- Příslovce *peut-être, là-bas* a *vice-versa* je počítáno jako jedno heslo.

L'introducteur

- Zde nebylo nutné stanovovat speciální kritéria.

Mot-phrase

- Francouzský zápor *non* považujeme za *mot-phrase*, stojí-li v textu samostatně jako jednočlenná věta. Je-li součástí věty, vnímáme

ho jako příslovce. Stejným způsobem postupujeme v případě *oui*.

Termíny *introduceur* a *mot-phrase* jsme ponechali v původní francouzské podobě. Jde o specifické slovní druhy, které nemají ekvivalentní pojmenování v české jazykovědné terminologii. Pod pojmem *introduceur* rozumíme neohebný slovní druh, který má schopnost uvádět slova, syntagmata či věty (např. *VOICI/VOILÀ votre journal. EST-CE QUE tu aimes cette maison? QUANT À moi, je ne sais pas.*). Odlišujeme ho od předložky či spojky, jelikož neslouží ke spojování výrazů, ale pouze k jejich uvozování. *Mot-phrase* je taktéž neohebný slovní druh, který sám o sobě může sloužit jako věta. Jedná se buďto o jednoslovné (*Bravo! Bonjour. Merci.* atd.) či víceslovné výrazy (*Tant mieux. Au revoir.* atd.). Při klasifikaci slov jako náležejících k *introduceur* a *mot-phrase* jsme vycházeli z *Le Bon usage* Maurice Grévisse.

3. Statistická analýza vybraných částí literárních děl

3.1 Lexikální statistika

3.1.1 Frekvenční seznamy: *Le Magasin des suicides* (scénář a román) a *L'Écume des jours* (scénář a román)

V této části práce představujeme frekvenční seznamy, které jsme vypracovali na základě jazykového materiálu převzatého ze scénářů k filmům *Le Magasin des suicides* (*Obchod pro sebevrahy*) a *L'Écume des jours* (*Pěna dní*) a z jejich románových předloh. Uvádíme vždy prvních 250 hesel. Celkový počet analyzovaných slov činil pro scénáře 1 205 a 2 007 slov a pro romány 3 323 a 4 362 slov. Pro úplnost uvádíme tabulku, která znázorňuje procentuální rozložení slov s frekvenčností 1, 2, 3, 4 a více než 5 v textech.

	Le Magasin des suicides Scénář	Le Magasin des suicides Román	L'Écume des jours Scénář	L'Écume des jours Román
Slova s frekvenčností 1	20 %	18,90 %	19,43 %	16,46 %
Slova s frekvenčností 2	11,46 %	10,53 %	6,78 %	8,34 %
Slova s frekvenčností 3	5,48 %	5,60 %	7,03 %	5,30 %
Slova s frekvenčností 4	2,65 %	4,94 %	4,58 %	2,39 %
Slova s frekvenčností > 5	60,41 %	60,03 %	62,18 %	67,51 %

V tabulce vidíme, že slova vyskytující se v textu pouze jedenkrát tvoří přibližně 16 až 20 % analyzovaného materiálu, oproti tomu slov opakujících se pětkrát a nalezneme v textu 60 až 68 %.

- **Le Magasin des suicides (Obchod pro sebevrahy) – scénář**

Table hiérarchique

71 être *verbe*
46 ce *pron.*
40 vous *pron.*
37 je *pron.*
36 de *prép.*
22 que *pron.*
22 pas *adv.*
21 ça *pron.*
18 avoir *verbe*
18 ne *adv.*
17 mais *conj.*
16 le *dét.*
16 on *pron.*
15 la *dét.*
15 un *dét.*
14 il *pron.*
14 et *conj.*
13 pour *prép.*
12 faire *verbe*
11 aller *verbe*
11 sourire *verbe*
11 bien *adv.*
11 à *prép.*
10 bon *adj.*

10 pouvoir *verbe*
9 une *dét.*
9 le *pron.*
9 me *pron.*
9 dire *verbe*
9 non *adv.*
8 les *dét.*
8 nous *pron.*
8 tu *pron.*
8 prendre *verbe*
8 plus *adv.*
8 que *conj.*
7 moi *pron.*
7 qui *pron.*
7 si *conj.*
7 ah *mot-phrase*
6 voir *verbe*
6 oui *adv.*
5 madame *subst.*
5 poison *subst.*
5 des *dét.*
5 du *dét.*
5 votre *dét.*
5 en *pron.*

5 quoi *pron.*
5 devoir *verbe*
5 savoir *verbe*
5 comme *adv.*
5 comment *adv.*
5 dans *prép.*
5 ou *conj.*
4 monsieur *subst.*
4 moche *adj.*
4 tout *dét.*
4 se *pron.*
4 vouloir *verbe*
4 alors *adv.*
4 pourquoi *adv.*
4 ben *mot-phrase*
3 Alan *subst.*
3 euros *subst.*
3 fois *subst.*
3 gens *subst.*
3 maman *subst.*
3 sœur *subst.*
3 temps *subst.*
3 belle *adj.*
3 sûr *adj.*

3 ta *dét.*
3 la *pron.*
3 tout *pron.*
3 y *pron.*
3 tenir *verbe*
3 jamais *adv.*
3 là *adv.*
3 puis *adv.*
3 que *adv.*
3 avec *prép.*
3 en *prép.*
3 voilà *introduceur*
3 heu *mot-phrase*
2 balle *subst.*
2 boutique *subst.*
2 effet *subst.*
2 famille *subst.*
2 chose *subst.*
2 kimono *subst.*
2 Marilyn *subst.*
2 nouvelles *subst.*
2 papa *subst.*
2 rêves *subst.*
2 seppuku *subst.*
2 sportif *subst.*
2 suicide *subst.*
2 tarif *subst.*
2 Tuvache *subst.*
2 vie *subst.*
2 beaux *adj.*
2 féminin *adj.*
2 deux *adj.*
2 cher *adj.*
2 joli *adj.*
2 même *adj.*
2 mort *adj.*
2 seule *adj.*
2 triste *adj.*
2 au *dét.*
2 ces *dét.*
2 cette *dét.*
2 notre *dét.*
2 quelque *dét.*
2 sa *dét.*
2 ton *dét.*
2 ils *pron.*
2 rien *pron.*
2 personne *pron.*
2 te *pron.*
2 toi *pron.*
2 arriver *verbe*
2 attendre *verbe*
2 excuser *verbe*
2 falloir *verbe*
2 hésiter *verbe*
2 montrer *verbe*
2 mourir *verbe*
2 préférer *verbe*
2 réfléchir *verbe*
2 regretter *verbe*
2 suffire *verbe*
2 toucher *verbe*
2 trouver *verbe*
2 tuer (se) *verbe*
2 venir *verbe*
2 autant que *adv.*
2 demain *adv.*
2 encore *adv.*
2 mieux *adv.*
2 peu *adv.*
2 si *adv.*
2 tard *adv.*
2 tout *adv.*
2 trop *adv.*
2 chez *prép.*
2 par *prép.*
2 parce que *conj.*
2 soit *conj.*
2 adieu *mot-phrase*
2 au revoir *mot-phrase*
2 bonjour *mot-phrase*
2 merde *mot-phrase*
1 accord *subst.*
1 affaire *subst.*
1 année *subst.*
1 assassins *subst.*
1 bas *subst.*
1 baume *subst.*
1 beurre *subst.*
1 bouche *subst.*
1 calme *subst.*
1 Calmel *subst.*
1 cas *subst.*
1 Célestin *subst.*
1 classe *subst.*
1 cliente *subst.*
1 cœur *subst.*
1 colique *subst.*
1 commerce *subst.*
1 commissures *subst.*
1 contact *subst.*
1 corde *subst.*
1 couple *subst.*
1 courage *subst.*
1 dames *subst.*
1 décès *subst.*
1 décision *subst.*
1 désespéré *subst.*
1 devise *subst.*
1 domicile *subst.*
1 dos *subst.*
1 eau *subst.*
1 enfant *subst.*
1 essence *subst.*
1 face *subst.*
1 fils *subst.*
1 frère *subst.*
1 gamme *subst.*
1 garçons *subst.*
1 grand-père *subst.*
1 grimaces *subst.*
1 gueule *subst.*
1 gym *subst.*
1 habitude *subst.*
1 hara-kiri *subst.*
1 heure *subst.*
1 histoire *subst.*
1 idiote *subst.*
1 immolation *subst.*
1 jour *subst.*
1 laboratoire *subst.*
1 légère *subst.*
1 linceul *subst.*
1 livraison *subst.*
1 Lucrèce *subst.*
1 lycée *subst.*
1 magasin *subst.*
1 messieurs *subst.*
1 métier *subst.*
1 mètres *subst.*
1 Mishima *subst.*
1 moment *subst.*
1 monde *subst.*
1 Mr *subst.*
1 napalm *subst.*
1 nœud *subst.*
1 nuit *subst.*
1 parfum *subst.*
1 passage *subst.*
1 peur *subst.*
1 pile *subst.*
1 Pitou *subst.*
1 plafond *subst.*
1 plaisir *subst.*
1 pli *subst.*
1 poches *subst.*
1 poil *subst.*
1 préau *subst.*
1 produits *subst.*
1 prof *subst.*
1 queue *subst.*
1 raison *subst.*
1 réussite *subst.*
1 révoluer *subst.*
1 samouraï *subst.*
1 service *subst.*
1 soirée *subst.*
1 somme *subst.*
1 sonnette *subst.*
1 sourires *subst.*
1 statistiques *subst.*
1 tentative *subst.*
1 terme *subst.*
1 tête *subst.*
1 tranchant *subst.*
1 truc *subst.*
1 truffe *subst.*
1 urgences *subst.*

• **Le Magasin des suicides (Obchod pro sebevrahy) – román**

Table hiérarchique

150 de <i>prép.</i>	9 ou <i>conj.</i>	5 vers <i>prép.</i>
109 le <i>dét.</i>	8 Alan <i>subst.</i>	5 parce que <i>conj.</i>
95 la <i>dét.</i>	8 cliente <i>subst.</i>	4 an <i>subst.</i>
90 être <i>verbe</i>	8 main <i>subst.</i>	4 chose <i>subst.</i>
66 et <i>conj.</i>	8 tête <i>subst.</i>	4 côté <i>subst.</i>
61 ce <i>pron.</i>	8 votre <i>dét.</i>	4 femme <i>subst.</i>
60 vous <i>pron.</i>	8 en <i>pron.</i>	4 feuille <i>subst.</i>
57 un <i>dét.</i>	8 y <i>pron.</i>	4 fils <i>subst.</i>
50 à <i>prép.</i>	8 pouvoir <i>verbe</i>	4 fois <i>subst.</i>
46 je <i>pron.</i>	8 bien <i>adv.</i>	4 landau <i>subst.</i>
41 se <i>pron.</i>	8 puis <i>adv.</i>	4 Mme <i>subst.</i>
40 que <i>pron.</i>	7 client <i>subst.</i>	4 néon <i>subst.</i>
40 ne <i>adv.</i>	7 gens <i>subst.</i>	4 œil <i>subst.</i>
39 il <i>pron.</i>	7 magasin <i>subst.</i>	4 papier <i>subst.</i>
35 les <i>dét.</i>	7 mère <i>subst.</i>	4 personne <i>subst.</i>
35 avoir <i>verbe</i>	7 me <i>pron.</i>	4 rayon <i>subst.</i>
33 une <i>dét.</i>	7 moi <i>pron.</i>	4 sabre <i>subst.</i>
32 des <i>dét.</i>	7 sortir <i>verbe</i>	4 seppuku <i>subst.</i>
31 pas <i>adv.</i>	7 voir <i>verbe</i>	4 suicide <i>subst.</i>
30 du <i>dét.</i>	7 tandis que <i>conj.</i>	4 tubes <i>subst.</i>
30 en <i>prép.</i>	6 beau <i>adj.</i>	4 deux <i>adj.</i>
29 qui <i>pron.</i>	6 ce <i>dét.</i>	4 mille <i>adj.</i>
27 ça <i>pron.</i>	6 devoir <i>verbe</i>	4 petit <i>adj.</i>
25 elle <i>pron.</i>	6 regarder <i>verbe</i>	4 aux <i>dét.</i>
25 on <i>pron.</i>	6 venir <i>verbe</i>	4 cette <i>dét.</i>
25 faire <i>verbe</i>	6 alors <i>adv.</i>	4 quelque <i>dét.</i>
24 dans <i>prép.</i>	6 aussi <i>adv.</i>	4 ton <i>dét.</i>
19 son <i>dét.</i>	6 oui <i>adv.</i>	4 quoi <i>pron.</i>
17 au <i>dét.</i>	6 que <i>adv.</i>	4 avaler <i>verbe</i>
17 dire <i>verbe</i>	6 entre <i>prép.</i>	4 donner <i>verbe</i>
16 lui <i>pron.</i>	6 par <i>prép.</i>	4 éteindre <i>verbe</i>
16 sur <i>prép.</i>	6 quand <i>conj.</i>	4 penser <i>verbe</i>
15 Tuvache <i>subst.</i>	6 si <i>conj.</i>	4 prendre <i>verbe</i>
15 sa <i>dét.</i>	5 air <i>subst.</i>	4 rendre <i>verbe</i>
15 que <i>conj.</i>	5 bouche <i>subst.</i>	4 tenir <i>verbe</i>
14 aller <i>verbe</i>	5 caisse <i>subst.</i>	4 vouloir <i>verbe</i>
14 mais <i>conj.</i>	5 commerçant <i>subst.</i>	4 ainsi <i>adv.</i>
12 nous <i>pron.</i>	5 dame <i>subst.</i>	4 comment <i>adv.</i>
12 sourire <i>verbe</i>	5 enfant <i>subst.</i>	4 non <i>adv.</i>
12 plus <i>adv.</i>	5 escalier <i>subst.</i>	4 plutôt <i>adv.</i>
12 devant <i>prép.</i>	5 fenêtre <i>subst.</i>	4 pourquoi <i>adv.</i>
12 pour <i>prép.</i>	5 Marilyn <i>subst.</i>	4 chez <i>prép.</i>
11 Mishima <i>subst.</i>	5 poison <i>subst.</i>	4 quand même <i>conj.</i>
11 où <i>pron.</i>	5 visage <i>subst.</i>	3 boutique <i>subst.</i>
11 sous <i>prép.</i>	5 bon <i>adj.</i>	3 comptoir <i>subst.</i>
11 ah <i>mot-phrase</i>	5 cent <i>adj.</i>	3 contact <i>subst.</i>
10 Lucrèce <i>subst.</i>	5 dernier <i>adj.</i>	3 crâne <i>subst.</i>
10 ses <i>dét.</i>	5 ces <i>dét.</i>	3 dernier <i>subst.</i>
10 le <i>pron.</i>	5 te <i>pron.</i>	3 dessin <i>subst.</i>
10 tu <i>pron.</i>	5 tout <i>pron.</i>	3 doigt <i>subst.</i>
10 savoir <i>verbe</i>	5 demander <i>verbe</i>	3 jour <i>subst.</i>
10 là <i>adv.</i>	5 expliquer <i>verbe</i>	3 kimono <i>subst.</i>
9 comme <i>adv.</i>	5 falloir <i>verbe</i>	3 lèvres <i>subst.</i>
9 avec <i>prép.</i>	5 près de <i>prép.</i>	3 M. <i>subst.</i>

3 madame *subst.*
3 maman *subst.*
3 parents *subst.*
3 porte *subst.*
3 produits *subst.*
3 prof *subst.*
3 service *subst.*
3 temps *subst.*
3 truc *subst.*
3 ventre *subst.*
3 Vincent *subst.*
3 cent *adj.*
3 âgée *adj.*
3 autre *adj.*
3 blanc *adj.*
3 féminin *adj.*
3 frais *adj.*
3 grand *adj.*
3 mort *adj.*
3 sûr *adj.*
3 chaque *dét.*
3 leur *dét.*
3 mes *dét.*
3 mon *dét.*
3 notre *dét.*
3 ils *pron.*
3 la *pron.*
3 rien *pron.*
3 appeler *verbe*
3 comprendre *verbe*

3 contempler *verbe*
3 dessiner *verbe*
3 glisser *verbe*
3 hésiter *verbe*
3 mener *verbe*
3 mettre *verbe*
3 ouvrir (s') *verbe*
3 passer (s') *verbe*
3 préparer (s') *verbe*
3 proposer *verbe*
3 regretter *verbe*
3 répondre *verbe*
3 respirer *verbe*
3 retourner (s') *verbe*
3 toucher *verbe*
3 haut *adv.*
3 jamais *adv.*
3 peu *adv.*
3 tout *adv.*
3 très *adv.*
3 trop *adv.*
2 âge *subst.*
2 amis *subst.*
2 appareil *subst.*
2 appartement *subst.*
2 ardoise *subst.*
2 blouse *subst.*
2 boîtes *subst.*
2 capote *subst.*
2 carton *subst.*

2 cheveux *subst.*
2 ciel *subst.*
2 ciment *subst.*
2 cocktail *subst.*
2 colère *subst.*
2 commerçante *subst.*
2 compte *subst.*
2 corde *subst.*
2 coup *subst.*
2 crémone *subst.*
2 désert *subst.*
2 dos *subst.*
2 éducation *subst.*
2 épaule *subst.*
2 exemple *subst.*
2 famille *subst.*
2 fille *subst.*
2 fioles *subst.*
2 fond *subst.*
2 garde *subst.*
2 gaz *subst.*
2 gelée *subst.*
2 genoux *subst.*
2 gondole *subst.*
2 gueule *subst.*
2 gym *subst.*
2 hara-kiri *subst.*
2 histoire *subst.*

• **L'Écume des jours (Pěna dní) – scénář**

Table hiérarchique

84 être *verbe*
78 de *prép.*
60 je *pron.*
46 la *dét.*
43 avoir *verbe*
35 le *dét.*
35 ce *pron.*
34 à *prép.*
29 il *pron.*
29 et *conj.*
27 une *dét.*
26 un *dét.*
25 non *adv.*
23 pas *adv.*
22 ça *pron.*
22 ne *adv.*
21 vous *pron.*
21 aller *verbe*
21 que *conj.*
19 elle *pron.*

18 mais *conj.*
17 les *dét.*
16 pouvoir *verbe*
15 monsieur *subst.*
14 on *pron.*
14 tu *pron.*
13 Nicolas *subst.*
13 des *dét.*
13 se *pron.*
13 te *pron.*
12 moi *pron.*
12 faire (se) *verbe*
12 en *prép.*
12 ah *mot-phrase*
11 dire *verbe*
10 la *pron.*
10 me *pron.*
10 aussi *adv.*
10 avec *prép.*
10 pour *prép.*

10 ben *mot-phrase*
9 Chloé *subst.*
9 mon *dét.*
9 que *pron.*
9 bien *adv.*
9 oui *adv.*
9 très *adv.*
9 voilà *introduceur*
8 Colin *subst.*
8 mon *dét.*
8 dans *prép.*
8 sur *prép.*
7 accord *subst.*
7 Chick *subst.*
7 bon *adj.*
7 y *pron.*
7 falloir *verbe*
7 touiller *verbe*
7 là *adv.*
7 chez *prép.*

7 merci *mot-phrase*
 6 deux *adj.*
 6 du *dét.*
 6 elles *pron.*
 6 rien *pron.*
 6 arrêter *verbe*
 6 vouloir *verbe*
 6 comment *adv.*
 6 par *prép.*
 6 si *conj.*
 5 anguille *subst.*
 5 Isis *subst.*
 5 soleil *subst.*
 5 au *dét.*
 5 celle-là *pron.*
 5 en *pron.*
 5 qui *pron.*
 5 attendre *verbe*
 5 regarder *verbe*
 5 revoir *verbe*
 5 trouver (se) *verbe*
 5 voir *verbe*
 5 oh *mot-phrase*
 4 Alise *subst.*
 4 anniversaire *subst.*
 4 attention *subst.*
 4 fait *subst.*
 4 fille *subst.*
 4 fois *subst.*
 4 Gouffé *subst.*
 4 heure *subst.*
 4 nièce *subst.*
 4 jolie *adj.*
 4 petite *adj.*
 4 ton *dét.*
 4 les *pron.*
 4 connaître *verbe*
 4 dormir *verbe*
 4 laisser *verbe*
 4 mettre *verbe*
 4 rencontrer *verbe*
 4 tenir *verbe*
 4 comme *adv.*
 4 parce que *conj.*
 4 non *mot-phrase*
 4 oui *mot-phrase*
 3 air *subst.*
 3 arc-en-ciel *subst.*
 3 boule *subst.*
 3 cadeau *subst.*
 3 caniche *subst.*
 3 conférence *subst.*
 3 contexte *subst.*
 3 Dupond *subst.*
 3 effet *subst.*
 3 goût *subst.*
 3 champagne *subst.*
 3 chose *subst.*
 3 lit *subst.*
 3 musique *subst.*
 3 note *subst.*
 3 Partre *subst.*
 3 personne *subst.*
 3 pianocktail *subst.*
 3 question *subst.*
 3 compliqué *adj.*
 3 seul *adj.*
 3 sûr *adj.*
 3 ce *dét.*
 3 certain *dét.*
 3 cette *dét.*
 3 chaque *dét.*
 3 ma *dét.*
 3 quelque *dét.*
 3 le *pron.*
 3 lui *pron.*
 3 toi *pron.*
 3 couper *verbe*
 3 croire *verbe*
 3 devoir *verbe*
 3 partir *verbe*
 3 plaire *verbe*
 3 savoir *verbe*
 3 tomber *verbe*
 3 venir *verbe*
 3 alors *adv.*
 3 enfin *adv.*
 3 évidemment *adv.*
 3 parfaitement *adv.*
 3 plus *adv.*
 3 tout *adv.*
 3 trop *adv.*
 3 sans *prép.*
 2 ami *subst.*
 2 ananas *subst.*
 2 bête *subst.*
 2 bigle moi *subst.*
 2 Duke Ellington *subst.*
 2 eau *subst.*
 2 expérience *subst.*
 2 foyer *subst.*
 2 gomme *subst.*
 2 idée *subst.*
 2 maître *subst.*
 2 morceau *subst.*
 2 partenaire *subst.*
 2 passion *subst.*
 2 peau *subst.*
 2 Ponteaupanne *subst.*
 2 route *subst.*
 2 technique *subst.*
 2 amoureux *adj.*
 2 autre *adj.*
 2 connu *adj.*
 2 heureux *adj.*
 2 lent *adj.*
 2 parfait *adj.*
 2 pharmaceutique *adj.*
 2 trois *adj.*
 2 mes *dét.*
 2 quelques *dét.*
 2 ses *dét.*
 2 son *dét.*
 2 vos *dét.*
 2 cela *pron.*
 2 dont *pron.*
 2 tout *pron.*
 2 appâter *verbe*
 2 arriver *verbe*
 2 boire *verbe*
 2 correspondre *verbe*
 2 danser *verbe*
 2 essayer *verbe*
 2 établir *verbe*
 2 faciliter *verbe*
 2 marcher *verbe*
 2 obtenir *verbe*
 2 prendre *verbe*
 2 présenter *verbe*
 2 rater *verbe*
 2 rendre (se) *verbe*
 2 ressembler *verbe*
 2 rester *verbe*
 2 retirer *verbe*
 2 revenir *verbe*
 2 sortir *verbe*
 2 surprendre *verbe*
 2 terminer (se) *verbe*
 2 aujourd'hui *adv.*
 2 déjà *adv.*
 2 ensemble *adv.*
 2 hier *adv.*
 2 maintenant *adv.*
 2 pourquoi *adv.*
 2 toujours *adv.*
 2 avant *prép.*
 2 ou *conj.*
 2 quand *conj.*
 2 voici *introduceur*
 2 bonjour *mot-phrase*
 2 hein *mot-phrase*
 1 accident *subst.*
 1 adepte *subst.*
 1 affaire *subst.*
 1 affection *subst.*
 1 agenda *subst.*
 1 alcool *subst.*
 1 Américaine *subst.*
 1 amour *subst.*
 1 apéritif *subst.*
 1 approche *subst.*
 1 aromate *subst.*
 1 articles *subst.*
 1 avocat *subst.*
 1 barreaux *subst.*
 1 beau-frère *subst.*
 1 Black and Tan Fantasy *subst.*
 1 boogie-woogie *subst.*
 1 cailloux *subst.*

1 calvados *subst.*
1 caravan *subst.*
1 cause *subst.*
1 centimètres *subst.*

1 collège *subst.*
1 comparaison *subst.*
1 complexité *subst.*
1 comprimés *subst.*

1 coup *subst.*
1 cousines *subst.*

• L'Écume des jours (Pěna dní) – román

Table hiérarchique

184 de *prép.*
126 la *dét.*
126 être *verbe*
125 dire *verbe*
110 et *conj.*
98 je *pron.*
96 le *dét.*
96 avoir *verbe*
78 Colin *subst.*
75 à *prép.*
74 il *pron.*
62 Nicolas *subst.*
62 un *dét.*
58 ce *pron.*
53 une *dét.*
48 se *pron.*
48 en *prép.*
46 Chick *subst.*
46 ne *adv.*
40 faire *verbe*
37 des *dét.*
37 pas *adv.*
35 elle *pron.*
34 vous *pron.*
33 les *dét.*
31 aller *verbe*
30 du *dét.*
28 le *pron.*
28 dans *prép.*
27 que *conj.*
25 on *pron.*
25 avec *prép.*
24 pour *prép.*
23 me *pron.*
22 monsieur *subst.*
22 y *pron.*
22 sur *prép.*
21 demander *verbe*
20 Chloé *subst.*
20 ça *pron.*
19 tu *pron.*
19 mettre *verbe*
18 qui *pron.*
18 pouvoir *verbe*
18 par *prép.*
16 Isis *subst.*
16 mais *conj.*

15 que *pron.*
14 au *dét.*
14 lui *pron.*
14 bien *adv.*
13 deux *adj.*
13 savoir *verbe*
12 ses *dét.*
12 en *pron.*
12 moi *pron.*
12 vouloir *verbe*
12 comme *adv.*
11 air *subst.*
11 son *dét.*
11 te *pron.*
11 venir *verbe*
11 alors *adv.*
11 là *adv.*
11 peu *adv.*
11 plus *adv.*
10 tout *adj.*
10 la *pron.*
10 passer *verbe*
9 tête *subst.*
9 sa *dét.*
9 très *adv.*
8 fille *subst.*
8 anguille *subst.*
8 ce *dét.*
8 prendre *verbe*
8 présenter (se) *verbe*
8 regarder *verbe*
8 oui *adv.*
7 biglemoi *subst.*
7 petit *adj.*
7 mon *dét.*
7 nous *pron.*
7 tout *pron.*
7 devoir *verbe*
7 tenir *verbe*
7 que *adv.*
7 trop *adv.*
7 si *conj.*
6 fois *subst.*
6 cuisine *subst.*
6 cette *dét.*
6 elles *pron.*
6 aimer *verbe*

6 revenir *verbe*
6 voir *verbe*
6 comment *adv.*
5 sauce *subst.*
5 pâté *subst.*
5 nièce *subst.*
5 moment *subst.*
5 famille *subst.*
5 eau *subst.*
5 ananas *subst.*
5 Alise *subst.*
5 joli *adj.*
5 quel *adj.*
5 ma *dét.*
5 quelques *dét.*
5 dont *pron.*
5 ils *pron.*
5 où *pron.*
5 arrêter (s') *verbe*
5 boire *verbe*
5 donner *verbe*
5 entrer *verbe*
5 falloir *verbe*
5 rester *verbe*
5 aussi *adv.*
5 encore *adv.*
5 ou *conj.*
5 parce que *conj.*
4 unité *subst.*
4 robe *subst.*
4 pâte *subst.*
4 partie *subst.*
4 mur *subst.*
4 main *subst.*
4 lit *subst.*
4 idée *subst.*
4 cheveux *subst.*
4 glace *subst.*
4 garçon *subst.*
4 façon *subst.*
4 grand *adj.*
4 gros *adj.*
4 mobile *adj.*
4 trois *adj.*
4 connaître *verbe*
4 danser *verbe*
4 suivre *verbe*

4 beaucoup <i>adv.</i>	3 dur <i>adj.</i>	3 presque <i>adv.</i>
4 non <i>adv.</i>	3 frais <i>adj.</i>	3 seulement <i>adv.</i>
4 pourquoi <i>adv.</i>	3 chaque <i>adj.</i>	3 tout <i>adv.</i>
4 chez <i>prép.</i>	3 jaune <i>adj.</i>	3 dessus <i>prép.</i>
4 sans <i>prép.</i>	3 lent <i>adj.</i>	3 devant <i>prép.</i>
4 comme <i>conj.</i>	3 même <i>adj.</i>	3 sous <i>prép.</i>
4 bojnour <i>mot-phrase</i>	3 rouges <i>adj.</i>	3 vers <i>prép.</i>
3 vestiaire <i>subst.</i>	3 sec <i>adj.</i>	3 donc <i>conj.</i>
3 ventre <i>subst.</i>	3 vert <i>adj.</i>	3 eh <i>mot-phrase</i>
3 table <i>subst.</i>	3 ces <i>dét.</i>	3 merci <i>mot-phrase</i>
3 système <i>subst.</i>	3 leurs <i>dét.</i>	2 yeux <i>subst.</i>
3 raison <i>subst.</i>	3 mes <i>dét.</i>	2 voiture <i>subst.</i>
3 point <i>subst.</i>	3 ton <i>dét.</i>	2 vin <i>subst.</i>
3 personne <i>subst.</i>	3 votre <i>dét.</i>	2 vibration <i>subst.</i>
3 pédale <i>subst.</i>	3 cela <i>pron.</i>	2 verre <i>subst.</i>
3 passage <i>subst.</i>	3 l'autre <i>pron.</i>	2 tube <i>subst.</i>
3 Partre <i>subst.</i>	3 l'une <i>pron.</i>	2 terre <i>subst.</i>
3 œuvre <i>subst.</i>	3 ajouter <i>verbe</i>	2 tempo <i>subst.</i>
3 monde <i>subst.</i>	3 arriver <i>verbe</i>	2 souris <i>subst.</i>
3 lame <i>subst.</i>	3 avaler <i>verbe</i>	2 soleil <i>subst.</i>
3 Jean-Sol Partre <i>subst.</i>	3 choisir <i>verbe</i>	2 sol <i>subst.</i>
3 chose <i>subst.</i>	3 consister <i>verbe</i>	2 soie <i>subst.</i>
3 harmonie <i>subst.</i>	3 croire <i>verbe</i>	2 sexe <i>subst.</i>
3 goût <i>subst.</i>	3 dormir <i>verbe</i>	2 sauternes <i>subst.</i>
3 Gouffé <i>subst.</i>	3 enlever <i>verbe</i>	2 salle <i>subst.</i>
3 fin <i>subst.</i>	3 obtenir <i>verbe</i>	2 rythme <i>subst.</i>
3 dinde <i>subst.</i>	3 ouvrir <i>verbe</i>	2 route <i>subst.</i>
3 Dieu <i>subst.</i>	3 permettre <i>verbe</i>	2 résultat <i>subst.</i>
3 crème <i>subst.</i>	3 porter <i>verbe</i>	2 reste <i>subst.</i>
3 cocktail <i>subst.</i>	3 poursuivre <i>verbe</i>	2 registre <i>subst.</i>
3 boutons <i>subst.</i>	3 réduire <i>verbe</i>	2 rangée <i>subst.</i>
3 ans <i>subst.</i>	3 remettre (se) <i>verbe</i>	2 question <i>subst.</i>
3 autre <i>adj.</i>	3 reprendre <i>verbe</i>	2 quadruple <i>subst.</i>
3 beau <i>adj.</i>	3 revoir <i>verbe</i>	2 principe <i>subst.</i>
3 blanc <i>adj.</i>	3 sentir <i>verbe</i>	2 porte <i>subst.</i>
3 compliqué <i>adj.</i>	3 trouver <i>verbe</i>	
3 doré <i>adj.</i>	3 mal <i>adv.</i>	

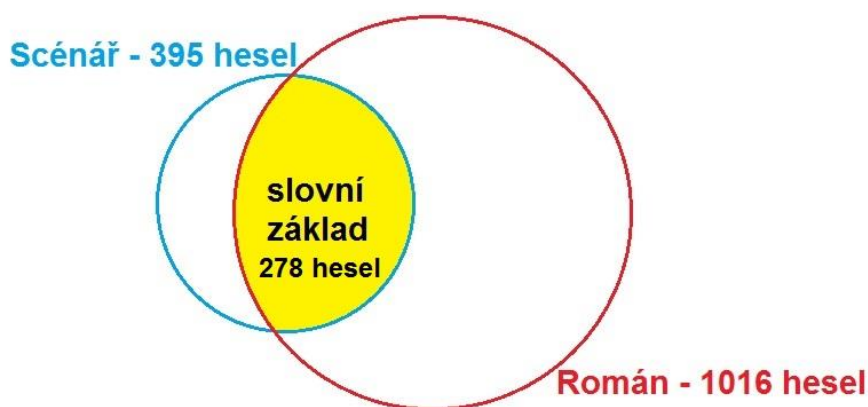
3.1.2 Lexikální komparace scénářů s románovou předlohou

Náš soupis hesel (tj. různých slov) vznikl na základě zpracování vlastního korpusu za podmínek, které jsme si již představili. Před každým heslem nalezneme údaj o absolutní frekvenci, za vlastním heslem poté následuje informace o slovním druhu daného slova.

Následující analýza se zabývá mimo jiné také komparací slovní zásoby obou druhů textů. Jak již bylo řečeno dříve, komparační způsob naší analýzy není vystavěn na porovnávání stejně dlouhých úseků z dialogové listiny

a románu, ale opírá se o lexikální a gramatické srovnávání stejných scén⁶⁵ pocházejících z obou zdrojů. Proto jsou pro nás zajímavější především procentuální hodnoty, tedy zjištění hodnoty části vůči celku, než konkrétní čísla, která jsou, vlivem různých rozsahů zdrojových textů, vzájemně neporovnatelná. Porovnáváním stejných scén se snažíme částečně odhalit lingvistickou složku adaptačního procesu.

Při převedení frekvenčních seznamů (scénáře a romány) do Excelu bylo možné za pomoci funkce odstranění duplikátů zjistit, jaká je vzájemná shoda ve slovní zásobě obou druhů textu. Nyní nám nezáleží na porovnávání frekvenčnosti, ale čistě na tom, jestli se ve scénáři (který se vyznačuje menším počtem slov) vyskytovala stejná slova, jako v románu, a v jakém procentuálním rozsahu. Porovnání ukázalo, že v případě scénáře *Obchodu pro sebevrahy* vychází 70,37 % jeho textu ze slovní zásoby románu a v případě *Pěny dní* vychází scénář z románu z 52,46 % své slovní zásoby. Znamená to tedy, že více než polovina textu scénářů je obsažena v románech. Z pohledu románové předlohy naopak scénář v případě *Obchodu* využil 27,36 % románového textu a scénář *Pěny* využil 32,17 % textu z románové předlohy. Pro větší názornost se podívejme na náčrtek vytvoření pro *Obchod pro sebevrahy*:



V náčrtu vidíme slovní základ společný oběma druhům textu. Slovní základ odpovídá hodnotě 278 hesel. Jedná se o kvantum slov využitých v rámci adaptačního procesu – transpozice syžetu z jednoho média do druhého. Při

⁶⁵ Jedná se vždy o 4–5 scénů vyskytujících se paralelně v románu a ve scénáři.

absenci tohoto slovního základu obsahujícího lexikální pole spojená s námětem literárního díla by adaptace významně oslabil lexikální i tématický charakter své předlohy.

I když naším záměrem nebylo porovnávat scénáře a romány v jejich plné délce a soustředili jsme se pouze na vybrané scény, je tato více než 70 % inspirace scénáře románovou předlohou⁶⁶ a současně necelých 30 % využití původního textu scénářem důvodem k obecnějšímu zamyšlení. Vezmeme-li v úvahu květnatost románového textu⁶⁷, který je prezentován lineárně na stránkách knihy, a současně s tím si uvědomíme dialogický charakter scénáře omezeného pouze na proslovy, přičemž líčení je ztvárněno pomocí mnohvrstevnaté audiovize, jsou tyto hodnoty logickým důsledkem adaptačního procesu. Scénář zcela přirozeně vychází v lexikální skladbě svého textu z dialogických částí románu (a přidává si i slova, která se v předloze nevyskytla) a román se naopak vyznačuje jazykově obohacenou stavbou, kdy dialogy jsou doprovázeny textem, který je ve filmu nahrazen audiovizuálními prostředky. Scénář tak dokáže postihnout asi jen třetinu původního textu. Vypovídá-li procentuální hodnota odpovídající množství textu převzatého z románu do scénáře o věrnosti adaptace, to je otázka ke zvážení, na jejíž zodpovězení se v této práci nesoustředíme. Čistě pro zajímavost: novela Jeana Teulého *Obchod pro sebevrahy* má délku 157 stran. Podklady k jeho filmové adaptaci jsou rozepsány na 111 stran. Vlivem nedlouhého textu novely a četných inscenačních popisů se délka scénáře blíží necelým 71 % délky novely. Tento údaj doplňujeme proto, že měli-li bychom se zabývat otázkou adaptování, musíme si uvědomit, že věrnost adaptace rozhodně nemůže být posuzována na základě grafických údajů a vyžaduje narativní analýzy či komplikovanější komparační postupy.

⁶⁶ V námi vybraných scénách *Obchodu pro sebevrahy*.

⁶⁷ Skládajícího se jak z popisů a líčení, tak z dialogů.

3.1.3 Nejfrekventovanější slova

Nyní se podívejme na prvních 25 nejčastěji se vyskytujících slov tak, jak byla použita ve scénářích a v jejich literárních předlohách.

- **Obchod pro sebevrahy**

Scénář:

1.	71	<i>être</i>	<i>verbe</i>
2.	46	<i>ce</i>	<i>pron.</i>
3.	40	<i>vous</i>	<i>pron.</i>
4.	37	<i>je</i>	<i>pron.</i>
5.	36	<i>de</i>	<i>prép.</i>
6.	22	<i>que</i>	<i>pron.</i>
7.	22	<i>pas</i>	<i>adv.</i>
8.	21	<i>ça</i>	<i>pron.</i>
9.	18	<i>avoir</i>	<i>verbe</i>
10.	18	<i>ne</i>	<i>adv.</i>
11.	17	<i>mais</i>	<i>conj.</i>
12.	16	<i>le</i>	<i>dét.</i>
13.	16	<i>on</i>	<i>pron.</i>
14.	15	<i>la</i>	<i>dét.</i>
15.	15	<i>un</i>	<i>dét.</i>
16.	14	<i>il</i>	<i>pron.</i>
17.	14	<i>et</i>	<i>conj.</i>
18.	13	<i>pour</i>	<i>prép.</i>
19.	12	<i>faire</i>	<i>verbe</i>
20.	11	<i>aller</i>	<i>verbe</i>
21.	11	<i>sourire</i>	<i>verbe</i>
22.	11	<i>bien</i>	<i>adv.</i>
23.	11	<i>à</i>	<i>prép.</i>
24.	10	<i>bon</i>	<i>adj.</i>
25.	10	<i>pouvoir</i>	<i>verbe</i>

Román:

1.	150	<i>de</i>	<i>prép.</i>
2.	109	<i>le</i>	<i>dét.</i>
3.	95	<i>la</i>	<i>dét.</i>
4.	90	<i>être</i>	<i>verbe</i>
5.	66	<i>et</i>	<i>conj.</i>
6.	61	<i>ce</i>	<i>pron.</i>
7.	60	<i>vous</i>	<i>pron.</i>
8.	57	<i>un</i>	<i>dét.</i>
9.	50	<i>à</i>	<i>prép.</i>
10.	46	<i>je</i>	<i>pron.</i>
11.	41	<i>se</i>	<i>pron.</i>
12.	40	<i>que</i>	<i>pron.</i>
13.	40	<i>ne</i>	<i>adv.</i>
14.	39	<i>il</i>	<i>pron.</i>
15.	35	<i>les</i>	<i>dét.</i>
16.	35	<i>avoir</i>	<i>verbe</i>
17.	33	<i>une</i>	<i>dét.</i>
18.	32	<i>des</i>	<i>dét.</i>
19.	31	<i>pas</i>	<i>adv.</i>
20.	30	<i>du</i>	<i>dét.</i>
21.	30	<i>en</i>	<i>prép.</i>
22.	29	<i>qui</i>	<i>pron.</i>
23.	27	<i>ça</i>	<i>pron.</i>
24.	25	<i>elle</i>	<i>pron.</i>
25.	25	<i>on</i>	<i>pron.</i>

Při porovnání hesel umístěných na prvních místech FS vidíme shodu v 68 % užitých slov. Pozice zbylých 32 % slov ve FS se liší. V případě scénáře se do prvních 25 nejfrekventovanějších slov dostalo více základních sloves typu *être*⁶⁸, *faire*, *aller*, *pouvoir*. Románový text obsahuje velké množství zájmen, determinantů a předložek. Můžeme si také povšimnout, že v obou případech (i když na různých pozicích frekvenčnosti) jsou prvními osobními zájmeny *vous* a *je* poukazující na charakter námětu, který se vyznačuje komunikací se zákazníky obchodu (oslovování: *vous* a mluva v první osobě: *je*). Zatímco v knižní podobě se do první pětadvacítky dostalo také zájmeno *il* a *elle*, ve frekvenčním seznamu náležejícím scénáři se setkáváme pouze s *il*. Tato

⁶⁸ Které dokonce obsadilo pozici nejfrekventovanějšího slova.

redukce zájmena *elle* může být částečným důsledkem hovorového charakteru mluveného projevu, kdy bývají objekty bez ohledu na rod označovány zájmenem *ça*, které se v případě scénáře nachází na vyšší příčce frekvenčnosti než u románu. Stejná situace se opakuje se zájmenem *on*. Z povahy francouzštiny, která svá podstatná jména doprovází členy určitými a neurčitými v antepozici a svá slovesa osobními zájmeny (s výjimkou slovesných tvarů v infinitivu, imperativu, přechodníku atd.) vyplývá, že na prvních místech by měly figurovat kromě zájmen také členy, což se při pohledu na seznamy potvrzuje.

- **Pěna dní**

Scénář

1. 84 *être* *verbe*
2. 78 *de* *prép.*
3. 60 *je* *pron.*
4. 46 *la* *dét.*
5. 43 *avoir* *verbe*
6. 35 *le* *dét.*
7. 35 *ce* *pron.*
8. 34 *à* *prép.*
9. 29 *il* *pron.*
10. 29 *et* *conj.*
11. 27 *une* *dét.*
12. 26 *un* *dét.*
13. 25 *non* *adv.*
14. 23 *pas* *adv.*
15. 22 *ça* *pron.*
16. 22 *ne* *adv.*
17. 21 *vous* *pron.*
18. 21 *aller* *verbe*
19. 21 *que* *conj.*
20. 19 *elle* *pron.*
21. 18 *mais* *conj.*
22. 17 *les* *dét.*
23. 16 *pouvoir* *verbe*
24. 15 *monsieur* *subst.*
25. 14 *on* *pron.*

Román

1. 184 *de* *prép.*
2. 126 *la* *dét.*
3. 126 *être* *verbe*
4. 125 *dire* *verbe*
5. 110 *et* *conj.*
6. 98 *je* *pron.*
7. 96 *le* *dét.*
8. 96 *avoir* *verbe*
9. 78 *Colin* *subst.*
10. 75 *à* *prép.*
11. 74 *il* *pron.*
12. 62 *Nicolas* *subst.*
13. 62 *un* *dét.*
14. 58 *ce* *pron.*
15. 53 *une* *dét.*
16. 48 *se* *pron.*
17. 48 *en* *prép.*
18. 46 *Chick* *subst.*
19. 46 *ne* *adv.*
20. 40 *faire* *verbe*
21. 37 *des* *dét.*
22. 37 *pas* *adv.*
23. 35 *elle* *pron.*
24. 34 *vous* *pron.*
25. 33 *les* *dét.*

V případě *Pěny dní* je shoda v otázce slovní frekvenčnosti jen drobně rozdílná. Hesla se zde shodují z 64 %. Zbýlých 36 % představuje charakteristickou rozdílnost obou médií. Stejně jako u *Obchodu pro sebevrahy* se do popředí dostalo sloveso *être*, zájmena *je*, *vous*, *il* a *elle*, členy a předložky. Za zmínku však stojí sloveso *dire* hojně využívané v románu. Sloveso *dire* (říct) se totiž pojilo téměř výhradně s uvozováním přímé řeči. Toto

uvozování se ve scénáři přirozeně nevyskytovalo. V literární předloze adaptace také často figurovala jména hlavních postav – *Colin*, *Chick* a *Nicolas* – a to často právě ve spojení se slovesem *dire*. Na spisovnost a literární hodnotnost Vianova románu také upozorní fakt, že v jeho scénáristickém protějšku se na prvních místech umístila hovorová slova *on* a *ça*, která na prvních místech FS románu nenajdeme.

Stejně jako v naší předchozí diplomové práci se i nyní podíváme na prvních 25 nejčastěji užívaných slov pocházejících z frekvenčního seznamu lexikologa Étiennea Bruneta (vlevo) a porovnejme je s našimi výsledky z románu *Obchod pro sebevrahy* (vpravo). Korpus vytvořený z románu obsahoval 3 323 slov a získali jsme z něj 1 016 hesel (různých slov).

1.	le <i>dét.</i>	1050561	1.	150 de <i>prép.</i>
2.	de <i>prép.</i>	862100	2.	109 le <i>dét.</i>
3.	un <i>dét.</i>	419564	3.	95 la <i>dét.</i>
4.	être <i>verbe</i>	351960	4.	90 être <i>verbe</i>
5.	et <i>conj.</i>	362093	5.	66 et <i>conj.</i>
6.	à <i>prép.</i>	293083	6.	61 ce <i>pron.</i>
7.	il <i>pron.</i>	270395	7.	60 vous <i>pron.</i>
8.	avoir <i>verbe</i>	248488	8.	57 un <i>dét.</i>
9.	ne <i>adv.</i>	186755	9.	50 à <i>prép.</i>
10.	je <i>pron.</i>	184186	10.	46 je <i>pron.</i>
11.	son <i>dét.</i>	181161	11.	41 se <i>pron.</i>
12.	que <i>conj.</i>	176161	12.	40 que <i>pron.</i>
13.	se <i>pron.</i>	168684	13.	40 ne <i>adv.</i>
14.	qui <i>pron.</i>	148392	14.	39 il <i>pron.</i>
15.	ce <i>dét.</i>	141389	15.	35 les <i>dét.</i>
16.	dans <i>prép.</i>	139185	16.	35 avoir <i>verbe</i>
17.	en <i>prép.</i>	143565	17.	33 une <i>dét.</i>
18.	du <i>dét.</i>	127384	18.	32 des <i>dét.</i>
19.	elle <i>pron.</i>	126397	19.	31 pas <i>adv.</i>
20.	au <i>dét.</i>	123502	20.	30 du <i>dét.</i>
21.	de <i>dét.</i>	119106	21.	30 en <i>prép.</i>
22.	ce <i>pron.</i>	107074	22.	29 qui <i>pron.</i>
23.	le <i>pron.</i>	105873	23.	27 ça <i>pron.</i>
24.	pour <i>prép.</i>	104779	24.	25 elle <i>pron.</i>
25.	pas <i>adv.</i>	103083	25.	25 on <i>pron.</i>

Kompletní seznam E. Bruneta čítá 1 500 hesel a končí slovem s frekvenční hodnotou 412. Náš seznam má pouhých 1016 hesel. Přestože náš výchozí korpus je výrazně menší, na první pohled jsou mezi oběma seznamy patrné četné podobnosti. Pozice jednotlivých slov v každém ze seznamů se většinou mírně liší, obecně lze však říci, že prvních 25 slov označovaných za nejfrekventovanější se v obou seznamech shoduje ze 72 %.

To považujeme za poměrně pozoruhodné, zvláště přihlédneme-li k naší skromné materiálové základně, která svým rozsahem nemůže výzkumu E. Bruneta konkurovat.

„Při posuzování poměru, v jakém jsou zastoupeny jednotlivé druhy slov v užitém materiálu, musíme rozlišovat frekvenci hesel a frekvenci slov.“ (Jelínek; Bečka; Těšitelová 1961, s. 30). Poměr mezi frekvencí slov (všech užitých) a frekvencí hesel (užitých různých slov) nazýváme *indexem opakování*. Víme-li tedy, že v našem textu je užito 1 016 různých slov (frekvence hesel) a celkem je v něm 3 323 slov (frekvence slov), můžeme určit, že heslo se tedy průměrně opakuje 3,3 x (index opakování). Index opakování je tedy do jisté míry ukazatelem bohatosti slovníku zkoumaného jazykového projevu. Ze zkušeností autorů při tvorbě *Frekvenčního slovníku češtiny* vyplývá, že se největším indexem opakování vyznačují mluvené projevy. Naopak nejmenší index opakování nalezneme například u básnických sbírek. Prohlédněme si tabulku indexu opakování pro jednotlivé zdroje:

INDEX OPAKOVÁNÍ	Filmový scénář	Románová předloha
<i>L'Écume des jours</i>	2,9x	3,9x
<i>Le Magasin des suicides</i>	3x	3,3x

Budeme-li se snažit postihnout index opakování v procentuální podobě, můžeme uvažovat například takto: považujeme-li za 100 % počet všech užitých slov (tedy v případě románu *Obchod pro sebevrahy* 3 323 slov), tvoří v našem případě počet různých slov 30,57 % (1 016 hesel). Naproti tomu 69,43 % označuje procento slov, která se v textu vyskytují více než jednou (tedy dvakrát a více). V rámci našeho seznamu hesel (tj. frekvenčního seznamu) se 18,90 % slov (tj. současně 628 hesel) vyskytuje v textu pouze jednou. Zbýlých 2 695 slov (388 jedinečných hesel) má četnost výskytu v textu větší než jedna, přičemž nejvíce se při frekvenci 150 opakuje předložka *de*.

3.2 Gramatická statistika

3.2.1 Zastoupení slovních druhů v textu

Komentář

Níže uvedená tabulka charakterizuje zastoupení jednotlivých slovních druhů v románech a scénářích. Hodnotíme procentuální zastoupení každého slovního druhu vzhledem k počtu slov v tom kterém druhu textu. Seřadíme-li slovní druhy podle jejich výskytu v celých textech, dostaneme tato pořadí:

Obchod pro sebevrahy – román

1.	substantiva	21,61 %
2.	verba	17,55 %
3.	pronomina	15,89 %
4.	determinanty	15,64 %
5.	prepozice	10,74 %
6.	adverbia	7,01 %
7.	adjektiva	6,77 %
8.	konjunkce	3,73 %
9.	mot-phrase	0,90 %
10.	l'introducteur	0,15 %

Obchod pro sebevrahy – scénář

1.	pronomina	23,32 %
2.	verba	21,16 %
3.	substantiva	13,86 %
4.	adverbia	12,20 %
5.	determinanty	9,13 %
6.	prepozice	6,47 %
7.	adjektiva	6,22 %
8.	konjunkce	4,98 %
9.	mot-phrase	2,41 %
10.	l'introducteur	0,25 %

V porovnání románu a scénáře vidíme změny v obsazenosti slovních druhů v textu. Zatímco u románu je nejčastějším slovním druhem podstatné jméno (substantivum), ve scénáři došlo k nahrazení substantiva zájmenem (pronominem). Souvisí to s větší úsporností hovorového jazyka, který nemá tendenci konkrétně pojmenovávat a objekty a osoby označuje spíše zájmeny.

Pěna dní – román

1.	substantiva	22,32 %
2.	verba	21,32 %
3.	pronomina	14,74 %
4.	determinanty	12,61 %
5.	prepozice	10,67 %
6.	adjektiva	7,08 %
7.	adverbia	6,63 %
8.	konjunkce	4,24 %
9.	mot-phrase	0,36 %
10.	l'introducteur	0,02 %

Pěna dní – scénář

1.	verba	20,48 %
2.	substantiva	18,04 %
3.	pronomina	17,14 %
4.	determinanty	10,96 %
5.	adverbia	9,47 %
6.	prepozice	9,22 %
7.	adjektiva	6,93 %
8.	konjunkce	4,63 %
9.	mot-phrase	2,59 %
10.	l'introducteur	0,54 %

Vidíme, že nejpočetnější skupinou jsou u románu opět substantiva, která pokrývají o něco více než 1/5 textu. S nimi je spojená četnost determinantů, které substantiva v podobě členů určitých a neurčitých ve většině případů doprovázejí. Slovesné časy ve francouzštině bývají zase provázeny osobními zájmeny (*je, il, elle* apod.) zastupujícími substantiva. Zatímco v češtině je osoba vyjádřena tvarem slovesa (*řekla jsem, jdu*), ve francouzštině (ale také například v angličtině a němčině) jsou zájmena anteponována před slovesné tvary (*j'ai dit, je vais*), což má zákonitě vliv na frekvenci osobních zájmen v těchto jazycích. Také četnost předložek má spojitost s výskytem sloves či substantiv. Jsou to právě předložkové vazby, které umožňují velkému počtu sloves a substantiv syntaktické zapojení do textu.

Porovnáme-li sekvence románu a scénáře mezi sebou, všimneme si u románu vyšší procentuální hodnoty u kategorie substantiv, adjektiv a determinantů, což v zásadě odpovídá povaze textu – písemného projevu, a snaze autora situace podrobně popisovat a hodnotit. V případě scénáře *Pěny dní* nám neunikne ani vysoká frekvence sloves, která může být do jisté míry ovlivněna velkým počtem imperativů v textu. Šéfkuchař Gouffé radí Nicolasovi, jak si počínat v kuchyni při přípravě gastronomických specialit. Opakují se příkazy typu: okořeňte, zalijte omáčkou, osolte, oloupejte atd. Po takovýchto slovesných příkazech následuje podstatné jméno předznamenané determinanem, čímž se vysvětluje jejich zvýšená frekvenčnost oproti scénáři k *Obchodu pro sebevrahy*.

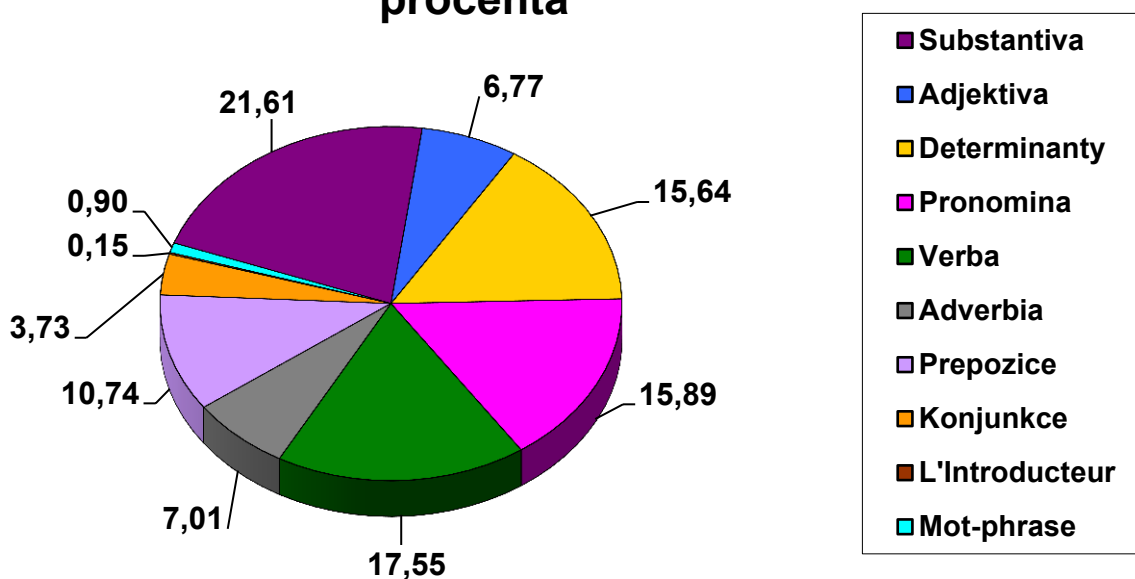
Početní a procentuální vyjádření slovních druhů zastoupených ve vybraných textech

	Počet slov ROMÁN Obchod pro sebevrahy	% určitěho slovního druhu v R / OPS	Počet slov SCÉNÁŘ Obchod pro sebevrahy	% určitěho slovního druhu v S / OPS	Počet slov ROMÁN Pěna dní	% určitěho slovního druhu v R / PD	Počet slov SCÉNÁŘ Pěna dní	% určitěho slovního druhu v S / PD
1. substantiva	718	21,61%	167	13,86%	974	22,32%	362	18,04%
2. adjektiva	225	6,77%	75	6,22%	309	7,08%	139	6,93%
3. determinanty	520	15,64%	110	9,13%	550	12,61%	220	10,96%
4. pronomina	528	15,89%	281	23,32%	643	14,74%	344	17,14%
5. verba	583	17,55%	255	21,16%	930	21,32%	411	20,48%
6. adverbia	233	7,01%	147	12,20%	289	6,63%	190	9,47%
7. prepozice	357	10,74%	78	6,47%	465	10,67%	185	9,22%
8. konjunktce	124	3,73%	60	4,98%	185	4,24%	93	4,63%
9. l'introducteur	5	0,15%	3	0,25%	1	0,02%	11	0,54%
10. mot-phrase	30	0,90%	29	2,41%	16	0,36%	52	2,59%
celkem	3323	100%	1205	100%	4362	100%	2007	100%

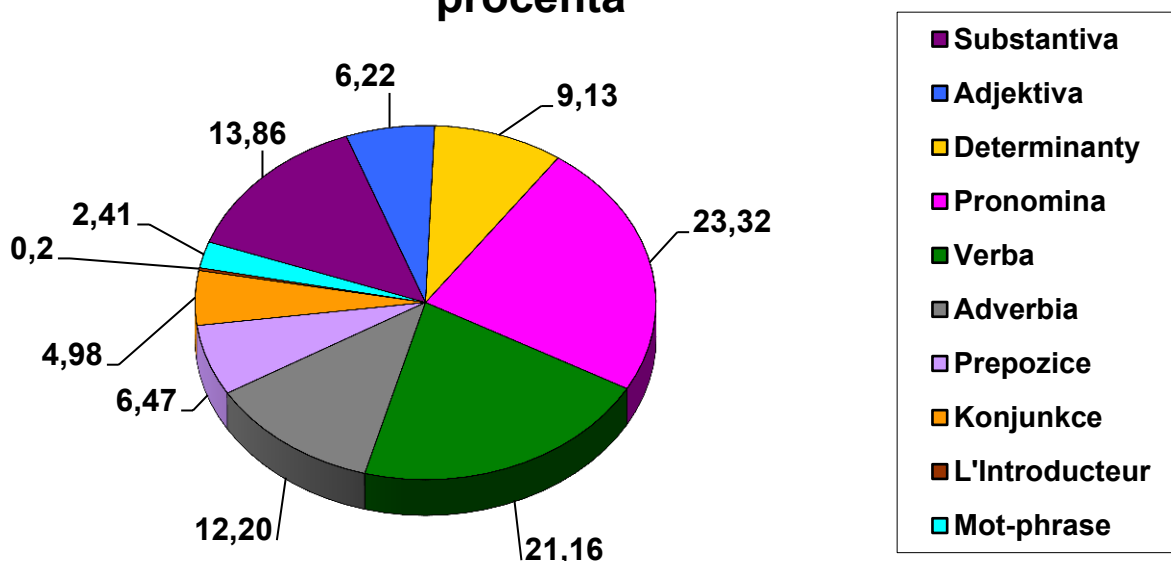
Ve scénářích vidíme jistou záměnu části substantiv za pronomina. Jak jsme již zmínili, větší počet zájmen v případě mluvených projevů je přirozený, protože v nich je nejvíce užito přímého oslovení a dialogu. V rámci dialogových promluv ve scénářích předpokládáme, že z důvodů spontánnosti či úspornosti vyjadřování jsou některá vlastní nebo obecná jména nahrazena zájmeny *il, elle* atd., což celou komunikaci mezi mluvčími a posluchači zjednodušuje, případně zpřehledňuje. Logickým důsledkem jakési nenucenosti v mluveném projevu (ve scénářích) je také nepatrně vyšší četnost slov u kategorie *introduceur* a *mot-phrase*.

Výsledky jsme zapracovali do tabulky a pro větší názornost také do koláčových grafů. Jelikož se grafy nijak významně neliší, uvádíme pouze výstup z románu a scénáře *Obchodu pro sebevrahy*.

Rozložení slovních druhů v románu - procenta



Rozložení slovních druhů ve scénáři - procenta



3.2.2 Kvantifikace kategorie rodu a čísla u jména

Při statistickém zpracování údajů v kategorii rodu a čísla u substantiv jsme vycházeli ze seznamu všech podstatných jmen užitých v textu (tedy nikoliv ze seznamu hesel, ale ze seznamu slov). Na základě analýzy seznamů jsme vypočítali následující hodnoty:

- **Obchod pro sebevrahy román (R) a scénář (S)** – 718 a 167 substantiv

	Femininum R	Femininum S	Maskulinum R	Maskulinum S
singulár	244	69	355	80
plurál	56	3	63	15

- **Pěna dní román (R) a scénář (S)** – 974 a 362 substantiv

	Femininum R	Femininum S	Maskulinum R	Maskulinum S
singulár	321	146	523	169
plurál	61	25	69	22

Sledujeme-li frekvenci rodů a čísel u substantiv z procentuálního hlediska, zjišťujeme tento poměr:

Obchod pro sebevrahy – scénář:

maskulina	56,89 %
feminina	43,11 %
singulár	89,22 %
plurál	10,78 %

Pěna dní – scénář:

maskulina	52,76 %
feminina	47,24 %
singulár	87,02 %
plurál	12,98 %

Obchod pro sebevrahy – román:

maskulina	58,22 %
feminina	41,78 %
singulár	83,43 %
plurál	16,57 %

Pěna dní – román:

maskulina	60,78 %
feminina	39,22 %
singulár	86,65 %
plurál	13,35 %

Vidíme, že procento singuláru ve všech případech výrazně převyšuje plurál a také že rod mužský v singuláru je nejpočetnější. Marie Těšitelová ve své příručce uvádí, že v textech uměleckého stylu (alespoň tedy v českém prostředí) mívá rod mužský výrazně větší frekvenci než například v textech odborných (Těšitelová 1977, s. 84). My se v naší analýze soustředíme pouze na umělecké texty, avšak už na množství vzorků, které máme k dispozici⁶⁹, můžeme pozorovat, že ona hodnotová hierarchie mezi maskulinem a femininem a singulárem a plurálem není náhodná a že je pravděpodobně vlastní každému francouzskému textu.

⁶⁹ A to jak z této, tak z předchozí diplomové práce.

3.2.3 Kvantifikace slovesných tvarů

Kvantifikovat můžeme také slovesné kategorie. Zaměřili jsme se tedy na frekvenci slovesných tvarů. Podívejme se na následující tabulky:

- **Obchod pro sebevrahy – román a scénář**

	Román		Scénář	
	počet	%	počet	%
INDICATIF				
présent	363	64,82 %	151	62,40 %
passé composé	16	2,86 %	4	1,66 %
imparfait	17	3,04 %	6	2,48 %
passé récent	0	0 %	0	0 %
plus-que-parfait	1	0,18 %	2	0,83 %
passé simple	0	0 %	0	0 %
futur simple	19	3,39 %	7	2,89 %
futur proche	6	1,07 %	7	2,89 %
SUBJONCTIF				
présent	1	0,18 %	1	0,41 %
passé	0	0 %	0	0 %
imparfait	0	0 %	0	0 %
CONDITIONNEL				
présent	12	2,14 %	12	4,96 %
passé	0	0 %	0	0 %
IMPÉRATIF				
présent	22	3,93 %	17	7,02 %
PARTICIPE				
présent	13	2,32 %	0	0 %
passé	0	0 %	0	0 %
INFINITIF				
présent	78	13,93 %	35	14,46 %
passé	0	0 %	0	0 %
GÉRONDIF				
présent	12	2,14 %	0	0 %

- **Pěna dní – román a scénář**

	Román		Scénář	
	počet	%	počet	%
INDICATIF				
présent	262	31,30 %	175	47,43 %
passé composé	66	7,89 %	32	8,68 %
imparfait	83	9,92 %	12	3,26 %
passé récent	0	0 %	1	0,27 %
plus-que-parfait	7	0,84 %	2	0,54 %
passé simple	187	22,34 %	0	0 %
futur simple	13	1,55 %	4	1,08 %
futur proche	17	2,03 %	10	2,71 %
SUBJONCTIF				
présent	6	0,71 %	4	1,08 %
passé	2	0,24 %	0	0 %
imparfait	0	0 %	1	0,27 %
CONDITIONNEL				
présent	16	1,91 %	6	1,63 %
passé	1	0,12 %	1	0,27 %
IMPÉRATIF				
présent	38	4,54 %	53	14,36 %
PARTICIPE				
présent	7	0,84 %	1	0,27 %
passé	2	0,24 %	2	0,54 %
INFINITIF				
présent	120	14,34 %	62	16,80 %
passé	0	0 %	1	0,27 %
GÉRONDIF				
présent	10	1,19 %	2	0,54 %

V tabulkách můžeme vysledovat, že vůbec nejfrekventovanějším slovesným tvarem je v obou případech indikativ přítomného času. Poté se četnost užívání slovesných časů liší. Na druhém místě následuje téměř ve všech druzích textu infinitiv přítomného času. Výjimkou je však románová podoba *Pěny dní*,

kteřá hned po indikativu prézentu vyniká častým užíváním *passé simple*, což je jednoduchý⁷⁰ minulý čas, který nalezneme především v literatuře. Zatímco Boris Vian tento čas ve svém románu užívá zpravidla vždy, píše-li v minulém čase, Jean Teulé dává ve své novele přednost prézentu⁷¹ a budoucímu času *futur simple*. Minulost pak Teulé vyjadřuje za pomoci *passé composé* a *imparfait*. Absence *passé simple* u současné francouzské novely publikované roku 2008 může být jedním ze znaků moderní literatury, která se stává více konverzační, jazykově méně komplikovaná a současně má tendence zapojovat více hovorových prvků jazyka. Jak je patrné z výsledku analýzy slovesných tvarů, frekvence slovesných tvarů a jejich kategorií je podmíněna nejen vlastnostmi jazyka, ale závisí také na druhu projevu (psaného či mluveného) a na funkčním stylu či stylových postupech autora.

Nyní se podívejme na tabulku zaznamenávající četnost kategorie osoby a čísla u slovesných tvarů:

- **Obchod pro sebevrahy – román a scénář**

	Román		Scénář	
	počet	%	počet	%
1. os. sg	49	10,68 %	39	18,84 %
2. os. sg	23	5,01 %	12	5,80 %
3. os. sg	293	63,83 %	106	51,21 %
1. os. pl.	2	0,44 %	4	1,93 %
2. os. pl.	57	12,41 %	38	18,36 %
3. os. pl.	35	7,63 %	8	3,86 %

⁷⁰ Ve smyslu „ne složený“.

⁷¹ Kterým vyjadřuje jak přítomnost, tak minulost a budoucnost.

- **Pěna dní – román a scénář**

	Román		Scénář	
	počet	%	počet	%
1. os. sg	96	13,85 %	69	22,92 %
2. os. sg	33	4,76 %	31	10,30 %
3. os. sg	470	67,82 %	133	44,19 %
1. os. pl.	6	0,87 %	4	1,33 %
2. os. pl.	49	7,07 %	48	15,95 %
3. os. pl.	39	5,63 %	16	5,31 %

Na první pohled si všimneme, že nejčastěji se ve všech analyzovaných textech vyskytovala 3. osoba singuláru a to s poměrně značným náskokem oproti ostatním osobám. Na druhém místě se ve většině textů vyskytuje 1. osoba singuláru následovaná 2. osobou plurálu. Výjimkou je pouze novela *Obchod pro sebevrahy*, kde 2. osoba pl. předstihla 1. os. sg, což může být opět dáno námětem a tématem díla, kde významnou část komunikace tvoří rozhovory se zákazníky obchodu, kterým se vyká.

Závěr

Jak je patrné z naší diplomové práce, kvantitativní zkoumání jazyka je poměrně náročnou činností. Je-li analýza navíc prováděna bez pomoci specializovaných programů, situace se tím značně komplikuje. S vědomím této slabiny jsme zjištěná data podrobovali o to větší kontrole, aby nedocházelo ke zbytečným chybám, které by zkreslovaly náš výzkum. Zpracovávání textů vyžaduje systematickosti a promyšlené postupy, které mohou usnadnit realizaci následných dílčích analýz. Chybně načasovaným krokem totiž může dojít ke ztrátě dat potřebných pro analýzu zařazenou dále do plánu. I přes to, že kvantitativní výzkumy jazyka si vyžadují větší tým pracovníků, podařilo se nám v rámci této práce analyzovat 5 a ½ krát více textu než v naší práci předchozí.

Úkolem teoretické části práce bylo podat látku zábavným a srozumitelným způsobem. Soustředili jsme se na představení francouzštiny psané i mluvené, na její lexikální a morfosyntaktickou charakteristiku, na její historické kořeny i současný stav. S ohledem na statistické zaměření naší práce jsme 4. a 5. kapitole věnovali výkladu o kvantitativní lingvistice a její terminologii a neopomněli jsme uvést přehled frekvenčních slovníků a mluvených korpusů francouzského jazyka.

Do výkladu jsme měli snahu zapojit také aktuální zajímavosti týkající se vztahu Francouzů k rodnému jazyku a zabývali jsme se i moderními trendy jako je soutěž v pravopise, zákony a nařízení týkající se spisovné podoby francouzského jazyka či jazyk SMS, kde jsme představili základní principy jeho užívání a doložili tak významný vliv mluvené podoby jazyka v užívání novodobých technologií a v každodenním životě. V části o kvantitativní lingvistice jsme odhalili souvislost grafematické statistiky s počítačovou klávesnicí QWERTY a podívali se na praktické využití lexikální statistiky při analýze politických projevů. Právě těmito kapitolami, které se mírně vzdalují od hlavního tématu naší práce, se snažíme čtenáři francouzský jazyk co nejvíce zpřítomnit a ukázat, že neexistuje jen v podobě frekvenčních seznamů, grafů či na stránkách učebnic, ale že jde, tak jako v případě každého jazyka, o živoucí jazykový systém, který se neustále vyvíjí, obměňuje, modernizuje a který je

nástrojem aktivní komunikace. Právě možnost aktivně komunikovat francouzsky je dle našeho názoru cíl, který si mnoho studentů francouzštiny, kteří pronikají do tajů složité francouzské gramatiky a lexika, musí neustále připomínat a posilovat tak svou motivaci. Právě statistická analýza jazyka může být jedním z nástrojů, který jim ulehčí první kroky při seznamování se s jazykovou strukturou, neboť díky ní je možné vysledovat, které gramatické jevy a která slova se v jazyce vyskytují nejčastěji a co je tedy vhodné osvojit si v jazyce jako první. Pro pokročilejšího znalce francouzského jazyka se pak naše práce může stát inspirací k dalšímu bádání v oblasti matematické lingvistiky a může pomoci probudit zájem o matematický pohled na jazyk, který dokáže přinést zajímavé a překvapivé výsledky.

Frekvenční analýza, rozebírající francouzský text z perspektivy lexikální a gramatické statistiky, vznikala na základě zpracování vybraných částí filmových scénářů a jejich literárních předloh. K analýze jsme záměrně vybrali jedny z nejnovějších filmových adaptací románových děl francouzské literatury. Především adaptace románu Borise Viana je důkazem nesmrtelnosti jeho uměleckého odkazu.

Oba druhy textů jsme vzájemně porovnali a vyjádřili se k lišícím se procentuálním hodnotám rozložení slovních druhů a slovesných kategorií v textu. Dialogy, na kterých je scénář postaven, dokážou do jisté míry simulovat mluvený projev. Tento fakt se stal hlavním rozlišujícím prvkem mezi textem filmového scénáře a románem. Z analýzy jsou patrné rozdíly obou médií. Tyto rozdíly plynou z jejich odlišných povah. Zatímco psaný text má za úkol rozvíjet čtenářovu představivost prostřednictvím květnatého písemného projevu, scénář se v adaptačním procesu soustředí spíše na zachycení dialogových promluv z románu a popis a líčení vyjadřuje prostřednictvím filmových prostředků audiovizuálního charakteru. Výsledkem je dialogová listina (scénář), která se skládá přibližně z jedné třetiny textu užitého v románu, jak je patrné z naší analýzy vybraných scén.

Ve výkladu v analytické části naší práce se dotkneme také problematiky adaptace, o které se začalo v badatelských kruzích přemýšlet poměrně

nedávno. Filmové adaptace se dlouhou dobu považovaly za nedokonalou náhražku literárních předloh a často se zdůrazňovala jejich neschopnost zachytit věrný obraz fikčního světa tak, jak to umožňuje literatura. Z frekvenční analýzy můžeme těžko vyvozovat zobecňující závěry týkající se věrnosti adaptace své literární předloze. Doufáme však, že jsme inspirovali čtenáře ke zhlédnutí a přečtení analyzovaných děl. Až tak učiní, budou mít možnost sami posoudit, nakolik filmová adaptace poškozuje či pozdvihuje literární dílo, na které se odkazuje. Dle našeho názoru jsou obě námi vybrané adaptace ukázkou zdařilé práce s původním textem, se kterým se potýkají velmi kreativně a osobitě a přitom zůstávají poměrně věrné charakteru původního zpracování. *Obchod pro sebevrahy* i *Pěnu dní* můžeme jediné doporučit.

Résumé

Notre thèse est consacrée à la linguistique quantitative appliquée au langage parlé en français. Son objectif est de présenter la problématique de la discipline de manière compréhensible et amusante et de prouver la vigueur des connaissances acquises dans la pratique. Pour cela, nous divisons la thèse en deux parties: la première théorique et la seconde analytique.

La partie théorique s'occupe de la présentation du français écrit et parlé, de son caractère lexical et morphosyntaxique, de ses racines historiques et de la situation d'aujourd'hui. Elle met en évidence la terminologie fondamentale liée au sujet donné et elle pose un certain nombre de problèmes particuliers. Nous mentionnons les diverses disciplines faisant partie de la linguistique quantitative et nous nous concentrons plus particulièrement sur la statistique lexicale et grammaticale qui constitue le thème majeur de la deuxième partie de cette thèse. Nous avons ajouté des particularités intéressantes concernant la relation des Français à leur langue maternelle, les lois régulant le registre de langue soutenu du français ou le langage SMS dont nous avons présenté les principes essentiels et nous démontrons qu'il existe une grande influence de l'oral sur l'écrit.

Dans la partie analytique, nous nous sommes décidés à appliquer l'analyse statistique aux scénarios de films (plusieurs scènes choisies) et à leurs modèles littéraires. Nous avons choisi une adaptation du roman bien connu écrit par Boris Vian *L'Écume des jours* publié pour la première fois en 1947. L'autre roman, écrit par Jean Teulé, a été publié en 2008 et s'intitule *Le Magasin des suicides*. Les adaptations cinématographiques ont été réalisées par Michel Gondry et Luc Bossi pour *L'Écume des jours* et par Patrice Leconte pour *Le Magasin des suicides*. Il s'agit de films réalisés tout récemment. Nous avons choisi les scénarios en voulant les substituer au langage parlé et les analyser du point de vue de la fréquence. En outre, nous ajoutons une comparaison de ces deux médias (les romans et les scénarios). Nous avons créé une liste de mots les plus fréquents et nous avons aussi ajouté les analyses grammaticales. Même si le caractère compliqué de l'élaboration manuelle de ces textes ne nous a permis de générer qu'un corpus plutôt modeste (10 897 mots), nous avons étudié un nombre suffisant de données intéressantes et fiables.

Notre thèse permettra au lecteur d'appréhender les principes de la statistique linguistique et d'évaluer les résultats issus de notre recherche.

Bibliografie

- ANDRIEUX, Véronique. 1993. *L'effet de la fréquence d'usage et de la familiarité du mot sur la reconnaissance visuelle de mots de tailles différentes*. Diplomová práce, U.F.R. Sciences Humaines. Dijon : Université de Bourgogne, 1993, 85 s. Vedoucí diplomové práce Daniel Zagar.
- ARGYLE, Michael, ALKEMA, Florisse, GILMOUR, Robin. 1971. „The communication of friendly and hostile attitudes by verbal and non-verbal signals“. *European Journal of Social Psychology*, 1971, č. 1, s. 395–402.
- BLANCHE-BENVENISTE, Claire. 1990. *Le français parlé : Études grammaticales*. Paris : Centre National de la Recherche Scientifique, 1990. 292 s.
- BLANCHE-BENVENISTE, Claire. 2010. *Approches de la langue parlée*. Paris : Ophrys, 2010. 171 s.
- BLUESTONE, George. 2003. *Novels into Film*. Baltimore : Johns Hopkins University Press, 2003. 251 s.
- BUBENÍČEK, Petr. 2010. „Filmová adaptace: Hledání interdisciplinárního dialogu“, *Illuminace* 22, 2010, č. 1, s. 7–21.
- COSSETTE, André. 1994. *La richesse lexicale et sa mesure*. Paris : Honoré Champion, 1994. 187 s.
- ČERMÁK, František, KŘEN, Michal, BLATNÁ, Renata, et al. 2004. *Frekvenční slovník češtiny*. Praha : Nakladatelství Lidové noviny, 2004. 595 s.
- ČERNÝ, Jiří. 1996. *Dějiny lingvistiky*. Olomouc : Votobia, 1996. 517 s.

- ČERNÝ, Jiří. 2008. *Úvod do studia jazyka*. 2. vyd. Olomouc : Rubico, 2008. 248 s.
- ENGSTRØM, Line. 2005. *Les anglicismes dans le français hexagonal contemporain, Analyse d'un corpus de presse*. Diplomová práce, Institutt for litteratur, områdestudier og europeiske språk. Oslo : Universitetet i Oslo, 2005, 107 s. Vedoucí diplomové práce Marianne Hobæk Haff.
- GADET, Françoise. 1992. *Le français populaire*. Paris : Presses Universitaires de France, Que sais-je ?, 1992. 128 s.
- GERMAIN, Claude. 1993. „La méthode SGAV“. *Évolution de l'enseignement des langues : 5 000 ans d'histoire*. Paris : CLE International, 1993. s. 153–169.
- GREENBERG, Joseph. 1960. „A Quantitative Approach to the Morphological Typology of Language“, *International Journal of American Linguistics* 26, 1960, s. 178–194.
- HAYESOVÁ, Nicky. 2011. *Základy sociální psychologie*. 6. vyd. Praha : Portál, 2011. 166 s.
- JELÍNEK, Jaroslav, BEČKA, Josef Václav, TĚŠITELOVÁ, Marie. 1961. *Frekvence slov, slovních druhů a tvarů v českém jazyce*. Praha : Státní pedagogické nakladatelství, 1961. 585 s.
- MOUGEON, Françoise. 1995. *Quel français parler? Initiation au français parlé au Canada et en France*. Toronto : Éditions du Gref, 1995. 197 s.
- NEW, Boris, BRYLSBAERT, Marc, VERONIS, Jean, PALLIER, Christophe. 2007. „The use of film subtitles to estimate word frequencies“, *Applied Psycholinguistics* 28, 2007, s. 661–667.

- PERRET, Michèle. 2008. *Introduction à l'histoire de la langue française*. 3. vyd. Paris : Armand Colin, 2008. 204 s.
- QUENEAU, Raymond. 1959. *Zazie dans le métro*, Paris : Gallimard, 1959. 253 s.
- RIEGEL, Martin, PELLAT, Jean-Christophe, RIOUL, René. 2009. *Grammaire méthodique du français*. 4. vyd. Paris : Quadriga / PUF, 2009. 1107 s.
- ROSKAM, Heaven-Lee. 2010. *Zazie dans le métro et la traduction du langage de Queneau*. Diplomová práce. Utrecht : Université d'Utrecht, 2010, 101 s. Vedoucí diplomové práce prof. dr. M. B. van Buuren.
- ŠONKOVÁ, Jitka. 2008. *Morfologie mluvené češtiny: Frekvenční analýza*. Praha : Nakladatelství Lidové noviny : Ústav Českého národního korpusu, 2008. 356 s.
- TĚŠITELOVÁ, Marie. 1974. *Otázky lexikální statistiky*. Praha : Československá akademie věd, 1974. 289 s.
- TĚŠITELOVÁ, Marie. 1977. *Lingvistické příručky: Kvantitativní lingvistika*. 1. vyd. Praha : Státní pedagogické nakladatelství, 1977. 177 s.
- TĚŠITELOVÁ, Marie. 1980. *Využití statistických metod v gramatice*. 1. vyd. Praha : Československá akademie věd, 1980. 219 s.
- WALTEROVÁ, Henriette. 1993. *Francouzština známá i neznámá*. Praha: Jan Kanzelsberger, 1993. 323 s.

Elektronické zdroje

- ABOU EL KHAIR, Catherine, DESHAYES, Benoit. 2011. „Lexique politique: En savoir plus“ [online]. L'Internaute [citováno 11. 3. 2011]. Dostupné z WWW: <<http://www.linternaute.com/actualite/politique/lexique-politique/en-savoir-plus.shtml>>.
- ABOU EL KHAIR, Catherine, DESHAYES, Benoit. 2011. „Lexique politique: Les mots de Ségolène Royal“ [online]. L'Internaute [citováno 11. 3. 2011]. Dostupné z WWW: <<http://www.linternaute.com/actualite/politique/lexique-politique/segolene-royal.shtml>>.
- ABOU EL KHAIR, Catherine, DESHAYES, Benoit. 2011. „Lexique politique: Les mots de François Hollande“ [online]. L'Internaute [citováno 20. 1. 2014]. Dostupné z WWW: <<http://www.linternaute.com/actualite/politique/lexique-politique/segolene-royal.shtml>>.
- ABOU EL KHAIR, Catherine; DESHAYES, Benoit. 2011. „Lexique politique: Les mots de Nicolas Sarkozy“ [online]. L'Internaute [citováno 11. 3. 2011]. Dostupné z WWW: <<http://www.linternaute.com/actualite/politique/lexique-politique/nicolas-sarkozy.shtml>>.
- ACADÉMIE FRANÇAISE. 2013. „Dire, ne pas dire : Néologismes & anglicismes – Digital“ [online]. Académie française [citováno 23. 1. 2014]. Dostupné z WWW: <<http://www.academie-francaise.fr/dire-ne-pas-dire/neologismes-anglicismes>>.

- BRANCA-ROSOFF, Sonia, FLEURY, Serge, LEFEUVRE, Florence, PIRES, Matthew. 2012. „Discours sur la ville : Corpus du Français Parlé Parisien des années 2000 (CFPP2000)“. Paris : Université Sorbonne Nouvelle Paris 3, 2012 [citováno 14. 1. 2011]. Dostupné z WWW: <<http://cfpp2000.univ-paris3.fr/CFPP2000.pdf>>.
- BRUNET, Étienne. 2011. „Liste des mots classée par fréquence décroissante“ [online]. Eduscol [citováno 21. 2. 2011]. Dostupné z WWW: <<http://eduscol.education.fr/cid47916/liste-des-mots-classee-par-frequence-decroissante.html>>.
- CABITEN, Mariana. 2013. „80% des Français s'estiment bons en orthographe, et vous ?“ [online]. France Bleu [citováno 26. 1. 2014]. Dostupné z WWW: <<http://www.francebleu.fr/infos/orthographe/80-des-francais-s-estiment-bons-en-orthographe-et-vous-995362>>.
- CORPUS DE LA PAROLE. „Comment on parle en France aujourd'hui“ [online]. Corpus de la parole du ministère de la culture et de la communication [citováno 14. 1. 2011]. Dostupné z WWW: <<http://corpusdelaparole.tge-adonis.fr/>>.
- ČESKÝ NÁRODNÍ KORPUS. „Dostupné korpusy“ [online]. Český národní korpus [citováno 14. 1. 2011]. Dostupné z WWW: <<https://korpus.cz/struktura.php>>.
- DICTIONNAIRE SMS. „Dictionnaire SMS“ [online]. Dictionnaire SMS [citováno 3. 1. 2014]. Dostupné z WWW: <<http://www.dictionnaire-sms.com/>>.
- D'URSEL, Laurent. 2013. „Le langage SMS ou L'impossible révolution démocratique : Qui est bien, tchathe bien“ [online video]. YouTube [citováno 28. 12. 2013]. Dostupné z WWW: <http://www.youtube.com/watch?v=ghF1_JDtvzQ>.

- EQUIPE DELIC. „Présentation du Corpus de référence du français parlé“ [online]. Université de Provence [citováno 21. 1. 2011]. Dostupné z WWW: <<http://sites.univ-provence.fr/~veronis/pdf/2004-presentation-crfp.pdf>>.

- FRANÇAIS FACILE/FRENCH FOR BEGINNERS. 2014. „Les caractéristiques du français familier“ [online]. Français facile / French for beginners [citováno 2. 2. 2014]. Dostupné z WWW: <<http://www.prononcer.net/les-caracteristiques-du-francais-familier>>.

- GENERAL DELEGATION FOR THE FRENCH AND THE LANGUAGES OF FRANCE. „« the language of the Republic is French» (art. 2 of the French Constitution)“ [online]. General Delegation for the French language and the languages of France [citováno 20. 1. 2014]. Dostupné z WWW: <http://www.dglf.culture.gouv.fr/dglf_presentation_anglais.htm>.

- HOUSER, Pavel. 2004. „Glottochronologie: Pohled na evoluci jazyků“ [online]. Science world [citováno 12. 2. 2011]. Dostupné z WWW: <<http://scienceworld.cz/lingvistika/glottochronologie-pohled-na-evoluci-jazyku-2501>>.

- KASÍK, Pavel. 2006. „Proč se na klávesnici blbě píše? Byl to záměr“ [online]. Technet.cz [citováno 22. 12. 2013]. Dostupné z WWW: <http://technet.idnes.cz/proc-se-na-klavesnici-blbe-pise-byl-to-zamer-fka-/hardware.aspx?c=A061208_170724_hardware_dno>.

- LE COURRIER DES ECHOS. 2013. „L'académie française propose une traduction pour les acronymes «WTF et OMG»“ [online]. Le Courrier des Echos [citováno 2. 2. 2014]. Dostupné z WWW: <http://www.lecourrierdesechos.fr/index.php/culture/lacademie-francaise-propose-une-traduction-pour-les-acronymes-wtf-et-omg/?ccm_paging_p_b231=15#.Uqyk5pqA3IU>.

- LES TIMBRÉS DE L'ORTHOGRAPHE. „Un nouveau label pour la langue française“ [online]. Les Timbrés de l'Orthographe [citováno 2. 1. 2014]. Dostupné z WWW: < http://timbresdelorthographe.com/?page_id=2>.

- MOULART, Caroline. 2000. „Recherches sur le français parlé - Revue de l'équipe DELIC : Présentation“ [online]. Recherches sur le français parlé [citováno 24. 1. 2011]. Dostupné z WWW: <<http://sites.univ-provence.fr/delic/rsfp/>>.

- PASTEBIN. „Simpsons Frequency Dictionary: Frequency Dictionary based on OpenSubtitles of all Simpson Episodes (in english)“ [online]. Pastebin [citováno 28. 1. 2011]. Dostupné z WWW: <<http://pastebin.com/anKcMdvk>>.

- QUILLET, Lucile. 2012. „Attachiant, phonard... les nouveaux mots préférés des jeunes“ [online]. Le Figaro [citováno 24. 1. 2014]. Dostupné z WWW: < <http://etudiant.lefigaro.fr/vie-etudiante/news/detail/article/attachiant-phonard-les-nouveaux-mots-preferes-des-jeunes-810/>>.

- SEIJIDO, Magali, CAPPEAU, Paul. 2006. „Les corpus oraux en français (inventaire 2005 v.1.0)“ [online]. Projet DGLFLF [citováno 16. 1. 2011]. Dostupné z WWW: <<http://ticsenfle.blogspot.cz/2012/11/voix-et-images-de-la-france-1962.html>>.

- VERONIS, Jean. 2012. „Débat: Moi, François Hollande“ [online]. Technologies du Langage [citováno 19. 12. 2013]. Dostupné z WWW: < <http://blog.veronis.fr/2012/05/debat-moi-francois-hollande.html>>.

PŘÍLOHA

Příloha 1

Ukázka vybrané části scénáře k filmu *Les Magazines de suicides* – text analyzovaný v praktické části naší práce (Leconte, 2012). Pozn. analyzovány byly pouze dialogové promluvy bez scénických poznámek.

LE MAGASIN DES SUICIDES

Un film de

PATRICE LECONTE

Adaptation, dialogues et paroles des chansons

PATRICE LECONTE

d'après le livre de

JEAN TEULE

(Editions Julliard)

produit par

Gilles Podesta

Diabolo Films

06 07 17 28 22

info@diabolofilms.com

www.diabolofilms.com

Grand Prix du Scénario du Festival des Antipodes 2010

Avec le soutien du CNC

du Département de la Charente

de la Région Poitou-Charentes

du Fonds d'Action Sacem

Version tournage : 10 septembre 2010

/.../

8 – BOUTIQUE DES SUICIDES, JOUR

Retour à la boutique où la vie a repris comme avant, et où tout est pareil. Sauf ce véhicule, parfaitement inattendu dans cet univers pour le moins morbide, à savoir un landau qui, de temps en temps, oscille en fonction des soubresauts du nourrisson qu'il abrite, lequel nourrisson émet divers sons plutôt gais, allant de areu à blougouchlabadagoubi, ce qui énerve copieusement le fils Vincent qui affûte, la fille Marilyn qui lit, les deux levant de concert les yeux au plafond.

Une dame âgée est penchée sur le landau d'où sortent ces borborygmes nourrissonnesques.

DAME AGEE

Oh, il sourit !

LUCRECE TUVACHE

Comment ça, mon fils sourit ? Mais non, il ne sourit pas. Ca doit être un pli de bouche. Pourquoi il sourirait ?

Elle se replonge dans ses comptes de caisse, tandis que la vieille dame insiste.

DAME AGEE

On dirait pourtant bien qu'il sourit.

Madame Tuvache se pousse du col, se démanche, se tortille, pour considérer l'enfant par-dessus sa caisse.

LUCRECE TUVACHE

Ca m'étonnerait, personne n'a jamais souri dans la famille Tuvache !

Elle dévisage le bébé, inquiète.

LUCRECE TUVACHE

Mishima, viens voir !

La trappe s'ouvre à toute volée, envoyant à nouveau gicler Marilyn, son tabouret et son magazine, elle vient atterrir à côté de son frère Vincent qui du coup se coupe le doigt avec la lame qu'il affûtait. Monsieur Tuvache apparaît, couvert de ciment.

MISHIMA TUVACHE

Voir quoi ?

LUCRECE TUVACHE

La cliente prétend qu'Alan sourit.

MISHIMA TUVACHE

C'est quoi, c't'histoire ?

Il époussette la poudre de ciment sur ses manches et ses épaules, disparaissant provisoirement dans un nuage blanchâtre, réapparaît pour s'approcher à son tour du landau. Les voilà tous les trois penchés sur le bébé qui, sans nul doute, leur sourit.

MISHIMA TUVACHE

Il a sûrement la colique. On peut parfois confondre avec des sourires, mais ça n'en est pas. C'est des grimaces.

Puis il glisse ses mains sous la capote du landau, pose ses pouces à chaque bout de la bouche de l'enfant, et, prenant l'aïeule à témoin :

MISHIMA TUVACHE

Regardez. Si je pousse les commissures vers le bas, il ne sourit pas. Il fait la gueule, comme son frère et sa soeur dès qu'ils sont nés.

Il désigne le frère et la soeur en question, Marilyn qui, après sa chute de tabouret, se relève mollement, faisant la gueule, tandis que Vincent entortille d'une bandelette son doigt coupé, faisant la gueule lui aussi, pile poil la même pour être précis.

DAME AGEE

Relâchez, pour voir.

Mishima Tuvache s'exécute, le bébé se remet à sourire.

DAME AGEE

Ah ! Vous voyez bien qu'il sourit.

Monsieur Tuvache se redresse, agacé, bombe le torse, et demande à la vieille :

MISHIMA TUVACHE

Bon, ça va, qu'est-ce que vous vouliez, vous ?

DAME AGEE

Une corde pour me pendre.

Il l'entraîne vers le rayon, tandis que Lucrece Tuvache reste penchée sur le berceau, effarée.

MISHIMA TUVACHE

C'est haut de plafond, chez vous ? Vous ne savez pas ? Tenez, prenez ça : deux mètres devraient suffire, le noeud coulant est déjà fait, vous avez plus qu'à passer votre tête, tenez, je vous la donne, si si ça me fait plaisir.

Il pousse la vieille dame vers la sortie, lui mettant la corde dans les bras, la bousculant, s'en débarrassant. Avant d'être mise dehors, elle a le temps de se retourner vers le landau.

DAME AGEE

En tout cas, ça met du baume au coeur de voir un enfant qui sourit.

MISHIMA TUVACHE

Oui, oui, c'est ça, allez hop, rentrez chez vous, vous avez mieux à faire, maintenant, là-bas.

Il l'a foutue dehors, lui claque la porte au cul, et revient illico vers sa femme, le landau et l'enfant.

MISHIMA TUVACHE

Comment ça, il sourit ?

Plantés de chaque côté, les parents, catastrophés, constatent l'évidence aveuglante : cet enfant est souriant.

MISHIMA et LUCRECE TUVACHE

Merde...

Příloha 2

Ukázka vybrané části scénáře k filmu *L'Écume des jours* – text analyzovaný v praktické části naší práce (Gondry, Bossi, 2013). Pozn. analyzovány byly pouze dialogové promluvy bez technických poznámek.

L'ECUME DES JOURS

BOBINE 2 AB

BOBINES 2A – 2B

2.10

Première et deuxième parties

plan : 16.11

Première image : 0.00

plan : 23.10

1^{er} changement de plan :

2^{ème} changement de

3^{ème} changement de

N° ST	PIETAGE IN	PIETAGE OUT	NBRE CAR.	PERSONNAGE	DIALOGUE
1.	35.01	36.12	17c	COLIN	Bon, ça suffit.
2.	39.02	40.03	11c	COLIN	Attends !
3.	53.11	55.02	15c	ISIS	Bonjour, merci pour les cadeaux. Merci.
4.	56.04	57.15	17c	ISIS COLIN	- Bonsoir, Colin. - Bonjour.
5.	59.00	60.06	14c	ISIS	Vous allez bien ?
6.	61.03	63.01	19c	COLIN	Bon anniversaire.
7.	63.05	66.00	27c	ISIS	Mais c'est pas mon anniversaire! C'est celui de Dupond !
8.	66.09	68.07	19c	COLIN	Ah oui, votre caniche, bien sûr.
9.	68.11	70.11	21c	ISIS	Non mais c'est pas un caniche, c'est un Jack Russel.
10.	70.15	75.01	42c	ISIS	En fait, ce matin, on l'a amené chez le tondeur pour qu'il soit beau pour son anniversaire
11.	75.11	77.09	19c	ISIS	et on l'a fait peigner pendant des heures, et tout,
12.	77.14	81.09	37c	ISIS	et tout à l'heure, trois de ses amis à lui sont venus avec un vieil os pourri et ils l'ont emmené.
13.	83.05	87.12	45c	ISIS	Il va revenir dans un état affreux, ça va être catastrophique. Il est parti depuis longtemps déjà.
14.	88.00	89.00	11c	COLIN	Voilà.
15.	90.15	92.10	17c	ISIS	Mais qu'est-ce que c'est ?
16.	94.03	96.02	20c	ISIS	Ah, mais c'est super !
17.	96.07	98.05	19c	ISIS	Mais il va adorer, merci beaucoup, merci !
18.	100.09	102.02	16c	ISIS	Mais il est là en plus ! Mais il est là, il est là !
19.	104.13	107.13	31c	ISIS	Mon Dupond d'amour,
20.	108.13	111.09	28c	ISIS	regarde, regarde ce que Colin t'a apporté.
21.	112.14	115.00	22c	ISIS	Dis-lui merci, allez !

22.	115.05	117.08	22c	COLIN	De rien, de rien, de rien.
23.	119.04	120.14	17c	COLIN	Vous avez une robe ravissante.
24.	121.03	123.11	26c	COLIN	On peut passer la main à travers les barreaux sans se faire mordre ?
25.	124.00	126.02	22c	ISIS	Oh ben essayez, pour voir.
26.	132.12	135.12	31c	ISIS	Désolée, ça me fait ça à chaque fois.
27.	136.05	139.00	27c	ISIS COLIN	- Vous voulez du champagne ? - Oui.
28.	174.06	179.09	52c	CHICK	C'est le seul exemplaire, l'original, et il m'appartient.
29.	180.13	183.05	26c	COLIN CHICK	- Salut. - Colin ! Tu vas bien ?

Příloha 2

Frekvenční seznam francouzských slov vytvořený lexikologem Étienne
Brunetem (Brunet, 2011)

Mots les plus fréquents de la langue écrite française (XIXe et XXe siècles) Table hiérarchique

le dét. 1050561	grand adj. 25388	moment subst. 12274
de prép. 862100	celui pron. 24270	rester verbe 12155
un dét. 419564	si conj. 24024	répondre verbe 12063
être verbe 351960	notre dét. 23883	tout dét. 12051
et conj. 362093	devoir verbe 22703	tête subst. 11999
à prép. 293083	là adv. 22695	père subst. 11854
il pron. 270395	jour subst. 22232	fille subst. 11842
avoir verbe 248488	prendre verbe 20489	mille numér. 11758
ne adv. 186755	même adv. 19994	premier adj. 11731
je pron. 184186	votre dét. 19942	car conj. 11695
son dét. 181161	tout adv. 19915	entendre verbe 12009
que conj. 176161	rien pron. 19379	ni conj. 11640
se pron. 168684	petit adj. 19008	bon adj. 11483
qui pron. 148392	encore adv. 19176	trois numér. 11372
ce dét. 141389	aussi adv. 18311	cœur subst. 11312
dans prép. 139185	quelque dét. 17953	ainsi adv. 11296
en prép. 143565	dont pron. 17797	an subst. 11274
du dét. 127384	tout pron. 17486	quatre numér. 10970
elle pron. 126397	mer subst. 17166	un pron. 10941
au dét. 123502	trouver verbe 16833	terre subst. 10786
de dét. 119106	donner verbe 16795	contre prép. 10692
ce pron. 107074	temps subst. 16785	dieu subst. 10661
le pron. 105873	ça pron. 16494	monsieur subst. 10489
pour prép. 104779	peu adv. 16251	voix subst. 10469
pas adv. 103083	même adj. 16081	penser verbe 10358
que pron. 99412	falloir verbe 16078	quel adj. 10343
vous pron. 89623	sous prép. 15944	arriver verbe 10288
par prép. 82277	parler verbe 15814	maison subst. 10287
sur prép. 80180	alors adv. 15639	devant prép. 9995
faire verbe 77608	main subst. 15540	coup subst. 9991
plus adv. 75499	chose subst. 15524	beau adj. 9870
dire verbe 72134	ton dét. 15513	connaître verbe 9769
me pron. 71086	mettre verbe 15339	devenir verbe 9759
on pron. 70246	vie subst. 15241	air subst. 9755
mon dét. 70121	savoir verbe 15102	mot subst. 9752
lui pron. 65988	yeux subst. 14981	nuit subst. 9694
nous pron. 62554	passer verbe 14976	sentir verbe 9585
comme conj. 59902	autre adj. 14688	eau subst. 9603
mais conj. 57690	après prép. 14606	vieux adj. 9515
pouvoir verbe 55394	regarder verbe 14604	sembler verbe 9482
avec prép. 55081	toujours adv. 14336	moins adv. 9472
tout adj. 47221	puis conj. 14257	tenir verbe 9312
y pron. 46031	jamais adv. 14255	ici adv. 9098
aller verbe 41702	cela pron. 14253	comprendre verbe 9037
voir verbe 39659	aimer verbe 14138	oui adv. 9005
en pron. 38935	non adv. 14039	rendre verbe 9002
bien adv. 37171	heure subst. 13940	toi pron. 8997
où pron. 36089	croire verbe 13881	vingt numér. 8920
sans prép. 35915	cent numér. 13798	depuis prép. 8907
tu pron. 35774	monde subst. 13737	attendre verbe 8851
ou conj. 34897	donc conj. 13562	sortir verbe 8768
leur dét. 33950	enfant subst. 13348	ami subst. 8744
homme subst. 33202	fois subst. 13191	trop adv. 8686
si adv. 32024	seul adj. 13104	porte subst. 8649
deux numér. 30211	autre pron. 13063	lequel pron. 8574
mari subst. 30082	entre prép. 13684	chaque dét. 8419
moi pron. 30053	vers prép. 12781	amour subst. 8283
vouloir verbe 29435	chez prép. 12698	pendant prép. 8202
te pron. 28542	demander verbe 12597	déjà adv. 8170
femme subst. 26148	jeune adj. 12593	pied subst. 8040
venir verbe 26023	jusque prép. 12465	tant adv. 7960
quand conj. 25592	très adv. 12432	gens subst. 7944

parce que conj. 7824
 nom subst. 7795
 vivre verbe 7625
 reprendre verbe 7544
 entrer verbe 7614
 porter verbe 7499
 pays subst. 7451
 ciel subst. 7433
 avant prép. 7425
 frère subst. 7415
 regard subst. 7399
 chercher verbe 7304
 âme subst. 7255
 côté subst. 7245
 mort subst. 7182
 revenir verbe 7114
 noir adj. 7038
 maintenant adv. 7024
 nouveau adj. 7019
 ville subst. 6983
 rue subst. 6923
 enfin adv. 7126
 appeler verbe 6892
 soir subst. 6877
 chambre subst. 6835
 mourir verbe 6785
 pas subst. 6751
 partir verbe 6726
 cinq numér. 6723
 esprit subst. 7031
 soleil subst. 6692
 dernier adj. 6650
 jeter verbe 6610
 dix numér. 6609
 roi subst. 6588
 état subst. 6489
 corps subst. 6425
 beaucoup adv. 6399
 suivre verbe 6397
 bras subst. 6304
 écrire verbe 6256
 blanc adj. 6246
 montrer verbe 6195
 tomber verbe 6182
 place subst. 6178
 ouvrir verbe 6169
 ah interj 6138
 parti subst. 6102
 assez adv. 6090
 leur pron. 6078
 cher adj. 6059
 voilà prép. 6054
 année subst. 6004
 loin adv. 5996
 point adv. 5961
 visage subst. 5954
 bruit subst. 5946
 lettre subst. 5946
 franc subst. 5922
 fond subst. 5861
 force subst. 5835
 arrêter verbe 5812
 perdre verbe 5786
 commencer verbe 5783
 paraître verbe 5779
 aucun dét. 5774
 marcher verbe 5747
 milieu subst. 5706
 saint subst. 5702
 idée subst. 5686
 presque adv. 5662
 ailleurs adv. 5629
 travail subst. 5623
 lumière subst. 5622
 long adj. 5562
 seulement adv. 5544
 mois subst. 5527
 fils subst. 5520
 neuf numér. 5508
 tel dét. 5505
 lever verbe 5494
 raison subst. 5481
 effet subst. 5829
 gouvernement sub 5470
 st.
 permettre verbe 5467
 pauvre adj. 5434
 asseoir verbe 5417
 point subst. 5416
 plein adj. 5413
 personne subst. 5391
 vrai adj. 5385
 peuple subst. 5349
 fait subst. 5343
 parole subst. 5295
 guerre subst. 5273
 toute adj. 5258
 écouter verbe 5216
 pensée subst. 5214
 affaire subst. 5179
 quoi pron. 5152
 matin subst. 5145
 pierre subst. 5127
 monter verbe 5088
 bas adj. 5087
 vent subst. 5002
 doute subst. 4977
 front subst. 4969
 ombre subst. 4939
 part subst. 4932
 maître subst. 4916
 aujourd'hui adv. 4915
 besoin subst. 4908
 question subst. 4908
 apercevoir verbe 4904
 recevoir verbe 4891
 mieux adv. 4881
 peine subst. 4859
 tour subst. 4836
 servir verbe 4806
 oh interj 4766
 autour adv. 4764
 près prép. 4731
 finir verbe 4709
 famille subst. 4705
 pourquoi conj. 4700
 souvent adv. 4665
 rire verbe 4662
 dessus adv. 4657
 madame subst. 4653
 sorte subst. 4635
 figure subst. 4618
 droit subst. 4595
 peur subst. 4574
 bout subst. 4571
 lieu subst. 4554
 silence subst. 4541
 gros adj. 4537
 chef subst. 4503
 eh interj 4584
 six numér. 4463
 bois subst. 4460
 mari subst. 4457
 histoire subst. 4451
 crier verbe 4449
 jouer verbe 4447
 feu subst. 4429
 tourner verbe 4371
 doux adj. 4355
 longtemps adv. 4355
 fort adv. 4350
 heureux adj. 4332
 comme adv. 4324
 garder verbe 4272
 partie subst. 4271
 face subst. 4236
 mouvement subst. 4231
 fin subst. 4217
 reconnaître verb 4198
 quitter verbe 4180
 personne pron. 4164
 comment adv. 4163
 route subst. 4155
 dès prép. 4141
 manger verbe 4127
 livre subst. 4097
 arbre subst. 4070
 courir verbe 4059
 cas subst. 4058
 huit numér. 4052
 lorsque conj. 4041
 mur subst. 4034
 ordre subst. 4028
 continuer verbe 4022
 bonheur subst. 3978
 oublier verbe 3965
 descendre verbe 3955
 haut adj. 3953
 intérêt subst. 3922
 cacher verbe 3920
 l'un pron. 3910
 chacun pron. 3890
 profond adj. 3878
 argent subst. 3876
 cause subst. 3856
 poser verbe 3841
 autant adv. 3834
 est subst. 3994
 travers subst. 3825
 grand subst. 3809
 instant subst. 3807
 façon subst. 3784
 d'abord adv. 3783
 oeil subst. 3783
 tirer verbe 3778
 forme subst. 3763
 présenter verbe 3757
 ajouter verbe 3755
 agir verbe 3753
 retrouver verbe 3717
 chemin subst. 3711
 cheveu subst. 3704
 offrir verbe 3671
 surtout adv. 3669
 certain dét. 3667
 plaisir subst. 3656
 suite subst. 3639
 apprendre verbe 3616
 malgré prép. 3612
 tuer verbe 3598
 rouge adj. 3576
 sang subst. 3571
 retourner verbe 3559
 rencontrer verbe 3556
 sentiment subst. 3548
 fleur subst. 3516
 cependant adv. 3508
 service subst. 3498
 plusieurs dét. 3480
 table subst. 3472
 vite adv. 3465
 paix subst. 3446
 envoyer verbe 3482
 moyen subst. 3444
 dormir verbe 3438
 pousser verbe 3422
 lit subst. 3410
 humain adj. 3393
 voiture subst. 3387
 rappeler verbe 3362
 être subst. 3345

lire verbe 3340
général adj. 3337
nature subst. 3335
or subst. 3326
pouvoir subst. 3309
nouveau subst. 3307
français adj. 3299
joie subst. 3292
sept numér. 3289
tard adv. 3281
président subst. 3272
pourtant adv. 3271
bouche subst. 3266
changer verbe 3258
petit subst. 3256
froid adj. 3250
compter verbe 3248
occuper verbe 3245
sens subst. 3245
cri subst. 3240
cheval subst. 3237
loi subst. 3236
sombre adj. 3234
ci adv. 3223
sûr adj. 3211
espèce subst. 3238
voici prép. 3191
ancien adj. 3190
tandis que conj. 3189
frapper verbe 3161
ministre subst. 3145
puisque conj. 3139
selon prép. 3134
travailler verbe 3133
expliquer verbe 3238
propre adj. 3125
obtenir verbe 3124
rentrer verbe 3099
mal adv. 3097
pleurer verbe 3096
essayer verbe 3254
répéter verbe 3079
société subst. 3079
parfois adv. 3076
politique subst. 3071
oreille subst. 3063
payer verbe 3056
politique adj. 3054
apporter verbe 3053
fenêtre subst. 3046
derrière prép. 3019
possible adj. 3013
fortune subst. 3010
compte subst. 3002
champ subst. 2979
manier subst. 2961
vraiment adv. 2960
immense adj. 2948
action subst. 2942
boire verbe 2937
public adj. 2929
garçon subst. 2914
pareil adj. 2914
bleu adj. 2906
sourire verbe 2904
couleur subst. 2890
coucher verbe 2889
papier subst. 2875
d'autres dét. 2865
mal subst. 2861
fort adj. 2851
bientôt adv. 2839
causer verbe 2825
pièce subst. 2820
montagne subst. 2818
sol subst. 2812
oeuvre subst. 2811
partout adv. 2810
trente numér. 2809
exister verbe 2944
cours subst. 2797
raconter verbe 2794
serrer verbe 2792
songer verbe 2792
désir subst. 2790
manquer verbe 2787
cour subst. 2776
nommer verbe 2754
bord subst. 2753
douleur subst. 2749
conduire verbe 2735
salle subst. 2732
saisir verbe 2729
premier subst. 2722
comment conj. 2721
projet subst. 2715
demeurer verbe 2709
simple adj. 2704
étude subst. 2701
remettre verbe 2700
journal subst. 2699
geste subst. 2697
disparaître verb 2689
battre verbe 2678
toucher verbe 2670
situation subst. 2664
oiseau subst. 2661
nécessaire adj. 2654
exemple subst. 2906
siècle subst. 2647
apparaître verbe 2645
souffrir verbe 2635
million subst. 2633
prix subst. 2616
groupe subst. 2612
centre subst. 2610
malheur subst. 2603
honneur subst. 2602
fermer verbe 2590
accepter verbe 2585
garde subst. 2575
mauvais adj. 2570
tendre verbe 2569
naître verbe 2555
sauver verbe 2554
entier adj. 2717
parmi prép. 2547
problème subst. 2547
larne subst. 2546
avancer verbe 2544
chien subst. 2539
peau subst. 2534
reste subst. 2530
traverser verbe 2522
nombre subst. 2517
debout adv. 2515
mesure subst. 2514
social adj. 2510
souvenir verbe 2508
article subst. 2507
vue subst. 2502
couvrir verbe 2491
âge subst. 2490
gagner verbe 2485
système subst. 2483
long subst. 2482
former verbe 2481
plaire verbe 2477
embrasser verbe 2458
rêve subst. 2455
oser verbe 2454
afin de prép. 2452
passion subst. 2448
auquel pron. 2440
rapport subst. 2426
refuser verbe 2420
important adj. 2416
décider verbe 2415
produire verbe 2401
soldat subst. 2398
lèvre subst. 2397
signe subst. 2397
vérité subst. 2390
charger verbe 2389
mariage subst. 2386
mêler verbe 2385
certain adj. 2380
plan subst. 2365
cesser verbe 2349
ressembler verbe 2349
dos subst. 2348
marche subst. 2341
souvenir subst. 2334
dame subst. 2333
chanter verbe 2332
plutôt adv. 2328
conseil subst. 2318
sou subst. 2314
triste adj. 2307
coin subst. 2306
jardin subst. 2303
joli adj. 2301
soit conj. 2297
empêcher verbe 2427
doigt subst. 2289
objet subst. 2288
fer subst. 2284
lendemain subst. 2281
lentement adv. 2281
combien adv. 2280
approcher verbe 2272
prier verbe 2265
train subst. 2259
espérer verbe 2371
papa subst. 2256
différent adj. 2254
valeur subst. 2252
jeu subst. 2247
échapper verbe 2240
glisser verbe 2239
secret subst. 2234
haut subst. 2233
vieillard subst. 2226
briller verbe 2225
docteur subst. 2222
brûler verbe 2218
terrible adj. 2218
placer verbe 2214
ton subst. 2213
jambe subst. 2204
juger verbe 2202
suffire verbe 2201
endroit subst. 2194
minute subst. 2192
atteindre verbe 2190
nuage subst. 2190
présence subst. 2187
fou adj. 2165
épaule subst. 2163
léger adj. 2158
feuille subst. 2155
liberté subst. 2155
journée subst. 2149
libre adj. 2147
annoncer verbe 2146
avenir subst. 2143
sourire subst. 2142
hier adv. 2141
résultat subst. 2136
élever verbe 2135
acheter verbe 2134

mener verbe 2133
préparer verbe 2133
pourquoi adv. 2131
hôtel subst. 2125
semaine subst. 2124
forêt subst. 2122
assurer verbe 2118
pur adj. 2118
qualité subst. 2116
prince subst. 2108
bien subst. 2101
également adv. 2100
deviner verbe 2095
médecin subst. 2093
considérer verbe 2092
volonté subst. 2087
seigneur subst. 2086
effort subst. 2468
quelque adv. 2083
vert adj. 2081
art subst. 2077
moindre adj. 2077
demain adv. 2076
quarante numér. 2073
cinquante numér. 2072
foule subst. 2070
appartenir verbe 2069
aussitôt adv. 2068
ligne subst. 2068
représenter verb 2067
tromper verbe 2065
intérieur subst. 2061
vendre verbe 2056
beauté subst. 2054
riche adj. 2048
craindre verbe 2047
étrange adj. 2046
emporter verbe 2035
ensuite adv. 2032
soin subst. 2025
naturel adj. 2020
hasard subst. 2017
puis adv. 2013
condition subst. 2006
quinze numér. 2000
classe subst. 1997
voyage subst. 1996
auprès prép. 1955
présent subst. 1955
caractère subst. 1953
attention subst. 1952
gris adj. 1952
or conj. 1940
rouler verbe 1939
faible adj. 1934
posséder verbe 1931
scène subst. 1925
difficile adj. 1921
français subst. 1921
réveiller verbe 1921
foi subst. 1920
aider verbe 1918
découvrir verbe 1918
odeur subst. 1913
choisir verbe 1912
musique subst. 1912
oncle subst. 1909
événement subst. 1906
prononcer verbe 1905
village subst. 1905
taire verbe 1904
envie subst. 1903
midi subst. 1902
ensemble adv. 1953
expression subst 1990
herbe subst. 1896
vieux subst. 1896
pluie subst. 1895
près adv. 1894
bas subst. 1892
rêver verbe 1886
appuyer verbe 1884
étendre verbe 1884
après adv. 1882
général subst. 1882
lutte subst. 1880
trembler verbe 1880
réponse subst. 1877
grâce subst. 1873
espace subst. 1872
habitude subst. 1866
défendre verbe 1864
mémoire subst. 1861
créer verbe 1856
grave adj. 1856
maintenir verbe 1853
verre subst. 1845
campagne subst. 1840
quelqu'un pron. 1838
juge subst. 1832
genou subst. 1827
impossible adj. 1818
fête subst. 1816
indiquer verbe 1814
prêt adj. 1813
promettre verbe 1812
relever verbe 1810
abandonner verbe 1809
ignorer verbe 1797
large adj. 1792
parent subst. 1792
colère subst. 1790
exprimer verbe 1951
étoile subst. 1788
devoir subst. 1787
conscience subst 1784
existence subst. 1861
accompagner verb 1769
immobile adj. 1769
adresser verbe 1763
observer verbe 1757
juste adj. 1756
puissance subst. 1756
matière subst. 1755
sable subst. 1754
séparer verbe 1753
marier verbe 1752
prévoir verbe 1751
vivant adj. 1751
accord subst. 1746
hiver subst. 1745
retour subst. 1744
autrefois adv. 1740
genre subst. 1736
d'autres pron. 1734
vif adj. 1733
amener verbe 1731
obliger verbe 1729
acte subst. 1725
image subst. 1724
horizon subst. 1722
éclairer verbe 1720
poursuivre verbe 1719
danger subst. 1717
livrer verbe 1717
rôle subst. 1716
escalier subst. 1711
goût subst. 1708
bête subst. 1706
ceci pron. 1706
recherche subst. 1705
membre subst. 1704
pain subst. 1700
phrase subst. 1697
contenir verbe 1696
rire subst. 1692
fuir verbe 1688
couler verbe 1687
terme subst. 1687
eaux subst. 1680
moyen adj. 1679
police subst. 1678
rocher subst. 1678
proposer verbe 1676
tranquille adj. 1676
unique adj. 1675
éprouver verbe 1673
retenir verbe 1667
type subst. 1667
vin subst. 1656
supérieur adj. 1649
attacher verbe 1645
voler verbe 1642
sec adj. 1638
justice subst. 1636
époque subst. 1635
passage subst. 1635
somme subst. 1635
science subst. 1634
surprendre verbe 1633
côte subst. 1626
doucement adv. 1620
gauche subst. 1617
faute subst. 1613
école subst. 1612
bon subst. 1603
ensemble subst. 1603
rayon subst. 1602
briser verbe 1601
sujet subst. 1598
imaginer verbe 1596
diriger verbe 1593
douze numér. 1591
en adv. 1680
l'une pron. 1587
dernier subst. 1585
avis subst. 1582
parvenir verbe 1581
ouvert adj. 1578
pénétrer verbe 1574
poète subst. 1573
meilleur adj. 1571
paysan subst. 1570
remarquer verbe 1569
chair subst. 1568
éviter verbe 1568
soudain adv. 1568
succès subst. 1561
île subst. 1558
établir verbe 1556
réussir verbe 1553
pencher verbe 1550
habiter verbe 1547
entourer verbe 1546
déclarer verbe 1544
détail subst. 1544
arme subst. 1543
réalité subst. 1543
confiance subst. 1539
masse subst. 1539
crise subst. 1537
étonner verbe 1535
poste subst. 1535
dresser verbe 1528
durer verbe 1528
depuis adv. 1527
faux adj. 1527
fixer verbe 1527
énorme adj. 1526
principe subst. 1524
direction subst. 1517

taille subst. 1514
 désirer verbe 1512
 santé subst. 1512
 ventre subst. 1511
 marché subst. 1508
 puissant adj. 1506
 simplement adv. 1505
 environ adv. 1504
 tellement adv. 1504
 arracher verbe 1503
 entraîner verbe 1636
 soutenir verbe 1501
 couper verbe 1499
 trou subst. 1498
 inconnu adj. 1497
 pont subst. 1495
 lune subst. 1494
 dehors adv. 1491
 certes adv. 1490
 beaux adj. 1489
 robe subst. 1489
 douter verbe 1488
 retirer verbe 1487
 cesse subst. 1486
 brusquement adv. 1485
 entrée subst. 1507
 source subst. 1482
 camarade subst. 1471
 dent subst. 1470
 quant à prép. 1470
 connaissance subst. 1469
 cou subst. 1469
 but subst. 1466
 promener verbe 1460
 vague subst. 1460
 élément subst. 1459
 fil subst. 1457
 voie subst. 1457
 nez subst. 1453
 forcer verbe 1447
 particulier adj. 1446
 discours subst. 1443
 maladie subst. 1443
 chaleur subst. 1442
 gloire subst. 1440
 vide adj. 1438
 examiner verbe 1497
 revoir verbe 1436
 aide subst. 1434
 début subst. 1432
 ennemi subst. 1432
 second adj. 1431
 aile subst. 1426
 flamme subst. 1426
 chaise subst. 1422
 lourd adj. 1422
 sein subst. 1422
 véritable adj. 1422
 toit subst. 1421
 remplir verbe 1420
 terminer verbe 1419
 vaste adj. 1419
 nu adj. 1418
 poussière subst. 1413
 nord subst. 1411
 tenter verbe 1397
 émotion subst. 1393
 hors prép. 1390
 un numér. 1390
 remonter verbe 1389
 révolution subst 1388
 théâtre subst. 1388
 armée subst. 1386
 court adj. 1386
 noir subst. 1385
 appartement subs 1384
 semblable adj. 1384
 installer verbe 1383
 haine subst. 1382
 jeune subst. 1382
 position subst. 1381
 seconde subst. 1381
 frais adj. 1379
 appel subst. 1378
 soulever verbe 1375
 espoir subst. 1475
 allumer verbe 1373
 imposer verbe 1373
 avant adv. 1372
 respirer verbe 1371
 arrière subst. 1370
 baisser verbe 1370
 droite subst. 1370
 poitrine subst. 1370
 mort adj. 1369
 jeunesse subst. 1368
 bureau subst. 1367
 sac subst. 1367
 étranger adj. 1366
 courage subst. 1363
 souffler verbe 1363
 jaune adj. 1360
 page subst. 1360
 étranger subst. 1359
 etc adv. 1356
 miser subst. 1353
 passé subst. 1352
 rapide adj. 1351
 digne adj. 1350
 chaud adj. 1349
 propos subst. 1349
 attirer verbe 1348
 prêter verbe 1344
 clair adj. 1336
 amuser verbe 1329
 occasion subst. 1327
 voile subst. 1325
 éclater verbe 1323
 importance subst 1322
 quartier subst. 1322
 soi pron. 1322
 auteur subst. 1318
 religion subst. 1316
 palais subst. 1314
 réunir verbe 1314
 traiter verbe 1310
 flot subst. 1309
 intelligence subst. 1309
 tantôt adv. 1307
 voisin subst. 1307
 carte subst. 1305
 secret adj. 1301
 animal subst. 1296
 été subst. 1293
 traîner verbe 1293
 cabinet subst. 1292
 morceau subst. 1292
 employer verbe 1290
 capable adj. 1287
 souffrance subst 1286
 marquer verbe 1285
 prouver verbe 1285
 importer verbe 1284
 titre subst. 1284
 désert subst. 1282
 facile adj. 1280
 spectacle subst. 1280
 exiger verbe 1279
 reposer verbe 1277
 départ subst. 1276
 fier adj. 1276
 danser verbe 1275
 demande subst. 1268
 saluer verbe 1268
 leur subst. 1267
 joue subst. 1266
 saint adj. 1265
 accorder verbe 1264
 prière subst. 1264
 achever verbe 1262
 avouer verbe 1262
 distinguer verbe 1261
 emmener verbe 1261
 fonction subst. 1260
 durant prép. 1256
 haut adv. 1253
 aspect subst. 1251
 sommeil subst. 1251
 éclat subst. 1249
 moitié subst. 1248
 demi adj. 1247
 calme adj. 1246
 contraire subst. 1244
 colline subst. 1239
 agiter verbe 1238
 hésiter verbe 1232
 terrain subst. 1223
 rare adj. 1222
 poids subst. 1221
 sonner verbe 1221
 changement subst 1219
 charge subst. 1218
 davantage adv. 1218
 composer verbe 1217
 enlever verbe 1215
 poche subst. 1211
 rejoindre verbe 1210
 son subst. 1208
 intérieur adj. 1205
 veille subst. 1204
 ramener verbe 1203
 fruit subst. 1200
 complet adj. 1199
 étudier verbe 1199
 partager verbe 1198
 croix subst. 1192
 suivant adj. 1191
 chasser verbe 1187
 interrompre verb 1186
 éloigner verbe 1185
 trésor subst. 1185
 compagnie subst. 1184
 étroit adj. 1181
 cuisine subst. 1180
 réduire verbe 1177
 engager verbe 1309
 égal adj. 1175
 empire subst. 1174
 nation subst. 1170
 éteindre verbe 1169
 recommencer verb 1169
 sauter verbe 1169
 plaindre verbe 1168
 conversation subst. 1167
 soirée subst. 1166
 violent adj. 1166
 impression subst 1164
 trait subst. 1164
 devant adv. 1163
 préférer verbe 1162
 révéler verbe 1162
 sien pron. 1162
 magnifique adj. 1156
 désespoir subst. 1154
 témoin subst. 1153
 visite subst. 1152
 respect subst. 1148
 solitude subst. 1142
 subir verbe 1140
 delà adv. 1136
 prochain adj. 1136

anglais subst. 1135
 rapporter verbe 1135
 coûter verbe 1128
 réfléchir verbe 1128
 officier subst. 1127
 remercier verbe 1127
 déposer verbe 1126
 fauteuil subst. 1125
 fumer verbe 1124
 tôt adv. 1124
 affirmer verbe 1119
 relation subst. 1118
 fumée subst. 1117
 convenir verbe 1116
 branche subst. 1115
 malade adj. 1115
 circonstance subst. 1113
 ouvrage subst. 1113
 compagnon subst. 1108
 vêtir verbe 1108
 expérience subst. 1106
 port subst. 1106
 accomplir verbe 1105
 avec adv. 1105
 résoudre verbe 1103
 plonger verbe 1100
 goutte subst. 1099
 mien pron. 1099
 chant subst. 1098
 détruire verbe 1095
 combat subst. 1087
 personnage subst. 1086
 aventure subst. 1085
 intéresser verbe 1085
 disposer verbe 1084
 absence subst. 1081
 machine subst. 1079
 aucun pron. 1078
 grâce prép. 1078
 chaîne subst. 1077
 honte subst. 1076
 fait adj. 1075
 laisser verbe 1075
 faim subst. 1072
 plaine subst. 1072
 verser verbe 1071
 pointe subst. 1066
 obéir verbe 1065
 preuve subst. 1065
 éternel adj. 1063
 lutter verbe 1062
 prétendre verbe 1061
 bataille subst. 1060
 construire verbe 1060
 énergie subst. 1060
 victime subst. 1055
 sauvage adj. 1053
 soumettre verbe 1052
 usage subst. 1052
 peser verbe 1050
 double adj. 1046
 tache subst. 1045
 guère adv. 1044
 hauteur subst. 1043
 troubler verbe 1043
 tendre adj. 1039
 beau subst. 1037
 curiosité subst. 1037
 répandre verbe 1032
 glace subst. 1031
 résister verbe 1031
 froid subst. 1027
 prison subst. 1024
 étage subst. 1023
 billet subst. 1022
 droit adj. 1022
 sérieux adj. 1022
 protéger verbe 1021
 pauvre subst. 1019
 rose subst. 1019
 enfermer verbe 1016
 attitude subst. 1014
 dur adj. 1014
 mode subst. 1012
 neuf adj. 1010
 crainte subst. 1006
 creuser verbe 1006
 grandir verbe 1006
 enfoncer verbe 1004
 vêtement subst. 1003
 envelopper verbe 1001
 vague adj. 1000
 prévenir verbe 999
 violence subst. 998
 inspirer verbe 996
 inutile adj. 993
 content adj. 992
 courant subst. 992
 folie subst. 992
 pitié subst. 992
 intention subst. 988
 ramasser verbe 988
 endormir verbe 987
 inventer verbe 986
 trace subst. 985
 toile subst. 980
 presser verbe 977
 confier verbe 976
 effacer verbe 976
 reculer verbe 973
 user verbe 973
 blanc subst. 972
 nourrir verbe 971
 dangereux adj. 970
 poésie subst. 967
 sommet subst. 962
 remplacer verbe 961
 souhaiter verbe 960
 avance subst. 956
 autorité subst. 955
 épais adj. 954
 inquiétude subst. 953
 choix subst. 952
 tombe subst. 951
 marchand subst. 950
 nombreux adj. 949
 muet adj. 948
 signer verbe 947
 absolument adv. 946
 cercle subst. 944
 interroger verbe 944
 dominer verbe 942
 défaut subst. 940
 enfance subst. 938
 faveur subst. 937
 réel adj. 937
 commander verbe 934
 supposer verbe 934
 dépasser verbe 933
 sourd adj. 930
 cruel adj. 929
 dimanche subst. 928
 erreur subst. 928
 cerveau subst. 927
 accuser verbe 926
 arrivée subst. 924
 rapidement adv. 924
 vol subst. 922
 habiller verbe 921
 condamner verbe 920
 lors adv. 918
 menacer verbe 918
 seuil subst. 918
 écraser verbe 916
 perte subst. 915
 troisième adj. 914
 chance subst. 913
 vieil adj. 906
 même pron. 905
 céder verbe 902
 douceur subst. 901
 droite adj. 901
 vide subst. 900
 autrement adv. 899
 drôle adj. 899
 ruine subst. 899
 écarter verbe 898
 rang subst. 898
 réclamer verbe 897
 chiffre subst. 896
 voisin adj. 892
 militaire adj. 890
 roche subst. 884
 distance subst. 883
 apparence subst. 882
 dessiner verbe 881
 conclure verbe 877
 françois subst. 877
 lier verbe 877
 discussion subst. 876
 admettre verbe 873
 banc subst. 873
 terreur subst. 873
 attaquer verbe 872
 vers subst. 872
 respecter verbe 871
 rose adj. 869
 silencieux adj. 869
 anglais adj. 868
 course subst. 868
 portier subst. 865
 chat subst. 861
 pendre verbe 860
 supporter verbe 859
 tempête subst. 856
 parfaitement adv. 855
 paysage subst. 855
 quart subst. 855
 figurer verbe 853
 profiter verbe 851
 accrocher verbe 847
 calmer verbe 843
 satisfaire verbe 843
 public subst. 842
 race subst. 838
 valoir verbe 838
 barbe subst. 837
 signifier verbe 836
 couche subst. 835
 inquiéter verbe 834
 colon subst. 833
 désormais adv. 833
 fidèle adj. 831
 assister verbe 829
 rideau subst. 829
 inviter verbe 828
 déchirer verbe 827
 fatigue subst. 824
 risquer verbe 824
 règle subst. 823
 gauche adj. 822
 parcourir verbe 822
 présent adj. 822
 rejeter verbe 821
 naissance subst. 820
 loup subst. 818
 renoncer verbe 816
 complètement adv. 815
 extraordinaire adj. 814
 veiller verbe 810
 transformer verb 806

tracer verbe 805
chute subst. 804
divers adj. 803
résistance subst 802
contenter verbe 801
chemise subst. 800
mince adj. 800
naturellement ad 800
siège subst. 799
as subst. 797
patron subst. 797
calme subst. 796
mériter verbe 795
printemps subst. 795
angoisse subst. 793
précipiter verbe 791
rompre verbe 791
habitant subst. 788
plein prép. 788
caresser verbe 784
métier subst. 784
étouffer verbe 783
animer verbe 782
note subst. 782
passé adj. 781
fine adj. 779
fixe adj. 779
casser verbe 778
fusil subst. 777
rond adj. 774
agent subst. 771
fonder verbe 763
roman subst. 760
franchir verbe 759
plante subst. 759
abattre verbe 754
discuter verbe 754
fatiguer verbe 748
humide adj. 746
réflexion subst. 746
consentir verbe 745
accent subst. 740
curieux adj. 738
repas subst. 735
étendue subst. 731
regretter verbe 731
joindre verbe 730
profondément adv 730
secours subst. 729
commencement subst. 728
corde subst. 724
secrétaire subst 720
vaincre verbe 720
meilleur subst. 412

saison subst. 718
précieux adj. 715
précis adj. 714
consulter verbe 712
hair verbe 709
repousser verbe 701
paupière subst. 700
certainement adv 696
tapis subst. 695
noire adj. 694
chasse subst. 691
exécuter verbe 690
nerveux adj. 690
nul dét. 690
commun adj. 689
exposer verbe 689
clef subst. 686
claire adj. 684
voyager verbe 683
haute adj. 680
renverser verbe 680
sueur subst. 677
âgé adj. 676
ferme subst. 675
rassurer verbe 675
retomber verbe 674
décrire verbe 672
mentir verbe 670
instinct subst. 669
armer verbe 667
paquet subst. 667
drame subst. 666
absolu adj. 664
savoir subst. 661
mine subst. 660
vision subst. 660
étaler verbe 659
sentier subst. 658
demain subst. 657
beau adv. 655
blond adj. 652
essuyer verbe 651
planche subst. 650
précéder verbe 650
dehors subst. 649
salut subst. 647
tâche subst. 646
désigner verbe 644
fin adj. 643
abri subst. 642
détacher verbe 641
recueillir verbe 638
rencontre subst. 636

croiser verbe 634
entretenir verbe 633
rouge subst. 633
professeur subst 627
surveiller verbe 627
visible adj. 627
perdu adj. 625
réserver verbe 622
bas adv. 620
lien subst. 618
queue subst. 615
bande subst. 608
confondre verbe 608
grain subst. 605
mensonge subst. 603
dégager verbe 602
probablement adv 592
illusion subst. 591
incapable adj. 589
parer verbe 589
épreuve subst. 586
immédiatement adv 581
attente subst. 578
visiter verbe 572
instrument subst 571
évidemment adv. 569
auparavant adv. 556
détourner verbe 554
explication subs 553
régulier adj. 551
reproche subst. 549
souci subst. 547
plier verbe 546
extrême adj. 544
accueillir verbe 540
juif subst. 538
leçon subst. 538
redevenir verbe 538
approuver verbe 537
parfait adj. 536
emparer verbe 535
aborder verbe 532
heurter verbe 529
tel pron. 523
noyer verbe 519
semmer verbe 515
ferme adj. 514
manche subst. 494
rage subst. 492
gré subst. 491
guider verbe 450
piquer verbe 449