UNIVERZITA PALACKÉHO V OLOMOUCI

Filozofická fakulta

Katedra romanistiky

Lenka Bičová

Lexikální analýza vybraných částí scénáře filmu Amélie z Montmartru

(úvod do kvantitativní lingvistiky)

Bakalářská diplomová práce

Vedoucí práce: PhDr. Ondřej Andrys

Olomouc 2011

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

V Olomouci dne 21. dubna 2011

Lenka Bičová

Poděkování

Mé velké poděkování patří vedoucímu práce PhDr. Ondřeji Andrysovi za příjemnou spolupráci, ochotu a cenné rady při tvorbě mé Bakalářské diplomové práce.

Obsah

Úvod

Druhá polovina 20. století znamenala v oblasti vědy výraznou změnu. Vznikala řada pomezních disciplín zabývajících se specifickými otázkami, které vycházely z propojení dvou či více tradičních vědních oborů. Výjimkou není ani lingvistika, která zahájila spolupráci s vědními odvětvími, jako je například psychologie, sociologie, etnologie, neurologie nebo matematika. Právě spojení lingvistiky a matematiky, jímž se zabývá naše práce, je v dnešní době využíváno k mnoha vědeckým účelům a dopomáhá k exaktnějšímu popisu jazykovědy jako takové. Významně se podílí také na stanovení jazykových zákonitostí.

Naši práci s názvem *Lexikální analýza vybraných částí scénáře filmu Amélie z Montmartru* jsme rozdělili do dvou hlavních částí - teoretické a praktické.

V první, teoretické části, se pokoušíme nastínit základní pilíře matematické a kvantitativní lingvistiky. Členěníme ji do tří stěžejních kapitol. V první kapitole se zaměřujeme na obecné informace o vzniku pomezních disciplín. Ve druhé kapitole přesouváme naši pozornost k matematické lingvistice, v rámci níž si představíme její podobory. Jedním z podoborů je kvantitativní lingvistika, která je tématem třetí kapitoly. I zde se soustředíme na charakteristiku jednotlivých oblastí kvantitativní lingvistiky, s důrazem na lexikální a gramatickou statistiku, a přidáváme zajímavosti týkající se jejího využití v praxi. Důležitou součástí kapitoly je objasnění některých pojmů, které jsou s disciplínou neodmyslitelně spjaty. Teoretickou část doplňuje i podkapitola představující nejdůležitější díla lexikální statistiky, a to s důrazem na české a francouzské práce.

Ve druhé části práce jsme se rozhodli aplikovat naše poznatky v praxi. Jako výchozí text k lexikální analýze posloužily vybrané pasáže ze scénáře k filmu *Le Fabuleux Destin d'Amélie Poulain* (2001; *Amélie z Montmartru*). Autory scénáře jsou Guillaume Laurant a Jean-Pierre Jeunet. Výběrem tohoto textu chceme do jisté míry nahradit autentický mluvený projev a soustředit se

tak na nejméně propracovanou oblast ve statistickém výzkumu lexikální problematiky. Cílem praktické části je matematická analýza vlastního korpusu a vytvoření frekvenčního seznamu slov užitých v textu. Dílčím úkolem je také podívat se na analyzovaný text skrze perspektivu gramatické statistiky a definovat četnost jednotlivých gramatických jevů v textu.

Kvantitativní lingvistika je oborem dosti specifickým. V českém prostředí se touto problematikou velmi aktivně zabývá 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. Velmi poučný a zajímavý zdroj informací představují *Dějiny lingvistiky* (1996) Jiřího Černého, ve kterých najdeme samostatnou kapitolu o kvantitativní lingvistice a teorii informace. Peirre Giraud je autorem knihy *Problèmes et méthodes de la statistique linguistique* (1959), se kterou považujeme za vhodné se seznámit, byť její téma je již natolik specifické, že poznatků načerpaných z této knihy nebylo možné v naší práci více využít. 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.

Naše práce, věnovaná kvantitativní lingvistice, má za cíl přehledně vyložit problematiku dané disciplíny a prokázat platnost teoretických poznatků v praxi. Doufáme, že její četba a výsledky našeho zkoumání budou pro čtenáře užitečným a zajímavým zdrojem informací a případně také podnětem pro další studium v oblasti matematické lingvistiky.

TEORETICKÁ ČÁST

Vymezení úkolu teoretické části práce

V první části naší práce se budeme soustředit zejména na představení kvantitativní lingvistiky jako moderní pomezní disciplíny. Pojednáme o okolnostech jejího vzniku a charakterizujeme různá její odvětví zabývající se odlišnými otázkami z oblasti lingvistiky. Uvedeme též základní pojmy, týkající se statistické analýzy textu, a teoreticky popíšeme jejich praktické využití. Na závěr se seznámíme s některými nejdůležitějšími pracemi lexikální statistiky.

1. Pomezní disciplíny

1.1 Vznik disciplín

Ve druhé polovině 20. století došlo ve vědě ke značné změně v oblasti dalšího rozvoje dosavadních oborů. Do té doby poměrně izolovaná vědní odvětví začala postupně přecházet na nový systém, pro nějž je typická kombinace těchto oborů mezi sebou. To se týká jak exaktních přírodních věd (např. chemická biologie), tak věd humanitních (sociální psychologie). Vznikaly nové vědní disciplíny stojící na pomezi dvou nebo více disciplín tradičních, které tak v sobě syntetizovaly jejich základní otázky a taktéž jejich výzkumné metody. Tyto interdisciplinární tendence dopomohly po druhé světové válce k mnoha objevům, kterých bylo často dosaženo právě v rámci těchto pomezních oborů.

K utváření interdisciplinárních oborů docházelo i v jazykovědě. Otázky lingvistické se tak prolínají s problematikou dalších tradičních vědních odvětví jako je matematika, logika, neurologie, psychologie, sociologie apod. Vzniklo množství nových disciplín (např. matematická lingvistika, neurolingvistika, etnolingvistika, psycholingvistika), které si v současné lingvistice postupně vydobyly dominantní postavení.

„Několik dalších lingvistických disciplín vzniklo v 70. letech 20. století v souvislosti s tzv. pragmatickým obratem.“ (Černý 2005, s. 196). Doposud se jazykovědci zabývali pouze praktickými otázkami jazykového systému jako

celku (langue[1]) a o promluvy (paroles[2]), realizované určitými jedinci v určitý čas na určitém místě, nejevili příliš zájem. Pragmatický obrat spočíval v tom, že si lingvisté uvědomili důležitost pragmatických faktorů, které komunikaci doprovázejí. Tak se zrodila pragmalingvistika a teorie řečové činnosti, které studují právě tyto faktory, mezi něž můžeme zařadit např. okolnosti, za jakých se komunikační proces realizuje, vztahy mezi komunikujícími účastníky, jejich počet, jejich vzdálenost atd.

V souvislosti s pragmatickým obratem se zformovala také paralingvistika. Zabývá se jevy, které nemají s jazykovým systémem nic společného, ale přesto doprovázejí komunikační proces nebo ho doplňují. Mezi paralingvistické prostředky patří zejména mimika a gesta. Tyto neverbální signály, často nazývány jako „řeč těla“, mají schopnost zdůraznit a posílit to, co říkáme. Někdy může naopak řeč těla popírat náš slovní projev, aniž si to uvědomujeme. Tak jako každý jazyk má i řeč těla svůj systém, který je možné se do určité míry naučit napodobováním. Na druhé straně je způsob takové komunikace značně ovlivněn naší kulturou. Příkladem může být haptický kontakt (tedy kontakt hmatem). Každý jedinec (patrně v souvislosti se svou kulturou a také svým osobním naturelem) je zvyklý vymezovat si své vlastní hranice týkající se toho, co je pro něj v rámci hovoru a v dané situaci přijatelné, ať už se jedná o doteky či vzdálenost mezi osobami. Na téma řeči těla vznikla již řada publikací nabízejících výklad teorie, popisy modelových situací či procvičování řeči těla v praxi. Porozumění či vědomé ovládání řeči těla pak slibuje větší úspěchy v pracovní či vztahové oblasti (Černý 2008, s. 197).

Tolik tedy malý přehled o vývoji pomezních disciplín. Slovo „pomezní“ je třeba spojovat s druhem disciplíny stojící na pomezí dvou či více oborů, které jsou provázány společnými zájmy. Nejedná se v žádném případě o synonymum pro okrajové disciplíny, které se od lingvistické problematiky vzdalují kamsi na její nevýznamný okraj. Stojí naopak v samém centru její pozornosti.

[1] Termín Ferdinanda de Saussure
[2] Termín Ferdinanda de Saussure

V naší práci se budeme soustředit na disciplínu kombinující lingvistické a matematické postupy, tedy na lingvistiku matematickou. Přesněji na její podobor označovaný jako „kvantitativní lingvistika“.

2. Matematická lingvistika a její oblasti

2.1 Pozice kvantitativní lingvistiky mezi ostatními podobory

Kvantitativní lingvistika představuje speciální oblast lingvistiky matematické. Termínem „matematická lingvistika“ se dnes označují disciplíny na pomezí lingvistiky a matematiky, které při svém zkoumání jazyka využívají matematické metody. Za počátek matematické lingvistiky se obvykle považuje rok 1957, tzn. rok VIII. mezinárodního kongresu lingvistů v Oslo. Nelze ovšem říci, že před polovinou 20. století se matematika v lingvistice nevyskytovala. Matematické metody byly uplatňovány již ke konci 19. století. Jednalo se ovšem o postupy, které bychom dnes označili za kvantitativní či statistické. Z dnešního pohledu je to tedy právě kvantitativní lingvistika, která má v oblasti matematické práce s jazykem nejdelší tradici. Dokonce se dá říct, že *„určité náznaky užití kvantitativních metod v jazykovědě najdeme například už ve starověké Indii.“* (Černý 1996, s. 248).

Další odvětví matematické lingvistiky, tj. lingvistika algebraická a počítačová (či strojová), vznikla teprve koncem 50. let v souvislosti s rozvojem moderní logiky a výpočetní techniky. Tehdejší vznik matematické teorie informace do jisté míry ovlivnil také samotnou kvantitativní lingvistiku. *„V lingvistice se v té době začalo pracovat např. s pojmy a termíny bit, entropie, šum, redundance apod.“* (Černý 2005, s. 197).

Přestože kvantitativní lingvistika je v dnešní době běžně užívaným pojmem, často se také setkáváme s termínem *statistická lingvistika*. Toto alternativní označení je explicitním poukázáním na metodu výzkumu, která se v této lingvistické oblasti nejčastěji využívá, a někteří jazykovědci mu tudíž dávají přednost. Marie Těšitelová považuje však tento termín za příliš úzký ve

vztahu k celé oblasti kvantitativních metod a jejich aplikací v lingvistice vůbec. Za relativně nejvýhodnější řešení považuje termín kvantitativní lingvistika, který u nás zavedl představitel pražské školy a průkopník v oblasti fonologické statistiky B. Trnka (Těšitelová 1977, s. 8).

Kvantitativní a algebraická lingvistika jsou považovány za podobory převážně teoretické, které za pomoci svých specifických metod vytváří nové poznatky v oblasti jazyka, které jsou počítačovou lingvistikou využity v praxi, zejména při strojovém překladu.

2.2 Algebraická lingvistika

Vedle kvantitativní lingvistiky se začala rozvíjet, jak již bylo řečeno výše, také lingvistika algebraická, jejímž hlavním úkolem je tvorba matematických modelů jazyka. Označení „algebraická“ nás upozorňuje především na to, že v rámci matematické lingvistiky jde o využití jiných než kvantitativních metod. Takto vytvořené modely si kladou za cíl vyjádřit a popsat prostřednictvím rovnic a symbolů jisté zákonité vlastnosti, kterými disponuje zkoumaný jazyk nebo některé jeho části. *„Dosud nejpropracovanějším a velmi rozšířeným modelem je v této souvislosti Chomského generativní gramatika.“* (Černý 2005, s. 201). Jeho jazykové modely se soustředí především na zpracovávání syntaktické části jazyka, kde jsou užitečnou pomůckou zvláště u jazyků s pevným slovosledem, tedy např. u angličtiny a jiných jazyků analytického typu. V zásadě jsou však jeho modely použitelné pro všechny typy jazyka.

2. 3 Počítačová lingvistika

Jak jsme již zmínili, počítačová lingvistika je především praktickou aplikací poznatků z oblasti kvantitativní a algebraické lingvistiky, ale také svými potřebami významnou měrou ovlivňuje vývoj obou těchto disciplín. Nejčastějším posláním počítačové lingvistiky je strojový překlad. Již od 50. let pracuje na vývoji této oblasti řada specializovaných týmů, které se s většími či menšími

úspěchy snaží neustále vylepšovat proces překladu z výchozího jazyka do jazyka cílového. Navzdory velkým snahám, pramenícím z přehnaných očekávání, se však dosud nepodařilo vyvinout natolik všestranný program, který by umožnil počítačům vytvářet alespoň přijatelné překlady např. krásné literatury. Asi největších úspěchů bylo dosud dosaženo v překladu odborných textů s omezenou slovní zásobou.

Bylo zjištěno, že veškeré nedostatky pramení z toho, že vstupní a cílový jazyk zatím nebyly patřičně popsány. Týká se to hlavně jazykových složek, které se na srozumitelnosti překladu podílejí především. Poměrně dobře popsána už byla složka syntaktická, problémem však zůstává složka sémantická, kterou se ještě nepodařilo matematicky správně zachytit.

3. Úvod do kvantitativní lingvistiky

3.1 Počátky a struktura kvantitativní lingvistiky

Jak jsme již uvedli výše, 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. Jedná se o soubor znaků a pravidel, který, je-li správně aplikován, zprostředkovává mezi jazykově kompetentními jedinci vzájemné předávání a sdělování informací. Například v případě cizích jazyků se dorozumívací proces výrazně usnadní a zdokonalí, když si mluvčí dovede osvojit gramatická pravidla daného jazyka a dbá na znalost jeho slovní zásoby. 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 nestejnou 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

3.2 Lexikální statistika

Lexikální statistika (dále LS) je jednou z nejpropracovanějších a nejstarších disciplín kvantitativní lingvistiky. Cílem statistiky bývá *„širší kvantifikace slovní zásoby jako celku, dále pak jejích 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 matematiky. 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*.[3] Marie Těšitelová vymezuje tuto problematiku třemi termíny: *slovo* (či *tvar slova*) pro jakoukoliv 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.

3.3 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

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

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.

3.4 Ostatní oblasti

3.4.1 Fonologická statistika

Fonologická statistika kvantifikuje v textu segmentální jevy, tzn. fonémy[4] 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. Výzkum může nabídnout nový pohled na problematiku fonologických jevů, což je hlavním úkolem tohoto typu statistiky, vedle cílů, které má společné s kvantitativní lingvistikou. Podle G. Herdana se fonologické jevy vyznačují značnou stabilitou. Pomocí statistických metod zkoumáme rozložení fonémů v textech, kvantitativní poměry uvnitř fonologického systému, distribuční poměry uvnitř fonémů a tak dále (Těšitelová 1977, s. 102). Máme-li určit jednotku fonémové analýzy, máme na výběr z několika řešení. Marie Těšitelová uvádí ve své příručce stručný přehled odlišných přístupů ke stanovení takovéto jednotky[5]. V téže kapitole o fonologické statistice se můžeme seznámit s nejdůležitějšími pracemi z této oblasti. Zde zmíníme pouze S. Gsella, který pro francouzštinu zkoumal frekvenci fonémů v próze a poezii autorů 19. a 20. století.

[4] 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

[5] Pro více informací o stanovení jednotky fonologické analýzy: TĚŠITELOVÁ, Marie. *Lingvistické příručky : Kvantitativní lingvistika*. 1. vyd. Praha : Státní pedagogické nakl., 1977. s. 102

3.4.2 Grafematická statistika

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).

3.4.3 Stylistická statistika

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

3.4.4 Typologická statistika

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.

3.4.5 Glottochronologie

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. Zkoumá změny jazyka v průběhu staletí, kdy se jazyky oddělí od společného základu a vyvinou se z nich dialekty nebo přímo další samostatné jazyky, které stále patří do stejné jazykové rodiny. Glottochrolonologická 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 diferenciace, že lze mluvit o různých jazycích, i když příbuzných.“ (Těšitelová 1977, s. 121). Glottochronologií se od 50. let 20. století zaobírali 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).

3.5 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 prézentním a perfektním tvaru, v didaktickém systému tvoří tedy tvary základní. Potom následuje imperativ, kondicionál prézentní a pasívní 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).

Lettre	Code	Mot associé	International
A	·-	Allo	Alpha
B	-···	Bonaparte	Bravo
C	-·-·	Coca Cola	Charlie
D	-··	Docile	Delta
E	·	Eh	Echo
F	··-·	Farandole	Fox-trot
G	--·	Goldorak	Golf
H	····	Hilarité	Hôtel
I	··	Ici	India
J	·---	jablonovo	Juliet
K	-·-	Koalo	Kilo
L	·-··	Limonade	Lima
M	--	Moto	Mike
N	-·	Noel	November
O	---	Oporto	Oscar
P	·--·	Philosophe	Papa
Q	--·-	Quocoriko	Quebec
R	·-·	Ricoré	Romoe
S	···	Sardine	Sierra
T	-	Thon	Tango
U	··-	Union	Uniform
V	···-	Valparéso	Victor
W	·--	Wagon long	Whisky
X	-··-	Xtrocadéro	X-Ray
Y	-·--	Yoshimoto	Yankee
Z	--··	Zoro est la	Zoulou
é	··-··		
è	·-··-		
à	·--·-		
ö	---·		
ç	-·-··		

Co se týká grafematické statistiky, Jiří Černý ve svých *Dějinách 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.

Pro zajímavost se podívejme na tabulku s Morseovou abecedou a pomocnými výrazy (viz výše) tak, jak je užívá francouzština. V prvním sloupci vidíme písmeno, které má být převedeno, v druhém sloupci příslušný symbol a ve třetím potom slovo, které se k písmenu váže a funguje jako mnemotechnická pomůcka. Každá slabika obsahující písmeno „o“ se počítá jako čárka, slabika, která „o“ nemá, je počítána jako tečka. Čtvrtý sloupec obsahuje mezinárodní výraz používaný údajně ke snadnějšímu zapamatování symbolu (Maridat, 2008).

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[6], 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 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).

[6] 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.

3.6 Kritéria pro výběr materiálu k analýze

Výběr jazykového materiálu pro statistickou analýzu provádíme z hlediska:

- kvalitativního
- kvantitativního

3.6.1 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:

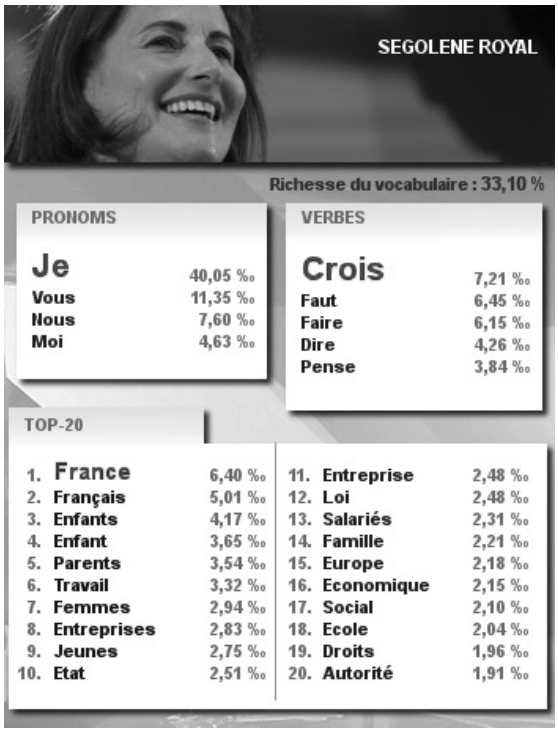
- lingvistické
- psychologické, sociologické
- tematické
- sémiotické

Některá kritéria si stručně představíme.

Lingvistické měřítko se uplatňuje tehdy, je-li cílem analýzy zkoumání jazyka vůbec, jazyka jednotlivých funkčních stylů (styl umělecký, publicistický, běžně dorozumívací) 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 jednotlivým každému z nich připisovat. K dosažení objektivity při zkoumání jazyka je vhodné zaměřit pozornost také na formu jazykových projevů, nejen psaných, ale také mluvených.

Velmi poučné je sledovat distribuci 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“, „Français“ atd. Je ovšem třeba si uvědomit, že vysoká frekvence zájmen ve francouzštině je dána specifickým vztahem osobních zájmen a slovesných tvarů.

Ani v současné době se ve Francii neupouští od analyzování promluv předních politiků. Výsledky zatím posledního výzkumu byly představeny na začátku tohoto roku. 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. Z korpusu byla eliminována slova, která jsou považovaná 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. S jejich politickými ambicemi úzce souvisí tematicky zaměřená sekce „Top-20“.

Pro představu se podívejme na výsledky analýzy u dvou známých francouzských politiků. U Ségolène Royalové, členky Socialistické strany, která v roce 2007 kandidovala na funkci prezidenta republiky, a u Nicolase Sarkozyho, současného francouzského prezidenta. Ten je v porovnání s jinými politiky rekordmanem v užívání první osoby. Zájmen „je“ a „moi“ užívá 2x více než jeho političtí kolegové, slovesný tvar „veux“ 4x, slova jako „famille“ a „police“ 3,5x a termín „immigration“ užívá 2,5x častěji než ostatní. Ségolène Royalová je šampionkou ve výběru slov z oblasti rodiny jako „mère“ (20x častější než u ostatních), „père“, „enfant“ a "parents" (16x častější). Z pohledu slovesného u ní byla zaznamenána snaha o přesvědčivost pomocí slov „crois“, „veux“ a „pense“, zmiňovaných 2x až 3x častěji. Kromě toho je Royalová Sarkozymu téměř rovnocenným soupeřem co se týká záliby v nadměrném využívání zájmena „je“, což lze bezpochyby připsat její aktivní účasti ve volební kampani (Abou El Khair; Deshayes, 2011).

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

3.6.2 Kvantitativní hledisko

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 nějž 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.

3.7 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.

3.7.1 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 Étiennem Brunetem (viz příloha) a zpracovává celkem 1500 slov, jež se ve francouzštině vyskytují nečastěji. Začátek frekvenčního seznamu slov uspořádaných podle klesající tendence vypadá takto:

1.	*le*	1050561	6.	*à*	293083
2.	*de*	862100	7.	*il*	270395
3.	*un*	419564	8.	*avoir*	248488
4.	*être*	351960	9.	*ne*	186755
5.	*et*	362093	10.	*je*	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.	*umělec*	50	58
51.	*chtíti*	51	57
52.	*hráti*	52	56
53.	*u*	53	55
54.	*až*	54	54
55.	*nějaký*		54
56.	*od(e)*		54
57.	*všechen*		54
58.	*říkati*	55	53
59.	*vel(i)ký*		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 (1996, s. 255*) „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.“*

3.7.2 Zipfovy zákony

Problematiky vztahu mezi absolutní frekvencí slova a jejich rankem se týkají tzv. Zipfovy zákony. Jejich formulaci vytvořil George Kingley Zipf, 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.[7] Zajímal se také o psychologické

[7] 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.*“ (1996, s. 253)

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*.[8] Jedna má tendenci produkovat slova s co největší frekvencí a snížit tak jejich množství. Proti ní působí druhá síla, která se snaží 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:

$$r \,.\, f = k$$

r znamená rank slova, který je zde používán bez přihlédnutí 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

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

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 taktéž 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 nepříliš 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).

3.7.3 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 1961[9] 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.

3.7.4 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.

[9] JELÍNEK, Jaroslav; BEČKA, Josef V. ; TĚŠITELOVÁ, Marie. *Frekvence slov, slovních druhů a tvarů v českém jazyce*, 1. vyd. Praha : Státní pedagogické nakl., 1961

3.8 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 pracovat[10]
- 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).

[10] 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)

3.8.1 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, 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.[11] Ve třetí verzi od tohoto systému upustil. Také Thorndikovy slovníky se staly zdrojem pro další lingvistické práce.

[11] 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).

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 výučování 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 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 přinášejí, 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.

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.[12] 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 miliónu 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 proporčně 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 zpracováváním psaných pramenů a rozsahem korpusu, který se pohyboval průměrně okolo jednoho miliónu slov. Proto je Kädingův slovník z roku 1897 tolik ceněn. Se svou materiálovou základnou okolo 10 miliónů 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*

[12] *Frekvence slov, slovních druhů a tvarů v českém jazyce*, Praha 1961 (zkr. FSČ)

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.

PRAKTICKÁ ČÁST

Vymezení úkolu praktické části práce

V návaznosti na náš teoretický úvod do kvantitativní lingvistiky jsme se rozhodli aplikovat v druhé části práce dosud získané vědomosti v praxi. Záměrem je matematicky analyzovat vlastní korpus získaný z francouzského textu tak, aby při práci bylo v co největší míře využito dosavadních poznatků. Hlavním cílem je vytvořit frekvenční seznam slov užitých ve vybraném textu, dále stanovit procentuální rozložení jednotlivých slovních druhů v textu a procentuální rozložení jednotlivých gramatických jevů v daném textu.

Na základě získaných údajů se pokusíme vysledovat určité zákonitosti či souvislosti mezi textovým stylem a jeho lexikální vybaveností a využitím vybraných gramatických jevů.

1. Výběr textu

Při výběru textu k analýze jsme si stanovili několik kritérií, přičemž určujícím bodem byla dostupnost daného textu v originálním francouzském znění.

Jelikož naše práce má v zásadě charakter pouhého úvodu do problematiky, nekladli jsme si nyní za cíl zpracovávat příliš rozsáhlý materiál. S ohledem na poměrně úzký korpus (1962 slov) nelze tedy výsledky naší práce prezentovat jako jevy obecně platné pro francouzský jazyk jako celek. „*Pravděpodobnost obecné platnosti zjištěného pořadí slova vzrůstá s rozsahem materiálu užitého k tabulaci.*“ (Jelínek; Bečka; Těšitelová 1961, s. 19). Naše analýza může nicméně posloužit svému základnímu účelu, to jest pokusu o prezentování jazyka z matematického pohledu.

Jako materiál k analýze jsme poněkud netradičně zvolili dialogovou listinu. Jedná se o realizovanou podobu scénáře k francouzskému filmu *Le*

Fabuleux Destin d'Amélie Poulain (2001, R: Jean-Pierre Jeunet)[13], jenž čeští diváci znají pod názvem *Amélie z Montmartru*. Film byl celosvětově úspěšný, získal řadu ocenění (mezi nimi i Českého lva v kategorii Nejlepší zahraniční film roku 2001) a nominací a byl také pět krát nominován na Oscara, mimo jiné právě za scénář, jehož autory jsou Guillaume Laurant a Jean-Pierre Jeunet.[14]

Jak jsme již nastínili v teoretické části, obvykle bývá hlavním pramenem podobných analýz soubor několika textů pokrývajících více stylových vrstev, mezi nimiž často najdeme beletrii, poezii, dramata, literaturu pro mládež, odbornou literaturu, žurnalistiku, vědeckou literaturu, transkribované mluvené projevy atd. Za nejproblematičtější a nejobtížněji postižitelný zdroj bývají považovány mluvené projevy, proto je také většina současných korpusových projektů zaměřena na psaný jazyk. Mluvené projevy mohou být z části nahrazovány dialogy, dramaty, či například redigovanými projevy politiků. Také charakter současné beletrie se svým jazykem více přibližuje mluveným projevům.

Za korpus nám poslouží výběr jednotlivých scén a dialogů z filmového scénáře, kterým chceme do jisté míry nahradit autentický mluvený projev současnosti a soustředit se tak na nejméně propracovanou oblast ve statistickém výzkumu lexikální problematiky. Výhodou tohoto scénáře je také přítomnost vypravěče, který uvádí diváky do děje, poskytuje komentáře k jednotlivým situacím a vysvětluje některé podrobnosti vztahující se k příběhu. Tento fakt nám umožňuje bezprostředně porovnat jazykovou rozdílnost dialogových scén s vypravěčovými proslovy, které se na první pohled vyznačují větší prozaičností a popisností. Předpokládáme, že tyto skutečnosti se mohou projevit jak na lexiku, tak na četnosti určitých gramatických jevů v textu.

[13] JEUNET, Jean-Pierre; LAURANT, Guillaume. 2003. *Le Fabuleux destin d´Amélie Poulain : Le Scénario*. Stuttgart : Ernst Klett Verlag, 2003. 80 s. Scénář byl vydán ve Stuttgartu pro studijní účely.
[14] Pro podrobný přehled ocenění a nominací, viz The Internet Movie Database (http://www.imdb.com/title/tt0211915/awards)

2. Metodologie

V naší praktické 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.

V první fázi bylo potřeba vybrat ze scénáře pasáže určené k tabulaci.[15] Z textu byly záměrně vytříděny přibližně stejně dlouhé úryvky z promluv vypravěče a také z některých dialogových scén. Scény byly zvoleny čistě na základě naší osobní preference.

Druhým krokem bylo spočítání slov v textu, protože statistické údaje o počtu slov automaticky vygenerované textovým editorem[16] se ukázaly jako nepřesné. Následovalo určování slovních druhů. Jako nejspolehlivější variantu ručního zpracovávání jsme zvolili vizuální, barevné odlišování jednotlivých sl. 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. Za užitečný považujeme v tomto směru také frekvenční seznam uveřejněný na webových stránkách francouzského ministerstva školství, jež sestavil lexikolog Étienne Brunet.[17] Nezbytnou součástí postupu bylo následné rozdělení textu do deseti diferencovaných seznamů (dle počtu slovních druhů) a sečtení slov v rámci každého z nich. Tyto seznamy nám umožnily přehlednější práci s jednotlivými kategoriemi. Výsledkem byla charakteristika materiálu z hlediska procentuálního a početního rozložení slov v celém textu i v jeho jednotlivých částech. Získaná data jsme zanesli do tabulky a většinu z nich převedli i do koláčových grafů.

---

[15] Materiál vybraný k analýze je součástí přílohy. Étienne Brunet je mimojiné autorem několika kvantitativních studií. Uveďme *Le vocabulaire de Proust* (1983), *Le vocabulaire de Zola* (1985) či *Le vocabulaire de Victor Hugo* (1988)

[16] Microsoft Office Word 2003

[17] Frekvenční seznam je k dispozici v příloze.

Při zjišťování lexikální frekvenčnosti jsme v každém ze seznamů našli opakující se slova a četnost opakování zapsali. Poté, co byl takovému postupu podroben každý slovní druh, spojili jsme všechny dílčí seznamy do jednoho. Slovní druhy se nám tak promíchaly a za pomocí 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. Klíčem k zacházení se získanými daty nám byla následující kritéria:

Podstatná jména (substantiva)

- Vyskytuje-li se v textu samostatné heslo, které je uvedeno jak v singuláru, tak v plurálu (*une fille; des filles*), do frekvenčního seznamu je slovo zařazeno pouze ve své singulární podobě.
- Pokud se substantivum nachází v textu pouze v plurálu, ve FS je zaneseno jako plurálový tvar (tedy nikoli jako singulár).
- Názvy měst chápeme jako jedno slovo (*Enghien-les-Bains*, *Pas-de-Calais*), názvy ulic rozdělujeme na jména obecná a vlastní, pokud to jde (například *Havre Caumartin* počítáme jako dvě hesla, *Boulevard de Strasbourg* jako 3 hesla atd.).
- Výjimku tvoří slova složená se spojovníkem - slova typu *micro-ondes*, *cache-pot* a *grand-mère* uvádíme jako jedno slovo, značku vozu *Citroën G7* počítáme taktéž jako jedno slovo.
- Co se týká následného kvantitativního zpracovávání 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 *année* 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 femininu (*bon; bonne*), do frekvenčního seznamu je slovo zařazeno pouze ve tvaru maskulina.
- Pokud se adjektivum nachází v textu pouze ve femininu či pouze v maskulinu, ve FS je zaneseno jako tvar feminina (či masulina).
- Adjektivum *ex-taulard* je počítáno jako jedno slovo.

Determinanty

- U determinantů respektujeme čísla i rody, u hesel je tedy nijak neupravujeme a zaznamenáváme je tak, jak se vyskytují v textu.
- Složenou číslovku *dix-huit* počítáme jako jedno slovo.

Zájmena (pronomina)

- U zájmen respektujeme čísla i rody a hesla uvádíme ve tvaru, ve kterém se vyskytují v textu.

Slovesa (verba)

- Slovesa uvádíme v infinitivu.
- Slovesné tvary typu *il a écrit* chápeme v naší lexikální statistice jako tři slova.
- Zvratnou podobu sloves typu *se lier* chápeme jako 2 slova
- 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* či *quant à* jsou počítány jako jedno slovo.

Spojky (konjunkce)

- Spojka *parce que* je počítána jako jedno slovo.

Pro příslovce (adverbia), l´introducteur a mot-phrase nebylo z našeho pohledu nutné stanovovat speciální kritéria.

Termíny *introducteur* a *mot-phrase* jsme ponechali v původní francouzské podobě. Jde o specifické slovní druhy, které nemají ekvivalentní pojmenování v české jazykovědné terminologii. Pod pojmem *introducteur* rozumíme neohebný slovní druh, který má schopnost uvádět slova, syntagmata či věty (např. *VOICI votre journal*.). 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.).

3. Statistická analýza vybraných částí filmového scénáře

3.1 Lexikální statistika

3.1.1 Frekvenční seznam

Table hiérarchique

être *verbe*	72	**tu** *pron.*	9	**enfin** *adv.*	3
de *prép.*	66	**là** *adv.*	8	**falloir** *verbe*	3
le *dét.*	54	**sa** *dét.*	8	**fille** *subst.*	3
avoir *verbe*	53	**bien** *adv.*	7	**frein** *subst.*	3
la *dét.*	50	**dire** *verbe*	7	**heure** *subst.*	3
il *pron.*	34	**tout** *pron.*	7	**chose** *subst.*	3
un *dét.*	34	**arriver** *verbe*	6	**instant** *subst.*	3
à *prép.*	33	**de** *dét.*	6	**main** *subst.*	3
ce *pron.*	33	**les** *pron.*	6	**mal** *adv.*	3
elle *pron.*	28	**temps** *subst.*	6	**même** *adj.*	3
je *pron.*	28	**ce** *dét.*	5	**mère** *subst.*	3
ne *adv.*	28	**devant** *prép.*	5	**minutes** *subst.*	3
les *dét.*	26	**enfant** *subst.*	5	**papa** *subst.*	3
pas *adv.*	24	**homme** *subst.*	5	**parler** *verbe*	3
qui *pron.*	24	**la** *pron.*	5	**passer** *verbe*	3
dans *prép.*	22	**me** *pron.*	5	**père** *subst.*	3
que *conj.*	22	**mon** *dét.*	5	**prendre** *verbe*	3
des *dét.*	21	**non** *adv.*	5	**regarder** *verbe*	3
le *pron.*	21	**ou** *conj.*	5	**rien** *pron.*	3
et *conj.*	20	**oui** *adv.*	5	**sentir** *verbe*	3
une *dét.*	19	**par** *prép.*	5	**seul** *adj.*	3
au *dét.*	16	**pouvoir** *verbe*	5	**tenir** *verbe*	3
en *prép.*	16	**ses** *dét.*	5	**tomber** *verbe*	3
on *pron.*	16	**ans** *subst.*	4	**toutes** *dét.*	3
aimer *verbe*	15	**appeler** *verbe*	4	**trois** *dét.*	3
faire *verbe*	15	**aux** *dét.*	4	**trouver** *verbe*	3
pour *prép.*	15	**croire** *verbe*	4	**vie** *subst.*	3
du *dét.*	14	**elles** *pron.*	4	**10** *dét.*	2
que *pron.*	14	**mettre** *verbe*	4	**15** *dét.*	2
ça *pron.*	13	**Poulain** *subst.*	4	**achever** *verbe*	2
se *pron.*	13	**plus** *adv.*	4	**ange** *subst.*	2
Amélie *subst.*	12	**sonner** *verbe*	4	**année** *subst.*	2
voir *verbe*	12	**te** *pron.*	4	**bain** *subst.*	2
avec *prép.*	11	**voilà** *introducteur*	4	**baskets** *subst.*	2
comme *conj.*	11	**vouloir** *verbe*	4	**béchamel** *subst.*	2
lui *pron.*	11	**air** *subst.*	3	**boucherie** *subst.*	2
son *dét.*	11	**alors** *adv.*	3	**cabine** *subst.*	2
sur *prép.*	11	**Amandine** *subst.*	3	**café** *subst.*	2
y *pron.*	11	**avant** *prép.*	3	**camion** *subst.*	2
ce *pron.*	10	**boîte** *subst.*	3	**Camus** *subst.*	2
mais *conj.*	10	**bon** *adj.*	3	**caniveau** *subst.*	2
petit *adj.*	10	**Bredoteau** *subst.*	3	**cet** *dét.*	2
si *conj.*	10	**celui** *pron.*	3	**cette** *dét.*	2
vous *pron.*	10	**eau** *subst.*	3	**cinquante** *dét.*	2
aller *verbe*	9	**en** *pron.*	3	**côté** *subst.*	2
moi *pron.*	9	**encore** *adv.*	3	**couper** *verbe*	2
quand *conj.*	9	**enfance** *subst.*	3	**dame** *subst.*	2

depuis *prép.* 2
descentes *subst.* 2
deux *dét.* 2
droite *adj.* 2
drôle *adj.* 2
empêcher *verbe* 2
endives *subst.* 2
escalier *subst.* 2
état *subst.* 2
exactement *adv.* 2
finir *verbe* 2
fois *subst.* 2
garder *verbe* 2
gardien *adj.* 2
gare *subst.* 2
gâteux *adj.* 2
genoux *subst.* 2
Gina *subst.* 2
gratin *subst.* 2
chat *subst.* 2
jour *subst.* 2
journée *subst.* 2
lauriers *subst.* 2
leur *dét.* 2
marcher *verbe* 2
mauvais *adj.* 2
mes *dét.* 2
nerf *subst.* 2
nettoyer *verbe* 2
nous *pron.* 2
outils *subst.* 2
parce que *conj.* 2
partir *verbe* 2
Philomène *subst.* 2
plaire *verbe* 2
plis *subst.* 2
préférer *verbe* 2
quelqu´un *pron.* 2
question *subst.* 2
quoi *pron.* 2
ranger *verbe* 2
Raphaël *subst.* 2
rendre *verbe* 2
rester *verbe* 2
ricochets *subst.* 2
rire *subst.* 2
Rodrigue *subst.* 2
rouges *adj.* 2
rue *subst.* 2
sac *subst.* 2
sans *prép.* 2
sortir *verbe* 2
Suzanne *subst.* 2
tard *adv.* 2
tête *subst.* 2
toujours *adv.* 2
tout *adv.* 2
type *subst.* 2
venir *verbe* 2
viande *subst.* 2
vider *verbe* 2
vieille *adj.* 2
virages *subst.* 2
volant *subst.* 2
vrai *adj.* 2
vraiment *adv.* 2
108 *dét.* 1
120 *dét.* 1
12h15 *pron.* 1
22 *dét.* 1
26ème *pron.* 1
37 *dét.* 1
50 *dét.* 1
5ème *pron.* 1
88 *dét.* 1
accident *subst.* 1
actionner *verbe* 1
afghans *adj.* 1
âge *subst.* 1
ah *mot-phrase* 1
aligner *verbe* 1
amant *subst.* 1
ancien *adj.* 1
anguille *subst.* 1
anomalie *subst.* 1
aperçu *subst.* 1
assis *adj.* 1
attention *subst.* 1
attrapes *subst.* 1
aujourd'hui *adv.* 1
automne *subst.* 1
autre *adj.* 1
avenante *adj.* 1
aventuriers *subst.* 1
banque *subst.* 1
baraque *subst.* 1
battre *verbe* 1
bavette *subst.* 1
bébé *subst.* 1
ben *introducteur* 1
béni *adj.* 1
bistrot *subst.* 1
blonde *adj.* 1
boire *verbe* 1
bol *subst.* 1
bord *subst.* 1
bortsch *subst.* 1
bouche *subst.* 1
boulevard *subst.* 1
bouleversée *adj.* 1
braqueurs *subst.* 1
bras *subst.* 1
briller *verbe* 1
briser *verbe* 1
Brossard *subst.* 1
brosse *subst.* 1
bruit *subst.* 1
brûlées *adj.* 1
cabane *subst.* 1
cagoules *subst.* 1
cache-pot *subst.* 1
calmer *verbe* 1
camarades *subst.* 1
cardiaque *adj.* 1
Caumartin *subst.* 1
cavale *subst.* 1
celle *pron.* 1
ces *dét.* 1
cigarette *subst.* 1
cinglés *subst.* 1
cirer *verbe* 1
Citroën G7 *subst.* 1
cœur *subst.* 1
coin *subst.* 1
coincer *verbe* 1
collectionner *verbe* 1
collègue *subst.* 1
coller *verbe* 1
comportement *subst.* 1
comptabilité *subst.* 1
compter *verbe* 1
connaître *verbe* 1
contact *subst.* 1
container *subst.* 1
contempler *verbe* 1
continuer *verbe* 1
contrariété *subst.* 1
copain *subst.* 1
cordonnier *subst.* 1
cotillons *subst.* 1
couloir *subst.* 1
cours *subst.* 1
crabes *subst.* 1
craquer *verbe* 1
crèmes *subst.* 1
créneau *subst.* 1
crever *verbe* 1
croûte *subst.* 1
cuillère *subst.* 1
cuisiner *verbe* 1
cultiver *verbe* 1
cuvette *subst.* 1
daim *subst.* 1
danseuse *subst.* 1
déception *subst.* 1
découvrir *verbe* 1
déguisements *subst.* 1
dentifrice *subst.* 1
dents *subst.* 1
départ *subst.* 1
dépendre *verbe* 1
déranger *verbe* 1
dernière *adj.* 1
derrière *prép.* 1
dès *prép.* 1
descendre *verbe* 1
deuxièmement *adv.* 1
devenir *verbe* 1
Dieu *subst.* 1
digérer *verbe* 1

digestion *subst.* 1
disponible *adj.* 1
dix-huit *dét.* 1
doigts *subst.* 1
domicile *subst.* 1
dominant *adj.* 1
dont *pron.* 1
dos *subst.* 1
draps *subst.* 1
éconduit *adj.* 1
écouter *verbe* 1
écrivain *subst.* 1
effleurée *adj.* 1
eh *introducteur* 1
emballages *subst.* 1
emmerder *verbe* 1
emprunter *verbe* 1
encorner *verbe* 1
Enghien-les-Bains *subst.* 1
entendre *verbe* 1
entrailles *subst.* 1
entrer *verbe* 1
envie *subst.* 1
épaulettes *subst.* 1
épicerie *subst.* 1
équestre *adj.* 1
escalator *subst.* 1
espérances *subst.* 1
espionner *verbe* 1
essayer *verbe* 1
est *subst.* 1
établissements *subst.* 1
étudier *verbe* 1
eux *pron.* 1
évoquer *verbe* 1
examen *subst.* 1
exceptionnelle *adj.* 1
exécuter *verbe* 1
explications *subst.* 1
ex-taulard *adj.* 1
Fabien *subst.* 1
face *subst.* 1
faite *adj.* 1
fanfare *subst.* 1
farces *subst.* 1
femme *subst.* 1
fenêtre *subst.* 1
filet *subst.* 1
fillette *subst.* 1
fils *subst.* 1
fleuriste *subst.* 1
flics *subst.* 1
fortune *subst.* 1
Fouet *subst.* 1
fourgonnette *subst.* 1
francs *subst.* 1
frontière *subst.* 1
fruit *subst.* 1
fumer *verbe* 1
gamin *subst.* 1
garçon *subst.* 1
gens *subst.* 1
Georgette *subst.* 1
gorge *subst.* 1
goût *subst.* 1
grand-mère *subst.* 1
grimée *adj.* 1
guérisseuse *subst.* 1
Gueugnon *subst.* 1
hasard *subst.* 1
hauteur *subst.* 1
havre *subst.* 1
herbe *subst.* 1
heureusement *adv.* 1
heureux *adj.* 1
Hipolito *subst.* 1
histoires *subst.* 1
hm *mot-phrase* 1
hôtesse *subst.* 1
humaine *adj.* 1
humilié *adj.* 1
chacun *pron.* 1
chamade *subst.* 1
champs *subst.* 1
chaude *adj.* 1
chaussures *subst.* 1
chercher *verbe* 1
cheval *subst.* 1
chevaline *adj.* 1
chien *subst.* 1
ils *pron.* 1
imaginaire *adj.* 1
imaginer *verbe* 1
immense *adj.* 1
imprimés *adj.* 1
instable *adj.* 1
institutrice *subst.* 1
intimité *subst.* 1
Istanbul *subst.* 1
jaloux *adj.* 1
jamais *adv.* 1
jeter *verbe* 1
jeune *adj.* 1
Joseph *subst.* 1
joue *subst.* 1
juste *adv.* 1
justement *adv.* 1
laquelle *pron.* 1
lendemain *subst.* 1
leur *pron.* 1
lier *verbe* 1
lilas *subst.* 1
longtemps *adv.* 1
lors *adv.* 1
loterie *subst.* 1
Ludovic *subst.* 1
mademoiselle *subst.* 1
magasin *subst.* 1
machine *subst.* 1
maillot *subst.* 1
maintenant *adv.* 1
malade *subst.* 1
mâle *subst.* 1
malice *subst.* 1
manger *verbe* 1
manière *subst.* 1
maquillée *adj.* 1
Martys *subst.* 1
matin *subst.* 1
médecin *subst.* 1
médical *adj.* 1
Médrano *subst.* 1
mensuel *adj.* 1
mètres *subst.* 1
métro *subst.* 1
micro-ondes *subst.* 1
mieux *adv.* 1
mignon *adj.* 1
migraine *subst.* 1
militaire *adj.* 1
militant *adj.* 1
mine *subst.* 1
miracle *subst.* 1
missiles *subst.* 1
Mme *subst.* 1
moineaux *subst.* 1
mois *subst.* 1
monaco *subst.* 1
monde *subst.* 1
montagnards *subst.* 1
motif *subst.* 1
Moudjahidin *subst.* 1
musique *subst.* 1
nain *subst.* 1
naître *verbe* 1
nature *subst.* 1
néon *subst.* 1
nerveuse *adj.* 1
nettoyage *subst.* 1
Nino *subst.* 1
nom *subst.* 1
Normandie *subst.* 1
nostalgique *adj.* 1
noter *verbe* 1
nouveau *adj.* 1
nuits *subst.* 1
obligée *adj.* 1
observer *verbe* 1
obstiner *verbe* 1
offrir *verbe* 1
orgasme *subst.* 1
originaire *adj.* 1
os *subst.* 1
otage *subst.* 1
où *adv.* 1
oublier *verbe* 1
ouvrir *verbe* 1
paraître *verbe* 1
par-dessus *prép.* 1
pardon *mot-phrase* 1

pareil *adj.* 1
parfait *adj.* 1
parfum *subst.* 1
parquet *subst.* 1
partance *subst.* 1
particulier *adj.* 1
Pas-de-Calais *subst.* 1
patins *subst.* 1
patron *subst.* 1
patronne *subst.* 1
payer *verbe* 1
pécari *subst.* 1
pédale *subst.* 1
pendant *prép.* 1
penser *verbe* 1
personne *pron.* 1
peu *adv.* 1
pfff *mot-phrase* 1
photo *subst.* 1
photomaton *subst.* 1
physique *adj.* 1
pinces *subst.* 1
pisser *verbe* 1
plaisirs *subst.* 1
planquer *verbe* 1
plastique *subst.* 1
plein *adj.* 1
pleurer *verbe* 1
plissés *adj.* 1
plonger *verbe* 1
poinçonner *verbe* 1
poinçonneur *subst.* 1
poissonnerie *subst.* 1
portatif *adj.* 1
porter *verbe* 1
poser *verbe* 1
possibles *adj.* 1
poulets *subst.* 1
pourquoi *adv.* 1
pourquoi *conj.* 1
pourtant *adv.* 1
poussette *subst.* 1
première *adj.* 1
premièrement *adv.* 1
présent *adj.* 1
profond *adv.* 1
promotion *subst.* 1
proposer *verbe* 1
provoquer *verbe* 1
puis *adv.* 1
pustules *subst.* 1
quant à *prép.* 1
quart *subst.* 1
quartier *subst.* 1
quel *dét.* 1
quinze *dét.* 1
quitter *verbe* 1
raconter *verbe* 1
raison *subst.* 1
ramoneur *subst.* 1
rangé *adj.* 1
raté *adj.* 1
rebord *subst.* 1
réconcilier *verbe* 1
reconstituer *verbe* 1
reconstitution *subst.* 1
recueillir *verbe* 1
reflet *subst.* 1
regard *subst.* 1
relever *verbe* 1
remplaçant *subst.* 1
renverser *verbe* 1
répéter *verbe* 1
répondre *verbe* 1
reste *subst.* 1
résultat *subst.* 1
retard *subst.* 1
retraite *subst.* 1
retrouver *verbe* 1
réussir *verbe* 1
revanche *subst.* 1
revenir *verbe* 1
revernir *verbe* 1
riz *subst.* 1
rognons *subst.* 1
rôtir *verbe* 1
roue *subst.* 1
rouillée *adj.* 1
rouler *verbe* 1
routier *subst.* 1
salut *mot-phrase* 1
sauter *verbe* 1
sciatique *adj.* 1
scientifique *adj.* 1
scotchées *adj.* 1
second *pron.* 1
seconde *subst.* 1
secret *subst.* 1
secrète *adj.* 1
Seine *subst.* 1
semer *verbe* 1
sens *subst.* 1
serrer *verbe* 1
serveuses *subst.* 1
servir *verbe* 1
sexuel *adj.* 1
si *adv.* 1
six *dét.* 1
soin *subst.* 1
soir *subst.* 1
sol *subst.* 1
soudain *adv.* 1
souvenir *subst.* 1
souvenir *verbe* 1
souvent *adv.* 1
soviétiques *adj.* 1
sportifs *subst.* 1
station *subst.* 1
stop *subst.* 1
Strasbourg *subst.* 1
stupide *adj.* 1
supporter *verbe* 1
survivant *subst.* 1
ta *dét.* 1
tabac *subst.* 1
Tadjikistan *subst.* 1
teinturerie *subst.* 1
télé *subst.* 1
téléviseur *subst.* 1
thé *subst.* 1
thermaux *adj.* 1
tic *subst.* 1
tituber *verbe* 1
toilette *subst.* 1
toit *subst.* 1
ton *dét.* 1
toréador *subst.* 1
touffe *subst.* 1
tour *subst.* 1
tourner *verbe* 1
tous *dét.* 1
tout *dét.* 1
toute *adj.* 1
train *subst.* 1
traîner *verbe* 1
travailler *verbe* 1
traverser *verbe* 1
très *adv.* 1
trésor *subst.* 1
triste *adj.* 1
troisièmement *adv.* 1
trottoir *subst.* 1
trous *subst.* 1
truc *subst.* 1
tunique *subst.* 1
valeur *subst.* 1
véritablement *adv.* 1
verre *subst.* 1
vers *prép.* 1
victime *subst.* 1
vieillir *verbe* 1
village *subst.* 1
vingt *dét.* 1
visage *subst.* 1
visite *subst.* 1
vivante *adj.* 1
voler *verbe* 1
vos *dét.* 1
votre *dét.* 1
voyage *subst.* 1
WC *subst.* 1
yeux *subst.* 1

Komentář

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

Nyní se podívejme na prvních 25 nejčastěji užívaných slov pocházejících z frekvenčního seznamu lexikologa Étienna Bruneta (vlevo), o kterém jsme se již zmínili, a porovnejme je s našimi výsledky (vpravo).

1.	**le** *dét.*	1050561	1.	**être** *verbe*	72
2.	**de** *prép.*	862100	2.	**de** *prép.*	66
3.	**un** *dét.*	419564	3.	**le** *dét.*	54
4.	**être** *verbe*	351960	4.	**avoir** *verbe*	53
5.	**et** *conj.*	362093	5.	**la** *dét.*	50
6.	**à** *prép.*	293083	6.	**il** *pron.*	34
7.	**il** *pron.*	270395	7.	**un** *dét.*	34
8.	**avoir** *verbe*	248488	8.	**à** *prép.*	33
9.	**ne** *adv.*	186755	9.	**ce** *pron.*	33
10.	**je** *pron.*	184186	10.	**elle** *pron.*	28
11.	**son** *dét.*	181161	11.	**je** *pron.*	28
12.	**que** *conj.*	176161	12.	**ne** *adv.*	28
13.	**se** *pron.*	168684	13.	**les** *dét.*	26
14.	**qui** *pron.*	148392	14.	**pas** *adv.*	24
15.	**ce** *dét.*	141389	15.	**qui** *pron.*	24
16.	**dans** *prép.*	139185	16.	**dans** *prép.*	22
17.	**en** *prép.*	143565	17.	**que** *conj.*	22
18.	**du** *dét.*	127384	18.	**des** *dét.*	21
19.	**elle** *pron.*	126397	19.	**le** *pron.*	21
20.	**au** *dét.*	123502	20.	**et** *conj.*	20
21.	**de** *dét.*	119106	21.	**une** *dét.*	19
22.	**ce** *pron.*	107074	22.	**au** *dét.*	16
23.	**le** *pron.*	105873	23.	**en** *prép.*	16
24.	**pour** *prép.*	104779	24.	**on** *pron.*	16
25.	**pas** *adv.*	103083	25.	**aimer** *verbe*	15

Kompletní seznam E. Bruneta čítá 1500 hesel a končí slovem s frekvenční hodnotou 412. Náš seznam má pouhých 673 hesel, přičemž většina z nich mají frekvenci 1. 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 76%. 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 673 různých slov (frekvence hesel) a celkem je v něm 1962 slov (frekvence slov), můžeme určit, že heslo se tedy průměrně opakuje 2,9 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.

Budeme-li se snažit postihnout index opakovaní v procentuální podobě, můžeme uvažovat takto: považujeme-li za 100% počet všech užitých slov (tedy 1962 slov), tvoří v našem případě počet různých slov 34,30% (673 hesel). Naproti tomu 65,70% označuje procento slov, která se v textu vyskytují více než jednou (tedy dvakrát a více). V rámci našeho seznamu hesel (tj. frekvenčního seznamu) se 69, 24% hesel (466 hesel) vyskytuje v textu pouze jednou. Zbylých 207 slov má četnost výskytu v textu větší než jedna.

3.2 Gramatická statistika

3.2.1 Zastoupení slovních druhů v textu

Komentář

Tabulka charakterizuje zastoupení jednotlivých slovních druhů jak v celém textu, tak v jeho dvou částech - v sekci Vypravěč a v sekci Scény.[18]

[18] Scénami nazýváme dialogy mezi postavami, které jsou oproti Vypravěči méně popisné a více spontánní.

Hodnotíme procentuální zastoupení každého slovního druhu vzhledem k počtu slov v té které sekci (2. a 5. sloupec), ale také vzhledem k počtu slov v celém textu (3. a 6. sloupec). Na závěr jsme zařadili údaj o celkovém počtu daného slovního druhu v celém textu (7. sloupec) a procentuální vyjádření této hodnoty (8. sloupec). Seřadíme-li slovní druhy podle jejich výskytu v celém textu, dostaneme toto pořadí:

1.	substantiva	21,25%
2.	verba	18,45%
3.	pronomina	17,43%
4.	determinanty	16,21%
5.	prepozice	10,04%
6.	adverbia	6,42%
7.	adjektiva	5,15%
8.	konjunkce	4,49%
9.	l´introducteur	0,31%
10.	mot-phrase	0,25%

Vidíme, že nejpočetnější skupinou jsou substantiva, která pokrývají o něco více něž 1/5 textu. S nimi je spojená četnost determinantů, které substantiva v případě členů určitých a neurčitých ve většině případů doprovázejí. Slovesné časy ve francouzštině bývají zase provázeny osobními zájmeny (*je, il, elle*,..), zastupujícími substantiva. Zatímco v češtině je osoba vyjádřena tvarem slovesa (*řekla jsem*), ve francouzštině (ale také například v angličtině a němčině) jsou zájmena součástí složených tvarů slovesných (*j´ai dit*), 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 Vypravěč a Scény mezi s sebou, všimneme si u vypravěče vyšší procentuální hodnoty u kategorie substantiv, adjektiv a determinantů, což v zásadě odpovídá povaze textu, připomínajícímu více

písemný projev, a snaze vypravěče situace podrobně popisovat a hodnotit.

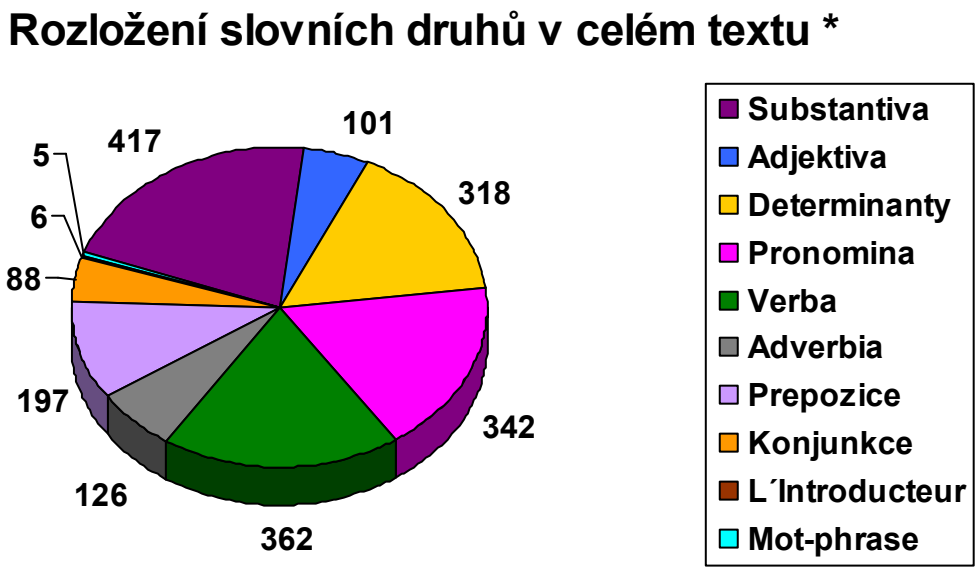
Ve scénách vidíme jistou záměnu části substantiv za pronomina. Větší počet zájmen v případě mluvených projevů či beletrie je přirozený, protože v nich je nejvíce užito přímého oslovení a dialogu. V rámci scén předpokládáme, že z důvodů spontánnosti či úspornosti vyjadřování jsou některá vlastní nebo 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 scénách) je také nepatrně vyšší četnost slov u kategorie *introducteur* a *mot-phrase*.

Početní a procentuální vyjádření slovních druhů zastoupených ve vybraném textu

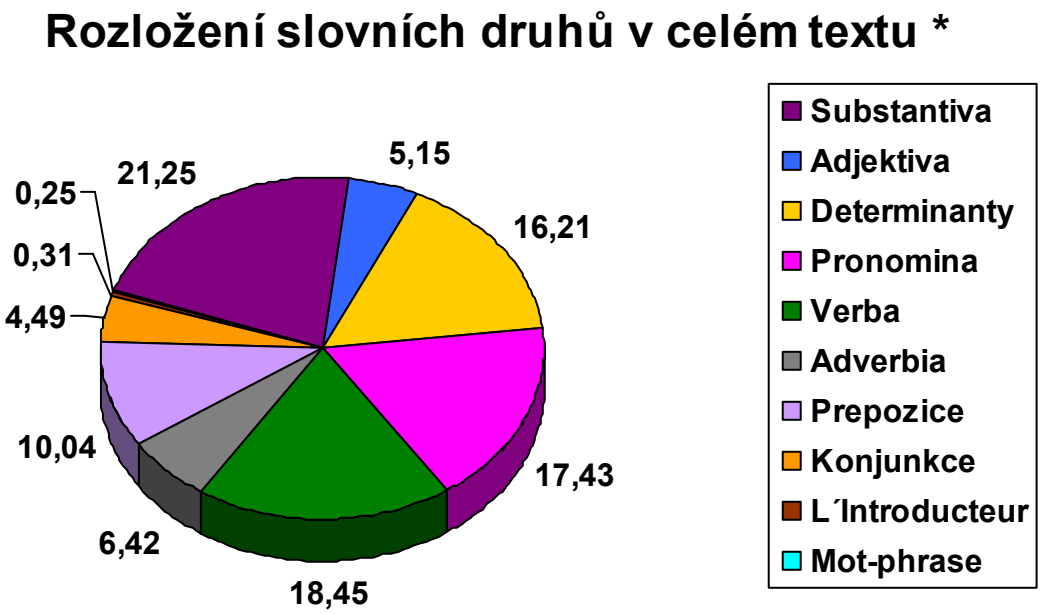
počet slov počítán bez scénáristických poznámek	počet slov v sekci „vypravěč“	% určitého slovního druhu v sekci „vypravěč“	% z celkového počtu slov za sekci „vypravěč“	počet slov v sekci „scény“	% určitého slovního druhu v sekci „scény“	% z celkového počtu slov za sekci „scény“	součet počtu slov scény+vypravěč	% z celkového počtu slov ze součtu v + s
1. substantiva	245	24,62%	12,49%	172	17,79%	8,77%	417	21,25%
2. adjektiva	54	5,43%	2,75%	47	4,86%	2,39%	101	5,15%
3. determinanty	178	17,89%	9,07%	140	14,48%	7,13%	318	16,21%
4. pronomina	115	11,56%	5,86%	227	23,47%	11,57%	342	17,43%
5. verba	173	17,39%	8,82%	189	19,54%	9,63%	362	18,45%
6. adverbia	70	7,03%	3,57%	56	5,79%	2,85%	126	6,42%
7. prepozice	121	12,16%	6,17%	76	7,86%	3,87%	197	10,04%
8. konjunkce	38	3,82%	1,96%	50	5,17%	2,55%	88	4,49%
9.1 introducteur	1	0,10%	0,05%	5	0,52%	0,25%	6	0,31%
10. mot-phrase	0	0%	0%	5	0,52%	0,25%	5	0,25%

celkový počet slov v textu (součet v+s): 1962
počet slov v sekci „vypravěč“: 995
počet slov v sekci „scény“: 967

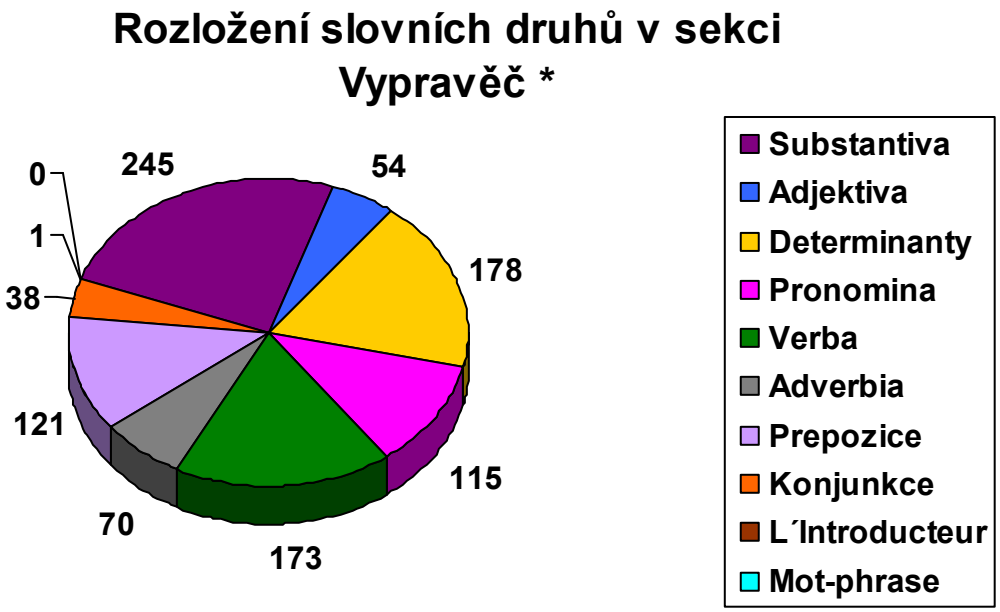
Výsledky jsme zapracovali do tabulky a pro větší názornost také do koláčových grafů:



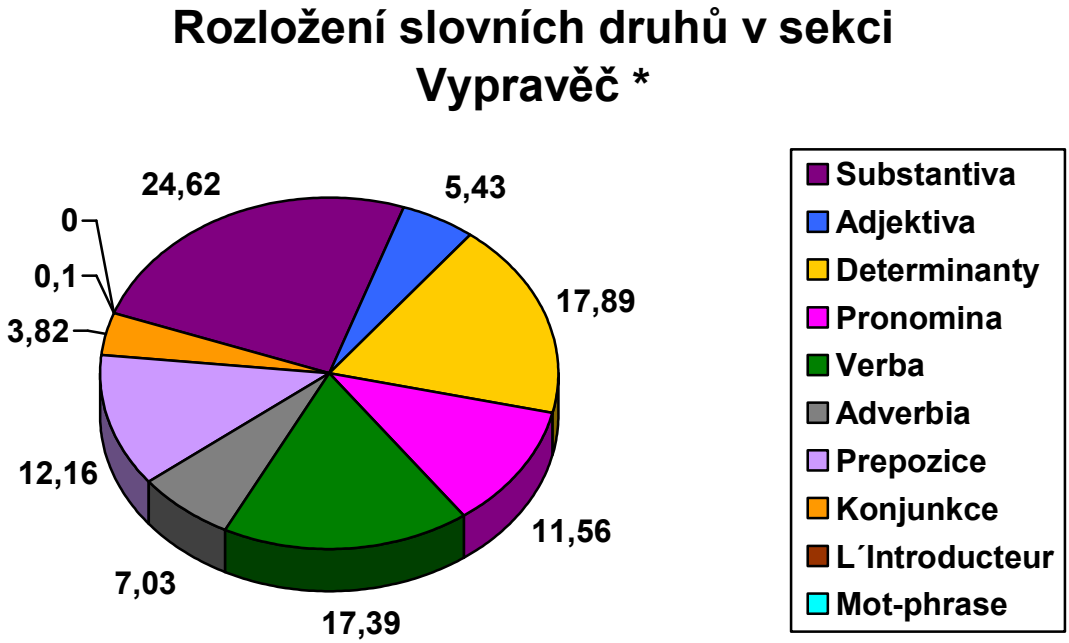
* Uvedené hodnoty představují užitý počet slov z daného slovního druhu.



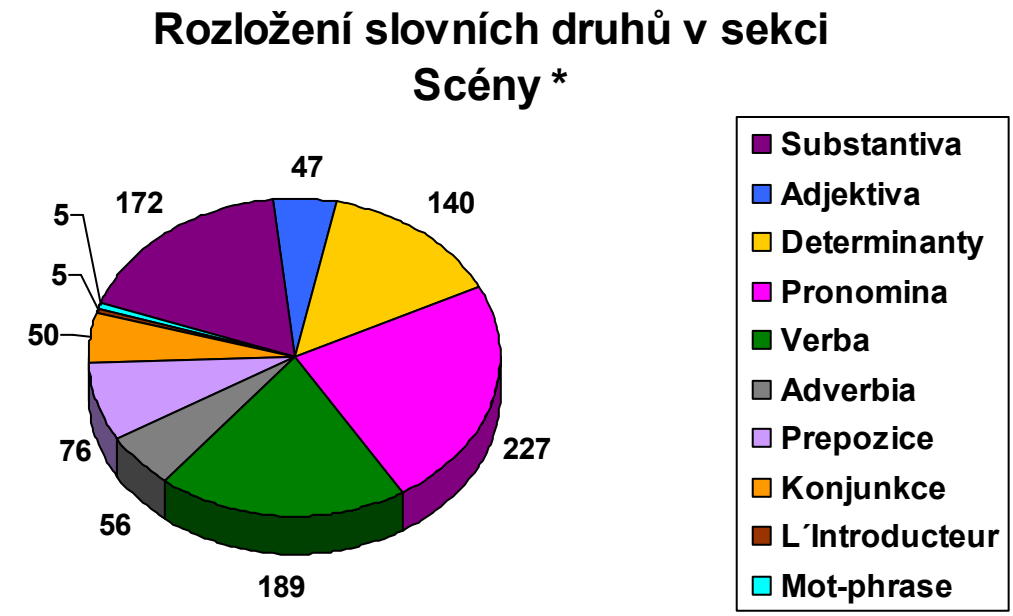
* Hodnoty jsou uvedeny v procentech.



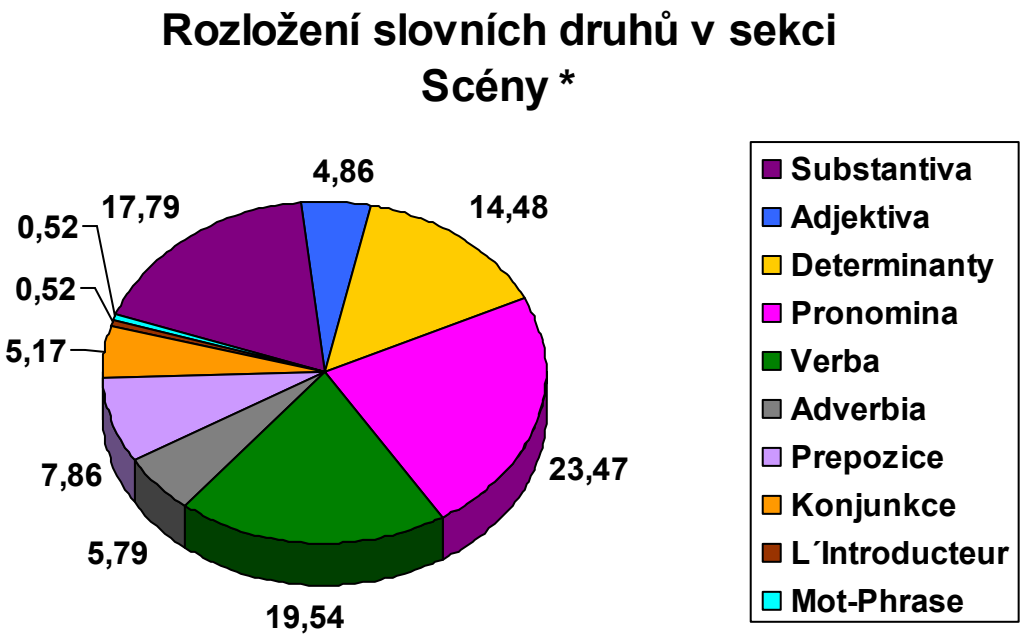
* Uvedené hodnoty představují užitý počet slov z daného slovního druhu.



* Hodnoty jsou uvedeny v procentech.



* Uvedené hodnoty představují užitý počet slov z daného slovního druhu.



* Hodnoty jsou uvedeny v procentech.

3.2.2 Kvantifikace kategorie rodu a čísla u jména

Při statistickém zpracovávání 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 seznamu obsahujícího 358 jmen jsme vypočítali následující hodnoty:

	femininum	maskulinum
singulár	119	157
plurál	31	51

Jinak řečeno v textu se vyskytuje celkem 119 feminin v singuláru a 31 feminin v plurálu. Dále 157 maskulin v singuláru a 51 maskulin v plurálu. Obecně řečeno nalezneme v textu 276 substantiv ve tvaru singuláru, zatímco ve tvaru plurálu jsou přítomna v počtu 82 slov. Maskulin je v textu 208 a feminin 150.

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

maskulina	58,10%	singulár	77,10%
feminina	41,90%	plurál	22,90%

Vidíme, že procento singuláru výrazně převyšuje plurál a také že rod mužský v singuláru má největší frekvenci. Vzhledem k tomu, že materiál k našim analýzám pochází z jednoho zdroje, nemůžeme porovnávat výskyt rodů vzhledem k jiným stylům. Marie Těšitelová ovšem 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).

3.2.3 Kvantifikace slovesných tvarů

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

	Vypravěč (počet jednotek: 150)		**Scény (počet jednotek: 172)**		**Celý text (celkový počet jednotek: 332)**	
	počet	%	počet	%	počet	%
INDICATIF						
présent	60	40%	102	59,30%	162	50,31%
passé composé	20	13,33%	9	5,23%	29	9,01%
imparfait	10	6,67%	20	11,63%	30	9,32%
passé récent	0	0	1	0,58%	1	0,31%
plus-que-parfait	1	0,66%	0	0	1	0,31%
passé simple	1	0,67%	0	0	1	0,31%
futur simple	0	0	1	0,58%	1	0,31%
futur proche	1	0,67%	5	2,91%	6	1,86%
SUBJONCTIF						
présent	1	0,67%	3	1,75%	4	1,24%
CONDITIONNEL						
présent	3	2%	5	2,91%	8	2,49%
passé	0	0	1	0,58%	1	0,31%
IMPÉRATIF						
présent	0	0	4	2,32%	4	1,24%
PARTICIPE						
présent	3	2%	1	0,58%	4	1,24%
INFINITIF						
présent	48	32%	20	11,63%	68	21,12%
GÉRONDIF						
présent	2	1,33%	0	0	2	0,62%

V tabulce můžeme vysledovat, že vůbec nejfrekventovanějším slovesným tvarem je indikativ prézentu, který tvoří celkem polovinu všech tvarů užitých v textu (50,31%). Následuje infinitiv s 21,12%. Na třetím a čtvrtém místě v četnosti výskytu nalezneme tvary minulé - imparfait (9,32%) a také minulý tvar složený - passé composé (9,01%). 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:

	Vypravěč		**Scény**		**Celý text (celkový počet jednotek: 248)**	
	počet	%	počet	%	počet	%
1. os. sg	0	0	30	12,10%	30	12,10%
2. os. sg	0	0	9	3,63%	9	3,63%
3. os. sg	90	36,30%	95	38,30%	185	74,60%
1. os. pl.	0	0	0	0	0	0
2. os. pl.	0	0	9	3,63%	9	3,63%
3. os. pl.	7	2,82%	8	3,22%	15	6,04%

Nejčastěji se v celém textu vyskytovala 3. osoba singuláru (74,60%) a také 1. osoba singuláru (12,10%). Ve scénách, kde jde o dialogy (tedy o oslovování, o sebeprosazení atd.), se 1. os. sg. vyskytuje z 12,09% a 2. os sg. z 3,63%, zato vypravěč první ani druhou osobu nepoužívá vůbec.

Závěr

Ruční vytváření frekvenčního seznamu je poměrně náročným úkolem. Nejenže je zapotřebí všechna zjištěná data opakovaně kontrolovat, aby nedocházelo ke zbytečným chybám, ale práce navíc vyžaduje systematičnost a promyšlené postupy, které mohou usnadnit realizaci následných dílčích analýz.

Zájmem teoretické práce bylo mimo jiné podat látku srozumitelným způsobem, přehledně ji rozčlenit do kapitol, které na sebe vzájemně navazují, a představit základní terminologii vztahující se k dané problematice. Některá témata jsme nastínili spíše obrysově. Důvodem je účel a rozsah naší práce, která si neklade za cíl stát se vyčerpávajícím a detailním průvodcem tohoto lingvistického odvětví. Její ambicí je inspirovat čtenáře k dalšímu bádání v oblasti matematické lingvistiky a probudit zájem o matematický pohled na jazyk, který dokáže přinést zajímavé a překvapivé výsledky.

Praktická část, analyzující francouzský text z perspektivy lexikální a gramatické statistiky, vznikala na základě zpracování vybraných částí filmového scénáře. Dialogy, na kterých je scénář postaven, dokážou do jisté míry simulovat současný mluvený projev. S ohledem na to, že převážná část dnešních lexikografických prací, zabývajících se slovní frekvenčností, je zaměřena na psaný projev, stává se tak pro nás filmový scénář zajímavým materiálem k analýze. Výsledky 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. Patrně ve spojitosti s úzkým korpusem se nám zatím nepodařilo vysledovat v četnosti slov výraznější příznaky mluvených projevů. Ty jsou spíše patrné ve frekvenčnosti gramatických jevů, na jejichž kvantifikaci jsme se zaměřili v druhé fázi našeho zkoumání.

Z obecného hlediska patří k nedostatkům naší práce nepřítomnost většího množství mluvených textů, které by dodaly našemu zkoumání objektivnější výsledky. Také rozsah analyzovaného korpusu je z důvodů komplikovaného ručního zpracovávání poměrně skromný. Máme ovšem za to, že by si naše práce zasloužila rozpracování v podobě rozšíření korpusu o další mluvené projevy.

Résumé

Notre thèse est consacrée à la linguistique quantitative. Son objectif est de présenter la problématique de la discipline de manière compréhensible et de prouver la vigueur des connaissances acquises dans la pratique. Pour cette raison, nous divisons la thèse en deux parties: la première théorique et la deuxième pratique.

La partie théorique met en évidence la terminologie fondamentale liée au sujet donné et elle regroupe 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 présentent le thème majeur de la deuxième partie de cette thèse.

Dans la partie pratique, nous avons decidé d'appliquer l'analyse statistique au scénario (quelques scènes choisies) du film bien connu *Le Fabuleux Destin d'Amélie Poulain* (2001 ; R : J.-P. Jeunet) dont les auteurs sont Guillaume Laurant et le réalisateur lui-même. Nous avons choisi ce texte en voulant le substituer au langage parlé et l'analyser du point de vue de fréquence ce qui n'est pas, même aujourd'hui, si courant. Nous avons créé la liste de mots les plus fréquents et nous avons aussi ajouté les analyses grammaticales. Même que le caractère compliqué de l'élaboration manuelle de ce scénario ne nous a permis que de générer le corpus plutôt petit, nous avons révélé assez de données intéressantes.

Notre thèse rend possible au lecteur de faire connaissance des principes de la statistique linguistique et d'évaluer les résultats issus de notre petite recherche.

Bibliografie

ČERMÁK, František; KŘEN, Michal; BLATNÁ, Renata, et al. 2004. *Frekvenční slovník češtiny*. 1. vyd. Praha : Nakladatelství Lidové noviny, 2004. 595 s.

ČERNÝ, Jiří. 1996. *Dějiny lingvistiky*. 1. vyd. Olomouc : Votobia, 1996. 517 s.

ČERNÝ, Jiří. 2005. *Malé dějiny lingvistiky*. 1. vyd. Praha : Portál, 2005. 239 s.

ČERNÝ, Jiří. 2008. *Úvod do studia jazyka*. 2. vyd. Olomouc : Rubico, 2008. 248 s.

GREENBERG, Joseph: *A Quantitative Approach to the Morphological Typology of Language*, International Journal of American Linguistics 26, 1960, 178-194.

GREVISSE, Maurice; GOOSSE, André. 2008. *Le bon usage*. 14. éd. Bruxelles : De Boeck, 2008. 1600 s.

GUIRAUD, Pierre. 1959. *Problèmes et méthodes de la statistique linguistique*. Dordrecht : D. Reidel Publishing Company, 1959. 145 s.

GUIRAUD, Pierre. 1963. *Etudes de linguistique appliquée*. 1963. La mécanique de l´analyse quantitative en linguistique, s. 35-46.

JEUNET, Jean-Pierre; LAURANT, Guillaume. 2003. *Le Fabuleux destin d´Amélie Poulain : Le Scénario.* Stuttgart : Ernst Klett Verlag, 2003. 80 s.

JELÍNEK, Jaroslav; BEČKA, Josef Václav; TĚŠITELOVÁ, Marie. 1961. *Frekvence slov, slovních druhů a tvarů v českém jazyce*. 1. vyd. Praha : Státní pedagogické nakl., 1961. 585 s.

LONSDALE, Deryle; LE BRAS, Yvon. 2009. *A Frequency Dictionary of French : Core Vocabulary for Learners* . USA, Canada : Routledge , 2009. 320 s.

TĚŠITELOVÁ, Marie. 1977. *Lingvistické příručky : Kvantitativní lingvistika*. 1. vyd. Praha : Státní pedagogické nakl., 1977. 177 s.

TĚŠITELOVÁ, Marie. 1980. *Využití statistických metod v gramatice*. 1. vyd. Praha : Československá akademie věd, 1980. 219 s.

TĚŠITELOVÁ, Marie. 1974. *Otázky lexikální statistiky*. 1. vyd. Praha : Československá akademie věd, 1974. 289 s.

ŠABRŠULA, Jan. 1983. *Základy francouzské lexikologie*. 1. vyd. Praha : SPN, 1983. 304 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 Nicolas Sarkozy* [online]. L´Internaute [citováno 11. 3. 2011]. Dostupné z WWW: <http://www.linternaute.com/actualite/politique/lexique-politique/nicolas-sarkozy.shtml>.

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>.

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>.

MARIDAT, Olivier. 2008. *Tableau du code morse* [[online]. Larousse [citováno 20. 3. 2011]. Odstupné z WWW: <http://www.larousse.fr/encyclopedie/media/Tableau_du_code_morse/11003951>.

PŘÍLOHA

Příloha 1

Vybrané části scénáře k filmu *Le Fabuleux Destin d'Amélie Poulain* - text analyzovaný v praktické části naší práce (Jeunet; Laurant, 2003)

Vypravěč

1

Le père d'Amélie, ancien médecin militaire, travaille aux Etablissements thermaux d'Enghien-les-Bains
Raphaël Poulain n'aime pas :
- Pisser à côté de quelqu'un
- Tituber en marchant dans le couloir d´un train
- Sortir de l'eau et sentir coller son maillot de bain.
Raphaël Poulain aime :
- Couper les virages dans les descentes sans les freins
- Aligner toutes ses chaussures et les cirer avec soin
- Vider sa boîte à outils, bien la nettoyer, et tout ranger enfin

La mère d'Amélie, Amandine Fouet, institutrice originaire Gueugnon, a toujours été d'une
nature instable et nerveuse.
Amandine Poulain n'aime pas :
- Avoir les doigts plissés par l'eau chaude du bain
- Etre, par quelqu'un qu'elle n'aime pas, effleurée de la main.
- Avoir les plis des draps imprimés sur la joue le matin
Amandine Poulain aime :
- Couper les virages dans les descentes sans les freins...
- Faire briller le parquet en marchant toute la journée avec des patins...
- Vider son sac à main, bien le nettoyer, et tout ranger enfin.

Amélie a six ans. Comme toutes les petites filles, elle aimerait que son père la serre dans ses bras de temps en temps. Mais il n'a de contact physique avec elle qu'au cours de l'examen médical mensuel.
La fillette, bouleversée par cette intimité exceptionnelle, ne peut empêcher son cœur de battre la chamade. Dès lors, son père la croit victime d'une anomalie cardiaque.

/.../

2

/.../

Suzanne, la patronne, boite un peu mais elle n'a jamais renversé un verre. Quand elle était jeune, elle était danseuse équestre à Médrano.
Elle aime: les sportifs qui pleurent de déception. Elle n'aime pas: voir dans son café un homme être humilié devant son enfant.

Au tabac, c'est Georgette, la malade imaginaire. Quand elle n'a pas de migraine, c'est le nerf sciatique qui coince. Celle là n'aime pas entendre : « le fruit de vos entrailles est béni. »

Gina, la collègue d'Amélie, est née en Normandie. Sa grand-mère était guérisseuse. Ce qu'elle aime, c'est faire craquer les os. On la voit qui sert un monaco à Hipolito, l'écrivain raté. Lui, ce qu'il aime par-dessus tout, c'est voir à la télé un toréador se faire encorner.

Le type qui les observe l'air mauvais, c'est Joseph, un amant jaloux éconduit par Gina. Il passe ses journées à l'espionner pour voir s'il a un remplaçant. La seule chose que celui-là aime, c'est crever les pustules des emballages en plastique.

Et enfin, voilà Philomène, l'hôtesse de l'air. C'est Amélie qui garde son chat « Rodrigue », quand elle part en voyage. Philomène aime le bruit que fait le bol d'eau du chat, quand on le pose sur le sol. Rodrigue quant à lui, aime être présent quand on raconte des histoires aux enfants.

On peut voir que, toujours disponible et avenante, tout le monde l'aime bien. Mais on sent qu'elle reste très secrète et ne se lie véritablement avec personne.

/.../

Amélie n'a pas d'homme dans sa vie. Elle a bien essayé une fois ou deux, mais le résultat n'a pas été à la hauteur de ses espérances.

En revanche, elle cultive un goût particulier pour les tout petits plaisirs : plonger la main au plus profond d'un sac de riz, faire des ricochets au bord de la Seine ou briser la croûte des crèmes brûlées avec le dos de la petite cuillère...

/.../

14

/.../

Un quart d'heure plus tard, au 22 Boulevard de Strasbourg, Amélie entre dans un magasin de farces et attrapes, déguisements et cotillons.

Au même instant, au 108 de la rue des Martys, un homme quitte son domicile.

A 10 heures 15 minutes, Amélie emprunte l'escalier roulant de la station Havre Caumartin. Elle ne peut s'empêcher de jeter un regard d'envie vers des enfants qui prennent l'escalator dans le mauvais sens.

Pendant ce temps, rue des Petits Champs, l'homme aux baskets rouges, au volant de sa fourgonnette Citroën G7, actionne la pédale de frein pour mieux contempler une femme blonde, au volant derrière lui.

Dix-huit minutes plus tard, Amélie arrive au photomaton de la gare de l'Est.

A la même seconde, l'homme aux baskets rouges exécute un créneau parfait, juste devant la gare, dérangeant deux moineaux qui faisaient leur toilette dans le caniveau. A cet instant, il est exactement 10 heures 37 minutes...

/.../

16

/.../

Nino est en retard. Pour Amélie, il n'y a que trois explications possibles :
Premièrement, il n'a pas encore trouvé la photo.
Deuxièmement, il l'a trouvée, mais n'a pas réussi à la reconstituer.
Troisièmement, il n'a pas eu le temps d'achever la reconstitution, parce que trois braqueurs avec des cagoules qui sortaient d'une banque l'ont pris en otage. Ils ont semé les flics, mais lui a provoqué un accident. Quand il est revenu à lui, il ne se souvenait de rien. Un routier ex-taulard l'a pris en stop et, le croyant en cavale, l'a planqué dans un container en partance pour Istanbul. Là, il est tombé sur des aventuriers afghans qui lui ont proposé de partir avec eux pour voler des têtes de missiles soviétiques. Mais leur camion a sauté sur une mine à la frontière du Tadjikistan, et il est le seul survivant. Et comme il a été recueilli dans un village de montagnards et qu'il est devenu militant Moudjahidin, Amélie ne voit vraiment pas pourquoi elle se mettrait dans un état pareil pour un type qui va passer le reste de ses jours à manger du bortsch avec un stupide cache-pot sur la tête.

/.../

18

/.../

Ce tic de contrariété, Amélie le découvrait pour la première fois. Pourtant, elle le connaissait bien, elle l'avait vu si souvent sur le visage de sa mère. Pour elle, ce fut comme un aperçu soudain de la manière dont elle pourrait mal vieillir. C'était exactement ce qu'il lui fallait à cet instant pour l'achever.

/.../

<u>Scény</u>

3

/.../

LE VIEUX
Bredoteau !

AMELIE
Pardon ?!?

LE VIEUX
C'est le nom que vous cherchez. Mais si c'est moi qui vous le dit, ça ne compte pas. *(baissant la voix)* : Je suis gâteux.

LA VIEILLE
Ne l'écoutez pas, il est gâteux. Vous avez vu dans quel état il a mis mes lauriers ?

Elle saisit une branche, toutes les feuilles sont pleines de petits trous ronds.

LA VIEILLE
Avant qu'on ait l'épicerie, lui, il était poinçonneur dans le métro… eh bien depuis trois mois, voilà qu'il se relève toutes les nuits avec sa machine à poinçonner pour aller faire des trous dans mes lauriers…

LE VIEUX
J'aurais préféré des lilas… la vie est mal faite. *(Se penchant vers Amélie)* Chacun son truc pour se calmer les nerfs…

AMELIE
(souriante) Moi, c'est les ricochets.

LA VIEILLE
(feuilletant les livres) Je vais vous trouver ça. Je note tout! Heureusement que je suis là. Quand je pense que mon fils a cinquante ans et que c'est moi qui suis obligée de faire sa comptabilité…

LE VIEUX
Faut dire qu'à quinze ans, tu lui mettais encore le dentifrice sur sa brosse à dents. Tout se paye.

LA VIEILLE
(le nez dans les registres) Alors, Camus… Camus… non ça c'était au second... Les Brossard... c'est l'escalier B... ça y est, j'y suis. Bredoteau, 5ème droite !
C'étaient des gens du Pas-de-Calais.

LE VIEUX
Bredoteau… c'est bien ce que je disais…

/.../

4

/.../

AMELIE
Salut papa. Je vois que tu t'es fait un nouveau copain.

LE PERE
(sérieux) Non, je l'avais depuis longtemps, mais je l'avais oublié. Comme ta mère ne le supportait pas, il était rangé dans la cabane à outils. Allez viens, on va aller les réconcilier..!

LE PERE
Pas mal, non ?

AMELIE
Hm hmm… Dis-moi papa, si tu retrouvais par hasard une chose de ton enfance, à laquelle tu tenais comme à un trésor… ça te rendrait heureux ou triste ou...nostalgique ? Ça te ferait quoi?

LE PERE
Si tu veux parler du nain, je ne l'avais pas quand j'étais petit… Ce sont les camarades du 26ème qui me l'ont offert pour mon départ à la retraite.

AMELIE
Mais non, papa, je te parlais de ces choses de l'enfance qu'on garde en secret comme si elles avaient une immense valeur.

LE PERE
Ah oui oui... *(ne l'écoutant pas)* Faudra que je le revernisse avant l'automne…

AMELIE
(déçue, filant vers la maison) Je vais faire du thé, tu en veux ?

/.../

LE CLIENT
Qu'est-ce qu'elle nous a fait de bon, aujourd'hui, Mme Suzanne ?

GINA
Les endives au gratin !

SUZANNE
(sortant de la tête de la cuisine) Vous allez voir, elles sont à tomber à genoux...

LE CLIENT
(à son copain) Ça veut dire qu'elles sont bonnes ?

LE COPAIN
Tout dépend où tu tombes à genoux…

LE PREMIER CLIENT
C'est vrai… si c'est devant la cuvette des WC…

LE COPAIN
Là, ça veut dire qu'elles sont pas bonnes !

Gina rit de bon coeur.
A sa table, Joseph sort un petit distaphone. Il appuie sur une touche et on réentend le rire de Gina qu'il vient d'enregistrer.

/.../

JOSEPH
(dans son dictaphone) 12h15. Rire de gorge évoquant l'orgasme. Motif: Plaire au mâle dominant.

Gina repose brutalement une carafe sur le bar.

GINA
(à Suzanne) Lui, s'il continue à m'emmerder, je réponds plus de rien !

SUZANNE
(à Joseph) C'est vrai quoi ! Pourquoi il s'obstine ? Il y a un bistrot tous les vingt mètres dans le quartier.

JOSEPH
(à un voisin de table) Raison scientifique. J'étudie le comportement sexuel des serveuses.

GINA
(*à Suzanne)* On dirait vraiment que je les attire moi, les cinglés.

SUZANNE
C'est pas que tu les attires, tu les collectionnes.

GEORGETTE
(changeant de sujet) Dans les endives au gratin, y'a de la béchamel, non ?

SUZANNE
Oui, et alors ?

GEORGETTE
Je ne digère pas la béchamel, c'est comme vous la viande de cheval.

SUZANNE
Moi, c'est pas une question de digestion, c'est une question de souvenir. Je préférerais cuisiner de la viande humaine.

/.../

5

/.../

A l'intérieur du café, Amélie est en train de boire un vin blanc sec lorsque la porte s'ouvre sur Bretodeau. Il pose la boîte sur le comptoir.

BRETODEAU
Un café s'il vous plaît.

Tandis que le patron le sert, Amélie, à l'autre bout du comptoir, se fait toute petite dans son coin.

BRETODEAU
Vous ne pouvez pas imaginer ce qui vient de m'arriver ! Si je croyais en Dieu, j'appellerais ça un miracle. Ou alors c'est mon ange gardien…

La patronne qui essuie les verres derrière le comptoir désigne une photo de l'équipe de France de football.

LA PATRONNE
Pour moi l'ange gardien, c'est Fabien.

BRETODEAU *(l'ignorant)*
Cette cabine ... je faisais que passer devant. C'est elle qui s'est mise à sonner toute seule... comme pour attirer mon attention. C'est comme si la cabine m'appellait, elle sonnait, elle sonnait, elle sonnait...

LE PATRON
Oui oui, c'est ça…

A ce moment, on entend une sonnette.

LE PATRON
Tenez, ben justement y'a le micro-ondes qui m'appelle.

BRETODEAU
(se tournant vers Amélie) C'est drôle, la vie… quand on est gamin, le temps n'en finit pas de se traîner, et puis du jour au lendemain, on a cinquante ans, comme ça pfff… Et l'enfance, tout ce qu'il en reste, ça tient dans une petite boîte rouillée… Vous avez des enfants, Mademoiselle ?

Amélie trop émue pour parler fait signe que non sans même le regarder.

BRETODEAU
Moi, j'ai une fille de votre âge. Ça fait des années qu'on ne s'est pas parlé. Il paraît qu'elle a eu un enfant, un garçon, l'année dernière. Ludovic il s'appelle.

Il sort avec soin le sucre de son emballage, le pose délicatement dans la tasse, et remue durant un long moment. Il boit son café, et repose la cuillère sur le comptoir.

BRETODEAU
Je crois qu'il est temps que je leur rende visite... avant de finir à mon tour dans une petite boîte…

6

/.../

Amélie s'arrête. Devant elle, un aveugle se tient au bord du trottoir, hésitant à traverser. Amélie le prend par le bras l'entraîne…

AMELIE
On descend dans le caniveau... on y voit le reflet du néon de la boucherie chevaline... Sur le toit en face il y a un ramoneur assis qui fume une cigarette. Voilà une vieille dame qui traverse. Elle a l'air maquillée comme une petite fille grimée en vieille dame... On arrive de l'autre côté. Sur le rebord du trottoir, il y a une petite touffe d'herbe qui n'a rien à faire là... Ce rire, c'est celui du cordonnier, il a plein de petits plis de malice au coin des yeux... là, vous sentez, c'est la poissonnerie...les crabes ont les pinces scotchées et il y a une anguille encore vivante qui ouvre la bouche... A la teinturerie, il y a promotion sur le nettoyage du daim et du pécari... Dans une poussette, on voit un bébé qui regarde un chien à la fenêtre d'un camion, qui lui-même regarde les poulets qui rôtissent... Ce parfum, c'est celui de la fleuriste qui porte une tunique avec des épaulettes parce que ce soir elle va répéter avec la fanfare... Nous voilà devant la boucherie , rognons 15 francs, filet mignon 120, bavette 88,50 et là, la musique, c'est la baraque de loterie, c'est drôle, le patron regarde « la roue de la fortune » sur son téléviseur portatif... et moi, maintenant, je tourne à droite...

Amélie disparaît dans la foule, laissant l'aveugle ébloui.

/.../

Příloha 2

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

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

le dét. 1050561
de prép. 862100
un dét. 419564
être verbe 351960
et conj. 362093
à prép. 293083
il pron. 270395
avoir verbe 248488
ne adv. 186755
je pron. 184186
son dét. 181161
que conj. 176161
se pron. 168684
qui pron. 148392
ce dét. 141389
dans prép. 139185
en prép. 143565
du dét. 127384
elle pron. 126397
au dét. 123502
de dét. 119106
ce pron. 107074
le pron. 105873
pour prép. 104779
pas adv. 103083
que pron. 99412
vous pron. 89623
par prép. 82277
sur prép. 80180
faire verbe 77608
plus adv. 75499
dire verbe 72134
me pron. 71086
on pron. 70246
mon dét. 70121
lui pron. 65988
nous pron. 62554
comme conj. 59902
mais conj. 57690
pouvoir verbe 55394
avec prép. 55081
tout adj. 47221
y pron. 46031
aller verbe 41702
voir verbe 39659
en pron. 38935
bien adv. 37171
où pron. 36089
sans prép. 35915
tu pron. 35774
ou conj. 34897
leur dét. 33950
homme subst. 33202
si adv. 32024
deux numér. 30211
mari subst. 30082
moi pron. 30053
vouloir verbe 29435
te pron. 28542
femme subst. 26148
venir verbe 26023
quand conj. 25592
grand adj. 25388
celui pron. 24270
si conj. 24024
notre dét. 23883
devoir verbe 22703
là adv. 22695
jour subst. 22232
prendre verbe 20489
même adv. 19994
votre dét. 19942
tout adv. 19915
rien pron. 19379
petit adj. 19008
encore adv. 19176
aussi adv. 18311
quelque dét. 17953
dont pron. 17797
tout pron. 17486
mer subst. 17166
trouver verbe 16833
donner verbe 16795
temps subst. 16785
ça pron. 16494
peu adv. 16251
même adj. 16081
falloir verbe 16078
sous prép. 15944
parler verbe 15814
alors adv. 15639
main subst. 15540
chose subst. 15524
ton dét. 15513
mettre verbe 15339
vie subst. 15241
savoir verbe 15102
yeux subst. 14981
passer verbe 14976
autre adj. 14688
après prép. 14606
regarder verbe 14604
toujours adv. 14336
puis conj. 14257
jamais adv. 14255
cela pron. 14253
aimer verbe 14138
non adv. 14039
heure subst. 13940
croire verbe 13881
cent numér. 13798
monde subst. 13737
donc conj. 13562
enfant subst. 13348
fois subst. 13191
seul adj. 13104
autre pron. 13063
entre prép. 13684
vers prép. 12781
chez prép. 12698
demander verbe 12597
jeune adj. 12593
jusque prép. 12465
très adv. 12432
moment subst. 12274
rester verbe 12155
répondre verbe 12063
tout dét. 12051
tête subst. 11999
père subst. 11854
fille subst. 11842
mille numér. 11758
premier adj. 11731
car conj. 11695
entendre verbe 12009
ni conj. 11640
bon adj. 11483
trois numér. 11372
coeur subst. 11312
ainsi adv. 11296
an subst. 11274
quatre numér. 10970
un pron. 10941
terre subst. 10786
contre prép. 10692
dieu subst. 10661
monsieur subst. 10489
voix subst. 10469
penser verbe 10358
quel adj. 10343
arriver verbe 10288
maison subst. 10287
devant prép. 9995
coup subst. 9991
beau adj. 9870
connaître verbe 9769
devenir verbe 9759
air subst. 9755
mot subst. 9752
nuit subst. 9694
sentir verbe 9585
eau subst. 9603
vieux adj. 9515
sembler verbe 9482
moins adv. 9472
tenir verbe 9312
ici adv. 9098
comprendre verbe 9037
oui adv. 9005
rendre verbe 9002
toi pron. 8997
vingt numér. 8920
depuis prép. 8907
attendre verbe 8851
sortir verbe 8768
ami subst. 8744
trop adv. 8686
porte subst. 8649
lequel pron. 8574
chaque dét. 8419
amour subst. 8283
pendant prép. 8202
déjà adv. 8170
pied subst. 8040
tant adv. 7960
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
larme 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
lueur 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
lisser 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
saison subst. 718
précieux adj. 715
précis adj. 714
consulter verbe 712
haïr 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
semer verbe 515
ferme adj. 514
manche subst. 494
rage subst. 492
gré subst. 491
guider verbe 450
piquer verbe 449
meilleur subst. 412