

Univerzita Palackého v Olomouci
Filozofická fakulta
Katedra psychologie

OSOBNOSTNÍ RYSY V PÍSEMNÉM
PROJEVU:
PSYCHOLINGVISTICKÁ STUDIE
SOUVISLOSTI MEZI TEXTOVÝMI ZNAKY A
OSOBNOSTNÍMI SKÓRY AUTORA NA ŠKÁLE
EXTRAVERZE

PERSONALITY TRAITS IN WRITTEN DISCOURSE:
A PSYCHOLINGUISTIC STUDY OF THE RELATIONSHIP
BETWEEN TEXTUAL FEATURES AND WRITER'S
PERSONALITY SCORE IN THE SCALE OF
EXTRAVERSION



Bakalářská diplomová práce

Autor: **PhDr. Bohumila Mia Chocholoušová Fagertun, M.A.**
Vedoucí práce: **PhDr. Daniel Dostál, Ph.D.**

Olomouc
2024

Poděkování:

Bez velkorysé pomoci, odborného vedení a veliké trpělivosti vedoucího této bakalářské práce – PhDr. **Daniela Dostála**, Ph.D. – by tato práce vůbec nemohla vzniknout. S jeho laskavou podporou se mi dostalo možnosti vybrat si téma, které mě zajímá, a zároveň získat přístup k datové matici, kterou v této práci využívám.

Zvlášť ocenit bych chtěla svého životního partnera, který mi byl v těžkých chvílích oporou a bez jehož podané ruky bych nebyla schopná tuto práci napsat.

Hotové dílo bych chtěla věnovat svým dětem, abych jim ukázala cestu dopředu a naučila je, že se těsně pod vrcholem nemají vzdávat.

Místopřísežně prohlašuji, že jsem bakalářskou diplomovou práci na téma: „Osobnostní rysy v písemném projevu: Psycholingvistická studie souvislosti mezi textovými znaky a osobnostními skóry autora na škále Extraverze“ vypracovala samostatně pod odborným dohledem vedoucího diplomové práce a uvedla jsem všechny použité podklady a literaturu.

V Olomouci dne 30.3.2024

Podpis

OBSAH

Číslo	Kapitola	Strana
	Obsah	3
	Úvod	5
	Teoretická část	7
1	Myšlení	8
1.1	Myšlení a osobnost	9
1.2	Myšlení a jazyk	10
2	Jazyk	11
2.1	Lingvistický determinismus	11
2.2	Lingvistický relativismus	13
3	Osobnost	16
3.1	Teorie osobnosti	16
3.1.1	Teorie osobnosti z pohledu psychoanalýzy	17
3.1.2	Teorie osobnosti z pohledu behaviorismu	18
3.1.3	Teorie osobnosti z pohledu humanistické psychologie	19
3.1.4	Strukturální teorie osobnosti	20
3.1.4.1.	Teorie osobnosti založená na somatotypu	20
3.1.4.2.	Teorie osobnosti založená na charakterových rysech	21
3.1.4.3.	Extraverze jako psychologický koncept	22
3.2	Metody výzkumu osobnosti	23
3.2.1	Pozorování	24
3.2.2	Projektivní techniky	24
3.2.3	Neurobiologický výzkum mozkových funkcí	25
3.2.4	Psychometrické inventáře	25
4	Vztah mezi jazykem a Osobností	27
4.1	Psychologická data na lingvistickém pozadí	27
4.1.1	Lexikální podstata psychometrických nástrojů	29
4.2	Lingvistická data na psychologickém pozadí	30
4.2.1	Psycholingvistické studie vztahu jazyka a osobnosti	30
4.2.2	Extraverze v psycholingvistickém výzkumu	32
4.2.3	Srovnatelné studie v českém prostředí	33
	Výzkumná část	36
5	Výzkumná otázka	37
6	Typ výzkumu a použité metody	39
6.1	Metoda zpracování přirozeného jazyka	39
6.2	Komparativní a korelační psycholingvistická analýza	39
6.2.1	Inventář BFI-2	40
6.3	Formulace hypotéz ke statistickému testování	42
7	Sběr dat a výzkumný soubor	44
7.1	Popis souboru	44
7.2	Psychologická data	44
7.3	Lingvistická data	45
7.3.1	Strojové zpracování jazyka UDPipe	46
7.3.2	Definice sledovaných lingvistických jevů	47

7.3.3	Postup extrakce sledovaných lingvistických jevů	49
7.4	Etické hledisko a ochrana soukromí	51
8	Práce s daty a její výsledky	53
8.1	Validita a reliabilita	53
8.2	Jazyk ve vztahu k pohlaví	55
8.3	Jazyk ve vztahu k Extraverzi	60
8.4	Míra účinku pohlaví a Extraverze na frekvenci výskytu sledovaných jazykových jevů	66
8.5	Srovnání našeho souboru s výsledky předchozích studií	68
8.6	Výsledky ověření platnosti statistických hypotéz	74
9	Diskuze	78
10	Závěr	84
11	Souhrn	86
Literatura	90
Přílohy	99
	Seznam příloh:	99

ÚVOD

Impulzem k sepsání této diplomové práce pro mě byla výzva spojená se sběrem dat pro výzkum autorství řízená z Katedry obecné lingvistiky FF UPOL. Tato data obsahovala možnost spárovat eseje studentů prvního ročníku s výsledky jejich osobnostních charakteristik zjištěných pomocí vyhodnocení sebeposuzovací škály BFI-2. Výzkum doc. Mgr. Dana Faltýnka, Ph.D. a Mgr. Vladimíra Matlacha, Ph.D. se měl zabývat predikcí autorství (Faltýnek, Matlach, Owsiankova, 2020), což je v pojetí aplikované lingvistiky spjato se strojovým učením a lineární analýzou jazykových prostředků.

Fascinující pro mě byl ale přesah k osobnostním charakteristikám autorů zkoumaných textů, tedy odrazový můstek k psychologickému výzkumu. Navrhovaná kvantitativní metoda určování autorství přinesla kolegům z oboru aplikované lingvistiky nečekaně vysokou přesnost (Faltýnek, Matlach, 2023). Nabízela se tedy otázka, zda by výzkum autorství mohl v mezioborovém výzkumu přispět i k psychologické profilaci samotných autorů, nebo alespoň některých jejich osobnostních rysů.

Na toto téma bylo na pomezí lingvistiky a psychologie uskutečněno velmi mnoho výzkumů (viz kapitola Teoretická část práce). Řada publikovaných studií vykazuje jednoznačné výsledky, které opakovaně potvrzují souvislost mezi osobnostními charakteristikami autorů a způsobem využití jejich jazykových prostředků. Abychom byli konkrétní, jedná se například o četnosti užití různých slovních druhů, délku a komplexnost vět, užití podmiňovacího způsobu nebo určitého druhu osobních zájmen. (c.f. Pennebaker, Mehl, Niederhoffer, 2003)

Teoretická část práce vychází ze samotného vztahu mezi myšlením, osobností a jazykem (Fodor, 1975). V klasickém lingvistickém pojetí jde o vztah vzájemně podmíněný a nerozlučný, protože jazyk předurčuje myšlení stejnou měrou, jakou myšlení předurčuje samotný jazyk (Whorf, 2012). Tyto teoretické koncepty jsou v současnosti umocněny technologickými možnostmi výzkumu velkého objemu dat za pomoci strojového učení, data mining a umělé inteligence (AI), nejde tedy o pouhou korelaci předem definovaných jazykových prostředků s psychologickými proměnnými, ale o možnost proniknout do způsobu, jakým se myšlení a osobnost autora v jazyce odráží (Baayen, 2017).

Tato práce je ve svých ambicích mnohem skromnější. Výzkumná část se opírá o znalosti z předchozích psychologicko-lingvistických výzkumů a snaží se kvantitativně popsat míru závislosti předem definovaných jazykových prostředků v psaném projevu na osobnostním skóre Extraverze daného autora, a to pomocí sebeposuzovací škály inventáře BFI-2.

Oporou mi byla zejména studie týmu doc. PhDr. Dalibora Kučery, PhD z Katedry psychologie Jihočeské univerzity (Kučera, 2022), která obsahuje komplexní kvantitativní metodologii pro posouzení míry závislosti škál osobnostních rysů - skóre inventáře BFI (na 44 a 10-ti položkách) daného autora s relativní četností užití konkrétních jazykových prostředků, neboli textových markerů, v jeho písemném projevu (Kučera, Havigerová, 2015).

Ve svém výzkumu se budu zaměřovat zejména na souvislost užití **předem definovaných gramatických jevů** popsanych v kapitole XX v textech studentských esejí a budu pozorovat jejich vztah s mírou **Extraverze** naměřenou v inventáři BFI-2.

Cílem této studie je tedy zjistit, zda:

Existuje souvislost mezi mírou Extraverze a způsobem užití jazykových prostředků v písemném projevu?

Kromě konkrétních jazykových rysů a jejich souvislostí s mírou Extraverze se budu dotýkat i dalších otázek:

- Do jaké míry souvisí jazykový projev autora s jeho pohlavím nebo dokonce s mírou jeho Extraverze?
- Potvrdí se v našem souboru výsledky předchozích studií?
- Jeví se výzkum osobnosti skrze rozbor jazykového projevu jako teoreticky možný a pokud ano, tak za jakých podmínek?
- Lze pomocí rozboru písemného textu predikovat osobnostní rys jejího autora?

Právě tyto otázky jsou bez detailní analýzy konkrétních dat nezodpověditelné, protože jsou natolik obecné, že je bez praktického výzkumu nemáme čím potvrdit nebo vyvrátit. Zároveň jsou důležité jako dílčí krok ve výzkumu souvislosti mezi osobností autora a jeho způsobem myšlení, které můžeme sledovat v jeho písemném projevu. Tato práce o vztahu textových markerů s mírou Extraverze by tedy měla sloužit jako jeden další nepatrný prvek v nekonečné mozaice komplexních vztahů definujících souvislosti mezi jazykem, myšlením a osobností autora.

TEORETICKÁ ČÁST

1 MYŠLENÍ

Myšlení můžeme definovat jako komplexní kognitivní proces, který zahrnuje mentální aktivity spojené s manipulací, organizací a interpretací informací. Jedná se o vyšší kognitivní funkci, kterou jednotlivci využívají při zpracování informací, řešení problémů, rozhodování a generování myšlenek. Právě díky myšlení je jedinec schopen uspořádat vztahy mezi předměty a ději ve svém okolí a vytvářet si *`mentální reprezentace`*, které vedou k pochopení těchto vztahů v jeho bezprostředním okolí (Plháková, 2004).

Myšlenkový proces je souhrn komplexních operací, které zahrnují smyslové vnímání, kognitivní percepci, uchovávání a třídění informací, paměť, způsob zpracovávání myšlenkových konceptů, návaznost mentálních obsahů, vyhodnocování mentálních vzorců, kreativitu, představivost, proces rozhodování, schopnost abstrakce, logiku, kritické myšlení spojené s metakognicí a schopnost být si vědom sám sebe jako myslícího subjektu.

Základním principem myšlení je manipulace s psychickými obsahy za účelem řešení praktických a teoretických problémů (Burton, Radford, 2022, s. 68).

Konkrétní myšlenkové postupy můžeme z hlediska komplexnosti dělit na *`algoritmy`* a *`heuristiky`*, kde algoritmy, podobně jako v matematice, představují exaktní aplikaci určitých kroků, které vždy vedou k určitému výsledku (např. výpočet rovnice o dvou neznámých nebo odměření ingrediencí do kuchařského receptu), tyto operace jsou do určité míry neinovativní, rigidní postupy, které však vyžadují značné mentální úsilí. Heuristiky oproti tomu představují mentálně úspornější řešení, tedy jakousi kognitivní zkratku k řešení, a to i za cenu nesprávného nebo neúplného výsledku. V popředí zde vstupuje inovace, intuice a zkušenosti, které nám umožňují problém vyřešit nejjednodušším možným způsobem s velkou úsporou mentální energie (Plháková, 2004). Základním paradigmatem myšlení je potom schopnost *`abstrakce`*, tedy mentální proces, který nám umožní identifikovat základní vlastnosti objektu a vyčlenit z něho charakteristické rysy tak, abychom tento objekt mohli kriticky zkoumat a zařadit ho do systému jiných objektů, tedy *`kategorizovat`* jej v rámci systému jiných již existujících mentálních konstruktů. Tento proces stojí u samých kořenů sémiologie a sémantiky, tedy vědy o přiřazování významů a vztahů mezi nimi (Chandler, 2022).

V sémiotice je myšlení schopnost přiřazování významu neboli *`inference`*, kdy určitý objekt vnější reality přiřadíme k určitému významu, *`referenci`*, a ta nám umožní daný objekt zařadit do systému již existujících mentálních kategorií, podle kterých můžeme daný objekt popsat nebo pochopit. Po spatření objektu si tedy pomocí *`abstrakce`* vytvořím představu o jeho charakteristických rysech (např. živý tvor, dospělý samec, zvíře), potom mu na základě *`inference`*, tedy vnímání těchto charakteristických rysů přidělím určitý význam (např. pes) a díky přiřazení této *`reference`* je nám umožněno zařadit tento objekt do kategorií stávajících kognitivních vztahů a myšlenkových vzorců, které tvoří základ naší osobní *`mentální reprezentace`* tohoto světa (Eco, 1979).

K tvorbě, třídění a zpřesňování *`mentálních reprezentací`* používáme dva protisměrné myšlenkové procesy, a sice schopnost *`dedukce`* a *`indukce`*. Dedukci odpovídá analytické myšlení, kdy od obecných charakteristik odvozujeme konkrétní rysy nebo znaky a v myšlenkovém postupu nejprve zohledňujeme celek, od něhož se dostáváme k jednotlivým rysům, znakům nebo částem. Opakem dedukce je induktivní myšlenkový proces, který je založen na analogii a syntéze, a umožňuje díky detailní charakteristice jednotlivých částí, prvků a rysů odvozovat obecné vlastnosti týkající se daného

významového celku. Oba tyto procesy by měly probíhat simultánně a vzájemně symbioticky dotvářet mentální reprezentace o konkrétním světě (Plháková, 2004).

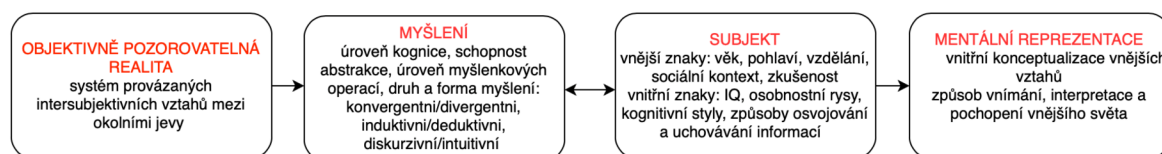
1.1 Myšlení a osobnost

Myšlení samotné, stejně jako citát „*Cogito ergo sum*“¹ stojí v samém základu západní filozofie. Ještě před tím než tuto myšlenku, která definovala racionalismus a moderní vědu, Descartes vyslovil, objevovala se u Aristotela i Platóna jako náznak jistoty o neoddělitelnosti myšlení a subjektu, který myslí (Sobotka, 2008). Pokud přijmeme toto základní filozofické paradigma jako neoddělitelnou spojitost mezi osobou, která myslí, a produktem jejího kognitivního úsilí – tedy myšlením, otevíráme zde logicky prostor pro přímou spojitost mezi osobnostními rysy myslícího subjektu, které se přirozeně promítají v produktu jeho myšlení.

Myšlení je neoddělitelně spjato s inteligencí, která do značné míry determinuje jeho kvalitu. V kvalitě myšlení se tedy odráží interindividuální intrapsychické rozdíly, patrné jsou především ve schopnosti myslícího subjektu správně uchopovat vztahy a v míře abstrakce, na které je subjekt o těchto vztazích schopen uvažovat (Nakonečný, 1997a). Je tedy patrné, že kvalita myšlení přímo a úzce souvisí s mentální kapacitou a osobnostními rysy myslícího subjektu.

Pomocí myšlenkových procesů si subjekt vytváří vlastní představu o vnější realitě, která může objektivně pozorovatelným okolnostem odpovídat do větší nebo menší míry v závislosti na kvalitě jeho myšlenkových procesů (Eysenck, Keane, 2008). Pro názornost je zařazeno následující schéma (Obrázek 1).

Obrázek 1: Schéma vzájemné závislosti myšlenkového procesu a subjektu myšlení



Zdroj: Vlastní nákres

Jak je tedy patrné, naše *mentální reprezentace* vnější reality nemusí s realitou vůbec souviset a je úzce závislá na našich osobních intrapsychických predispozicích a mentální kapacitě ve všech směrech a rovinách. Kognice začíná vnímáním, a pokud existuje nějaké percepční omezení na smyslové úrovni (např. omezení zraku nebo sluchu), těžko si bude subjekt vytvářet přesnou představu o vnějším světě. Další taková omezení jsou méně zřejmá, avšak stejnou měrou determinující pro výslednou přesnost myšlení (Eysenck, Keane, 2008).

Mezi intrapsychické faktory ovlivňující kvalitu myšlení patří vnímání a kognitivní funkce, schopnost abstrahovat, třídit a kategorizovat informace, paměť, způsob aplikace myšlenkových postupů a tvorba kognitivních vzorců a stop (Cakirpaloglu, 2012). Stejnou měrou je ale myšlení, jakožto i vlastní percepce, determinována osobnostní charakteristikou myslícího subjektu a jeho jak vnějšími (věk, pohlaví, sociální statut,

¹ Z lat. „*myslím, tedy jsem*“

vzdělání, kulturní kontext atd.), tak i vnitřními (IQ, emoční nastavení, osobnostní rysy, zkušenosti atd.) predispozicemi (Říčan, 2010).

Samotný výsledek mentálního procesu myšlení je tedy velkou měrou závislý na osobnosti myslícího subjektu. Osobnost myslícího subjektu se ve výsledném myšlenkovém produktu odráží stejně jako pozorovaná realita, o které myslící subjekt zrovna přemýšlí.

„Personality is the dynamic organization within the individual of those psycho-physical systems that determine their characteristic behaviour and thought.“ (Allport, 1955, str. 64).

1.2 Myšlení a jazyk

Myšlení ve své metafyzické podstatě zůstává abstraktním konceptem, který můžeme sice subjektivně cítit, nikoliv ovšem objektivně popsat nebo pozorovat. Nemá totiž jasně danou fyzickou formu ani podobu. Na metafyzické úrovni ho můžeme filozoficky popsat nebo prakticky zažít vlastní zkušeností, nemůžeme ho ale konkrétně zkoumat. K praktickému výzkumu myšlení je zapotřebí, aby myšlení získalo jasně definovanou formu. Za takto jasně definovanou, ohraničenou formu potom můžeme považovat jasný produkt myšlení (vypočítaný příklad, namalovaný obraz, převyprávěný obsah díla). Tento mentální produkt je pak nepopíratelným důkazem o myšlenkovém pochodu, který se uvnitř myslícího subjektu odehrál (Flower, Hayes, 1981).

Komplexnost myšlení jako kognitivního procesu je nejpatrnější ve verbálním projevu myslícího subjektu. Z behaviorálního hlediska je jazyk výlučně lidským projevem, i když zvířata nepochybně dokáží komunikovat, zůstává platným předpokladem, že myšlení zrcadlí se v lidském jazykovém projevu je nejvyšším důkazem tohoto mentálního úsilí. Vývoj myšlení také úzce souvisí s ontogenezí a stupněm vývoje myslícího subjektu (Piaget, 2005).

Podle Vygotského (2012) se vyšší úrovně myšlení vážou na jazyk. Určité druhy myšlení probíhají na verbálním základu a jsou úzce spjaty s `vnitřní` nebo `egocentrickou řečí`,² která je plně internalizována a probíhá bez použití mluvidel, jedná se tedy spíše o projev verbálního myšlení ve formě práce s pojmy a jejich významy. Vnitřní řeč je projevem vyššího stádia vývoje kognitivních operací spojených se schopností interní manipulace s `mentální reprezentací` vnějších objektů, a to nejprve konkrétních (vývojové stádium `konkrétních operací` - 7-11 let věku) a potom abstraktních konceptů (vývojové stádium formálních operací – od 11 let věku).

Na úrovni formálních operací je tedy člověk schopen `abstrakce`, `dedukce` i `indukce`, vytváří si pojmové kategorie a myšlenkové mapy, do kterých je schopen třdit a kategorizovat informace i abstraktní pojmy, je schopen odpoutat se od bezprostředně vnímané reality, zakládat komplexní myšlenkové stopy vedoucí k zpřesňování `mentálních reprezentací`, popisuje své mentální obsahy pomocí vnitřní i vnější řeči a je schopen metakognice, tedy uvědomění si vlastního vnímání a myšlení, stejně jako zaujmout a pochopit jiné stanovisko než svoje vlastní (Plháková, 2004).

² `la langue egocentrique` (Piaget, 2005, str. 84)

2 JAZYK

Obor, který zkoumá rozhraní mezi myšlením a jazykem, se nazývá psycholingvistika. V této kapitole bychom si měli ujasnit, jak úzce se myšlení v jazyce odráží a naopak, jak úzce může jazyk myšlení omezovat nebo zkreslovat. Odchýlíme se tedy od klasicky psychologických disciplín zkoumání mysli a budeme se soustředit na základní lingvistické paradigma, a sice do jaké míry je myšlení na jazyce závislé, a naopak do jaké míry jazyk myšlení ovlivňuje, umožňuje a limituje. Dotkneme se otázky, do jaké míry je vůbec myšlení bez určité koncepční jazykové reprezentace možné (Sternberg, 2009). A dozvíme se o počátcích výzkumu této souvislosti.

Myšlenka, že jazyk ovlivňuje nebo formuje způsob, jakým jedinec vnímá a chápe svět, se objevila na počátku 20. století. Franz Boas, zakladatel americké kulturní antropologie, byl jedním z prvních, kdo se zabýval faktem, že různé jazyky mohou poskytovat odlišné perspektivy na svět a že kulturní a jazykový kontext může formovat myšlení jednotlivce natolik, že si vytváří diametrálně odlišné *'mentální reprezentace'* o vnějším světě (Boas, 1911).

Výzkum, kterým se Boas (1911) zabýval, byl spojen s analýzou jazyka a kultury indiánských kmenů na pacifickém severozápadě. Zde poprvé jasně formuloval myšlenku, že různé kultury mají odlišné jazyky a odlišný způsob myšlení, který je úzce spjat s výrazovými prostředky v samotném jazyce obsaženými.

V období osidlování amerického středozápadu bylo pro antropology atraktivní zkoumat kulturu a jazyk původních obyvatelů. Během své interakce s natolik odlišným způsobem života, hodnotovým systémem a pochopením řádu světa byli často konfrontováni s nutností naučit se jazyk daného kmene, aby s původními obyvateli mohli komunikovat. Díky znalosti jazyka byli potom vědci často nuceni konfrontovat se se situací, že určité myšlenky byly nesdělitelné, protože pro ně v druhém jazyce neexistovaly výrazové prostředky, nebylo tedy možné plnohodnotně komunikovat mentální obsahy, protože v daném jazyce pro ně neexistoval způsob, jak je popsat.

Tento fakt odstartoval řadu empirických výzkumů (Gleitman, Papafragou, 2005) zaměřených na způsob, jakým jazyk formuje naše myšlení, jaký vliv má jazyk na náš pohled na svět, a zda existují univerzální myšlenkové vzory nebo zda jsou formovány jazykem, kterým mluvíme.

Edward Sapir (1921) klasifikoval domorodé atabaské jazyky a definoval myšlenku, že jazyk úzce souvisí s myšlením, osobností a kulturou mluvčího. Spolu s Benjaminem Lee Whorfem (1952) potom vyslovili hypotézu, ve které popsali, že jazyk podstatným způsobem formuje naše myšlení. Tato "**Sapir-Whorfova hypotéza**" se stala základním stavebním prvkem řady antropologických, psychologických a lingvistických výzkumů. Sapir-Whorfova hypotéza je spojována se dvěma variantami: slabou, umírněnou, oynačovanou jako *'lingvistický relativismus'*, a silnou, radikální, nazývanou *'lingvistický determinismus'*.

2.1 Lingvistický determinismus

Lingvistický determinismus je spojen s radikální verzí Sapir-Whorfovy hypotézy a představuje myšlenku, že jazyk přímo ovlivňuje nebo určuje způsob myšlení a vnímání

reality. Struktura jazyka, kterým mluvíme, tedy přímo determinuje nebo dokonce omezuje naše myšlení. Podle silné verze Sapir-Whorfovy hypotézy lidé, kteří mluví různými jazyky, mají odlišné způsoby vnímání a chápání světa a vytvářejí si zcela odlišné *`mentální reprezentace`* o vnějším světě, a ty jsou přímo závislé na struktuře jazyka a dostupných výrazových prostředcích.

Benjamin Lee Whorf, spoluautor této hypotézy, byl americký lingvista, který se zabýval studiem jazyka Hopi indiánů a publikoval myšlenky o tom, jak struktura jejich jazyka ovlivňuje jejich vnímání času, prostoru a událostí. Whorf (1956) pro svůj deterministický předpoklad, že myšlení je závislé na jazyku, použil přesvědčivý příklad, a sice že Hopiové v Arizoně nekonjugují slovesa do minulého, přítomného a budoucího času, jak to dělají mluvčí angličtiny, a že jejich slova pro časové jednotky jako např. "den" nebo "hodina" jsou spíše slovesa než podstatná jména. (Whorf, 1956, str 124-125). Závěrem tohoto zjištění bylo, že Hopiové nevnímají čas jako fyzický objekt, který lze spočítat v minutách a hodinách, jak to dělají mluvčí angličtiny, ale namísto toho Whorf (2012) tvrdil, že Hopiové vnímají čas jako neformální proces úzce spojený s dějem, který v daném období právě probíhá.

Toto zjištění, jak specifické lingvistické výrazy ovlivňují vnímání času, prostoru a konceptů ve společnosti Hopi indiánů, bylo následně přijato jako radikální formulace Sapir-Whorfovy hypotézy neboli *`lingvistický determinismus`*. V tomto pojetí jsme schopni vnímat realitu a uvažovat o vztazích a dějích kolem sebe pouze do té míry, do které nám vlastní jazyk tyto souvislosti pozorovat a popisovat umožňuje (Whorf, 1956). Pokud tedy v mém jazyce nemám výrazové prostředky popisující časovou posloupnost nebo kauzalitu, nedokážu přemýšlet v těchto intencích.

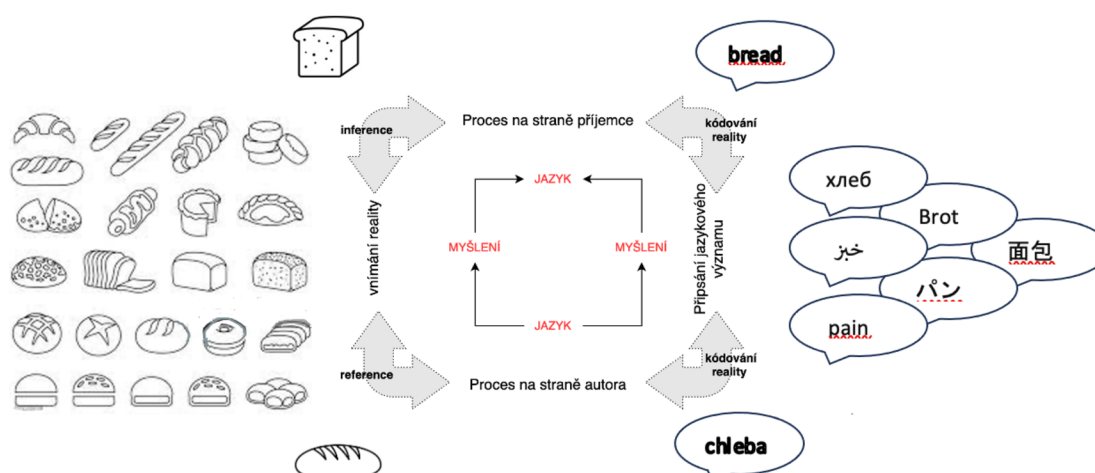
V radikálním pojetí Sapir-Whorfovy hypotézy je tedy *`lingvistický determinismus`* vzájemně propojený a zcela závislý vztah mezi myšlením a jazykem:

„Language shapes the way we think, and determines what we can think about“³
(Whorf, 2012, str. 94)

Tento předpoklad, že jazyk a mysl jsou jediným propojeným celkem, byl od té doby opakovaně vyvrácen (Kay, Kempton, 1984). Než se však dostaneme k modifikované verzi a umírněnější verzi Sapir-Whorfovy hypotézy, uvedu jeden platný předpoklad, na kterém můžeme názorně pochopit, jak jazyk naše vnímání reality může ovlivňovat. Na Obrázku č.2 vidíme, že percepce vnější reality může být determinována jak kulturními, tak lingvistickými předpoklady. Tyto procesy probíhají v mysli zcela nevědomě, a proto mohou významnou měrou ovlivnit způsob, kterým si tvoříme *`mentální reprezentace`* vnější reality nebo vybíráme mezi vztahy a objekty podle jazykového kódu, kterým je jejich význam přenesen.

³ „Jazyk utváří myšlení a předurčuje, o čem jsme schopni přemýšlet“ (zdroj: vlastní překlad, Whorf, 2012, str. 94)

Obrázek 2: Vztah jazyka a myšlení z pohledu lingvistického determinismu



Zdroj: Vlastní nákres

V levé části obrázku vidíme reálné objekty – druhy pečiva, které mohou být označeny termínem `chleba`. Jakmile ale použijeme konkrétní lingvistický termín určitého jazyka, budeme přirozeně referovat ke specifickému tvaru, chuti a formě, pro kterou je tento konkrétní lingvistický termín určen. Český `chleba` se anglickému `bread` svou formou, chutí a tvarem příliš nepodobá, i když formálně a logicky patří do stejné konceptuální kategorie `pečiva`. Díky procesu *inference* totiž ke slovu `chleba` automaticky přiřadíme jinou formu pečiva, která je pro český jazyk a českou kulturu příznačná, než když tento pojem přeložíme anglickým ekvivalentem `bread`, který v kulturním prostředí rodilých mluvčích anglického jazyka popisuje diametrálně odlišný druh pečiva.

Mentální reprezentace těchto dvou forem stejného objektu `bread` a `chleba` je tedy závislá na užití konkrétního lingvistického termínu, který v sobě obsahuje kulturní předpoklady i omezení vyplývající z prostředí a myšlení daného jazyka. Silná forma Sapir-Whorfovy hypotézy tedy předpokládá, že kategorie, pro které v daném jazyce nejsou dostupné výrazové prostředky, nemůžeme v realitě ani pochopit. Takto deterministický přístup se ve své radikální formě už nepoužívá, ve studiu vztahu jazyka a myšlení však má stále své opodstatnění.

2.2 Lingvistický relativismus

V pozdějších výzkumech byl deterministický pohled na podmíněnost vztahu mezi myšlením a jazykem výrazně modifikován. Lingvistický relativismus je myšlenka, že jazyk ovlivňuje nebo formuje způsob, jakým jedinec vnímá a chápe svět, ale nedeterminuje způsob myšlení. Tento koncept zdůrazňuje, že různé jazyky mohou poskytovat odlišné perspektivy na svět a že kulturní a jazykový kontext může myšlení jednotlivce formovat, nikoliv však omezovat. Franz Boas (1911), zakladatel americké kulturní antropologie, byl jedním z prvních, kdo se lingvistickým relativismem zabývali.

Relativismus vychází ze slabé verze Sapir-Whorfovy hypotézy a zaujímá umírněný postoj k míře závislosti jazyka a myšlení. Tento pohled je často spojován s dílem Edwarda Sapira (1921), který stál u samotných počátků formulace hypotézy a v relaci k Whorfovi (1952) byl jeho předchůdcem a učitelem. Jeho koncepce zdůrazňuje, že různé jazyky mají různé prostředky a možnosti pro formování *mentálních reprezentací* o vnějším světě.

Z pohledu relativismu mohou mít lidé mluvící různými jazyky odlišné perspektivy a zaměření myšlení či vnímání, ale nejsou plně uvězněni v lingvistických modelech svého

jazyka (Morency, 2017). V antropologii, lingvistice a kognitivních výzkumech byla ve prospěch této teorie provedena řada faktických šetření zaměřených na prokázání míry závislosti a podstaty vztahu mezi jazykem a způsobem chápání vnější reality.

Pro existenci této závislosti hovoří řada faktických důkazů, v jazycích s odlišnými sémantickými kategoriemi (např. rozdílná klasifikace barev a odstínů), odlišnou gramatickou strukturou (např. přítomnost nebo nepřítomnost určitých kategorií – slovesného vidu nebo času) strukturují mluvčí okolní realitu odlišným způsobem podle kategorií, které jsou pro něho v daném jazyce dostupné (Heider, 1972).

Nejtypičtějším důkazem lingvistického relativismu je rozdíl v kódování barev. V popsaných jazykových systémech se kategorie pro popsání barev mohou diametrálně rozlišovat. Jedním z klasických příkladů jazyka s omezeným rozlišením barev je jazyk Dani, kterým mluví domorodci v pohoří Papua Nová Guinea. Jazyk Dani tradičně obsahuje pouze dva základní barevné termíny: mili pro chladné nebo tmavé odstíny (zahrnující odstíny černé, modré a zelené) a mola pro teplé nebo světlé odstíny (zahrnující odstíny bílé, červené a žluté). Domorodý kmen jazyka Warlpiri, kterým mluví aboridžinci v australském pouštním regionu, rozlišuje naopak 12 základních barevných odstínů a kombinací (Heider, 1972).

Tyto lingvistické rozdíly neznamenaí nutně, že mluvčí těchto jazyků nejsou schopni vnímat různé barvy; spíše jejich jazyk může kategorizovat barvy odlišně než jazyky s větším množstvím barevných termínů (Heider, 1972; Wierzbicka, 2013). Tento závěr potvrzují také výzkumné týmy Marian a Neiser (2000) a Winawer, Witthoft, Frank, Wu, Wade a Boroditsky (2007) zkoumající rozdíly mezi angličtinou a ruštinou, kde byla prokázána nižší kognitivní pohotovost při identifikaci barevných odstínů, které nejsou v jazykovém systému daného jazyka dostupné (Winawer, Witthoft, Frank, Wu, Wade a Boroditsky, 2007, s. 783), a zároveň zlepšení paměťové kapacity spojené s bilingvismem a vyvoláním vzpomínek z paměti mluvčího v daném jazyce (Marian, Neiser, 2000, s. 365).

Dalším příkladem lingvistického relativismu je způsob popisu směru v daném jazyce. Mluvčí českého i anglického jazyka považují popisovaný objekt za středobod pro prostorovou orientaci a vyjadřují se pomocí systému předložek (`před`, `za`, `vedle`, `pod`, `nad`). U domorodého australského kmene Guugu Ymithirr se ale prostorová orientace i jazykový systém orientuje v intencích, zda se objekt nachází `na sever`, `jih`, `západ` nebo `východ`. Mluvčí tohoto jazyka tedy popisují polohu objektu podle světových stran (Boroditsky, 2011).

V psychologickém výzkumu se často setkáváme s výzkumem vnímání a chápání emocí u mluvčích, kteří disponují ve svém jazyce různými výrazovými prostředky k popisu toho, co cítí nebo prožívají. Stejně jako u jiných aspektů myšlení a vnímání, slovník a gramatická struktura, kterou máme k vyjádření našich pocitů, nevytváří naše pocity, ale formuje, jak je chápeme a do určité míry, jak je prožíváme (Boroditsky, 2011).

Slova nám pomáhají pojmenovat naše emoce. Například schopnost rozpoznat nelibost z obličeje je univerzální. Ale v jazyce, který má slova `hněv` a `smutek`, můžeme dále rozlišit, o jaký druh nelibosti jde, a lépe popsat výraz obličeje. To neznamenaí, že lidé nikdy předtím neprožívali hněv nebo smutek, než se pro ně objevila slova. Ale mohli se potýkat s obtížemi v porozumění nebo vysvětlení jemných rozdílů mezi různými dimenzemi nelibosti (Lindquist, Gendron, 2013).

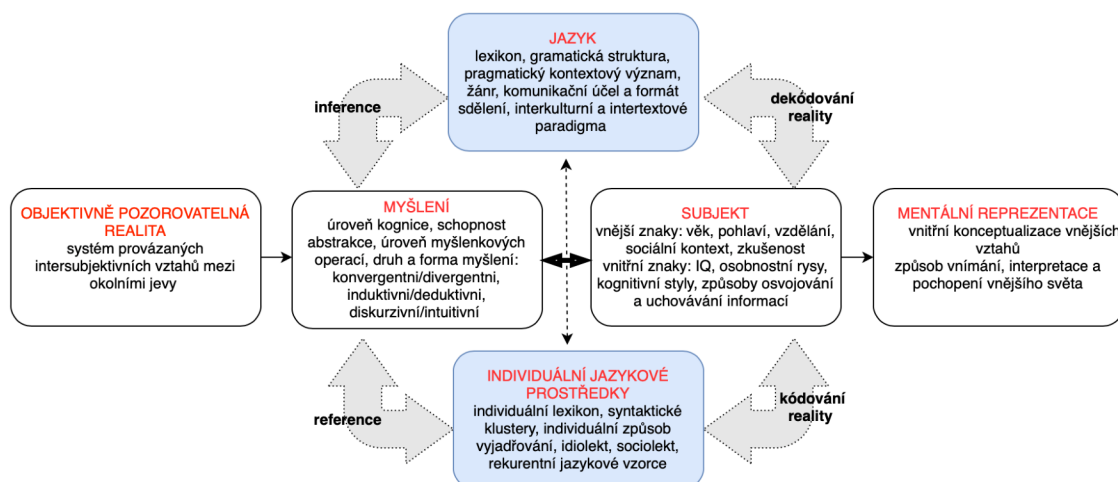
Experimentální skupině anglicky mluvících dětí předložil výzkumný tým Lindquist, MacCormack, Shablack (2015) obrázek osoby s rozčileným výrazem. Poté jim byl předložen soubor obrázků lidí vyjadřujících různé výrazy, včetně šťastných, smutných, překvapených, vystrašených, znechucených nebo rozčilených. Výzkumníci děti vyzvali, aby všechny obrázky, které odpovídaly prvnímu obrázku s rozčileným výrazem, umístily do krabice. Dvouleté děti v experimentu obvykle umístily všechny tváře - kromě šťastných

- do krabice. Čtyřletí byli selektivnější a často kromě šťastných tváří vynechávali i smutné nebo vystrašené tváře. To naznačuje, že s rozšiřováním našeho slovníku pro vyjadřování emocí roste i naše schopnost tyto emoce chápat a rozlišovat (Lindquista, MacCormack a Shablack, 2015, s. 444).

Potvrzuje se tedy opakovaně ontogenetický předpoklad, že schopnost popsat vlastní pocity souvisí nejen s dostupnými výrazovými prostředky daného jazyka, ale i stupněm vývoje myšlení a schopností extrakce abstraktních kategorií (Gazzinga, Ivry, Mangun, 2002).

Následující Obrázek 3 znázorňuje vztah mezi myšlením, jazykem, myslícím subjektem, včetně jeho osobnostního nastavení, a charakteristickým způsobem vyjadřování, které si mluvčí osvojil.

Obrázek 3: Vztah myšlení, osobnosti a jazyka



Zdroj: Vlastní náčrt

Z pohledu lingvistického relativismu je tedy evidentní, že se osobnostní znaky mluvčího, ať už vnější (pohlaví, věk, vzdělání a socio-kulturní kontext) nebo vnitřní (IQ, osobnostní rysy, emoční vyzrállost a kognitivní styl), se odrážejí nejen ve způsobu myšlení, ale i ve způsobu, jakým se vyjadřuje. To vše za předpokladu, že dostupné komunikační prostředky v daném jazyce umožňují odpovídajícím způsobem vytvářet mentální kategorie a vytvářet *'mentální reprezentace'* odpovídající realitě. Obecně se tedy podařilo nepopíratelně prokázat, že struktura jazyka ovlivňuje mentální procesy a kategorizaci myšlení (Gleitman, Papafragou, 2005).

Lingvistický výzkum dále naznačuje, že vliv jazyka není omezen pouze na rozvíjení širšího slovníku pro kategorizaci abstraktních či konkrétních pojmů. Jazyk může „také pomáhat vytvářet pochopení tím, že sloučí pojmy do konkrétních kategorií, jako jsou 'pečivo', 'chleba' nebo 'znechucení', čímž umožňuje rozeznání souvislostí a pochopení hlubších vztahů. Samotná existence nebo neexistence dostupného lingvistického výrazu neznamená, že myslící subjekt nemůže vnímat danou realitu (Grossman, Ellsworth, Hong, 2012).

3 OSOBNOST

Osobnost v psychologii je komplexní a unikátní soubor dlouhodobých charakteristik a chování, které odlišují jednotlivce od ostatních. Tato definice odráží jednotlivcův způsob myšlení, cítění a jednání, který je poměrně stabilní a vyvíjí se pomalu s časem. Osobnost ovlivňuje, jak jednatelce vnímá svět, jak reaguje na situace a jak interaguje s ostatními lidmi (Smékal, 1985).

V této kapitole si představíme několik teorií a přístupů charakteristických pro výzkum osobnosti v psychologii. Některé z nich se zaměřují na významné osobnostní rysy, které ovlivňují naše jednání a chování, jiné se snaží porozumět hlubším vrstvám osobnosti, jako jsou nevědomé motivace a konflikty. Z hlediska metodologie shrneme různé nástroje pro zkoumání struktury osobnosti, včetně dotazníků, interview, pozorování a psychometrických testů. Speciální pozornost budeme věnovat strukturálnímu modelu OCEAN, na kterém je založen sebeposuzovací inventář BFI-2, který slouží ke sběru psychologických dat v praktické části této práce. Z uvedeného modelu se zaměříme na popis dimenze Extraverze a jejích subškál, Energičnosti, Asertivity a Sociability, které tvoří jeden z hlavních psychologických rysů a patří k základním strukturálním složkám osobnosti.

3.1 Teorie osobnosti

V psychologii existuje několik definic osobnosti, protože různí teoretici a školy myšlení přistupují ke studiu osobnosti různými způsoby. V této kapitole jsou shrnuty některé z hlavních definic a přístupů ke studiu osobnosti. Cílem této kapitoly je představit pojem osobnost v jeho komplexnosti a obsáhnout všechny složky, které k jeho pochopení nepochybně patří.

Osobnost je strukturálně dynamický systém (Říčan, 2010), ve kterém se promítají všechny tři složky psychologické triády osobnosti – tedy *`myšlení`*, *`citové prožívání`* a *`chování`* člověka. Tyto složky mohou v určitém okamžiku uvnitř psychiky působit symbioticky, tedy stejným směrem, zároveň však i proti sobě, což způsobuje intrapsychický rozpor nebo konflikt.

Utváření osobnosti je podmíněno (i) biologickými faktory, tedy genetickým potenciálem, vrozenými predispozicemi, charakterem a temperamentem, (ii) sociálními faktory, tedy kulturním a sociálním kontextem, vzděláním, statutem, morálkou a příslušností k určité sociální hodnotové skupině, a v neposlední míře (iii) sebeutvářením, tedy vůlí, motivací, disciplínou, vytrvalostí, touhou po seberealizaci a sebenaplnění (Cakirpaloglu, 2012).

Studium osobnosti můžeme spojovat s třemi základními teoretickými koncepty – tedy psychoanalýzou, humanistickou psychologií a behaviorismem (Smékal, 2004), nebo s vymezením jednotlivých složek osobnosti, tedy s její strukturou a dynamikou (Cakirpaloglu, 2012) a vymezením jednotlivých složek, tedy charakterových vlastností, temperamentem, motivací, vůlí, sebezpojetím, emočním prožíváním, morálním nastavením, inteligencí, mentální kapacitou, různými charakterovými rysy a úrovní kognice (Cattel, 1945; Eysenck 1984).

V psychologii osobnosti neexistuje univerzálně přijímaná teorie ani jednoznačné teoretické paradigma (Cakirpaloglu, 2012). Jde totiž o natolik komplexní systém, že každý přístup upřednostňuje jiný prvek nebo složku, a je těžké obsáhnout stejnou měrou všechny proměnné, které hrají svoji roli při formování a tvorbě osobnosti. V následném přehledu se budu snažit o stručnou rekapitulaci dostupných teoretických paradigmat v přístupu ke studiu osobnosti. Následující kapitoly shrnují různé úhly pohledu na osobnost a ukazují, jak se tento koncept vyvíjel a transformoval v průběhu času.

3.1.1 Teorie osobnosti z pohledu psychoanalýzy

Sigmund Freud byl nejen zakladatelem psychoanalýzy samotné, ale také jedním ze zakladatelů teorie osobnosti. Jako první vymezil složky, které tvoří jádro naší osobnosti. Podle Freuda je struktura osobnosti tvořená třemi úrovněmi vědomí (vědomí, předvědomí, podvědomí) a třemi částmi osobnosti (id, ego, superego). Freud zdůrazňoval roli nevědomí a význam dětství při formování osobnosti (Drapela, 1998)

Ve Freudově pojetí byl pojem osobnosti spíše negativní, chování bylo podle jeho názoru determinované pudovými pohnutkami a poháněné vnitřním konfliktem. Za základní motivační složku považoval Freud agresi a sexualitu. Drapela (1998, s. 21) uvádí:

„Freud vymezil tři subsystemy, řízené vzájemně rozporuplnými principy a cíly, které v osobnosti existují spolu a jsou mezi sebou v neustálém boji. Jsou to Id, Ego a Superego.“

Jak je z uvedeného citátu patrné, v tomto pojetí byla osobnost založena na vnitřním konfliktu. Tento postoj byl potom Freudovými následovníky dále rozpracováván. Carl Gustav Jung Freudovu tří složkovou teorii osobnosti rozšířil, podle jeho názoru se osobnost měla skládat ze čtyř složek: (i) Ego (vědomého Já), (ii) kolektivního nevědomí, (iii) osobního nevědomí a (iv) bytostného Já, které mělo vědomí a nevědomí spojuvat (Smékal, 2004).

Jung do studia osobnosti zařazuje už i vliv prostředí, který je zahrnut do konceptu kolektivního nevědomí. To obsahuje informace ze života našich předků, jejich přesvědčení a prožívání, je vrozené, dědí se a je společné všem jedincům dané společnosti. Tyto vrozené tendence nazývá Jung Archetypy, jsou spojené se silným emočním nábojem. V osobním nevědomí jsou uloženy pouze vlastní vzpomínky z našeho života a Ego je v pojetí Junga vědomou řídicí složkou, která rozhoduje o tom, které mentální obsahy vpustí do vědomí a vytváří tak identitu jedince. Katalyzátorem osobního růstu je potom bytostné já, které zajišťuje syntézu vědomých i nevědomých procesů. Jung zohledňoval i potřebu seberozvoje, tedy individuace. Z Jungova pojetí vychází i Eysenckova typologie osobnosti (Drapela 1998).

Mezi další představitele psychoanalytického směru ve výzkumu osobnosti řadíme Alfreda Adlera s jeho teorií o sourozeneckém pořadí a komplexem méněcennosti jako základním motivačním prvkem pro rozvoj osobnosti poháněné úsilím o nadřazenost. Dále pak Karen Hornyovou s její teorií neurotických potřeb a třemi základními pohyby motivovanými potřebou snížit neurotickou úzkost – tedy pohyb od lidí, k lidem, nebo proti nim (Říčan, 2010). Henry Stack Sullivan v teorii osobnosti zohledňoval interpersonální vztahy, rozvoj osobnosti byl spojen s ontogenezí a raným dětstvím (kojením), za základní motivační faktor pro rozvoj osobnosti považoval snahu o odstranění úzkosti a naplnění potřeby bezpečí (Drapela 1998). Erich Fromm zkoumal vztah osobnosti, přírody a

přirozenosti a rozlišil dvě základní pohnutky – potřebu socializace (vytváření vztahu) a potřebu asimilace (získávání a osvojování). Mezi těmito pohnutkami hledal rovnováhu tak, aby orientace osobnosti byla vyvážená a měla spíše produktivní charakter, na rozdíl od nevyvážené osobnosti s neproduktivním charakterem s maladaptivní orientací (receptivní, kořistnickou, hromadivou a tržní) (Smékal, 1985). V neposlední řadě sem patří i egocentrická teorie Erika H. Eriksona, pro kterého vývoj osobnosti nekončí dětstvím, ale probíhá celý život, dokud člověk úspěšně neprojde všechna epigenetická stádia. Dokončení prvního stádia je předpokladem pro vstup do druhého, posledním vývojovým stádiem je dosažení integrity osobnosti, moudrosti a lásky. Erikson identifikoval osm stadií psychosociálního vývoje, každý posun na další úroveň probíhá v reakci na sociální výzvy (Mikšík, 1999).

3.1.2 Teorie osobnosti z pohledu behaviorismu

V behaviorálním pojetí je při formování osobnosti kladen největší důraz na vliv prostředí. Konečná podoba naší psychiky je dána výchovou a naučeným způsobem chování, které je determinováno schopností přizpůsobit se vnějším podmínkám a zajistit si co nejlepší podmínky pro přežití. Vliv nevědomí je oproti psychoanalytickému pojetí odsunut do pozadí a vnitřní nevědomé pohnutky mají na osobnost minimální vliv v porovnání s vlivem prostředí. Učení a imitace a identifikace jsou považovány za hlavní hybné složky formování osobnosti, včetně maladaptivních mechanismů, které vznikají jako reakční postoje na patologii prostředí (Mikšík, 2007).

Pro J.B. Watsona, představitele klasického behaviorismu, se osobnost jedince přímo zrcadlí v jeho chování. Základní jednotkou chování je souhrn reflexů (podmíněných i nepodmíněných, které tvoří základní stavební kameny naší osobnosti). Jeho teorie byla ovlivněna výzkumem podmíněných reflexů I.P.Pavlova, osobnost by se proto měla zkoumat pomocí pozorování zákonitostí chování ve srovnání s podmínkami ve vnějším okolí (Mikšík, 1999). Watson se mimo jiné proslavil i následujícím výrokem:

„Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select – doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors“ (Watson, 1930, s.78)⁴

V behaviorálním přístupu hrají velkou roli i teorie učení, zejména Clarka L. Hulla, který lidské chování definoval jako závislou reakci na přicházející stimul (Mikšík, 1999). Tuto teorii učení a jeho vliv na formování lidské osobnosti dále rozpracovali John Dollard a Neal E. Miller, kteří oba vycházeli z psychoanalýzy. Ve svém pojetí se soustředili na popsání osobnosti z hlediska vývoje, prožívání a nevědomí a do teorie chování jako reakce na stimul přiřadili ještě pud jako motiv něčeho dosáhnout a potřebu odměny nebo odezvy v případě, že je toho dosaženo. Tento úspěšný výsledek lidského chování potom označili za *zpevnění*, což je hlavním motivačním faktorem ve vývoji lidského chování, a tudíž i formování lidské osobnosti (Nakonečný, 1997b).

⁴Český překlad: „Dejte mi tučet zdravých dětí a můj vlastní svět, ve kterém bych je mohl vychovat, a zaručuji, že každé vycvičím jako odborníka, v jakémkoliv povolání se mi zamane, bez ohledu na jeho talent, záliby, tendence či původ.“ Zdroj: https://www.lidovky.cz/domov/proc-jsou-nadane-deti-nadane.A100612_000077_In_noviny_sko [citováno z 18.3.2024].

Za předního představitele radikální formy behaviorismu je považován B.F. Skinner, který naopak odmítal pudy a vnitřní pohnutky v popisu lidské osobnosti a omezoval se pouze na pozorovatelné projevy lidské osobnosti, za které považoval právě lidské chování. Veškeré intrapsychické procesy považoval za objektivně neprokazatelnou a nezkoumatelnou proměnnou, tzv. `black box` a omezil se na zkoumání kauzality lidského chování, tedy vztahu mezi podmínkami vnějšího prostředí `stimuly` a výsledným chováním jedince, tedy `reakcí`. V tomto vztahu hledal kauzalitu a řád pomocí kontrolovaných experimentálních podmínek. Vyslovil teorii `operantního podmiňování` - tedy formování osobnosti se děje formou zpevnování žádoucí reakce na předložený stimul (Hall, Lindzey, Leohlin, Manosevitz, 1997).

Albert Bandura rozšířil behavioristický vzorec stimul = organismus = reakce a tvrdil, že osobnost je výsledek interakcí mezi kognitivními procesy, behaviorálními faktory a sociálním prostředím. Zdůrazňoval význam pozorování, modelování a sebedeterminace, což už je přesah k fenomenologii a potřebě sebeurčení. Osobnost je tedy podle Bandury zároveň produktem, ale i samotným tvůrcem vlastního prostředí (Nakonečný 1995).

3.1.3 Teorie osobnosti z pohledu humanistické psychologie

Humanistická psychologie přichází v teorii osobnosti s principem svobodné volby, člověk tedy není pouhým objektem tvarovaným prostředím nebo ovládaným pudově podvědomými pohnutkami, ale je tvůrcem vlastního osudu – neboli subjektem, který rozhoduje o povaze, formě a zaměření vlastní osobnosti (Drapela 1998). Do popředí se dostává motiv sebeaktualizace a seberealizace, potřeba uznání a smyslu života.

Gordon W. Allport je považován za amerického průkopníka psychologie osobnosti. Sjednocoval ve svém humanistickém přístupu řadu dosavadních poznatků, rozlišoval mezi temperamentem jako surovinou a charakterem coby zhodnocenou kultivovanou částí osobnosti. Obě tyto složky považoval za nerozdílné součásti jediného celku a zdůrazňoval dynamický vztah mezi nimi. Zavedl teorii rysů, které činí osobnost výrazově konzistentní. Rysy jsou vzájemně funkčně propojené s konkrétními situacemi a oboustranně se determinují. Navenek se rysy projevují jako vrozené a získané dispozice (Mikšík, 1999).

Carl R. Rogers, coby představitel fenomenologie, zdůrazňoval koncepci seberealizace a důvěry vůči sobě samému a možnost volby. Definoval osobnost jako proces, který se odehrává v průběhu celého života a je formován zkušenostmi, zejména zkušenostmi se sebou samým a s ostatními, osobnost v jeho pojetí tedy není hotovým produktem. Do studia osobnosti přinesl teorii Self (já) – rozlišoval mezi ideálním já (čím chci být) a reálným já (čím jsem). Hnacím motorem vývoje osobnosti je touha po seberealizaci. Do popředí se dostává pojem kongruence, tedy shody mezi ideálním a reálným já, psychopatologie je následkem rozporu, kdy člověk některé svoje pohnutky a stránky není schopen přijmout, dochází potom k sebepopření, asimilaci nebo zkreslené sebereprezentaci a člověk není schopen dosáhnout zdravého sebepojetí a dobrého života (Drapela, 1998).

Mezi další důležité prvky v teorii osobnosti patří logoterapie V. E. Frankla, která stojí v juxtapozici proti Freudově pojetí slasti a Adlerově pojetí moci jako základních hybných sil lidské psychiky. Ve Franklově pojetí je základní hybnou silou rozvoje osobnosti `vůle ke smyslu`. Do teorie studia osobnosti tedy přibyla další dimenze, a to hledání smyslu života. Smysl lidské existence ale není vrozený nebo univerzálně daný, naopak je úkolem každého člověka ho vlastním životem autenticky objevovat. Smysl musíme hledat altruisticky ve vykonání konkrétních činů ve prospěch obecného dobra,

nikoli sebestředně ve prospěch vlastní (Smékal, 2004). Vyjádřeno slovy samotného Frankla:

„Člověk by se neměl ptát po smyslu vlastního života. Měl by si uvědomit, že je to on, kdo je tázán. Život se každého ptá na smysl jeho života - a každý může odpovědět životu jen tím, co dělá; naši jedinou odezvou na život je chovat se odpovědně“ (Covey, 2009, s. 123).

Aby byl výčet kompletní, nesmíme v opomenout G. A. Kellyho, který přinesl do studia osobnosti `konstruktovou teorii`. Podle Kellyho je osobnost složena z jednotlivých psychologických konstruktů, které lidé používají k interpretaci světa. Samotný pojem osobnosti je pro Kellyho uměle vytvořeným psychologickým konstruktem. Kelly zdůrazňoval subjektivní povahu konstruktů, poznání osobnosti pro něho spočívalo v poznání, jak daný jedinec konstruuje osobní zkušenost. Zdůrazňoval individualitu (jedinečnost) a organizaci (řád) v utváření osobnosti – jednotlivé psychické složky pak tvoří dynamickou hierarchickou strukturu (Mikšík, 1999).

3.1.4 Strukturální teorie osobnosti

Zatímco jsme představili hlavní teoretické přístupy ke studiu osobnosti, došli jsme k bodu, kdy je zřejmé, že osobnost není nahodilá změn jednotlivých složek, ale že má pevnou strukturu, která je konzistentní v čase a odpovídá jednotlivým charakterovým a temperamentovým složkám, které spolu dynamicky interreagují (Cakirpaloglu, 2012).

Strukturální pojetí osobnosti v sobě zahrnuje pokusy o klasifikaci jednotlivých složek osobnosti do smysluplného celku, který můžeme třídít dle různých typologických hledisek. Nakonečný (1997a) rozděluje klasifikační modely osobnosti do dvou základních proudů: (i) klasifikace založené na vztahu psychických složek s fyzickou podstatou somatotypu a (ii) klasifikace zohledňující vztah mezi psychickými složkami jedince a jeho vztahu ke světu – tedy způsobu, jakým o světě přemýšlí, jakým jazykem ho popisuje a jak se k němu ve svém chování vztahuje.

3.1.4.1. Teorie osobnosti založená na somatotypu

Struktura osobnosti založená na somatotypu byla známá už od antiky. Hippokrates, proslulý starořecký lékař a filozof, přinesl myšlenku, že psychika člověka, jeho prožívání a chování, je závislá na fyziologické konstituci. Roztřídil psychologické nastavení jedince podle čtyř základních typů tělesných tekutin, které měly předurčit jeho temperament. Na základě této kategorizace rozlišujeme temperament (i) *sangvinický* – spojen s krví – charakteristickými vlastnostmi byla optimistická povaha, dostatek energie, veselost, Extraverze a sklon zapojovat se do sociálních interakcí; (ii) *flegmatický* – spojen s hlenem – značící klidnou povahu, vyrovnanost, mírumilovnost a sklon k introverzi, rezervovanost v chování a řešení konfliktů s chladnou rozvahou; (iii) *cholerický* – spojen s žlučí – tento druh temperamentu patřil k lidem vášnivým, energickým, impulzivním a extrovertním, náchylným k hněvu a zapojování se do konfliktů; (iv) *melancholický* – spojen s černou žlučí – byl temperament charakterizující hlubokou, introspektivní a citlivou povahu introvertních jedinců s náchylností ke smutku a úzkosti (Řičan, 2010).

Dle Řičana (2010, s. 66), prvním, kdo vyslovil myšlenku, že temperament, včetně psychopatologických projevů, souvisí s fyziologickou stavbou těla, byl E. Kretschmer. V jeho pojetí se určitý somatotyp pojil s psychickým stavem jedince, označil tak například

typ (i) *pyknický* – hrubé stavby kostí se sklonem k nadváze – pojící se s psychickou nestabilitou a sklonem k psychotismu; a typ (ii) *leptosomatický* – s dlouhými tenkými končetinami a útlým tělem – se sklonem k neuroticismu.

Dalším příkladem konstituční typologie osobnosti je W.H. Sheldonova temperamentová typologie rozlišující tři typy konstitucí dle dominantní složky hmoty, která tělo tvoří: (i) *ektomorf* je tělesný typ štíhlé postavy s vysokým metabolismem a slabým osvalením, temperamentově pozitivně korelující s druhem temperamentu zvaným `cerebrotonní` spojovaný s vědci a intelektuálně aktivními jedinci; (ii) *mezomorf* je typ atletické postavy s převahou svalů a sklonem k agilitě, silně korelující s temperamentovým typem `somatotonní` charakteristickým pro atlety a sportovce; (iii) *endomorf* má sklon k nadváze a ukládání tuku, charakterizuje emočně citlivé jedince temperamentově označované jako `viscemorfní` (Mikšík, 1999).

3.1.4.2. Teorie osobnosti založená na charakterových rysech

Tyto teorie mají společného jmenovatele v hledání vnitřního řádu uspořádání lidské osobnosti na základě vnitřních intrapsychických znaků, tělesné předpoklady a fyziologie pro tyto kategorizační systémy nejsou relevantní. Osobnost je tedy popsána jako systém charakterových vlastností a intrapsychických tendencí. Tyto strukturální modely se pojí s rozvojem psychometrických nástrojů pro měření osobnostních rysů a ověřování teoretických závěrů pomocí testování velkého vzorku respondentů.

Prvním významným představitelem byl Raymond Cattell, který svůj výzkum opíral právě o rozsáhlé testové baterie a řadu psychometrických proměnných. Jeho model je v teorii osobnosti zřejmě nejsložitější, protože popisuje osobnost jako mnohvrstevný model, ve kterém je několik rysových systémů hierarchicky propojených mezi sebou (Cakripaloglu, 2012). Drapela (2011) vyzdvihuje Cattellův záměr osobnost nejen dokonale poznat a popsat, ale zároveň i predikovat.

Cattell ke svému výzkumu sbíral data pomocí sebeposuzovacích dotazníků (*Questionary data*), pozorováním a sbíráním informací od třetích osob (*Life data*) a z výsledků vlastních testů (*Test data*). Pomocí faktorové analýzy postupně vydestiloval 16 primárních rysů, ve kterých se daly shrnout příčiny pozorovatelného chování a které se staly základem jeho osobnostního dotazníku **16PF**. Těchto 16 rysů dále vztahoval k (i) *temperamentu* (vrozenému základu determinujícímu množství energie, rychlost reakcí nebo emoční stabilitu); (ii) *schopnostem* (s měřitelným výkonem, např. IQ); a (iii) *dynamice osobnosti* (spojené s pudy a reflexy – nazývanými *ergy*, a s hodnotami a postoji – neboli *sentimenty*) (Drapela, 1999). Během dalších výzkumů těchto 16 primárních faktorů dále redukoval na 5 univerzálních faktorů: *Extraverze-introverze*, *úzkostnost – integrace*, *senzitivita – rozumovost*, *rezignace – vzdornost*, *inteligence – mentální tupost* (Noller et al., 1987).

Hans Eysenck, na rozdíl od Cattella, na základě teorie sestavil hierarchii rysů, a tu potom pomocí faktorové analýzy ověřoval. Stejně jako Cattell shromáždil data z různých zdrojů – objektivních testů, sebeposuzovacích dotazníků, životopisných faktů a skóre z pozorování respondentů. Na základě své analýzy dospěl k sestavení 3 základních osobnostních dimenzí: *intraverze - Extraverze*, *neuroticismus - stabilita* a *psychoticismus - síla superega*. Odtud také pochází označení Eysenckovy teorie jako **PEN modelu** – tedy *psychoticismus*, *Extraverze* a *neuroticismus* (Mikšík, 1999). V rámci těchto 3 dimenzí by mělo být možné vysvětlit veškeré pohnutky ovlivňující lidské chování. Později k jednotlivým dimenzím Eysenck přiřazuje i neuro-fyziologické průvodní jevy, např.

Extraverze souvisí s hladinou kortikální aktivity, která je měřitelná EEG; neuroticismus potom s limbickým a viscerálním systémem (Mikšik 1999, s. 189). Podle Plhákové (2003) byla platnost Eysenckovy teorie potvrzena i dalšími výzkumy (např. teorií Big Five).

Další strukturální teorií osobnosti, založenou na faktorové analýze, je právě Velká pětka, **Big Five Theory**. Jak už sám název napovídá, tento model osobnosti je založený právě na pěti dimenzích osobnosti, vychází z Cattelova šestnáctifaktorového modelu a odráží se v ní Eysenckovy tři strukturální rysy osobnosti. Samotný název této teorie stanovil v 80. letech L.R. Goldberg (1992) a potvrdil tak objev pěti faktorů v modelu osobnosti, který byl před ním opakovaně formulován a potvrzen řadou výzkumů, jako např. Fiske (1949), Norman (1963) a Tupes a Christal (1961), kteří se řadí k zakladatelům pětifaktorového modelu teorie osobnosti (Borkenau a Ostendorf, 1990).

Mezi těchto pět základních faktorů v teorii osobnosti patří: (i) *neuroticismus-emoční labilita*, která indikuje míru stability emočního prožívání a míru schopnosti adaptace jedince; (ii) *Extraverze – intraverze*, označující míru interpersonálních schopností a potřebu socializace; (iii) *otevřenost vůči zkušenosti*, která měří míru potřeby vyhledávání nových zkušeností a zážitků, ale i nekonvenčnosti, originality a kreativity; (iv) *přívětivost*, rozlišuje mezi jedinci orientovanými na potřeby vlastní a naopak na potřeby ostatních; a (v) *svědomitost*, měřící míru sebekontroly, péle a schopnosti dosahovat stanovených cílů (Hřebíčková, 2011).

Z pětifaktorového modelu osobnosti vycházeli i Costa a McCrae (1992), kteří pomocí statistické analýzy Cattelova 16 faktorového inventáře vyčlenili tři škály – *neuroticismus, extraverzi a otevřenost vůči zkušenosti* – a použili je pro svůj osobnostní inventář NEO-PI. Ve svém přístupu zohledňují ve vývoji osobnosti i vliv prostředí a podmínek, které jedince formují. Vybrané osobnostní rysy jsou pak univerzálně platné, vrozené formou genové dispozice a objevují se nejen u lidí, ale i u zvířat (Costa, McCrae, 1997). Díky lexikální analýze jednotlivých položek rozšířili autoři svůj inventář ještě o dvě další škály odpovídající pětifaktorovému modelu osobnosti, a sice o dimenze *přívětivost* a *svědomitost*, a tak vznikl teoretický model **OCEAN**, tvořící velkou pětku osobnostních dimenzí (Pervin a John, 1999).

3.1.4.3. Extraverze jako psychologický koncept

Pojem Extraverze je jedním ze základních stavebních prvků při teoretickém popisu osobnosti, svými kořeny sahá až k Jamesově (1907) koncepci osobnostních charakteristik (Lucas, Diener, Grob, Suth a Saho, 2000). Definicí Extraverze ve 20. letech 20. století poprvé detailněji rozpracoval Carl Gustav Jung (Jung, Beebe, 1997), který tento termín použil při svém popisu charakteristických tendencí osobnosti, které dělil na dva hlavní typy: introverzi a extroverzi. Extroverzi charakterizoval jako zaměření na vnější svět, společnost a objekty, zatímco introverzi spojoval s orientací směrem k vnitřnímu světu, vlastním myšlenkám a pocitům (Beebe, 2012).

Extraverze je jednou z hlavních dimenzí osobnosti a figuruje v různých modelech, nalezneme ji u G. Allporta jako jeden z aspektů behaviorálního naladění, u E.H. Eriksona jako vývojovou výzvu mezi potřebou intimity a izolace, nebo u J. Bowlbyho jako vztahovou potřebu sociálního kontaktu a adaptace na sociální prostředí (Nettle, 2005). Jako samostatnou dimenzi strukturálního modelu osobnosti ji shodně označili R. Cattel, Hans J. Eysenck i Costa a McCrae (Hřebíčková, 2011). V této kapitole se budeme zabývat psychickým obsahem, který se pojí k pojmu Extraverze.

V rámci stereotypních představ jsou osoby s extravertní osobností komunikativní, společenské, sebejisté, v kolektivu obsazují spíše vedoucí pozice, jsou vysoce energické a mají přátelskou a veselou povahu. Jedinci s vysokou mírou Extraverze mají silnější emoční

reakce, jsou pozitivně naladěni a nastavení na spolupráci (Nettle, 2005). Ohledně budoucnosti jsou velmi optimističtí, v sociálním kontaktu méně zranitelní a více otevření vůči soutěži. Jedinci s nízkou mírou extroverze, naproti tomu preferují být sami, méně se zapojují do aktivit, jsou klidní, raději zůstávají stranou kolektivu a mají tendenci vyhýbat se náročným situacím a stimulaci (Lucas a Diener, 2009).

V mnoha studiích byla Extraverze označena za jeden z nejrozeznatelnějších znaků lidské osobnosti, Funder (1995, s. 663) uvádí řadu výzkumů, které potvrzují, že extraverzi je možné detekovat mimo jiné v samotném verbálním projevu. Extravertní jedinci mluví rychleji (Gifford a Hine, 1994) a více nahlas (Scherer, 1978).

V mezikulturní studii základních osobnostních rysů Extraverze (Lucas et al., 2000) byla shrnuta řada dílčích vlastností, které do osobnostní dimenze Extraverze neodmyslitelně patří. Tyto povahové rysy jsou shrnuty v následující tabulce.

Tabulka 1: Psychologické aspekty osobnostního rysu Extraverze

Psychologická dimenze	Popisná adjektiva vysokého a nízkého skóre BFI-2	Popis daného rysu dle Hřebíčkové (2011)
Extraverze		
vzrušivost	+ nadšený - ostýchavý	touha po vyhledávání výzev, vzrušení, zážitků, dobrodružství, pozitivní postoj ke změnám
Sociabilita	+ společenský - uzavřený	touha někam patřit, být součástí kolektivu, pocit laskavosti a přátelskosti
náklonnost	+ povídavý - málomluvný	pocit radosti a nadšení, oceňování blízkých mezilidských vztahů, citová vřelost
Energičnost	+ aktivní - zdrženlivý	pocit živosti, aktivity a dostatku energie, vysoká výkonnost a nízká míra unavitelnosti
sebeprosazení	+ sebejistý - rezervovaný	pocit nadřazenosti, sociální dominance, exhibicionismu, asertivity, nastavení na výkon a splnění úkolu
cílevědomost	+ průbojný - samostatný	ambicióznost, oceňování úspěchu, pocit kompetence a orientace na dosažení cíle
impulsivita	+ smělý - zdrženlivý	výrazné emoční prožívání, vysoká míra vzrušivosti, sklon k reaktivitě
satisfakce	+ na odměnu orientovaný - soběstačný	orientace na externí validaci, potřeba uznání, motivace v dosahování odměn, záliba ve veřejném ocenění

Zdroj: Lucas et al., 2000

K popisu dílčích vlastností Extraverze ve spojitosti s jazykovými prostředky se budeme dále věnovat v kapitole 4.2.2. a 4.2.3.

3.2 Metody výzkumu osobnosti

Výzkum osobnosti započal už ve starověku (viz kapitola 3.1.4.1) a byl nejprve spojován s fyziologickými znaky, které měly určovat povahu a charakter daného člověka pomocí spojitosti s jeho tělesnou konstitucí (Říčan, 2010). Z historického hlediska můžeme tedy odvodit i metodologický vývoj v přístupu k popisu a zkoumání osobnosti. Každá z níže prezentovaných metod přispěla k porozumění komplexnímu fenoménu lidské osobnosti různými důkazy a přinesla poznání tohoto konceptu z různých stran a v různých stádiích vývoje lidského poznání.

Každá z těchto metod má své výhody i omezení, a volba vhodné metody závisí nejen na cílech výzkumu, ale i na charakteru zkoumané osobnosti a podmínkách realizovaného výzkumu. Kombinace různých metod může poskytnout komplexnější a vyváženější pohled na osobnost jednotlivce.

3.2.1 Pozorování

Metoda pozorování cílí na sledování reálného chování jedince v daném prostředí. Zaměřuje se na systematické a pečlivé vyhodnocování chování, emočních projevů a sociálních interakcí pozorovaného subjektu, včetně verbální komunikace, neverbálních signálů, gest, spontánních reakcí na určité podněty, řízených behaviorálních vzorců ve specifických situacích, mimiky a emočního doprovodu. Cílem je získat hlubší porozumění jedincovým myšlenkám, emocím, motivacím a chování, což je klíčové nejen pro diagnostiku a typologii osobnostních rysů, ale i léčbu psychologických problémů (Orel, 2016).

Mezi nepochybné výhody klinického pozorování patří přímý pohled na chování, které je přímo pozorovatelným důkazem o intrapsychickém procesu, emočním prožívání a mentálních procesech zkoumaného subjektu. Pozorování minimalizuje zkreslení způsobené sebeuposouzením a stylizací subjektu do určité role nebo podoby. Pozorování můžeme dělit na (i) *naturalistické* - tedy sledování v přirozeném prostředí bez zásahu nebo manipulace ze strany pozorovatele; nebo (ii) *strukturované a polostrukturované* - v prostředí předem připraveném, experimentálním, řízeném a manipulovatelném ze strany pozorovatele (Ferjenčík, 2008).

Přirozenou nevýhodou je možnost zkreslení nebo ovlivnění výsledků ze strany pozorovatele, čemuž může předcházet molární (zaměřeným na celek) nebo molekulární (zaměřeným na detail) způsobem kódování pozorovaných faktů, analýzou správného zakotvení v místě, času a teorii. Příkladem může být Flandersův systém interakční analýzy FIAS nebo klasifikační systém výrazů lidských tváří FACS P. Ekmana (Hendl, 2016).

3.2.2 Projektivní techniky

Sigmund Freud, zakladatel psychoanalýzy, vyvinul na přelomu minulého století teorii osobnosti založenou na nevědomí, konfliktech a psychosexuálním vývoji. Ke zkoumání vnitřních procesů osobnosti se v psychoanalytickém pojetí využívají metody jako jsou volná asociace, analýza snů a interpretace lapsů založených na práci s nevědomými obsahy (Říčan, 2010).

Na stejném principu byly vyvinuty i projektivní testy, které jsou založeny na principu, že respondenti reagují na nejednoznačné podněty, jako jsou obrázky nebo slova, a jejich odpovědi odhalují nevědomé aspekty osobnosti.

Mezi nejznámější projektivní metody patří Rorschachův test, ve kterém respondenti interpretují abstraktní inkoustové skvrny, což odhaluje jejich psychologické postoje a osobnostní charakteristiky. Další používanou projektivní metodou je TAT – Tematický apercepční test, kde respondenti vytvářejí příběhy k obrázkům postav, což umožňuje odhalit jejich způsob vnímání sociálních situací a rozkrýt jejich vnitřní postoje a subjektivní prožívání reality (Svoboda, 2011).

Nepopiratelným omezením projektivních technik je náročnost jejich standardizace, subjektivní způsob interpretace a náchylnost ke zkreslení, přinášejí však cenné informace o struktuře osobnosti, které jsou jinak těžko přístupné jinými testovými prostředky (Plhánková, 2004).

3.2.3 Neurobiologický výzkum mozkových funkcí

Tato metodologie výzkumu osobnosti patří v současnosti k nejmodernějším a technicky nejnáročnějším, cílí na měření a analýzu mozkových struktur a funkcí spojených s osobností. Výzkum různých částí mozku, které měly být spojeny s charakterovými aspekty osobnosti, se historicky objevil už v 19. stol., kdy frenologové předpokládali, že se ve struktuře lebky odrazí charakter jedince, z této pavědy se však díky moderní zobrazovací technice stal uznávaný vědecký obor přinášející seriózní a objektivně měřitelné výsledky ve výzkumu struktury osobnosti (Robinson, 2008).

Moderní přístupy k výzkumu osobnosti reflektují snahu získat hlubší porozumění mechanismům, které formují a ovlivňují osobnost. Kombinace genetických, neurovědních a interdisciplinárních přístupů poskytuje komplexní perspektivu o tom, jak osobnost vzniká a jakým způsobem se projevuje v chování jednotlivců. Výzkum neurologických procesů je spojen s metodami jako MRI, fMRI nebo EEG, které přímo sledují neurobiologické mozkové procesy a objektivním způsobem je dokáží vyhodnocovat. Díky nim mohou vědci sledovat strukturální složení, funkční aktivitu i intenzitu mozkových procesů. (Orel, 2016).

Neurobiologický výzkum osobnosti se dále zaměřuje např. na výzkum neurochemických procesů, zejména neurotransmiterů, které zásadním způsobem ovlivňují strukturu a dynamiku osobnosti. Další studie se soustředí na identifikaci genetických markerů, spojených s různými aspekty osobnosti, a na vliv dědičnosti (Grisby, 2000). K nejmodernějším metodám výzkumu osobnosti patří i výzkum neuronových sítí, neurobiologické korelace osobnostních markerů a jejich integrace s AI a moderní metodologií jako jsou big data analýzy a strojové učení. Všechny tyto kroky vedou k efektivnější a komplexnější analýze dat a detailnější identifikaci vzorců spojených s osobností (Reed a Miller, 2002).

3.2.4 Psychometrické inventáře

Ve 20. století se začaly vyvíjet metody, které se zaměřovaly na hodnocení osobnosti pomocí hodnotících škál nebo dotazníků (viz kapitola 3.1.4.2.). Za většinou těchto inventářů stál dlouhodobý výzkum a rozsáhlé psychometrické analýzy různorodých souborů dat za účelem ověření validity a vnitřní konzistence tak, aby nabízené psychometrické nástroje mohly v současné době sloužit jako základní metodické postupy při zkoumání struktury osobnosti jedince (Smékal, 2004).

Mezi nejznámější hodnotící škály patří soubor sebeposuzovacích inventářů založených na faktorové analýze odvíjející se od Cattelova 16-faktorového inventáře **16PF**, ten byl potom dále rozpracován a přeformulován jako **PEN** model, ve kterém Eysenck pojmenoval 3 hlavní biologicky podmíněné škály *Psychotismus – Extraverze – Neuroticismus*.

Z mnoha dotazníkových studií, např. Eysenckově PEN inventáři, vycházeli pomocí faktorové analýzy dva hlavní rysy Extraverze a Neuroticismus (Barret et al., 1998). Autoři Costa a McCrae (1992) na základě Cattelova (1945) dotazníku 16PN shlukovou analýzou identifikovali extraverzi a neuroticismus jako dva stabilní faktory a k nim připojili třetí – otevřenost vůči zkušenosti, což se stalo základem pro jejich **NEO** inventář (Aniaí, 2012).

Později byla opakovaným měřením zjištěna skutečnost, že další dva rysy – svědomitost a přívětivost, popisují rozdílné složky osobnosti, a empiricky bylo doloženo, že dispozičně patří do struktury inventáře NEO, touto revizí tedy vznikl dotazník **NEO-PI**, později **NEO-PI-R** (Costa a McCrae, 2014), který odpovídá modelu OCEAN a zahrnuje pět hlavních osobnostních faktorů: *Otevřenost, Svědomitost, Extraverzi, Přívětivost a Neuroticismus*. Každý z těchto hlavních faktorů je dělen do facet, které poskytují detailnější pohled na konkrétní aspekty osobnosti (Costa a McCrae, 1992).

V podstatě současně s NEO-PI (Costa, McCrae, 1992) vznikal další psychometrický nástroj založený na *pětifaktorovém modelu* – sebeposuzovací inventář **BFI** (John et al., 1991). Ve své podstatě byl mnohem kratší, obsahoval pouze 44 položek, oproti původním 240 položkám v NEO-PI. Původně nebyl hierarchicky uspořádán, avšak poslední revizí **BFI-2** byl rozšířen na 60 položek, které jsou uspořádány do 3 subškál u každé z pěti hlavních dimenzí osobnosti (Soto, John, 2017).

Dalším dnes diagnosticky a výzkumně využívaným psychometrickým dotazníkem je Minnesota Multiphasic Personality Inventory (**MMPI**) vyvinutý v 40. letech 20. století výzkumným týmem Univerzity v Minesotě. Tento sebeposuzovací inventář má 567 položek a pro svůj značný rozsah (oproti 60ti položkové verzi BFI-2 nebo NEO-FFI) je v psychologickém výzkumu osobnosti používán méně často (Dahlstrom, 1992).

Nespornou nevýhodou sebeposuzovacích inventářů v psychologickém výzkumu je možnost zkreslení díky snaze respondentů o pozitivní sebe prezentaci nebo naopak účelovému vyplnění položek týkajících se jejich postojů a hodnot. Mnoho výzkumníků se proto snaží o redukci takového zkreslení tím, že do výzkumu zapojí ještě hodnocení třetí osobou, což by mělo případné odchylky korigovat.

Zásadní výhodou sebehodnotících škál však zůstává jejich snadná administrace, efektivita sběru dat, jednoznačná kvantitativní analýza výsledků a široká možnost mezioborové a mezikulturní komparace díky standardizaci v mnoha jazycích (Blatný, 2006; Hřebíčková, 2011).

4 VZTAH MEZI JAZYKEM A OSOBNOSTÍ

Vliv jazyka na výzkum osobnosti je v psychologii důležitým tématem. Jazyk hraje klíčovou roli v procesu komunikace a vyjadřování myšlenek, emocí a zkušeností, a proto může být důležitým zdrojem informací pro porozumění osobnostním rysům jedince. A nejen to, v této kapitole si ukážeme podmíněnost vztahu mezi jazykovými rysy a výzkumem osobnosti a ukážeme si základní předpoklady, které musíme vzít v úvahu při formulaci výzkumných hypotéz. Mnohé z psychometrických nástrojů, které používáme k měření a profilaci osobnostních rysů, totiž stojí na lingvistickém základě (Hřebíčková, 1993).

Nejednou se tedy dostaneme do situace, kdy si musíme upřímně zodpovědět, zda se v případě výzkumu osobnosti v kontextu jazykových projevů nejedná o `sebenaplňující proroctví`⁵ (Biggs, 2013), připomínající notorickou otázku o prvotní existenci slepice nebo vejce. V případě jazykových prostředků je situace obdobná. Charakterové rysy teorie Velké pětky jsou založeny na slovech popisujících osobnost – nikoli obráceně – čímž se dostáváme do stavu podmíněnosti osobnostních rysů samotnou přítomností daných popisných jazykových termínů (slov) dostupných k vyjádření psychického obsahu. Jinými slovy popis osobnosti bez jazyka by nebyl možný, stejně jako by neexistoval jazyk bez osobnosti, která ho používá.

4.1 Psychologická data na lingvistickém pozadí

Výzkum osobnosti založený na jazykové deskripci, tedy lexikálním základě, je založen na předpokladu, že jazyk zrcadlí realitu (Hřebíčková et al, 1993), a v kontextu osobnosti tudíž přirozeně zachycuje individuální rozdíly v lidském chování, prožívání a celkovém nastavení osobnosti. Aby tyto osobnostní rysy mohl jazyk zrcadlit, musí být viditelné v běžném životě. Jestliže takové viditelné rysy osobnosti existují, budou kódovány v přirozeném jazyce.

Tento postulát vychází z lexikální teorie (Galton, 1949; Allport, 1937), která předpokládá, že nejvýraznější interpersonální odlišnosti mají odpovídající ekvivalent (zažitý termín, výraz, slovo, lexém) v běžném jazyce. Allport a Odbert (1936, s. 47) formulovali `lexikální hypotézu` takto:

„Those individual differences that are most salient and socially relevant in people's lives will eventually become encoded into their language; the more important such a difference, the more likely is it to become expressed as a single word“⁶.

Z lexikální hypotézy potom vychází celá řada studií. Přítomnost termínu pro popis psychologického rysu v přirozeném jazyce tedy značí objektivní podstatu takového rysu a

⁵ Termín `self-fulfilling prophecy` převzato z angl. orig. (Biggs, 2013)

⁶ Z angl. orig. „*Takové individuální rozdíly, které jsou v životě člověka nejzřetelnější a společensky nejpodstatnější, se nakonec zapíší do jeho jazyka; čím podstatnější takový rozdíl je, tím větší je pravděpodobnost, že bude popisován jediným slovem.*“ (Převzato z Allport a Odbert, 1936, s. 47) – vlastní překlad.

opodstatnění jeho zahrnutí do objektivního popisu lidské osobnosti. Takovýmto termínům se v lexikálním pojetí psychologické teorie říká *`osobnostní deskriptory`* (Hřebíčková, 1994).

Psychologové tohoto vztahu začali využívat pro popis charakteristických rysů osobnosti. Výhodou bylo jakési zjednodušení terminologie na základě kódování v přirozeném jazyce a unifikace přístupů ke studiu osobnosti (Galton, 1889; Allport, 1937; Goldberg, 1990).

Psychometrické inventáře založené na lexiku jsou dnes jednou z nejrozšířenějších metod sběru psychologických dat (viz kapitoly 3.2.4, 4.1.1 a 6.2.1). Psycholinguistický přístup k popisu osobnosti má však jistá omezení, která pramení ze samé podstaty přirozeného jazyka. Na tyto aspekty se zaměříme právě v této kapitole.

V přirozeném jazyce je přítomno velké množství hodnotících, expresivních a jinak zkreslujících termínů, které nejsou pro psychologický výzkum vůbec vhodné a na druhou stranu jazyk nemusí obsahovat všechny termíny, které jsou pro psychologickou profilaci osobnosti žádoucí (Hřebíčková, 1994).

První pokusy o sestavení lingvistických deskriptorů pro psychologický výzkum osobnosti byly sestavovány na taxonomickém základu (Hřebíčková, Ostendorf, 1995). Tedy pokusech o vytvoření hierarchického systému pro popis hlavních a podružných rysů osobnosti, které by mohly být systematicky tříděny do skupin a podskupin dle své podstaty a příbuznosti. Goldberg (1990) popisuje metodický postup k vytvoření rysové taxonomie, kde je třeba dodržet určité předpoklady s ohledem na fakt, že osobnost je abstraktním, neohrazeným pojmem a v jeho popisu je třeba předem omezit okruh psychických fenoménů, které chceme zkoumat. V tomto procesu omezování Goldberg (1990) zohledňuje dva přístupy: (i) *`idiografický`*, tedy postup od jednotlivosti k celku; a naopak (ii) *`nomotetický`*, tedy od celku k částem. Oba tyto přístupy by se měly vzájemně doplňovat a existence jednotlivých psychologických deskriptorů by měla být ověřena ještě překlady v druhém jazyce, aby byla zajištěna univerzální platnost (Saucier et al., 2014).

První kompletní taxonomii pro psychologický výzkum osobnosti sestavil dle Hřebíčkové (1993) Francis Galton (1889), jenž uspořádal hierarchický model lingvistických termínů, které následně klasifikoval jako deskriptory osobnosti. To se stalo teoretickým paradigmatem pro další lexikální výzkumy osobnosti. V praxi potom psychologové procházeli jazykové slovníky a snažili se destilovat signifikantní termíny k popisu osobnosti. Allport a Odbert (1936) z anglického výkladového slovníku Merriam-Webster utřídili 18 000 přídavných jmen, podstatných jmen a sloves ve tvaru přičestí, které potom roztřídili do čtyřech kategorií podle povahy osobnostních rysů, které popisují.

Na jejich práci navázal Cattell (1945) a provedl výraznou sémantickou redukci původního seznamu deskriptorů, redukoval zejména počet synonym, které potom třídil do sémantických klusterů (významově příbuzných celků) a hledal k nim protilehlý sémantický shluk antonym (slov opačného významu), ze kterých vytřídil jednotlivé faktorové škály. Slovníkový seznam psychologických deskriptorů ale nebyl kompletní, proto musel být doplněn o klinické termíny z odborné literatury, a to především v oblasti psychopatologie. Sem mimo jiné spadaly i lexikální deskriptory popisující neurotické a psychotické poruchy. Na základě tohoto lexikálního seznamu potom Cattell vytvořil svůj šestnáctifaktorový model osobnosti 16PF (Hřebíčková, 1993).

Redukcí tohoto slovníku deskriptorů pokračovali i Norman (1963) a Goldberg (1990), až postupně dospěli k formulování nové teorie *`Pětifaktorového modelu`* osobnosti známé jako *`The Big Five Model`* neboli model Velké pětky (Hřebíčková, 2008), jejich redukce byla mimo jiné založena na eliminaci určitých slovních druhů a nebyla zaměřena pouze sémanticky. Výrazné omezení počtu psychologických deskriptorů přineslo i omezení na přídavná jména a vyloučení jiné gramatické kategorie (Hřebíčková et al.,

1993). Tento *‘Pětifaktorový model’* se stal mezinárodně uznávaným nástrojem v oblasti měření osobnostních rysů a byl standardizován v řadě jazykových lokalizací (Hřebíčková, 2008).

4.1.1 Lexikální podstata psychometrických nástrojů

Jazyková podstata psychometrických studií byla mnohdy zkoumána i kritizována (Ashton et al. 2004). Mnoho výzkumů vychází z Goldbergovy (1990) rozsáhlé taxonomie adjektiv a je převáděna do moderního prostředí, kde je ověřována pomocí moderních nástrojů AI a strojového učení. Pomocí hierarchických modelů založených na frekvenční analýze deskriptivních sémantických znaků (slov) je neustále ověřována platnost stávajícího modelu pětifaktorové dimenze osobnosti (Mereu, 2021).

V moderním pojetí jde spíše o predikci osobnostních rysů na základě automatické analýzy jazykového produktu autora pomocí AI (Yuan, 2018; Pramodh a Vijayalata, 2016). Nesmíme však zapomínat, že k predikci osobnosti nepoužívá jazykovou analýzu pouze AI, ale jazyk byl už v samém základu u zrodu osobnostních inventářů a řada psychometrických nástrojů je založena na jazyku.

Podle Hřebíčkové (2008, 1993) je třeba rozlišit *‘Pětifaktorový model’*, *‘The Big Five Model’*, který Goldberg (1990) sestavil na základě **lexikální analýzy přídavných jmen**, z něž vychází pětifaktorový osobnostní inventář **NEO-FFI**, který Costa a McCrae (1992) sestavili na složité taxonomii jazykově podmíněných vztahů

A inventář **BFI-2**, který má sice podobnou strukturu (Soto, John, 2017), ale stojí na jiném teoretickém základu, a sice „*Pětifaktorové teorii*“ *‘Big Five Theory’*, je ve své podstatě tedy méně determinován hierarchií jazyka a více strukturálně podmíněn faktorovými rysy osobnosti (John et al. 2008)

Tyto dva přístupy se liší zejména ve svém přesahu na nižší a vyšší úroveň hierarchie popisu osobnosti a v rozsahu, do kterého jsou v něm kompletně zahrnuty všechny důležité psychologické deskriptory, což se odráží ve formulaci algoritmů pro extrakci nejmenšího počtu faktorů za předpokladu, že jejich výčet se v těchto dvou teoriích liší (Hřebíčková, 2011).

NEO-FFI je založen na systému šesti facet, které detailněji popisují každou z pěti dimenzí osobnosti, u BFI-2 se každá z pěti základních dimenzí osobnosti diferencuje na tři subškály, z nichž každou sytí čtyři různé položky inventáře (Hřebíčková et al., 2020; Hřebíčková a Urbánek, 2021). Ačkoli jsou oba inventáře svým rozsahem, standardizací a pětifaktorovým přístupem srovnatelné, stále mezi nimi existují určité rozdíly.

Inventář NEO-FFI je od samého počátku orientován více na psychometrické kvality a možnosti kvantitativní analýzy psychologických dat (John, Srivastava, 1999), zatímco BFI-2 je orientován na strukturální popis osobnosti. V praktické rovině je BFI-2 i lépe srozumitelný, protože je založený na krátkých tvrzeních, se kterými se respondent může nebo nemusí identifikovat a není podmíněn složitou taxonomií popisných adjektiv, jako je tomu u NEO-FFI (Hřebíčková, 2011)

Jak vyplývá z dosud posbíraných faktů, bez jazyka by charakterové rysy osobnosti šly jen těžko zkoumat. Od samého počátku je totiž jazyk používán jednak k popisu osobnosti a jednak k výzkumu jejich kvalit (Galton, 1889). Již v roce 1884 zformuloval Galton lexikální teorii, ze které potom Allport a Odbert (1936) **Obecnou lexikální hypotézu**, která vycházela z předpokladu, že:

„Největší individuální rozdíly jsou uchovány v jazyce“ (Hřebíčková, 2004, s.15)

4.2 Lingvistická data na psychologickém pozadí

Výzkum jazykových prostředků spadá do oblasti lingvistiky, sémantiky a sémiologie. Díky propojení jazyka s myšlením (tedy psychologickým obsahem a pozorovatelným projevem myšlenkových procesů), a zároveň kulturním prostředím (tedy sociálním kontextem, morálními paradigmaty a zažitými vzorci chování), se jazyk jako takový stal cenným nástrojem ve výzkumu psychologie a sociologie.

Věda, která se zasloužila o měřitelnost jazykového projevu, se nazývá korpusová lingvistika. Zabývá se strukturou přirozeného jazyka a jeho převodem na měřitelné jednotky, které potom mohou sloužit k další analýze nejrůznějších aspektů jazyka samotného, nebo jeho přesahu do jiných oborů, jak jsou nejen sociologie nebo psychologie, ale třeba i pedagogika.

Korpusová lingvistika zkoumá jazyk pomocí analýzy velkých datových souborů, obecně známých jako jazykové korpusy, což jsou strukturované sbírky textů v elektronické podobě. Tyto texty mohou pocházet z různých zdrojů, jako jsou psané dokumenty, transkripce mluveného jazyka, online fóra, sociální média a další. Obecně korpusy slouží jako reprezentativní vzorky určitého jazyka nebo jazyků, poskytující poznatky o variabilitě, užití a změnách jazyka v různých kontextech a časových obdobích (McEnery, 2019).

Díky velkému množství textů využívají výzkumníci v korpusové lingvistice automatické zpracování dat a statistické postupy k analýze jazykových vzorců, což může zahrnovat zkoumání frekvencí slov, kolokací, syntaktických struktur, sémantických vztahů a dalších. Kvantitativní zpracování jazykových dat umožňuje lingvistům zkoumat variabilitu jazykových jevů systematickým a empiricky ověřitelným způsobem (Biber, Conrad, 1998).

Přestože se kvantitativní analýza textu rozvíjí přímo úměrně s možnostmi výpočetní techniky, pokusy o frekvenční analýzu lingvistických znaků v textu a jejich souvislosti s osobnostními rysy autora se objevovaly už v samých počátcích rozmachu korpusové lingvistiky, nejprve v kvalitativním (Hamilton, 1957) a později v čistě kvantitativním pojetí (Scherer, 1978).

S rostoucím rozsahem jazykových korpusů – Big Data, rozvojem AI, strojového učení a neuronových sítí roste využitelnost výzkumu přirozeného jazyka a možnost predikce chování člověka na základě jeho jazykového projevu (Hui, 2021). Dále se podíváme na využitelnost lingvistických nástrojů pro popis osobnostních charakteristik mluvčího či autora textu.

4.2.1 Psycholingvistické studie vztahu jazyka a osobnosti

V této kapitole si představíme základní principy a paradigmaty současného výzkumu na poli psycholingvistiky. V předchozích kapitolách jsme si popsali variabilitu osobnostních rysů, jejich proměnlivost i dynamické vztahy, v této kapitole se budeme soustředit na variabilitu jazykových prostředků.

Z hlediska psycholingvistiky je důležité upozornit na žánrovou nejednotnost jazykového projevu. I samotnému laikovi musí být jasné, že existuje rozdíl mezi verbálním slovním projevem a psaným textem. Typ textu má totiž rozhodující vliv na povahu komunikátu (Husser, 2001). Musíme tedy předpokládat, že jazykové výzkumy se zaměřují na žánrově příbuzné celky, které potom mohou porovnávat mezi sebou nebo proti sobě.

Výsledky prezentovaných výzkumů tedy nelze srovnávat plošně, ale vždy jen s ohledem na žánrovou konzistenci.

V současné době můžeme rozlišit mezi dvěma přístupy k výzkumu vztahu mezi využívanými jazykovými prostředky a osobností autora. První je označován za *otevřený* a je založený na dolování nebo těžbě lingvistických dat, tzv. *data mining*, pomocí počítačové analýzy tedy strojově vytěžíme lingvistická data bez předem definovaných jazykových kategorií (Hearst, 1999). Tato metoda je nepochybně účinnější a při sběru dat efektivnější, avšak značně technicky náročná.

Otevřený přístup k jazykové analýze umožňuje nalézat jazykové vzorce, o kterých jsme dopředu ani nemuseli tušit, že existují. Pomocí strojového učení a automatické analýzy obsahu jsme schopni z textu vydestilovat lexikální, sémantické i strukturální jednotky, které jsou pro daný text nebo celou databázi textů typické (Grimmer et al., 2022).

Celá řada takových studií ověřuje souvislost mezi jazykovým projevem a osobním postojem mluvčího v široké škále kontextů, např. na poli politickém (Grimmer and Stewart, 2013), sociologickém (Bail, 2014) nebo psychologickém (Baayen, 2017). Druhým typem metodologie při analýze lingvistických proměnných je takzvaný *uzavřený* přístup. V něm je rozhodující systematická příprava a kompletní definice žádoucích proměnných, které chceme v textu sledovat.

Pomocí *frekvenční analýzy* potom počítačovými metodami zaznamenáme počet výskytů žádoucích proměnných a můžeme s nimi potom dále kvantitativně nebo kvalitativně pracovat. Tímto přístupem se proslavil James W. Pennebaker se svým vlastním počítačovým softwarem LIWC⁷, který je schopen pracovat v různém jazykovém rozhraní a zaznamenává frekvenční analýzu předem určených lexikálních, sémantických a gramatických kategorií (Pennebaker, 2007; Tausczik, Pennebaker, 2010, Pennebaker et al., 2015).

Pennebaker vedl celou řadu psycholingvistických výzkumů týkajících se vztahů mezi osobnostními rysy nebo psychologickými stavy a užívaným jazykem. Mezi jeho nejznámější studie patří výzkum detekce lži pomocí lingvistické analýzy politických textů a projevů (Jordan a Pennebaker, 2017; Pennebaker a Graybeal 2001). Ve svých studiích nachází mimo jiné statisticky významné vztahy mezi poklesem frekvence užívání 1. osoby jednotného čísla u politických leaderů a plánovaným vstupem do válečného konfliktu (Jordan et al., 2019;). Obdobná studie byla provedena i v českém prostředí v souvislosti s epidemií COVID-19 a změnou veřejného diskurzu (Kučera, 2021).

Podobná závislost mezi frekvencí osobních zájmen a osobnostním nastavením autorů se projevila i v korelačních analýzách písemných esejí studentů pod stresovou zátěží a s prokazatelně depresivním prožíváním (Rude et al., 2004). Psychická zátěž se mimo jiné projevuje zvýšeným výskytem referenčních jazykových prostředků směřujících k vlastní osobě (Pennebaker, 2013) a poklesem schopnosti analytického myšlení (Tackman et al., 2019).

Z frekvenčních psycholingvistických výzkumů mimo jiné vyplývá, že největší rozdíly nemůžeme očekávat v kategorii slov se sémantickým obsahem, jako jsou podstatná a přídavná jména nebo významová slovesa, ale naopak v kategorii funkčních slov, jako jsou předložky, spojky, prostředky pro referenci – osobní a ukazovací zájmena, pomocná slovesa a příslovce). Tyto závěry se potvrdily v řadě různých jazykových kontextů a napříč jednotlivými žánry (Peabody a Raad, 2002; Oberlander a Gill, 2004; Pennebaker a King, 1999).

⁷ LIWC = *Linguistic Inquiry and Word Count*, linguistic software developed by J. W. Pennebaker (Pennebaker et al., 2001)

Pokud vezmeme v úvahu sociální proměnné, které by měly ovlivňovat způsob užívání jazykových prostředků ve vztahu k osobnosti, musíme vzít v úvahu pohlaví a věk autora. S rostoucím věkem mají mluvčí používat více plurálu, zejména osobních zájmen v množném čísle, a více budoucího času u sloves (Pennebaker a Stone, 2003). V textech starších jedinců se také objevuje emočně více nabitých slov (Scheler et al., 2006) a výrazně se snižuje užití záporu a negace. U starších také ubývá výskytu referenčních výrazů (osobních a ukazovacích zájmen), které referují na osobnost autora, tedy kterými autor poukazuje sám na sebe (Chung a Pennebaker, 2007).

V případě pohlaví se prokázal statisticky významný rozdíl v užití 1. osoby singuláru, která je výrazně častější u žen než u mužů (Pennebaker a Stone, 2003, Pennebaker, 2013). Ženy dále užívají více referencí na sociální interakce (Argamon et al., 2007). Muži, vzhledem k reakcím na sociálních sítích a internetu, používají více determináčnických výrazů (členů určitých a neurčitých) a více formální styl vyjadřování (Mukherjee a Liu, 2010; Rao et al., 2010).

Pennebaker jako jeden z vůdčích osobností na poli psycholingvistického výzkumu opakovaně zdůrazňuje, že studium plnovýznamových slov se sémantickým obsahem (např. hustý les, běžet, politické uskupení) není pro frekvenční studium jazyka příliš vhodné a často nevede k plnohodnotným výsledkům, protože jejich strukturální kategorie jsou příliš otevřené a těžko kvantifikovatelné. Pro frekvenční psycholingvistickou analýzu se hodí slova jazykově funkční, která nenesou sémantický obsah, ale jejich smysl je v utváření jazykové struktury a obsahového smyslu textu pomocí kohese a koherence (Chung a Pennebaker, 2011).

4.2.2 Extraverze v psycholingvistickém výzkumu

V současnosti nejčastěji používaným psychometrickým nástrojem pro korelaci psychologických proměnných měřících strukturu osobnosti s lingvistickými daty je právě sebeposuzovací inventář NEO-BFI-2 (Kučera, 2022). Ze všech pěti faktorů je právě Extraverze nejčastěji předmětem psycholingvistických výzkumů (Funder, 1995).

Kučera (2020) uvádí pozitivní vztah Extraverze s častějším výskytem emočně pozitivně nabitých slov (např. překrásný, úžasný). Tato studie byla provedena v anglickém jazyce metodou uzavřeného přístupu k lingvistické extrakci dat právě za pomoci psycholingvistického softwaru LIWC (Pennebaker et al., 2001), zmiňovaného v předchozí kapitole 4.2.1. Autoři uvedené analýzy potvrzují v pozitivním směru statisticky významné korelace mezi vysokou mírou Extraverze a kategorií adjektiv s pozitivně emočním nábojem (Pennebaker, King, 1999).

K obdobným výsledkům se dostal výzkumný tým Sherer et al. (2013) z opačné strany, tedy pomocí otevřeného přístupu *‘open vocabulary’*, kdy pomocí logaritmu počítačový program sám definuje signifikantní nebo rekurentní lingvistické vzorce. Zvýšený počet pozitivních slov také uvádí Yarkoni (2010) ve své analýze veřejně publikovaných textů bloggerů.

Holtgraves (2011) uvádí statisticky významný pozitivní vztah mezi vysokou mírou Extraverze a užíváním delších slov a akronymů (tedy zkratk složených z počátečních písmen slov v názvech, které se nemusí hláskovat, ale většinou se čtou jako běžné slovo, např. NATO).

V jiných psycholingvistických studiích se objevily psychologické dimenze, které sice nebyly zjištěné pomocí inventáře BFI-2, ale svým obsahem odpovídají osobnostnímu rysu Extraverze tak, jak je definován v kapitole 3.1.4.3. Cegala (1989) popisuje osobnostní dimenzi *‘sociální angažovanost’*, která se projevuje interpersonální orientací a vysokou

měrou sociálního kontaktu. V pojetí Velké pětky tedy hrubě odpovídá osobnostnímu rysu Extraverze. Během výzkumu bylo zjištěno, že vysoce sociálně angažovaní jedinci ve svém jazykovém projevu používají více vztažných zájmen (*jenž, který*) a méně 1. osoby singuláru (Cegala, 1989). Naopak zvýšená frekvence výskytu 1. osoby singuláru pozitivně koreluje s anxiétou (Pennebaker a Stone, 2003)

Jiné jazykové prostředky se projeví u ostatních dimenzí Velké pětky, uvedeme je zde pouze pro případ srovnání s výsledky vlastní datové analýzy prezentované v kapitolách 8.2-8.4. Škála Neuroticismu pozitivně koreluje s 1. osobou singuláru u slovního druhu zájmen (*já, můj, mému*) a negativně s četností užití emočně pozitivně nabitých slov (*výborně, úžasný*) (Pennebaker a King, 1999). Nárůst výskytu 1. osoby singuláru (*já, můj, mně*) zároveň pozitivně koreluje i se škálou Přívětivost, stejně jako nárůst výskytu určitých a neurčitých členů (Summer et al, 2011). S osobnostní dimenzí Otevřenost vůči zkušenosti roste počet referencí na sociální procesy a počet citací a intertextových odkazů v textech autorů (Summer et al., 2011).

4.2.3 Srovnatelné studie v českém prostředí

Většina studií prezentovaných v kapitolách 4.2.1 - 4.2.3 je provedena na frekvenční analýze anglického jazyka. Pro srovnání s češtinou musíme vzít v úvahu strukturální a systemické rozdíly v obou jazycích. Zatímco angličtina má pevný slovosled a fixní syntaktickou větnou stavbu s jasně určeným pořadím větných členů, čeština je ve svém slovosledu flexibilní, řídí se spíše funkčním řazením větných členů a tuto volnost jí umožňuje systém ohýbání slovních druhů formou: (i) deklinace, skloňování podstatných, přídavných jmen a zájmen; a (ii) konjugace, časování u sloves (Hajič, 2004; Smolík, 2009).

Veškeré komparativní jazykové studie srovnávající jazykové prostředky mezi anglickým a českým jazykem musí vycházet z těchto strukturálních rozdílů v samém základu uvedených jazykových systémů (Vachek, 1976; Sgall, 1985). Tyto rozdíly znamenají určitá omezení – ne ve smyslu, že by v angličtině nemohl být vyjádřen stejný obsah jako v češtině, ale za účelem vyjádření stejného sémantického významu oba jazyky používají různé strukturální, syntaktické a gramatické prostředky (Chamonikolasová, 2005).

Pro příklad se můžeme podívat na funkci větného členu v angličtině (*a/an, the*). Každý začátečník ví, že určitý člen v angličtině označuje věci známé a neurčitý člen věci nové, neznámé. Tato poučka je pro mluvčí českého jazyka důležitá, ale nedostačující, protože v českém jazykovém systému tato mluvnická kategorie určitého a neurčitého členu není vůbec přítomná. Mimo jiné i z toho důvodu mluvčí češtiny, kromě předložek, v určitých a neurčitých členech v angličtině nejčastěji chybují (Luba, 2018).

Pro vyjádření reference na předměty známé (v angličtině *the*) a neznámé (v angličtině *a/an*) používají rodilí mluvčí českého jazyka právě slovosled. Umístěním slov na začátek věty označují referenci na věci v kontextu již známé nebo přítomné (*Na okně je kniha*) versus umístěním na konec věty referují na věci nové (*Kniha je na okně*). U prvního příkladu je známé okno a větná funkční perspektiva poukazuje na slovo *kniha* jako na nositele nové informace. U druhého příkladu je situace opačná, větná struktura naznačuje, že slovo *kniha* je v kontextu již známá informace, za novou informací větná funkční perspektiva označuje místo, kde se *kniha* nachází – tedy *na okně* (Firbas, 1964).

Podobná situace se vyskytuje v oblasti slovesného vidu (*dělat – udělat*), což je kategorie, která v angličtině není přítomna ve formě morfologického tvaru slovesa, ale je

vyjadřována systémem slovesných časů participia, tedy ve formě času předminulého a předpřítomného, jež zase naopak nenajdeme v češtině (Schmiedtová, 2003). Podobná situace je v gramatické kategorii přivlastňovacích a zvratných zájmen, jejichž užití se v češtině a angličtině značně odlišuje (Veselovská, 2014).

Aniž bychom zabíhali příliš hluboko do strukturálních rozdílů mezi češtinou a angličtinou, je nutné je zde zmínit, protože vstupují do hry během analýzy lingvistických dat v praktické části této práce (konkrétně v kapitole 7.3.2).

Pokud bychom měli zmínit několik psycholingvistických studií v českém prostředí, které jsou pro naši práci významné, museli bychom začít prací Havigerové a Kučery (2014), která shrnuje možnosti počítačnické lingvistiky v psychologickém výzkumu. Kučera, Havigerová (2015) se zabývali kvantitativní psycholingvistickou analýzou komunikátů v českém prostředí a postupně vybudovali pevnou metodologii pro frekvenční analýzu jazykových prostředků a jejich následnou korelaci s psychologickými daty.

Pomocí frekvenční analýzy jazykových prostředků byli Havigerová et al. (2019) schopni předpovědět riziko depresivity u svých respondentů. Kučera et al. (2020) zaměřili pozornost na stylistickou diferenciaci textů pomocí kvantitativní analýzy psycholingvistických dat. Tyto studie představují rozsáhlou platformu, která poskytuje solidní základ pro další výzkumy, zejména v oblasti kombinace frekvenční jazykové analýzy a psychometrických nástrojů BFI-2 (Kučera et al., 2022).

Praktická část této práce se opírá o rozsáhlou psycholingvistickou studii Dalibora Kučery (2022), který pomocí frekvenční analýzy jazykových znaků zkoumal souvislost mezi osobností autora (měřenou pomocí sebesposuzovací škály inventáře BFI-2 a zároveň pomocí posouzení druhou osobou) a užitými jazykovými prostředky v žánrově odlišných textech (formálních a neformálních dopisech).

Ve své studii došel Kučera (2022) mimo jiné k závěru, že největší vliv na užití jazykových prostředků má právě žánr zkoumaných textů (s.112). Další zajímavou proměnnou bylo pohlaví respondentů. Rozdíly v užití jazykových prostředků mezi muži a ženami byly statisticky významnější než jejich korelace s osobnostními rysy (Kučera, 2022, s. 164).

Za nejvýznamnější indikátor (*jazykový marker*) pohlaví Kučera (2022) považuje častější užívání 1. osoby a vyšší frekvenci výskytu sloves u žen, které zároveň vykazují vyšší dynamičnost textu a potvrzují tendence k častějšímu užívání přítomného času a negace, toto bylo prokázáno i v zahraničních studiích (Biber, 1991; Pennebaker a Stone, 2002; Argamon et al., 2007 a Kučera, 2022).

Muži oproti tomu vykazují nižší dynamiku textu s nižším podílem sloves a vzrůstající frekvenci podstatných jmen. Tento trend se nejvýrazněji projevuje na škále svědomitosti (Kučera, 2022, s. 103-104). U osobnostního rysu nejistoty se však muži projevují opačně. U žen se vzrůstajícím rysem svědomitosti klesá frekvence užití slovního druhu částic a s osobnostními rysy bezelstnosti a skromnosti naopak stoupá počet adjektiv a klesá užití 2. osoby singuláru (Kučera 2022).

Pokud se zaměříme na korelační vztahy osobnostních rysů, potvrzuje Kučera (2022) následující vztahy, které odpovídají zahraničním výzkumům (Pennebaker a King, 1999; Schwartz et al., 2013). S rostoucí svědomitostí narůstá počet afirmací, předložek a spojek. S nižší mírou přívětivosti vzrůstá počet negací a mluvnického záporu (Schwartz et al., 2013). Emočně labilní jedinci používají méně podstatných jmen (Du, 2023). Ostatní vztahy (Lee et al., 2007) v Kučerově studii nebyly statisticky významné.

Ve své studii formálních a neformálních dopisů došel Kučera (2022) k závěru, že nejdůležitějšími jazykovými znaky, které je dobré sledovat pro pozdější korelační analýzu s osobnostními rysy jsou tyto gramatické kategorie: osoba (1., 2., a 3.), stupeň (pozitiv,

komparativ a superlativ), spojky (souřadící a podřadící), slovesa, přivlastňovací způsob (zájmenný i jmenný) a interpunkce.

Kučera (2022) zjistil, že 1. osoba singuláru pozitivně koreluje s loajalitou mluvčího, 1. osoba plurálu se častěji vyskytuje u osob s vysokou sebedůvěrou a emoční stabilitou, naopak méně u osob pesimistických s nízkou měrou přívětivosti. Přídavná jména v základním tvaru se ve formálních textech objevují častěji u osob s vysokou měrou kontroly a sociální jistoty a méně u osob loajálních a podřízených. Komparativ je v textech formálních užíván častěji u osob sociálně zdatných s vysokou měrou sebekontroly, avšak pozitivně koreluje s nižší měrou ambiciosnosti v neformálních textech. Superlativ je užíván jedinci s nižší měrou sociální intuice a vyšší měrou optimismu.

Vyšší počet spojek je spojován s nižší měrou sociální kontroly i sebekontroly a zároveň nižší měrou interpersonálních dovedností, naopak pozitivně koreluje s vyšší měrou kritičnosti a nižší měrou přátelskosti v neformálních textech (Kučera 2022).

Zvýšená frekvence výskytu významových sloves je spojována s nižší měrou sociální kontroly a sebekontroly, menší měrou emoční stability a horšími interpersonálními schopnostmi. Slovesa také pozitivně korelují s osobnostní škálou rezervovanosti (Kučera 2022).

Gramatická kategorie mluvnického čísla v jednotném tvaru přivlastňovacích zájmen je spojována s vyšší měrou sebedůvěry a sociální jistoty na straně mluvčího, jedinci sociálně upozaděni referují na kolektivní formu vlastnictví jen zřídka a v malé míře (Kučera, 2022). Užití interpunkce negativně koreluje s pocity strachu a nejistoty.

Je důležité zde zmínit, že ve srovnatelných českých výzkumech (Kučera et al., 2020; Kučera, 2022 a Kučera et al. 2022) je opakovaně upozorňováno na potřebu zkoumat spíše funkční slovní druhy, jako jsou předložky, spojky, zájmena, která vykazují statisticky významné odchylky ve frekvenci užití při korelaci s psychologickými daty, než slova s plnohodnotným sémantickým obsahem, jako jsou slovesa, podstatná a přídavná jména, která ve frekvenčních analýzách nedosahují statisticky významných výsledků. Obecně však je nutné očekávat, že korelační koeficienty budou vykazovat spíše slabé hodnoty a nízkou míru závislosti mezi psychologickými proměnnými a lingvistickými jevy (Yarkoni, 2010; Kučera et al. 2022).

VÝZKUMNÁ ČÁST

5 VÝZKUMNÁ OTÁZKA

Na základě dostupných teorií a praktických výsledků předchozích psycholingvistických studií shrnutých v teoretické části práce jsme došli k závěru, že vztah mezi osobnostními rysy autora a způsobem užívání jazykových prostředků v jeho písemném projevu bude existovat.

Jak napovídá samotný název práce „*Psycholingvistická studie souvislosti mezi textovými znaky a osobnostními skóry autora na škále Extraverze*“ je cílem této práce definovat, jakým způsobem spolu osobnostní rysy a užití jazykové prostředky souvisí. Vzhledem k omezenému rozsahu předkládané práce se v praktickém výzkumu zaměříme pouze na jediný osobnostní rys, a to **Extraverzi**. V průběhu výzkumu jsme si pokládali následující otázky:

- Do jaké míry souvisí jazykový projev autora s jeho pohlavím nebo dokonce s mírou jeho Extraverze?
- Potvrdí se v našem souboru výsledky předchozích studií?
- Jeví se výzkum osobnosti skrze rozbor jazykového projevu jako teoretický možný a pokud ano, tak za jakých podmínek?
- Lze pomocí rozboru písemného textu predikovat osobnostní rys jejího autora?

Obecně bychom tedy výzkumnou otázku mohli definovat takto:

Tabulka 2: Výzkumná otázka

Existuje souvislost mezi mírou Extraverze a způsobem užití jazykových prostředků v písemném projevu?

Korelační analýza si klade za úkol rozpoznat stěžejní psychologické proměnné, které souvisí s frekvencí užití sledovaných jazykových prostředků, a popsat pozorované odchylky v jazykovém projevu (i) **v kontextu strukturálních možností češtiny** a (ii) **ve vztahu k psychologické podstatě sledovaného osobnostního rysu**.

Kvantitativní rozbor dat bude nasedat na výsledky předchozích studií, které jsou shrnuty v teoretické části práce (zejména v kapitolách 4.2.1 - 4.2.3.). Na základě sledovaných výsledků se dá náš výzkumný záměr rozdělit na 4 dílčí oblasti:

- **Výzkumný záměr 1:** Ověřit, zda se relativní četnost sledovaných lingvistických jevů mění v závislosti na pohlaví respondentů.
- **Výzkumný záměr 2:** Ověřit, zda se relativní četnost sledovaných lingvistických jevů mění v závislosti na míře Extraverze jednotlivých autorů textů.
- **Výzkumný záměr 3:** Ověřit míru účinku, zda na četnost užití sledovaných lingvistických jevů má silnější efekt pohlaví respondentů nebo jejich skóre na škále Extraverze.
- **Výzkumný záměr 4:** Ověřit, zda se v našem souboru potvrdí výsledky předchozích psycholingvistických studií.

Výzkumný záměr by však měl by poskytovat i určitý přesah do praktické či teoretické roviny (Willing, 2009). Kučera (2020, s. 117) podotýká:

„Jistě, sledování, popis, analýza a interpretace jsou klíčovými znaky většiny projektů základního výzkumu. Pokud však nemá být takový výzkum samoučelný, měl by nabízet i určitou přidanou hodnotu, praktický smysl, nebo přímo potenciál k využití v aplikační rovině.

Takový přesah v naší práci představuje teoretická otázka, do jaké míry je možné zkoumat jazykové prostředky ve vztahu k osobnostním rysům kódovaným právě pomocí jazykových prostředků. Nesnažíme se tady o nemožné? Netočíme se v bludném kruhu? Nebo se jen snažíme sami sebe zvednout za vlastní tkaničky?

Tyto a jiné otázky jsou bez základního výzkumu nezodpověditelné. Proto se budeme v této práci kromě frekvenční analýzy konkrétních jazykových rysů a jejich souvislostí s mírou Extraverze dotýkat i dalších otázek. A sice (i) **nakolik jazyk samotný vypovídá o struktuře a povaze lidské osobnosti**; a (ii) **nakolik jsou naše osobnostní rysy vnímány skrze konkrétní formu a strukturu užívaného jazyka**.

Ambicí této práce není zobecňovat nalezené vztahy směrem k populaci, ani zodpovědět si na tyto komplexní otázky v plné jejich šíři, ale začít si klást související otázky a nebát se zkoumat vztahy a souvislosti i tam, kdy se dostáváme do neprobádaných míst, protože zatím neexistuje konceptuálně vytvořená mapa, která by nás snadno dovedla k cíli.

6 TYP VÝZKUMU A POUŽITÉ METODY

Psycholingvistický výzkum se pohybuje na rozhraní lingvistické a psychologické metodologie. Pro zvolený výzkumný problém jsme zvolili **kvantitativní přístup zpracování dat pomocí korelační analýzy a strojového zpracování přirozeného jazyka**. V následujícím přehledu si představíme konkrétní metodologické principy aplikované v této práci, z hlediska metodologie však můžeme tuto práci označit za **komputační psycholingvistickou analýzu** (Kučera, 2020).

6.1 Metoda zpracování přirozeného jazyka

Pro zpracování lingvistických dat jsme zvolili metodu zpracování přirozeného jazyka, tzv. „*Natural Language Processing*“ (NLP)⁸, což je obor lingvistiky spojený s umělou inteligencí, který se zaměřuje na interakci mezi počítači a lidským jazykem. NLP zahrnuje vývoj algoritmů a technik umožňujících počítačům rozpoznávat, analyzovat, interpretovat a generovat lidský jazyk způsobem, který je pro lidi srozumitelný, smysluplný a užitečný (Sag, 1991; Straka et al., 2016).

Techniky NLP často spoléhají na strojové učení a statistické modely určené k analýze nestrukturovaných dat přirozeného jazyka. Tyto modely jsou trénovány na velkých souborech anotovaného textu, což jim umožňuje naučit se vzory a vztahy v jazyce. Navíc NLP systémy mohou zahrnovat lingvistická pravidla, gramatické kategorie a znalosti o syntaktické, morfologické i sémantické stavbě jazyka (Eisenstein, 2019). Pro zlepšení přesnosti a výkonnosti NLP se pracuje s trénovanými soubory dat, kde počítač zlepšuje svůj výkon na základě feedbacku lidských posuzovatelů, to přináší více než uspokojivé výsledky na poli nejen kategorizace textu, ale i predikce osobnostních rysů autora (Mereu, 2021; Jelínek, 2018; Hui, 2021).

V této práci jsme se omežili na **strojové zpracování přirozeného textu** pomocí nástroje UDpipe (viz kapitola 7.3.1), což obnáší automatické rozdělení textu na menší části, konkrétně na slova a věty, a extrakci konkrétních jazykových jevů nebo gramatických kategorií, které jsou v algoritmu softwaru přímo určeny (Straka et al., 2016).

Použitou metodologií pro zpracování lingvistických dat tedy můžeme označit za **frekvenční analýzu morfologických znaků**, která přináší informace o četnosti výskytu jednotlivých jazykových jevů s ohledem na jejich výskyt v konkrétní větě určitého textu daného autora (Kučera 2020).

6.2 Komparativní a korelační psycholingvistická analýza

Komparativní a korelační analýza patří k nejčastějšímu typu kvantitativní metodologie v psychologickém výzkumu (Frejencík, 2008). V první řadě jde o **komparativní analýzu** vztahu mezi dvěma **nezávislými výběry**: (i) **pohlavím** respondentů a (ii) **frekvencí výskytu určitého jazykového jevu**, tato metodologie je spojena s výzkumným záměrem 1

⁸ Metoda Natural Language Processing je blíže popsána v publikaci Eisenstein (2019).

(kapitola 8.2). Cílem komparativní analýzy je porovnání středních hodnot obou výběrů kardinálních proměnných.

V druhé řadě pak jde o **korelační analýzu dvou závislých kardinálních proměnných (i) frekvencí výskytu určitého jazykového jevu, a (ii) skóre osobnostního rysu Extraverze na sebezposuzovací škále autora**, tento metodologický postup aplikujeme při popisu vztahů popsanych ve výzkumném záměru 2 a 4 (kapitola 8.3 a 8.5). Cílem korelační studie je posoudit, zda existuje statisticky významný vztah mezi proměnnými a jaký je jeho charakter (pozitivní nebo negativní směr vztahu).

Je důležité zde uvést, že korelace neimplikuje příčinný vztah mezi proměnnými a nijak nevysvětluje povahu jejich vztahu. Pouze ukazuje, že proměnné spolu nějakým způsobem souvisejí – tedy jsou určitým způsobem na sobě závislé, a to buď v kladném nebo záporném směru, ale od korelační studie v žádném případě nemůžeme očekávat definici kauzality (Dostál, 2016). Pro potvrzení příčinného vztahu je třeba dalšího výzkumu, například longitudinální studie nebo experimentálního designu.

V našem statistickém souboru jde pouze o **lineární korelace**, které měří sílu a směr vztahu mezi dvěma spojitými proměnnými. Pokud existuje lineární vztah mezi dvěma proměnnými, znamená to, že jejich hodnoty se mění ve společném směru a v konzistentním poměru. Korelace se vyjadřuje pomocí koeficientu korelace, který nabývá hodnoty mezi -1 a 1a pro zjištění, zda je statisticky významný, je nutné provést **test statistické významnosti**.

Pro zvolení správných testů **komparativní analýzy** bylo nutné ověřit pomocí Shapiro – Wilkova testu normalitu rozdělení kardinálních proměnných. V případě splnění normality byl v komparativní analýze použitý parametrický t-test pro dva nezávislé výběry; v případě nesplnění normality byl použit neparametrický Man-Whitneyův U-test.

Korelační analýza byla z důvodu nesplnění normálního rozdělení zkoumaných proměnných popsána pomocí neparametrických Spearmanových korelačních koeficientů a jeho testu významnosti.

Kromě statistické významnosti jsme zjišťovali ještě **praktickou významnost** ve výzkumném záměru 3 (kapitola 8.4) pomocí ověření míry účinku u vztahů, u kterých se statistická významnost prokázala ($p < 0,05$) zároveň ve vztahu k pohlaví a k míře Extraverze. A to za účelem ověření, která ze sledovaných proměnných má vyšší míru účinku. Ukazatel míry účinku pro t-test dvou nezávislých výběrů je Cohenovo d , pro Man-Whitney U test je to η^2 a pro korelační koeficient je mírou účinku r^2 . Jednotlivé ukazatele míry účinku jsme pomocí kalkulátoru⁹ převedli na Cohenovo d tak, abychom jednotlivé míry účinku mezi sebou mohli srovnávat (Soukup, 2013)

Cohenovo d slouží k porovnání rozdílů mezi skupinami, protože bere v úvahu variabilitu v datech a standardizuje rozdíl mezi skupinami na jednotky směrodatné odchylky. Tím umožňuje porovnání efektů napříč různými soubory a měřeními. V našem datovém souboru jsme ho využili k ověření, zda má na frekvenci užití jazykových prostředků větší efekt pohlaví respondenta nebo jeho skóre na škále Extraverze měřené pomocí inventáře BFI-2.

6.2.1 Inventář BFI-2

Inventář BFI-2 jsme v teoretické části práce již mnohokrát zmiňovali (viz kapitoly 3.1.4.2, 3.2.4 a 4.1.1), proto se zde omezíme na stručný přehled jeho psychometrických kvalit. BFI-2 pracuje s 60-ti položkami vrstvicími se po 12-ti do pěti dimenzí osobnosti, z nichž

⁹ Odkaz na psychometrický kalkulátor: https://www.psychometrica.de/effect_size.html [citováno z 30.3.2024]

každá dimenze obsahuje 3 subškály. Pro znázornění podoby inventáře u osobnostního rysu Extraverze slouží Tabulka 3. Respondenti odpovídají na otázku: *Považuji se za někoho, kdo....* a skórují svoje odpovědi na škále likertova typu od 1 - zcela nesouhlasím, do 5 - zcela souhlasím (Hřebíčková et al., 2020).

Tabulka 3: Přehled položek osobnostního rysu Extraverze v inventáři BFI-2

škála	Subškály a jejich položky	popis škály a ukázka položek	Cronbachova α (koeficient reliability)
Extraverze	<p>Sociabilita (1, 16R, 31R, 46)</p> <p>Asertivita (6, 21, 36R, 51R)</p> <p>Energičnost (11R, 26R, 41, 56)</p>	<p>Vysoko skórující jedinec je aktivní, komunikativní, sebejistý a společenský, orientovaný na sociální interakci, je optimisticky laděný a přístupný</p> <p>+ ...je společenský, družný</p> <p>- ... bývá tichý</p>	0,86

Zdroj: Přejato a upraveno z Hřebíčková et al., (2020)

BFI-2 patří dnes k nejvyužívanějším nástrojům pro typologii a výzkum struktury osobnosti (Nakonečný, 1995), vzhledem k jeho standardizaci pro české prostředí, vysokou reliabilitu a lehkou zpracovatelnost se stal nástrojem ke sběru psychometrických dat v praktické části této práce. Hřebíčková et al. (2020) uvádí, že ve srovnání s průměrnou hodnotou reliability $\alpha = 0,76$ u pěti dimenzí NEO-FFI je průměrná vnitřní konzistence pěti škál BFI-2 dokonce vyšší ($\alpha = 0,85$).

6.3 Formulace hypotéz ke statistickému testování

Obecně formulovaná výzkumná otázka této práce zní: nakolik pohlaví respondenta a míra jeho Extraverze souvisí s četností užití sledovaných lingvistických jevů v jeho písemném projevu.

V kapitole 5 jsme definovali tři dílčí oblasti výzkumného záměru, ke kterým bude směřovat srovnávací a korelační analýza. Nyní přistoupíme ke stanovení alternativních hypotéz pro statistický rozbor dat.

- Výzkumný záměr 1 si klade za cíl pomocí **komparativní analýzy** ověřit, zda se relativní četnost sledovaných lingvistických jevů mění v závislosti na pohlaví respondentů, stanovené hypotézy jsou:

H1a: Existuje rozdíl v četnosti užívání slovních druhů mezi muži a ženami.

H1b: Existuje rozdíl v četnosti užívání sledovaných lingvistických jevů mezi muži a ženami.

- Výzkumný záměr 2 si klade za cíl pomocí **korelačních vztahů** ověřit, zda se relativní četnost sledovaných lingvistických jevů mění v závislosti na míře Extraverze jednotlivých autorů textů, stanovené hypotézy jsou:

H2a: Existuje statistická závislost mezi četností užívání slovních druhů a mírou Extraverze.

H2b: Existuje statistická závislost mezi četností užívání sledovaných lingvistických jevů a mírou Extraverze.

- Výzkumný záměr 3 si klade za cíl ověřit **míru účinku**, zda na četnost užití sledovaných lingvistických jevů má silnější vliv pohlaví respondentů nebo jejich skóre na škále Extraverze, stanovená hypotéza:

H3: Pohlaví respondentů má ve srovnání s mírou Extraverze vyšší míru účinku na četnost užívání sledovaných lingvistických jevů.

- Výzkumný záměr 4 si klade za cíl pomocí korelačních vztahů ověřit, zda se v našem souboru potvrdí **výsledky předchozích psycholingvistických studií** (Kučera, 2020 a 2022; Pennebaker a Stone, 2003; Pennebaker a King, 1999; Yarkoni, 2010; Lee et al., 2007; Cegala, 1989; Summer et al. 2011; Argamon et al., 2007; Rao et al. 2010; Mukherje a Liu, 2010; Holtgraves, 2011; Johnson-Laird, Byrne, 2002).

H4a: Jedinci skórující vysoko na škále Extraverze používají méně záporu v písemném projevu.

H4b: Jedinci skórující vysoko na škále Extraverze používají komplexnější delší a rozvitější souvětí.

H4c: Jedinci skórující vysoko na škále Extraverze používají méně trpného rodu.

H4d: Jedinci skórující nízko na škále Extraverze používají více podmíňovacího způsobu

H4e: Jedinci skórující nízko na škále Extraverze používají více komparativu, tedy 2. stupně přídavných jmen a příslovčí

H4f: Jedinci skórující vysoko na škále Extraverze používají více osobních zájmen v 1.os. singuláru.

V následujících kapitolách se budeme věnovat způsobu zpracování dat a konkrétním statistickým postupům použitým v této práci za účelem ověření platnosti předložených hypotéz.

7 SBĚR DAT A VÝZKUMNÝ SOUBOR

Tato kapitola se zaměří na popis datového souboru a proces získávání lingvistických a psychologických proměnných. Za základní datovou matici vděčím PhDr. Danielu Dostálovi, Ph.D., vedoucímu své práce, který materiál posbíral a ochotně poskytl k dalšímu statistickému zpracování. Sběr dat byl prováděn pro potřeby rozsáhlejší studie z katedry Obecné lingvistiky FF UPOL za účelem výzkumu doc. Mgr. Dana Faltýnka, Ph.D., a Mgr. Vladimíra Matlacha, Ph.D., kteří pomocí strojového učení predikují hlavní sémantické klustery s přesahem k možnosti predikce autorství (Faltýnek, Matlach, 2023). Za účelem přehlednosti bude v této kapitole zvlášť popsána akvizice lingvistické a psychologické části datového souboru.

7.1 Popis souboru

Našeho výzkumu se zúčastnilo celkem 357 respondentů, tedy 292 žen a 65 mužů. Všichni z 1. ročníku studia Psychologie na FF UPOL. K jejich věkovému rozložení nemáme dostupné informace, dá se ale předpokládat, že kromě studentů kombinovaného studia (kterých je minimum) se jedná o jednotnou věkovou kohortu a sociálně vyhraněnou skupinu, budoucí psychology, což je typologicky odlišuje od běžného průměru populace.

Veškeré zde publikované výsledky tedy nemůžeme zobecnit a musíme brát v úvahu, že výpočty jsou prováděny pouze na malém výběrovém souboru a značně specifickém výseku populace, v žádném případě z nich tedy není možné vyvozovat univerzálně platné závěry.

Data jsme zařadili do datové matice a vyřadili respondenty s neúplnými nebo chybějícími údaji nebo neúplnými daty. Do konečné analýzy bylo zařazeno celkem **59 mužů** a **277 žen**, celkový počet analyzovaných autorů, u kterých jsme sledovali vztah mezi mírou jeho Extraverze a užitými jazykovými prostředky byl **336**.

7.2 Psychologická data

Potřebné psychologické proměnné byly získány vyhodnocením psychometrického inventáře **BFI-2** (popsaného v kapitole 6.2.1) od 336 respondentů (59 mužů a 277 žen). Z pětifaktorového inventáře jsme se soustředili pouze na škálu Extraverze, včetně jejích třech subškál – Sociabilita, Asertivita a Energičnost.

Odpovědi respondentů byly skórovány na pětistupňové škále likertova typu, reverzní položky byly skórovány v opačném směru dle manuálu Hřebíčkové et al. (2020). Výsledky hrubých skórů byly kvůli možnosti další interpretace převedeny na **T-skóry** tak, aby byly dále srovnatelné s populačním průměrem a výsledky ostatních studií.

Ani na škále Extraverze, ani u jejích subškál neměly výsledky skórů normální rozdělení, proto jsme dále pracovali s neparametrickým testem pro dva nezávislé výběry, a to Man-Whitney U testem. Zjišťovali jsme, zda existuje statisticky významný rozdíl mezi dvěma nezávislými skupinami (muži a ženy) v hodnotách jedné spojité proměnné (skóre na škále Extraverze, případně na jejích subškálách).

Tabulka 4: Rozdíl v míře Extraverze (včetně jejích subškál) v podsouboru mužů a žen

	Normalita obou výběrů	Zvolený test	Medián muži	Medián ženy	p-hodnota
Extraverze	Ne	M-W test	52,68	51,86	0,985
Energičnost	Ne	M-W test	53,87	56,42	0,850
Asertivita	Ne	M-W test	52,47	51,76	0,185
Sociabilita	Ne	M-W test	50,73	50,40	0,940

Z výsledků uvedených v Tabulce 4 vyplývá, že p-hodnota ani v jednom osobnostním rysu neklesá pod statisticky významnou hladinu 0,05, **statisticky významný rozdíl mezi muži a ženami v míře Extraverze se tedy nepotvrdil.**

Dále ověříme schopnost testu rozlišovat mezi dvěma skupinami pomocí hodnoty AUC = 0,81 u Extraverze (u jednotlivých subškál se AUC pohybuje od 0,72 do 0,81). Vyšší hodnota AUC naznačuje lepší diskriminační schopnost testu. Hodnota AUC mezi 0,7 a 0,9 naznačuje, že test má slušnou diskriminační schopnost a má schopnost rozlišit mezi skupinami.

Hodnota testové statistiky pro náš soubor se pohybuje v rozmezí od $Z = -0,08$ do $Z = 1,3$. Použijeme ji k určení statistické významnosti rozdílu mezi dvěma nezávislými soubory (muži a ženy) v hodnotách t-skóru na škále Extraverze coby jedné spojité proměnné. Ani v jedné subškále, ani v celkovém skóre Extraverze se nepotvrdil statisticky významný rozdíl mezi muži a ženami. V další části práce budeme tedy vycházet z předpokladu, že neexistuje žádný statisticky významný rozdíl v míře Extraverze mezi muži a ženami v našem souboru.

7.3 Lingvistická data

Po vyčištění dat bylo celkem statisticky zkoumáno 613 031 slov ze 36 973 vět v 336 textech od 336 různých autorů (z toho 277 žen a 59 mužů), viz Tabulka 5. Statisticky se tedy jedná o nereprezentativní vzorek, a i z hlediska korpusové lingvistiky jde o tak zanedbatelný datový soubor, že z něj nelze vyvozovat obecně platné závěry. Pro statistické účely byly psané texty v datové matici anonymně spárovány přes ID autora s výsledky jeho osobnostního inventáře BFI-2.

Tabulka 5: Skladba lingvistického datového souboru

Rozsah lingvistického korpusu	Minimální	Maximální	průměrný	STD
Počet slov v textu	296	4180	1717	499,44
Počet vět v textu	14	238	103	34,50
Počet slov ve větě	11	28	17	2,73
Počet zkoumaných textů		336		

Jak vyplývá z předložených údajů, jednotlivé texty vykazují značnou variabilitu, co se týká délky vět i samotných textů, pro veškerou další manipulaci s lingvistickými daty byla četnost jednotlivých jazykových jevů vždy uváděna v poměru k délce textu, tedy ve své relativní četnosti.

Vzhledem k výsledkům předchozích studií (Kučera, 2020 a 2022), kde lingvistická data vykazovala nestabilitu vzhledem k žánrové rozmanitosti textů, jsme zkoumané texty žánrově omezili pouze na jediný typ – formální akademické eseje. Žánrově (stejně jako populačně) jde tedy o příliš homogenní soubor a ani zde nelze vzhledem k předchozím studiím výsledky zobecňovat na jiné než formální akademické texty. Jazykové rysy totiž vykazují značnou variabilitu v závislosti na žánrovém typu textu (Kučera, 2020 a 2022).

K další statistické analýze bylo však zapotřebí autorské texty dále zpracovat a rozdělit na jednotlivé segmenty, které by bylo možné dále kvantifikovat a podrobit frekvenční a korelační analýze. Za tímto účelem jsme použili strojové zpracování přirozeného jazyka pomocí rozhraní UDpipe, které je blíže popsáno v kapitole 7.3.1.

7.3.1 Strojové zpracování jazyka UDpipe

Pro strojové zpracování esejí byl využit nástroj **UDpipe**, který slouží k automatickému zpracování přirozeného jazyka. Jeho hlavní funkcí je analyzovat text a poskytovat informace o jeho gramatické struktuře a slovní morfologii, a to v různých jazycích včetně češtiny (Mikulová et al. 2020).

UDpipe¹⁰ je přístupný jako webová aplikace, nebo jako volně dostupné rozhraní pro operační systémy Linux, Windows a OS X a také jako knihovna pro programovací jazyky C++, Python, Perl, Java, C# nebo jako balíček třetí strany pro R.

Zkratka `Udpipe` je odvozena od názvu projektu, kterým je `Universal Dependencies pipeline` (Straka et al., 2016). `Universal Dependencies` je open-source platforma, která si klade za cíl vytvořit univerzální a konsistentní reprezentaci syntaktického a morfologického rozboru přirozeného jazyka v mnoho jazykových variantách (DeMarneffe et al., 2021).

`Pipeline` odkazuje na proces zpracování, který UDpipe provádí, tj. syntaktickou a morfologickou analýzu textu včetně závislostních reprezentací jednotlivých prvků. Mimo jiné UDpipe dokáže: (i) **tokenizovat**, tedy rozdělit vstupní text na jednotlivé tokeny (slova nebo znaky, včetně interpunkčních znamének); (ii) **lematizovat**, neboli převést slova na jejich základní formy (lemmata, jako je např. infinitiv); (iii) **morfologicky kategorizovat**, tedy určit např. pád, číslo, čas, způsob a další kategorie u jednotlivých slovních druhů; (iv) **syntakticky strukturovat**, neboli určit příslušnost jednotlivých větných členů včetně jejich závislostí a kontextových vztahů.

Výsledek automatického zpracování textu se objeví jako tabulka ve formátu CoNLL-U, přičemž každé slovo je uvedeno na vlastním řádku, ve sloupcích jsou potom řazeny veškeré údaje gramatického a syntaktického rozboru, včetně identifikace slova k danému textu, odstavci, větě nebo souvětí, ve které bylo nalezeno.

Jednotlivé sloupce tabulky jsou tedy řazeny takto:

[1] identifikační znaky: *doc_id, paragraph_id, sentence_id, sentence, token_id*,

[2] slovo v původním tvaru: *token*,

[3] slovo v základním tvaru: *lemma*,

[4] gramatické značky univerzální a specifické pro daný jazyk: *upos, xpos, feats*,

[5] syntaktická struktura závislostí větných členů: *head_token_id, dep_rel, deps, misc*

Pro ilustraci si ukážeme, jak vypadá automatický výstup z morfologického `označkování` slova – tedy tokenu – pomocí univerzálního kódování jazykových kategorií, výsledek se zobrazuje takto:

¹⁰ UDpipe je dostupný na <https://ufal.mff.cuni.cz/udpipe/1> [citováno z 27.3.2024]

Tabulka 6: Automatické tagování morfologické struktury přirozeného jazyka pomocí nástroje UDpipe

Token	Lemma	Upos	Xpos	Feats
ILUZE	iluze	NOUN	NNFS1-----A----	Case=Nom Gender=Fem Number=Sing Polarity=Pos

Zdroj: Vlastní datová matice, podrobný popis morfologických kategorií viz Mikulová et al. (2020)

Na základě tohoto výstupu jsme definovali jednotlivé lingvistické jevy pro komparační a korelační analýzu, které jsou detailně představeny v následující kapitole 7.3.2.

UDpipe je pro svou dostupnost a relativně snadnou využitelnost široce využíván v oblasti nejen lingvistického výzkumu, ale i při zpracování přirozeného jazyka pro další účely, jak je strojového učení, analýzy sentimentu, extrakce informací a strojového překladu (Straka, Straková, 2017).

7.3.2 Definice sledovaných lingvistických jevů

Tato kapitola je věnována definici lingvistických proměnných, se kterými budeme ve výzkumu pracovat. Volba jednotlivých jazykových jevů je podmíněná výstupem UDpipe ze strojového zpracování přirozeného jazyka (viz kapitola 7.3.1), které nabízí předem definované morfologické a syntaktické kategorie.

Zde je nutné podotknout, že tyto kategorie jsou v matematické lingvistice navrhovány univerzálně, tedy bez ohledu na specifika konkrétního jazykového systému (Straka, Straková, 2017). Uváděné lingvistické kategorie tedy nebudou přesně odpovídat mluvnickému a gramatickému řádu českého jazyka.

Následující Tabulka 7 uvádí veškeré lingvistické jevy, které jsme použili pro následnou statistickou analýzu.

Tabulka 7: Seznam jazykových jevů, u kterých byla sledována relativní četnost výskytu v psaných textech jednotlivých autorů

Slovní druh	Jazykový jev	UDPipe - feats	Příklad termínu z textů
Příd. jm.	pozitivní	ADJ Polarity=Pos	zaostřený, snadný
	Negativní	ADJ Polarity=Neg	nepříjemný, necelý
	2. stupeň	ADJ Degree=Com	lepší, jasnější
	3. stupeň	ADJ Degree=Sup	nejnovější, nejstarší
	1. stupeň	ADJ Degree=Pos	příjemný starý
	Pasivum	ADJ Voice=Pas	zahalen, odstraněny
Slovesa	přít. Čas	VERB Tense=Pres	chodí, poslouchají
	minulý čas	VERB Tense=Past	šel, poslechli
	budoucí čas	VERB Tense=Fut	půjde, zapíšeme
	rod činný	VERB Voice=Act	tvoří, ulehčuje
	Tvar záporný	VERB Polarity=Neg	nelze, neztotožňuji
	Tvar kladný	VERB Polarity=Pos	lze, zaručuji
	1. osoba pl/sg	VERB Person=1	jdu, půjdeme
	2. osoba pl/sg	VERB Person=2	zapamatujte si, zjistí
	3. osoba pl/sg	VERB Person=3	zanebávají, zapsal se
	Jednotné číslo	VERB Number=Sing	Určuji, determinuje
Množné číslo	VERB Number=Plur	Předurčuji, dáváme	

	Způsob oznamovací	VERB Mood=Ind	<i>Myslím, znázorňuje</i>
	Způsob rozkazovací	VERB Mood=Imp	<i>zapamatujte si, zjisti</i>
Pom sloveso `být`	Přítomný čas	AUX Tense=Pres	<i>je, jsou</i>
	Minulý čas	AUX Tense=Past	<i>byl, byli</i>
	Budoucí čas	AUX Tense=Fut	<i>budou, budeme</i>
	Rod činný	AUX Voice=Act	<i>jsme, je</i>
	Tvar záporný	AUX Polarity=Neg	<i>není, nejsme</i>
	Tvar kladný	AUX Polarity=Pos	<i>je, bylo</i>
	1. osoba sg/pl	AUX Person=1	<i>jsem, jsme, budu</i>
	2. osoba sg/pl	AUX Person=2	<i>jsi, jste, budete</i>
	3. osoba sg/pl	AUX Person=3	<i>jsou, je, byl</i>
	Číslo jednotné	AUX Number=Sing	<i>byl, budu</i>
	Číslo množné	AUX Number=Plur	<i>budeme, jsou</i>
	Způsob oznamovací	AUX Mood=Ind	<i>je, jsou</i>
	Způsob rozkazovací	AUX Mood=Imp	<i>buď, buďte</i>
	Způsob podmiňovací	AUX Mood=Cnd	<i>by, bychom</i>
Zájmeno	neurčité	PronType=Ind	<i>někdo, něco</i>
	záporné	PronType=Neg	<i>nikdo, nic</i>
	osobní/zvratné	PronType=Prs	<i>ji, nám, vás, jich, mu/se, si</i>
	vztažné	PronType=Rel	<i>což, něž, čímž</i>
	všeobecné/totalizátory	PronType=Tot	<i>ten sám, veškerý, samý</i>
	tázací	PronType=Int,Rel	<i>co, čím</i>
Determinující výrazy/zájmena	neurčité	DET PronType=Ind	<i>nějaký, jakýsi, jakékoli</i>
	záporné	DET PronType=Neg	<i>žádný, žádnému</i>
	osobní	DET PronType=Prs	<i>jej, jich, mu</i>
	všeobecné/totalizátory	DET PronType=Tot	<i>každý, všichni, veškerý</i>
	vztažné	DET PronType=Int,Rel	<i>který, které</i>
	1. osoba sg/pl	DET Person=1	<i>můj, naše</i>
	2. osoba sg/pl	DET Person=2	<i>tvoje, vašimi</i>
	3. osoba sg/pl	DET Person=3	<i>jejich, jeho</i>
	Číslo jednotné	DET Number[psor]=Plur	<i>vaše, jejich</i>
	Číslo množné	DET Number[psor]=Sing	<i>svůj, jeho</i>
	přivlastňovací	DET Poss=Yes	<i>můj, náš svůj, jejich</i>
Příslovce	2. stupeň	ADV Degree=Cmp	<i>lépe, déle</i>

	3. stupeň	ADV Degree=Sup	<i>nejhlouběji, nejjasněji</i>
	1. stupeň	ADV Degree=pos	<i>dobře, špatně</i>
Intertextové odkazy	Citace	Upos Punct „,“	<i>(Plháková, 2022)</i>

V přehledu vidíme, že uvedené kategorie nenesedají vždy přesně na gramatické kategorie tak, jak je uvádí Mluvnice českého jazyka (Kolářová et al., 2012). Největším odchylkou jsou zájmena, která mají v anglickém jazyce vlastní skupinu `determiners`, kam patří například určité a neurčitě členy (*a/an, the*), ve stejné kategorii se ale objevují i neurčitá zájmena (*každý, kdejaký*), včetně některých typů zájmen přivlastňovacích (*náš, jeho*). `Determiners` jsou tedy spíše ukazovací výrazy, které blíže určují (determinují) povahu podstatných jmen. Tato kategorie českému jazyku příliš neodpovídá a pro přesnější jazykovou analýzu bychom museli provádět nejen kvantitativní, ale i kvalitativní rozbor textů. Přestože se ve zřkumech objevuje i termín `determinátory` (Progovac, 1995), v této práci jsme zvolili pojmenování kategorií tak, aby co nejlépe odpovídaly jazykovému systému češtiny.

Pro začátek se spokojíme s frekvenční analýzou výše jmenovaných jevů, a to ze dvou důvodů: (i) poskytují dostatečně přesné kategorie k popisu jazykových změn závislých na pohlaví a míře Extraverze; a (ii) tyto mluvnické kategorie odpovídají sledovaným jazykovým rysům ve srovnatelných studiích na českém jazyce (Kučera, 2020 a 2022). Abychom tedy byli schopni komparace a další práce s našimi daty ve smyslu zpřesňování výsledků, budeme pracovat s takovými gramatickými kategoriemi, které automaticky vygeneroval program UDPipe.

7.3.3 Postup extrakce sledovaných lingvistických jevů

Nejprve musely být konkrétní jazykové jevy přiřazeny k autorovi daného textu, aby mohly být zkoumány v závislosti na jeho pohlaví nebo míře Extraverze. To jsme provedli pomocí **ID autora**. Pro každý sledovaný lingvistický jev, slovní druh nebo hypotézu jsme nejdříve vyfiltrovali požadované jazykové kategorie a k nim přiřadili kód autora.

Za účelem smysluplného výzkumu jsme museli automaticky vygenerované jazykové kategorie dále **třídít a kombinovat**, abychom získali kategorie, které lépe odpovídají gramatickému systému češtiny. Tento postup si představíme krok za krokem.

Nejprve bylo nutné spárovat respondenta a jeho t-skóry v inventáři BFI-2 s konkrétní esejí, aby bylo možné jeho konkrétní jazykový projev zkoumat (viz Tabulka 8).

Tabulka 8: Příklad extrakce trpného rodu z datové matice přes ID autora

Slov. druh	autor	ROD
ADJ	1	Act
ADJ	114	Pass
ADJ	114	Pass
ADJ	114	Pass
ADJ	114	Pass
ADJ	114	Pass
ADJ	114	Act
ADJ	114	Pass
ADJ	114	Pass

ADJ	117	Act
ADJ	117	Pass
ADJ	117	Pass
ADJ	117	Pass
ADJ	117	Pass
ADJ	117	Act
ADJ	117	Act
ADJ	117	Pass

Tabulka 8 znázorňuje extrakci trpného rodu (aktiva a pasiva) v kategorii přídavných jmen pomocí sledování tvaru přičestí (byl *tázán, uveden, dokončen*) přes ID autora. Tak je možné sledovat četnost výskytu uvedeného jazykového jevu v textu konkrétního jedince, u nějž budeme sledovat pohlaví a míru Extraverze.

Výskyt veškerých jazykových jevů v textech byl přepočten na celkový počet slov daného textu a tím zjištěna jeho **relativní četnost**. Relativní četnost je statistický ukazatel používaný k vyjádření podílu, který daný jazykový jev představuje vzhledem k celkovému počtu slov v daném textu. Relativní četnost je vyjádřena jako poměr mezi počtem pozorování lingvistického jevu a délkou textu, tedy celkovým počtem slov. S relativní četností nepracujeme přímo, ale dáváme ji do statistického vztahu s pohlavím nebo mírou Extraverze

Vzhledem k tomu, že univerzální gramatické kategorie, které UDpipe pomocí strojové analýzy jazyka vygeneruje, neodpovídají vždy systému českého jazyka, bylo zapotřebí určité kategorie vzájemně kombinovat tak, abychom se dopracovali žádoucích výsledků.

Příkladem může být třeba kategorie osoby a čísla u řady slovních druhů – podstatná jména, zájmena, významová slovesa i pomocné sloveso být. Pokud chceme sledovat konkrétní jazykové jevy – tedy. např. 1. osobu čísla jednotného (*já, byl, můj, slejuji*), nemohu se spolehnout na automatický výstup UDpipe, který mi nabízí pouze kategorii osoba (1,2,3) a číslo (singulár nebo plurál), v automatickém výstupu totiž 1. osoba znamená jak singulár, tak plurál (*já i my, byl i byli, můj i náš, sleduji i sledujeme*) a číslo jednotné zase znamená kombinaci všech osob (*já, ty, on, byl, byls i byla, sleduji, sleduješ i sleduje, můj, tvůj i jeho*).

Takových jevů byla celá řada, např. pasivum, trpný rod, podmiňovací způsob, zápor nebo užívání citací.

Tabulka 9 ukazuje extrakci dat pro ověření relativní četnosti trpného rodu. Byla zde použita automatická vyhledávací funkce v programu excel, která v tomto případě měla zachytit veškerý výskyt trpného rodu `Voice=Pas` v mluvnické kategorii přídavných jmen `ADJ`. Daná formule pro vyhledávání v textové databázi lingvistických jevů vypadala následovně:

Tabulka 9: Příklad vyhledávacího příkazu v programu Excel k extrakci trpného rodu z datové matice

```
=IFERROR(MID(J81596; SEARCH("Voice="; J81596) + LEN("Voice=");
IF(ISNUMBER(SEARCH("|"; J81596; SEARCH("Voice="; J81596)));
SEARCH("|"; J81596; SEARCH("Voice="; J81596)) - SEARCH("Voice="; J81596)
- LEN("Voice="); LEN(J81596) + 1)); "")
```


Takový příkaz obsahuje komplexní kombinaci funkcí, které mají za úkol extrahovat určený řetězec znaků z textového formátu uloženého v zadané buňce. Přesněji řečeno tento příkaz vyhledá `SEARCH` pozici prvního výskytu textu `Voice=` v textovém poli buňky J81596, a poté pomocí funkce `MID` vybere podřetězec (substring) začínající za textem `Voice=`. Vnitřní část příkazu využívá funkce `IF`, `ISNUMBER` a `SEARCH`, aby určila, zda se v textovém řetězci nachází znak "|" (svislá čára) za textem "Voice=". Pokud ano, vypočítá se délka podřetězce od textu "Voice=" do znaku "|". Pokud znak "|" není nalezen, délka podřetězce se počítá od textu "Voice=" do konce textového řetězce v buňce. Nakonec, pokud se při vyhledávání vyskytne chyba (např. pokud se text "Voice=" nenachází v buňce), funkce `IFERROR` vrátí prázdný řetězec ("").

Pomocí kontingenční tabulky byly následně přiřazeny vyhledané výrazy k jednotlivým autorům a následně vypočítána relativní četnost daného jevu v textu daného autora (viz Tabulka 10)

Tabulka 10: Příklad kontingenční tabulky pro párování výskytu konkrétních jazykových jevů s textem konkrétního autora

ROD	Popisky sloupců			Celkový součet
Popisky řádků	Act	Pass		
1	23	1		24
114	239	1	7	247
117	303	10	14	327
118	251	13	6	270
121	360	10	27	397
125	227	7	32	266
126	310	9	22	341
127	292	8	28	328
129	228	11	8	247
132	353	5	15	373
135	237	3	15	255

Takto připravené součty výskytu daného jazykového jevu – v případě Tabulky 10 se jedná o trpný rod – poskytly absolutní četnost sledovaného lingvistického jevu u daného autora. Z absolutních četností byla potom vypočítána četnost relativní pomocí přepočtu četnosti výskytu na délku daného textu. Požadované sloupce kontingenční tabulky byly pak vloženy do datové matice, ze které vycházejí výsledky celé korelační analýzy.

7.4 Etické hledisko a ochrana soukromí

Respondenti vyplňovali inventář online a byli osloveni z univerzitního e-mailu s žádostí o **anonymní** účast ve výzkumu. Identitu jednotlivých respondentů není možné dohledat ani nebyly zaznamenávány žádné další údaje kromě pohlaví. Údaje o věku a demografické příslušnosti nejsou známy a jediným identifikačním znakem je přidělené ID, které spojuje hrubý skór inventáře BFI-2 s autorem textu.

Účast ve výzkumu byla **dobrovolná** a respondenti mohli kdykoliv práci přerušit. Na konci vyplněného sebeposuzovacího inventáře získali krátkou automaticky generovanou zpětnou slovní vazbu shrnující krátce výsledky jejich výkonu vzhledem

k populaci. Zároveň dostali účastníci upozornění, že výsledky neposkytují klinický popis jejich osobnosti.

Všechny výsledky této práce jsou použity v souladu s **morálním imperativem** psychologického výzkumu a jediným záměrem je aplikovat je ve jménu obecného blaha a pro dobro člověka (Prilleltensky, Walsh-Bowers, 1993).

8 PRÁCE S DATY A JEJÍ VÝSLEDKY

V této kapitole přistoupíme k popisu veškerých úkonů, které jsme při práci s naším datovým souborem provedli. Začneme ověřením validity a reliability psychometrických dat. Zvláštní pozornost budeme věnovat identifikaci jednotlivých lingvistických jevů, které budeme během naší komparativní a korelační analýzy sledovat. Zjistíme, zda existují statistické rozdíly mezi podsouborem mužů a žen. Připomeneme si jednotlivé výzkumné záměry a stanovené hypotézy a pomocí standardních statistických postupů ověříme jejich platnost.

8.1 Validita a reliabilita

Validita a reliabilita jsou dva klíčové pojmy v oblasti psychometrie a týkají se spolehlivosti a přesnosti měření. Tyto dvě koncepce se vzájemně doplňují, my začínáme ověřením Validity, tedy přesností měření.

Validita ověřuje, zda měření skutečně měří to, co měřit má, a zda výsledek skutečně odpovídá jevu, který má být měřen. Existuje několik typů validity, včetně obsahové validity (zda obsah měření skutečně odráží zkoumaný konstrukt), kriteriální validity (zda měření koreluje s jinými známými proměnnými) a konstruktové validity (zda měření měří přesně ten konstrukt, který si klade za cíl).

Abychom ve svém výzkumu zabránili nesprávným závěrům, přistoupili jsme k ověření validity výsledků měření inventáře BFI-2 v našem souboru. Pomocí **faktorové analýzy** jsme zkoumali validitu všech položek inventáře BFI-2 na škále Extraverze, včetně jejích subškál (viz Tabulka 11).

Tabulka 11: Zátěže z analýzy hlavních komponent (rotace Varimax)

Položka	Negativní emocionalita	Svědomitost	Otevřenost mysli	Přívětivost	Extraverze
bfi1	0,196	-0,019	-0,072	0,176	0,630
bfi6	0,292	0,133	0,153	-0,175	0,526
bfi11	0,034	0,051	0,195	0,216	0,425
bfi16	0,094	-0,079	-0,057	0,025	0,692
bfi21	0,224	0,086	0,083	-0,375	0,588
bfi26	0,257	0,249	0,213	0,059	0,500
bfi31	0,231	0,003	-0,075	-0,068	0,643
bfi36	0,127	0,096	0,177	-0,111	0,289
bfi41	0,356	0,138	-0,098	0,183	0,619
bfi46	-0,032	-0,034	-0,047	0,073	0,663
bfi51	0,201	0,108	0,194	-0,380	0,478
bfi56	0,038	0,030	0,113	0,176	0,607

Faktorovou analýzu jsme provedli pomocí **metody hlavních komponent**, ta se snaží najít lineární kombinace původních proměnných, které vysvětlují co nejvíce

variability v datech. Jednotlivé komponenty jsou pak uspořádány podle toho, nakolik celkovou variabilitu dat vysvětlují. Po provedení analýzy hlavních komponent se někdy používají rotace, které transformují komponenty tak, aby co nejvíce oddělily proměnné, které mají vysoké a nízké váhy v každé komponentě. Toto slouží k dosažení lepší interpretovatelnosti. V našem případě jsme využili **rotaci Varimax**, která je citlivá na latentní faktory.

Výsledky jsou prezentovány pomocí **zátěže** z hlavních komponent a ukazují míru, do jaké každá původní proměnná sytí hlavní komponentu. Jedná se o korelační koeficienty (jejich hodnoty se pohybují v rozmezí od -1 do 1) a vyjadřují vztahy mezi původními proměnnými a nově vytvořenými hlavními komponentami.

Zátěže z naší analýzy potvrzují validitu pro naše data a jsou shrnuté v Tabulce 12. Nejvyšší míra zátěže označuje takové proměnné, které hlavní komponenty sytí nejvíce. Na našich datech se potvrzuje, že vybrané položky dle manuálu Hřebíčkové et al. (2020) skutečně sytí hlavní faktor Extraverze. Validitu pro naše data na základě této faktorové analýzy můžeme potvrdit.

Nyní přistoupíme k ověření reliability. **Reliabilita** se týká spolehlivosti měření a je nezbytná, protože nespolehlivé měření nemůže být považováno za relevantní nebo užitečné, ačkoli může být validní. Reliabilitu považujeme za prokázanou, pokud měření poskytuje konzistentní výsledky při opakovaném použití nebo v různých situacích.

Existuje několik typů reliability, včetně interní konzistence (například Cronbachova alfa), test-retest reliability (když jsou stejné osoby testovány dvakrát s určitým časovým odstupem) a inter-rater reliability (když různí hodnotitelé hodnotí stejné zkoumané jednotky).

Pro náš soubor jsme reliabilitu ověřovali pomocí **testu interní konzistence** - Cronbachova alfa. Cronbachova alfa umožňuje určit, do jaké míry je možné spoléhat na konzistenci odpovědí respondentů na jednotlivé položky dotazníku při měření daného psychologického konstruktů. Zaměřili jsme se na osobnostní rys Extraverze a jeho subškály (viz Tabulka 12).

Tabulka 12: Reliabilita položek inventáře BFI-2

Škála	Cronbachova alfa
Extraverze	0,86
Soucit	0,65
Uctivost	0,74
Důvěra	0,64

Cronbachova alfa se pohybuje v rozmezí od 0 do 1, kde vyšší hodnota indikuje větší interní konzistenci. Interpretace hodnot Cronbachovy alfy je následující: hodnota blízká 0 ukazuje na nízkou interní konzistenci a naznačuje, že položky dotazníku nejsou vzájemně korelované; hodnota blízká 1 ukazuje na vysokou interní konzistenci a naznačuje, že položky dotazníku jsou vzájemně silně korelované.

Na našich datech nižší reliabilitu vykazují pouze subškály soucit a důvěra. Celkově však naše data vykazují dostačující nebo vysokou spolehlivost měření osobnostního rysu Extraverze.

8.2 Jazyk ve vztahu k pohlaví

Tato kapitola nasedá na výzkumný záměr 1 a klade si za cíl ověřit, zda pohlaví respondentů souvisí s četností výskytu sledovaných jazykových rysů. V této práci budeme komparativní analýzou ověřovat zvláště jednotlivé slovní druhy (H1a) a k nim dodatečně zpracovávat gramatické kategorie (H1b), které chceme sledovat.

H1a: Existuje rozdíl v četnosti užívání slovních druhů mezi muži a ženami.

Nejdříve jsme ověřili rozdíly mezi muži a ženami v závislosti na četnosti výskytu jednotlivých slovních druhů. Pro správné zvolení statistického testu jsme ověřili normální rozdělení podílů užívání daných slovních druhů ve skupinách dle pohlaví pomocí Shapiro-Wilkova testu.

Shapiro-Wilkův test je statistický test normality, který se používá k určení, zda data pocházejí z normálního rozdělení. Jeho hlavním cílem je testovat nulovou hypotézu, která říká, že data pocházejí z normálně rozdělené populace. Pokud je p-hodnota získaná z Shapiro-Wilkova testu menší než vybraná hladina významnosti (my budeme pracovat s 0,05), nulová hypotéza o normálním rozdělení dat je zamítnuta, což znamená, že data nejsou normálně rozdělená.

Shapiro-Wilkův test je považován za jedno z nejlepších testů normality pro menší vzorky (obvykle méně než 50 pozorování), ale je také použitelný pro větší vzorky. Jeho výhodou je, že je poměrně citlivý na odchylky od normálního rozdělení v oblasti středních hodnot.

Následující Tabulka 13 obsahuje informaci o splnění normality v obou výběrech (muži/ženy), dále zvolený statistický test a porovnávané střední hodnoty.

Tabulka 13: Relativní četnost jednotlivých slovních druhů v závislosti na pohlaví

Slovní druh	Normalita obou výběrů	Zvolený test	Medián/průměr muži	Medián/průměr ženy	Testové kritérium	p-hodnota	Míra účinku
Přídavná jména	Ne	M-W test	0,138	0,131	Z=0,408	0,683	
Předložky	Ano	t-test	0,098	0,093	t=3,821	0,000	d=0,526
Příslovce	Ne	M-W test	0,076	0,074	Z=0,397	0,692	
Pomocné sloveso `být`	Ne	M-W test	0,037	0,041	Z=-0,609	0,543	
Spojky souřadící	Ano	t-test	0,046	0,048	t=-2,044	0,042	d=0,250
Determinující výrazy	Ano	t-test	0,055	0,059	t=-2,781	0,006	d=0,358
Slang/dialekt/slova převzatá	Ne	M-W test	0,000	0,000	Z=0,551	0,582	
Podstatná jména	Ne	M-W test	0,285	0,277	Z=0,498	0,618	

Číslovky	Ne	M-W test	0,026	0,027	Z=-0,683	0,495	
Částice	Ne	M-W test	0,005	0,005	Z=0,071	0,943	
Zájmena	Ano	t-test	0,047	0,052	t=-2,993	0,003	d=0,423
Vlastní jména/názvy	Ne	M-W test	0,029	0,029	Z=-0,482	0,630	
Interpunkce	Ne	M-W test	0,208	0,208	Z=-0,221	0,825	
Spojky podřadící	Ne	M-W test	0,031	0,033	Z=-1,235	0,217	
Symboly/znaky	Ne	M-W test	0,000	0,000	Z=0,329	0,742	
Slovesa	Ne	M-W test	0,120	0,123	Z=-1,358	0,175	

U slovních druhů, jejichž podíl používání splňoval u mužů i u žen normální rozdělení, byl zvolen parametrický t-test pro nezávislé výběry, a porovnávanými středními hodnotami jsou tedy aritmetické průměry.

Provedením t-testu jsme vypočetli hodnoty t, které se pak porovnávají s kritickou hodnotou z příslušného rozdělení (zde Studentovo t-rozdělení) na hladině významnosti 0,05. Výsledek t-testu ukazuje, zda je rozdíl mezi průměry skupin dostatečně velký na to, aby byl považován za statisticky významný, tyto hodnoty jsou pro lepší přehlednost uvedeny tučně.

Pokud je hodnota testové statistiky t velká, naznačuje to, že rozdíl mezi průměry skupin je pravděpodobně významný a nejpravděpodobněji není výsledkem náhodného kolísání. Naopak, pokud je hodnota testové statistiky t malá, naznačuje to, že rozdíl mezi průměry skupin je pravděpodobně výsledkem náhodného kolísání, a tím pádem ani ukazatel p-hodnoty nebude statisticky významný.

Statisticky významný vztah se potvrdil u slovního druhu zájmen a determinujících výrazů, což je vlastně v českém jazyce jeden a tentýž slovní druh, dále pak u předložek a spojek souřadících. U ostatních slovních druhů nebyl rozdíl v průměru mezi pohlavími statisticky významný.

U slovních druhů, jejichž podíl používání nespĺňoval u mužů i u žen normální rozdělení, byl použit neparametrický Mann-Whitney U test, což je neparametrická obdoba t-testu a porovnávanými středními hodnotami jsou mediány. Statistika z je spočtena na základě hodnot U, které jsou vypočteny nejprve nezávisle pro muže a ženy. V každé skupině jsou hodnoty nejprve setříděny podle velikosti, poté jsou jednotlivým hodnotám přiřazeny ranky, pokud jsou dvě hodnoty totožné, je jim přiřazen rank průměrný. Hodnota U je pak spočtena jako součet ranků v jedné ze skupin. Hodnota U sama o sobě ale není interpretovatelná pro určení statistické významnosti rozdílů mezi pohlavími, pracujeme tedy s testovým kritériem Z, které odráží jak hodnotu U, tak velikost vzorku a další faktory.

V našem souboru byla hodnota Z u všech slovních druhů menší než kritická hodnota hladiny významnosti (0,05). P-hodnoty zvolených statistických výpočtů jsou

uvedeny v posledním sloupci. U Man-Whitneyova U testu se tady neprokázal statisticky významný rozdíl mezi muži a ženami.

Statisticky významné rozdíly jsme prokázali pouze u parametrických dat s normálním rozdělením. Rozdíly, které byly nalezeny u relativní četnosti předložek (*na, pod, u*), spojek souřadících (*a, ale, nebo*) a zájmen (*jim, oni, který, všem*) můžeme považovat za statisticky signifikantní. Předložky používají významně častěji muži než ženy. Spojky souřadící a zájmena používají významně častěji ženy než muži.

H1b: Existuje rozdíl v četnosti užívání sledovaných lingvistických jevů mezi muži a ženami.

Dále jsme ověřili rozdíly v používání stanovených lingvistických jevů mezi muži a ženami, výsledky jsou zřehledněny v následující Tabulce 14.

Tabulka 14: Relativní četnost extrahovaných lingvistických jevů v závislosti na pohlaví

Slovní druh	Lingvistický jev	Normalita v obou výběrech	Zvolený test	Medián/průměr muži	Medián/průměr ženy	Testové kritérium	p-hodnota	Míra účinku
Příd. jm.	pozitivní	Ne	M-W test	0,129	0,123	Z=1,086	0,278	
	negativní	Ne	M-W test	0,003	0,002	Z=2,052	0,040	$\eta^2=0,012$
	2. stupeň	Ne	M-W test	0,005	0,005	Z=0,407	0,684	
	3. stupeň	Ne	M-W test	0,001	0,001	Z=0,536	0,592	
	1. stupeň	Ne	M-W test	0,108	0,107	Z=0,875	0,382	
	Pasivum	Ne	M-W test	0,007	0,007	Z=-0,916	0,360	
Sloveso	přít. čas	Ne	M-W test	0,071	0,076	Z=-2,474	0,013	$\eta^2=0,017$
	minulý čas	Ne	M-W test	0,017	0,018	Z=-0,605	0,545	
	budoucí čas	Ne	M-W test	0,000	0,000	Z=0,391	0,696	
	rod činný	Ne	M-W test	0,092	0,096	Z=-2,787	0,005	$\eta^2=0,022$
	tvar záporný	Ano	t-test	0,007	0,007	t=1,024	0,307	
	tvar kladný	Ne	M-W test	0,111	0,116	Z=-2,458	0,014	$\eta^2=0,017$
	1. osoba pl	Ne	M-W test	0,010	0,012	Z=-0,815	0,415	
	1. osoba sg	Ne	M-W test	0,004	0,004	Z=-1,553	0,120	
	2. osoba pl	Ne	M-W test	0,000	0,000	Z=0,216	0,829	
	2. osoba sg	Ne	M-W test	0,000	0,000	Z=0,826	0,409	
	3. osoba pl	Ne	M-W test	0,014	0,014	Z=-2,304	0,112	
	3. osoba sg	Ne	t-test	0,043	0,043	t=-0,300	0,764	
	způsob oznamovací	Ne	M-W test	0,071	0,076	Z=-2,426	0,015	$\eta^2=0,017$

	Způsob rozkazovací	Ne	M-W test	0,001	0,000	Z=2,015	0,044	$\eta^2=0,011$
Pomocné sloveso 'být'	Přítomný čas	Ne	M-W test	0,022	0,023	Z=-1,756	0,079	
	Minulý čas	Ne	M-W test	0,002	0,002	Z=-0,542	0,588	
	Budoucí čas	Ne	M-W test	0,002	0,002	Z=-0,447	0,655	
	Rod činný	Ne	M-W test	0,026	0,028	Z=-2,256	0,024	$\eta^2=0,014$
	Tvar záporný	Ne	M-W test	0,002	0,002	Z=1,133	0,257	
	Tvar kladný	Ne	M-W test	0,026	0,028	Z=-1,946	0,052	
	1. osoba pl	Ne	M-W test	0,012	0,003	Z=-1,701	0,089	
	1. osoba sg	Ne	M-W test	0,004	0,003	Z=0,378	0,705	
	2. osoba pl	Ne	M-W test	0,000	0,000	Z=0,278	0,781	
	2. osoba sg	Ne	M-W test	0,000	0,000	Z=0,260	0,795	
	3. osoba pl	Ne	M-W test	0,014	0,004	Z=-2,074	0,038	$\eta^2=0,015$
	3. osoba sg	Ne	M-W test	0,044	0,017	Z=-0,603	0,547	
	Způsob oznamovací	Ne	M-W test	0,023	0,026	Z=-1,998	0,046	$\eta^2=0,011$
	Způsob rozkazovací	Ne	M-W test	0,000	0,000	Z=0,022	0,983	
	Způsob podmiňovací	Ne	M-W test	0,008	0,008	Z=-1,198	0,231	
Zájmena	neurčité	Ne	M-W test	0,001	0,001	Z=0,381	0,703	
	záporné	Ne	M-W test	0,000	0,000	Z=-0,492	0,622	
	osobní	Ano	t-test	0,041	0,045	t=-3,085	0,002	d=0,377
	vztažné	Ne	M-W test	0,002	0,002	Z=0,377	0,706	
	Všeobecné/otalizátory	Ne	M-W test	0,000	0,000	Z=0,108	0,914	
	tázací	Ne	M-W test	0,002	0,002	Z=-1,802	0,072	
Determinující výrazy	neurčité	Ne	M-W test	0,004	0,004	Z=0,311	0,756	
	záporné	Ne	M-W test	0,000	0,000	Z=0,492	0,623	
	osobní	Ano	t-test	0,012	0,012	t=-0,072	0,942	
	všeobecné/otalizátory	Ne	M-W test	0,002	0,003	Z=-2,710	0,007	$\eta^2=0,021$
	vztažné	Ne	M-W test	0,011	0,012	Z=-1,494	0,135	
	1. osoba pl	Ne	M-W test	0,018	0,022	Z=-2,154	0,031	$\eta^2=0,013$
	1. osoba sg	Ne	M-W test	0,013	0,014	Z=-0,783	0,434	

	2. osoba pl	Ne	M-W test	0,000	0,000	Z=0,811	0,418	
	2. osoba sg	Ne	M-W test	0,000	0,000	Z=1,552	0,121	
	3. osoba pl	Ne	M-W test	0,019	0,021	Z=-2,304	0,021	$\eta^2=0,015$
	3. osoba sg	Ne	M-W test	0,064	0,067	Z=-1,767	0,077	
	přivlastňovací	Ano	t-test	0,012	0,012	t=-0,054	0,957	
Příslovce	2. stupeň	Ne	M-W test	0,005	0,004	Z=1,278	0,201	
	3. stupeň	Ne	M-W test	0,001	0,001	Z=-0,429	0,668	
	1. stupeň	Ne	M-W test	0,019	0,017	Z=2,625	0,009	$\eta^2=0,019$
	záporné	Ne	M-W test	0,000	0,000	Z=-1,136	0,256	
Intertextové odkazy	Citace	Ne	M-W test	0,008	0,009	Z=-2,150	0,032	$\eta^2=0,013$

Výsledky v Tabulce 14 můžeme opět interpretovat pomocí testových kritérií, které srovnávají rozdíl středních hodnot dvou nezávislých skupin, tedy mužů a žen. Pro data s normálním rozdělením jsme zvolili t-test a u dat sešikmených nebo špičatých pak neparametrickou obdobu Man-Whitney U testu. U obou testových kritérií platí, že čím jsou vyšší, tím větší je rozdíl středních hodnot mezi pohlavími.

P-hodnota poskytuje informaci o pravděpodobnosti, že pozorované rozdíly jsou výsledkem skutečného efektu a nikoli pouze výsledkem náhody. Čím nižší je p-hodnota, tím více důvodů máme pro zamítnutí nulové hypotézy a přijetí alternativní.

P-hodnoty menší, než zvolená hladina významnosti, jsou v tabulce označeny tučně. V případě, že jsou nižší než $p=0,05$, zamítáme nulovou hypotézu a konstatujeme, že existuje statisticky významný rozdíl mezi pohlavími v námi sledovaném lingvistickém jevu.

Statisticky významné rozdíly byly nalezeny u negativního tvaru přídavných jmen (*nemalý, nevelký*), přítomného času u sloves (*ověřujeme, lze*), činného slovesného rodu (*záleží, přísluší*), kladného tvaru sloves (*zasahuje, rozhodují*), oznamovacího způsobu (*chodí, přichází*) a rozkazovacího způsobu (*představte si, zdůrazněte*) u významových sloves, dále u pomocného slovesa `být` v rodě činném (*je, bude*), v třetí osobě množného čísla (*jsou, byli*), u oznamovacího způsobu (*je, budeme*), u osobních zájmen (*ji, vás, jim, si, se*), dále potom u determinujících výrazů všeobecných/totalizátorů (*veškerý, každý*), v 1. osobě množného čísla (*naše, našich*), ve 3. osobě množného čísla (*jejich, jim*), a u 1. stupně příslovcí (*dobře, rychle*).

Přídavná jména v negativním tvaru (*nemilý, nechtěný*), rozkazovací způsob u sloves (*určete, uvaž*), 3. osobu množného čísla pomocného slovesa být (*jsou, byli*), 1. stupeň u příslovcí (*dobře, rychle*) a 3. osobu množného čísla u determinujících výrazů (*jejich, jejím*) využívají významně častěji muži než ženy.

Přítomný čas u sloves (*nese, založíme*), činný rod (*jde, založili*), slovesa v kladném tvaru (*splňují, směřovali*), oznamovací způsob (*hraje, zkoumáme*), činný rod u pomocného slovesa `být` (*budou, je*), oznamovací způsob (*je, byl*), zájmena osobní (*ji, vás, jim, si, se*), a determinující výrazy všeobecné (*všechny, každý*) a v 1. osobě množného čísla (*naše, našimi*) využívají významně častěji ženy než muži. Ženy také používají více intertextových odkazů ve formě nepřímých citací a příslovcí ve tvaru komparativu (*lépe, rychleji*).

8.3 Jazyk ve vztahu k Extraverzi

Tato kapitola shrnuje výsledky statistické analýzy spojené s výzkumným záměrem 2, tedy s ověřením, zda se relativní četnost sledovaných lingvistických jevů mění v závislosti na míře Extraverze jednotlivých autorů. V této kapitole se tedy budeme zabývat korelačními vztahy. Nejprve zjistíme, nakolik míra Extraverze souvisí s frekvencí výskytu slovních druhů (H2a), a poté jednotlivých lingvistických jevů (H2b).

U slovních druhů a lingvistických jevů, kde se prokázala jak souvislost s pohlavím, tak s mírou Extraverze se budeme dívat na výsledky korelační analýzy zvlášť v podsouboru mužů a podsouboru žen, abychom zaznamenali souvislost korelačních vztahů mezi pohlavím a mírou Extraverze.

Míra Extraverze byla zjištěna pomocí výsledků inventáře BFI-2 a převedeny na t-skóry (viz kapitola 7.2). Výsledky korelační analýzy mezi četností sledovaných jazykových jevů a mírou Extraverze je sledována na všech subškálách, tedy Energičnost, Asertivita a Sociabilita.

Pro ověření statistického vztahu jsme zvolili Spearmanův korelační koeficient. Tento statistický ukazatel se používá k měření síly a směru vztahu mezi dvěma pořadovými proměnnými nebo spojitými proměnnými, které nesplňují normální rozdělení.

Spearmanův korelační koeficient se vypočítá na základě hodnot pořadí jednotlivých proměnných, namísto samotných hodnot. Nejprve jsou data seříděna podle velikosti, a poté jsou přiřazena pořadí k jednotlivým hodnotám. Spearmanův koeficient korelace pak měří, do jaké míry se pořadí obou proměnných shodují. Pokud je korelační koeficient blízký hodnotě +1, indikuje dokonalou pozitivní korelaci. Naopak, pokud je korelační koeficient blízký hodnotě -1, indikuje dokonalou negativní korelaci. Hodnota 0 pak znamená, že mezi proměnnými není žádný lineární vztah.

Spearmanův korelační koeficient má tu výhodu, že je robustní vůči odlehlým hodnotám a vyžaduje méně předpokladů o distribuci dat než Pearsonův korelační koeficient, což jej činí vhodným pro náš datový soubor, který nemá normální rozdělení.

Nejprve jsme opět zvolili ověření závislosti mezi četností výskytu jednotlivých slovních druhů a mírou Extraverze.

H2a: Existuje statistická závislost mezi četností užívání slovních druhů a mírou Extraverze.

Tabulka 15: Vztah mezi relativní četností slovních druhů a mírou extraverze ve všech subškálách inventáře BFI-2

Celý soubor Slovní druhy	Spearmanovy korelační koeficienty			
	Extraverze	Energičnost	Asertivita	Sociabilita
Přídavná jména	-0,048	-0,037	-0,032	-0,014
Předložky	0,113*	0,087	0,082	0,115*
Příslovce	0,004	-0,012	0,007	0,022
Slovesa pomocná	0,037	-0,027	0,048	0,047
Spojka souřadící	0,070	0,086	0,077	0,028
Determinující výrazy/zájmena	0,065	0,011	0,010	0,105
Slang/dialekt/slova převzatá	0,018	0,014	-0,071	0,082
Podstatná jména	-0,051	-0,037	-0,022	-0,064
Číslovky	-0,069	0,008	-0,064	-0,089

Částice	-0,005	-0,015	0,019	-0,027
Zájmena	0,029	0,042	-0,007	0,025
Vlastní jména	-0,057	0,020	-0,089	-0,076
Interpunkce	-0,006	-0,030	0,001	-0,013
Spojka podřadící	-0,032	-0,042	-0,010	-0,040
Symboly/znaky	0,048	0,014	0,047	0,048
Slovesa	0,060	0,030	0,033	0,062

*korelační koeficient je významný na 5% hladině významnosti

Tabulka 15 shrnuje výsledky korelační analýzy, hodnoty, které jsou statisticky významné, jsou označeny hvězdičkou na 5 % hladině významnosti. Z předložených výsledků vyplývá, že mezi četností užití jednotlivých slovních druhů a mírou Extraverze existují jen slabé vztahy. Jediný statisticky významný vztah se prokázal v četnosti užití předložek.

S rostoucí mírou Extraverze roste četnost výskytu slovního druhu předložek. Tento kladný korelační vztah je statisticky významný v subškále Sociability, kde je jeho p-hodnota ($p=0,036$). V ostatních subškálách se tento vztah neprojevil. Z korelační analýzy lze tedy vyvodit, že existuje statisticky významný korelační vztah v kladném směru mezi Sociabilitou a četností užití předložek. Ostatní korelační koeficienty jsou nízké a statisticky nevýznamné.

Nyní přistoupíme k diferenciaci výsledků mezi podsoubory mužů a žen. U slovních druhů, kde se prokázal rozdíl mezi pohlavími ve frekvenci výskytu sledovaných jazykových jevů (viz H1a a H2a), budeme ověřovat míru Extraverze u každého pohlaví zvlášť.

Nebudeme pracovat plošně se všemi slovními druhy a lingvistickými výrazy, ale pouze s těmi, u kterých se v H1a/b a H2a/b prokázal statisticky významný rozdíl mezi pohlavími.

Začneme u slovních druhů a využijeme opět Spearmanův korelační koeficient, jelikož budeme měřit lineární vztah mezi mírou Extraverze a frekvencí výskytu daného slovního druhu. Naše data nemají normální rozdělení, proto využíváme neparametrický korelační koeficient a jeho test statistické významnosti. Hladinu významnosti jsme zvolili $\alpha=0,05$.

Nejprve se podíváme na podsoubor žen (Tabulka 16) a ověříme četnost výskytu těch slovních druhů, kde se v H1a objevil rozdíl mezi pohlavími, a budeme sledovat jejich závislost na míře Extraverze (viz Tabulka 16). V tabulce jsou uvedeny hodnoty korelačního koeficientu r . Pokud se hodnota r blíží k +1, indikuje to silnou pozitivní lineární korelaci, pokud se blíží k -1, indikuje to silnou negativní lineární korelaci.

Tabulka 16: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností výskytu slovních druhů v podsouboru žen

Ženy Slovní druhy	Spearmanovy korelační koeficienty			
	Extraverze	Energičnost	Asertivita	Sociabilita
Předložky	0,107	0,075	0,064	0,127*
Spojka souřadící	0,115	0,137*	0,126*	0,045
Determinující výrazy/zájmena	0,060	-0,005	0,045	0,081
Zájmena	0,048	0,073	0,018	0,024

*korelační koeficient je významný na 5% hladině významnosti

Z uvedených výsledků vyplývá, že v podsouboru žen se souběžný vztah u pohlaví a Extraverze projevil u souřadících spojek, a to na subškálách Energičnosti a Aserivity.

Se vzrůstající mírou Sociability užívají ženy více předložek (což je slovní druh, který se v komparativní studii rozdílů mezi pohlavími statisticky významně prokázal jako častější u mužů (viz Tabulka 13). Nabízí se tady úvaha, co vede ženy s vysokou mírou Sociability k tomu, že používají více předložek a předložkových frází (viz kapitola 9).

Ženy s vyšší mírou Energičnosti a Aserivity používají více spojek souřadících (což je slovní druh, který se v komparativní studii rozdílů mezi pohlavími častěji vyskytuje u žen (viz Tabulka 13).

Dále se podíváme na četnost výskytu sledovaných slovních druhů v závislosti na míře Extraverze v podsouboru mužů (viz Tabulka 17). Opět budeme brát v úvahu pouze slovní druhy, které vykazují statisticky významné vztahy mezi frekvencí jejich výskytu a pohlavím autora (H1b a H2b).

Tabulka 17: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností výskytu slovních druhů v podsouboru mužů

Muži Slovní druhy	Spearmanovy korelační koeficienty			
	Extraverze	Energičnost	Aserivita	Sociabilita
Předložky	-0,031	-0,022	0,065	-0,106
Spojky souřadící	-0,113	-0,131	-0,157	-0,009
Determinující výrazy/zájmena	0,022	0,025	-0,141	0,148
Zájmena	-0,063	-0,097	-0,096	0,021

*korelační koeficient je významný na 5% hladině významnosti

Výsledky můžeme interpretovat pomocí Spearmanova korelačního koeficientu. Na rozdíl od žen, frekvence užívání spojek souřadících u mužů naopak klesá a vykazuje, ač statisticky nevýznamný, vztah v negativním směru. Záporný směr korelací se potvrzuje i u předložek a zájmen, avšak nedosahuje požadovanou hladinu významnosti. Ostatní předpoklady se v podsouboru mužů tedy nepotvrdily.

Zajímavé však je, že se u mužů - na rozdíl od žen - vyskytují se vzrůstající měrou Energičnosti a Aserivity negativní korelace. Zatímco u žen naznačují (ač statisticky nevýznamné korelační koeficienty) pozitivní závislost mezi extraverzí a užíváním předložek, spojek souřadících a zájmen v 1.osobě, u mužů je tato závislost negativní.

To může naznačovat, že se v četnosti užívání předložek a spojek souřadících projevuje v silnější míře efekt pohlaví než osobnostních rysů. S rostoucí extraverzí můžeme očekávat u žen rostoucí míru užívání předložek, spojek souřadících a zájmen, u mužů můžeme očekávat pravý opak, tedy snížení využívání těchto slovních druhů.

Teď přistoupíme k analýze jednotlivých lingvistických jevů, které slovní druhy dále definují. Opět sledujeme korelace Extraverze zvláště v podsouboru mužů a žen u jevů, kde byl zjištěna souvislost s pohlavím v H1b a H2b.

Stejně jako v případě slovních druhů jsme i u konkrétních lingvistických jevů použili Spearmanův korelační koeficient. Pomocí korelace četnosti výskytu jednotlivých lingvistických jevů s mírou Extraverze jsme zjistili následující souvislosti, které shrnuje Tabulka 18.

Tabulka 18: Vztah mezi relativní četností sledovaných lingvistických jevů a mírou Extraverze na všech subškálách BFI-2

Slovní druh	Celý soubor Lingvistický jev	Spearmanovy korelační koeficienty			
		Extraverze	Energičnost	Asertivita	Sociabilita
Příd. jm.	pozitivní	-0,046	-0,043	-0,038	-0,007
	negativní	-0,012	-0,012	0,036	-0,040
	2. stupeň	0,067	0,071	0,077	0,038
	3. stupeň	-0,034	-0,022	-0,015	0,006
	1. stupeň	-0,055	-0,058	-0,049	-0,013
	pasivum	0,030	0,046	0,067	-0,007
Sloveso	přít. čas	0,025	0,039	0,010	0,004
	minulý čas	0,017	-0,024	-0,004	0,046
	budoucí čas	0,030	-0,005	0,030	0,020
	rod činný	0,043	0,030	0,029	0,027
	Tvar záporný	0,065	0,011	0,105	0,039
	Tvar kladný	0,054	0,028	0,014	0,064
	1. osoba pl	0,065	0,095	0,026	0,019
	1. osoba sg	0,085	-0,019	0,121*	0,076
	2. osoba pl	0,087	0,092	0,041	0,090
	2. osoba sg	-0,056	-0,001	-0,095	-0,031
	3. osoba pl	0,012	-0,016	-0,030	0,054
	3. osoba sg	-0,015	0,018	-0,007	-0,030
	Způsob oznamovací	0,024	0,038	0,009	0,005
	Způsob rozkazovací	0,019	0,018	-0,020	0,032
	Pomocné sloveso `být`	Přítomný čas	0,086	0,035	0,111*
Minulý čas		-0,037	-0,019	-0,024	-0,020
Budoucí čas		0,090	0,066	0,089	0,086
Rod činný		0,085	0,042	0,094	0,056
Tvar záporný		-0,042	-0,039	-0,017	-0,029
Tvar kladný		0,082	0,048	0,089	0,056
1. osoba pl		0,062	0,064	0,051	0,033
1. osoba sg		0,080	-0,021	0,061	0,120*
2. osoba pl		0,119*	0,146**	0,000	0,134*
2. osoba sg		-0,039	0,019	-0,086	-0,022
3. osoba pl		0,041	0,029	0,059	-0,004
3. osoba sg		0,078	0,040	0,094	0,056
Způsob oznamovací		0,110*	0,056	0,127*	0,070
Způsob rozkazovací		0,068	0,040	0,063	0,048
Způsob podmiňovací		-0,013	-0,066	-0,020	0,037
Zájmena	neurčité	-0,027	0,045	-0,005	-0,064
	záporné	0,116*	0,062	0,119*	0,058
	osobní	0,014	0,022	-0,022	0,018

	vztažné	0,021	0,006	0,020	0,024
	všeobecné/totalizátory	0,014	0,005	-0,003	0,028
	tázací	0,009	0,031	0,002	0,007
Determinující výrazy/zájmena	neurčité	0,000	-0,014	-0,028	0,038
	záporné	-0,066	-0,085	-0,042	-0,037
	osobní	0,075	0,062	0,030	0,081
	ukazovací/zvratné	0,074	0,021	0,083	0,073
	vztažné	-0,001	-0,069	-0,015	0,059
	1. osoba pl	0,089	0,121*	0,035	0,039
	1. osoba sg	0,147**	0,047	0,129*	0,163**
	2. osoba pl	0,158**	0,137*	0,092	0,152**
	2. osoba sg	-0,039	0,019	-0,086	-0,022
	3. osoba pl	0,016	0,000	-0,024	0,043
	3. osoba sg	0,005	0,035	0,001	-0,008
	přivlastňovací	0,071	0,062	0,027	0,075
Příslovce	2. stupeň	-0,089	-0,117*	-0,050	-0,057
	3. stupeň	0,018	-0,011	-0,013	0,048
	1. stupeň	-0,013	-0,060	0,039	0,006
	záporné	-0,005	-0,065	0,014	0,009
Intertextové odkazy	Citace	-0,028	0,021	-0,036	-0,027

*korelační koeficient je významný na 5% hladině významnosti

**korelační koeficient je významný na 1% hladině významnosti

Z výsledků korelační analýzy vyplývá, že míra Extraverze má s četností výskytu konkrétních lingvistických jevů statisticky významný vztah, zejména v kategorii determinujících výrazů, kde s mírou Extraverze vzrůstá počet přivlastňovacích zájmen v 1. osobě singuláru (*můj, moje*) $p=0,007$ a 2. osobě plurálu (*vy, vaše*) $p=0,004$. Obdobný statisticky významný vztah v pozitivním směru se objevuje v četnosti výskytu pomocného slovesa `být` v 1. osobě singuláru $p=0,029$ (*jsem, budu*) na subškále Sociability; a také v v 2. os. plurálu (*jste, budete*), se statisticky nevýznamnější korelací na škále Extraverze $p=0,007$. Tyto statisticky významné vztahy v užití 1. osoby singuláru (*můj, jsem*) a 2. osoby plurálu (*váš, jste*) se v pozitivním směru statisticky nejvýznamněji projevují na subškále **Sociability**. To může naznačovat stereotypní předpoklad, že se Extravertní jedinci s vysokou mírou Sociability budou orientovat na mezilidské vztahy, což se v jejich písemném projevu projeví vyšší frekvencí 1.os.sg (*já*) a 2.os.pl (*vy*).

V 2. osobě plurálu se u pom. slovesa `být` (*jste, budete*) statisticky významně projevil pozitivní vztah na škále **Energičnosti** $p=0,029$). Stejná tendence se potvrdila i u slovního druhu determinujících výrazů v podobě přivlastňovacích zájmen 2. os. plurálu (*vaše, vašimi*) $p=0,004$. U Energičnosti se jako statisticky významný jeví vztah s četností užití komparativu u příslovčí (*lépe, strměji*) $p=0,032$, což může souviset s častější mírou srovnávání a relativizace formou lingvistického komparativu.

1. osoba singuláru u pomocného slovesa `být` (*jsem, budu*) má nejsilnější statistický významný vztah v subškále **Asertivity** $p=0,007$, což zase nasedá na stereotypní charakterový rys sebeprosazení. U asertivních jedinců se také projevil statisticky významný vztah v pozitivním směru v používání oznamovacího způsobu $p=0,019$

a přítomného času pom. slovesa `být` $p=0,042$ (*jdu, umíte, ověřuji*). Asertivita pozitivně koreluje s užíváním záporných zájmen $p=0,029$ (*nemalý, nelehký*).

Statistická významnost uváděných vztahů byla pozorována na 1% a 5% hladině, tyto hodnoty jsou značeny systémem hvězdiček přímo do tabulky.

Pro zpřesnění analýzy souvislosti mezi mírou Extraverze užitými jazykovými prostředky provedeme opět statistickou korelační analýzu konkrétních lingvistických jevů zvláště v podsouboru žen a mužů, a to v případech, kde se vyskytly statisticky významné rozdíly v závislosti na pohlaví u H1b a H2b.

Tabulka 19: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností užití jednotlivých lingvistických jevů v podsouboru žen

Slovní druh	Ženy		Spearmanovy korelační koeficienty			
	Lingvistické jevy	Extraverze	Energičnost	Asertivita	Sociabilita	
Příd. jm.	Tvar záporný	-0,014	-0,024	0,054	-0,047	
Sloveso	Čas přítomný	0,052	0,052	0,057	0,020	
	Rod činný	0,045	0,038	0,047	0,017	
	Tvar kladný	0,071	0,050	0,044	0,065	
	Způsob oznamovací	0,052	0,049	0,056	0,020	
	Způsob rozkazovací	0,055	0,029	0,021	0,057	
	Pomocné slov. `být`	Rod činný	0,087	0,022	0,126*	0,051
	3. os. číslo množné	0,048	0,019	0,068	0,002	
	Způsob oznamovací	-0,032	-0,080	-0,018	0,018	
Zájmena	osobní	0,023	0,046	0,002	0,007	
Determin. výrazy/zájmena	všeobecné/totalizátory	0,095	0,013	0,136*	0,065	
	1.os. číslo množné	0,105	0,138*	0,076	0,030	
	3.os. číslo množné	0,009	-0,010	-0,031	0,043	
Příslovce	1. stupeň pozitiv	-0,011	-0,064	0,038	0,005	
Intertext. odkazy	Citace	-0,206	-0,119	-0,166	-0,147	

*korelační koeficient je významný na 5% hladině významnosti

V podsouboru žen se v závislosti na vzrůstající míře Asertivy zvyšuje výskyt činného rodu u pomocného slovesa `být` (*jsem, je, jsi*), narůstá počet všeobecných zájmen/totalizátorů (*každý, všichni*) a na subškále Energičnosti vzrůstá frekvence výskytu 1. osoby množného čísla ukazovacích determinujících výrazů (*náš, našimi*).

Nyní se podíváme, jakou souvislost má míra Extraverze s jevy, kde se prokázal rozdíl mezi pohlavími H2b v podsouboru mužů. Vztahy budeme opět ověřovat pomocí Spearmanova koeficientu.

Tabulka 20: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností užití lingvistických jevů v podsouboru mužů

Slovní druh	Muži		Spearmanovy korelační koeficienty			
	Lingvistické jevy	Extraverze	Energičnost	Asertivita	Sociabilita	

Příd. jm.	Tvar záporný	0,028	0,068	-0,089	0,049
Sloveso	Čas přítomný	-0,135	-0,025	-0,188	-0,096
	Rod činný	-0,003	-0,056	-0,018	0,071
	Tvar kladný	-0,087	-0,115	-0,128	0,029
	Způsob oznamovací	-0,132	-0,018	-0,190	-0,092
	Způsob rozkazovací	-0,133	-0,041	-0,271*	-0,016
Pomocné slov. `být`	Rod činný	0,035	0,159	-0,082	0,030
	3. os. číslo množné	-0,035	0,092	-0,008	-0,123
	Způsob oznamovací	0,056	-0,018	-0,008	0,148
Zájmena	osobní	-0,043	-0,074	-0,124	0,068
Determin. výrazy/zájmena	všeobecné/totalizátor y	-0,034	-0,001	-0,123	0,082
	1.os. číslo množné	0,012	0,035	-0,098	0,041
	3.os. číslo množné	0,067	0,100	0,054	0,003
	1. stupeň	0,001	-0,027	0,026	0,034
Intertext. odkazy	Citace	0,012	0,056	0,007	-0,002

*korelační koeficient je významný na 5% hladině významnosti

V podsouboru mužů se na subškále Aseritivity prokázal opačný trend. Zatímco u žen korelační koeficienty ukazují statisticky nevýznamné, ale pozitivní vztahy, v podsouboru mužů se prokazují spíše vztahy negativní.

Jako statisticky významný se prokázal pouze jediný negativní vztah mezi Aseritivou a rozkazovacím způsobem (*porovnejme, představte si*). Čím vyšší míru Aseritivity muži vykazují, tím méně užívají rozkazovacího způsobu.

Zajímavé je, že s rostoucí mírou Aseritivity u žen rozkazovací způsob naopak roste. Tyto opačné trendy se projevují v řadě dalších vztahů, např. u záporu přídavných jmen (*nelehký, nezvyklý*), u sloves v kladném tvaru, u přítomného času a oznamovacího způsobu (*chodí, jde*), u osobních zájmen (*jej, ho, se*), u příslovčí v 1. stupni (*dobře, rychle*) i v počtu nepřímých citací. Přestože jsou uvedené korelační vztahy pod hladinou významnosti, přesto naznačují určitý trend, a sice že Aseritivita koreluje pozitivně s uvedenými jazykovými jevy v podsouboru žen a negativně v podsouboru mužů. Můžeme tedy předpokládat, že Aseritivita bude mít s četností výskytu těchto jevů větší souvislost než pohlaví respondentů.

Pro ověření tohoto předpokladu ověříme míru účinku mezi pohlavím a mírou Extraverze.

8.4 Míra účinku pohlaví a Extraverze na frekvenci výskytu sledovaných jazykových jevů

Tato kapitola se pojí s výzkumným záměrem 3 a budeme v ní pracovat s mírou účinku, tedy ověřením, zda na četnost užití sledovaných lingvistických jevů má silnější efekt pohlaví respondentů nebo jejich skóre na škále Extraverze.

Rozdíl mezi pohlavími se objevil u slovních druhů pomocí ověření H1a (viz jejich výčet v Tabulce 13), a u lingvistických jevů pomocí ověření H2a (výčet shrnuje Tabulka

15), více o souvislostech mezi pohlavím a četností užitých jazykových prostředků viz kapitola 8.2.

Nyní bude naším cílem u těchto jazykových jevů, kde se vyskytují statisticky významné vztahy zároveň mezi pohlavím i mírou Extraverze, zjistit míru účinku. Pomocí míry účinku ověříme, zda má na frekvenci výskytu uvedených jazykových prostředků větší efekt pohlaví nebo míra Extraverze. Pro tento výzkumný záměr jsme stanovili následující hypotézu:

H3: Pohlaví respondentů má ve srovnání s mírou Extraverze vyšší míru účinku na četnost užívání sledovaných lingvistických jevů.

U této hypotézy zkoumáme pouze slovní druhy a lingvistické jevy, u kterých byla prokázána souvislost u pohlaví i Extraverze zároveň (pomocí H1a a H2a u slovních druhů; a pomocí H1b a H2b u lingvistických jevů, které slovní druhy doplňují). Míru Extraverze jsme u proměnných, kde se zároveň ukázal korelační vztah u pohlaví a Extraverze, ověřovali zvláště v podsouboru mužů a žen (Tabulky 16 a 17 pro slovní druhy; Tabulky 19 a 20 pro lingvistické jevy).

Na základě těchto měření jsme určili lingvistické jevy, které ve frekvenci výskytu vykazují statisticky významné korelace jak ve vztahu k pohlaví, tak ve vztahu k míře Extraverze. V tabulce 21 je jejich výčet a jsou zastoupeny jak slovní druhy (předložky a spojky souřadící), tak lingvistické jevy (rozkazovací způsob u sloves, 1. osoba čísla množného u determináčních výrazů přivlastňovacích (*naš, našich*) nebo totalizátorů (*veškerý, každý*) atd.) Ve stejné Tabulce 21 jsou zaznamenány míry účinku pohlaví a Extraverze na užívání slovních druhů a lingvistických jevů dle zvoleného kritéria testování. Pro t-test je mírou účinku Cohenovo d , pro Mann-Whitney U test je to η^2 , který lze přepočíst na Cohenovo d (Soukup, 2013), a pro korelační koeficient je to korelační koeficient na druhou.

Tabulka 21: Míra účinku pohlaví a jednotlivých subškál osobnostního rysu Extraverze

Míra účinku	Pohlaví	Energičnost	Asertivita	Socibilita
Předložky	$d=0,526$			$r^2_{\zeta}=0,016$
Spojka souřadící	$d=0,25$	$r^2_{\zeta}=0,019;$	$r^2_{\zeta}=0,016;$	
Slovesa způsob rozkazovací	$\eta^2=0,011$ ($d=0,21$)		$r^2_M=0,073$	
Pomocné sloveso být rod činný	$\eta^2=0,014$ ($d=0,24$)		$r^2_{\zeta}=0,016$	
Det. výrazy totalizátory	$\eta^2=0,021$ ($d=0,29$)		$r^2_{\zeta}=0,018$	
Det. Výrazy 1. os množné číslo	$\eta^2=0,013$ ($d=0,23$)	$r^2_{\zeta}=0,019$		

$d=Cohenovovo\ d$

Tabulka 21 tedy znázorňuje míry účinku příslušející dvěma statistickým testům, neparametrický Man-Whitney U test (η^2) a parametrický t-test (Cohenovo d), které zjišťují praktické významnosti rozdílu mezi skupinami. Dále obsahuje míry účinku

příslušející Spearmanovým korelačním koeficientům (r^2), které zjišťují praktickou významnost souvislosti dvou metrických veličin.

Každý z těchto ukazatelů míry účinku je bezrozměrný a má jinou stupnici hodnocení, což může být náročné na interpretaci. Míra účinku pro **korelační koeficienty** je r^2 a vychází v hodnotách od 0 do 1. Nula znamená, že mezi veličinami neexistuje žádná závislost; 1 znamená 100 % lineární závislost.

Cohenovo d přibližně s hodnotou 0,2 značí malou úroveň účinku, s hodnotou přibližně 0,5 značí střední úroveň účinku a s přibližnou hodnotou 0,8 značí velkou úroveň účinku. Interpretace hodnot **eta²** je následující: η^2 přibližně 0,01 = malá úroveň účinku; η^2 přibližně 0,06 = střední úroveň účinku; a η^2 přibližně 0,14 = velká úroveň účinku.

Hodnoty eta² byly za účelem možnosti srovnání převedeny na Cohenovo d pomocí psychometrického kalkulátoru¹¹ tak, aby byly hodnoty jednotlivých ukazatelů vzájemně porovnatelné (Soukup, 2013).

V Tabulce 21 jsou zvláště uvedeny hodnoty pro muže (M) a hodnoty pro ženy (Ž) tam, kde se prokázal statisticky signifikantní vztah mezi pohlavím i mírou Extraverze u sledovaného jazykového jevu v obou podsouborech (viz H1a, b a H2a, b).

Z uvedených výsledků můžeme vysledovat pouze slabé hodnoty, které vesměs poukazují na malé úrovně efektů. Nejvýraznější střední míra účinku ($>0,5$) se prokázala u předložek (*na, u před*), které mírou středního efektu ovlivňuje více pohlaví autora než jeho míra Extraverze na subškále Sociability v podsouboru mužů (M).

Podobná situace je u spojek souřadících, kde se prokázal pouze malý efekt ($>0,2$) pohlaví v podsouboru žen. Efekty subškál Extraverze jsou velmi malé. Jejich hodnoty jsou velmi blízké nule.

U konkrétních lingvistických jevů je situace podobná jako u slovních druhů, míra Extraverze v každé z uvedených subškál vykazuje na sledovaných lingvistických jevech pouze nízkou míru účinku, její hodnoty jsou blízké nule. Silněji se zde projevuje efekt pohlaví, který se nejsilněji prokázal slabou měrou účinku ($>0,2$) u determinujících výrazů, jejichž kategorie v češtině nejlépe odpovídají zájmenům všeobecným (*každý, veškerý*) a přivlastňovacím v 1. osobě množného čísla (*naš, našimi, našich*).

Pohlaví má tedy jednoznačně silnější účinek na užívání předložek a spojek souřadících než Extraverze. Dle výsledků prezentovaných v Tabulce 13 můžeme připomenout, že předložky používají výrazně častěji muži než ženy, a naopak spojky souřadící využívají častěji ženy než muži.

U lingvistických jevů je účinek pohlaví i Extraverze slabý, avšak podobný. Dle naměřených hodnot měr účinku je o něco silnější efekt pohlaví. Z výsledků prezentovaných v Tabulce 14 můžeme vyčíst, že determinující výrazy všeobecné/totalizátory i 1. osobu množného čísla používají častěji ženy než muži.

Na závěr srovnáme specifické hypotézy s výsledky ostatních studií.

8.5 Srovnání našeho souboru s výsledky předchozích studií

Tato kapitola obsahuje konkrétní hypotézy spojené s výzkumným záměrem 4 a klade si za cíl pomocí korelačních vztahů: ověřit, zda se v našem souboru potvrdí výsledky předchozích psycholingvistických studií na českém jazyce, zejména Dalibora Kučery

¹¹ Odkaz na psychometrický kalkulátor: https://www.psychometrica.de/effect_size.html [citováno z 30.3.2024]

(2020 a 2022) a zahraničních výzkumů (Pennebaker, 2013; Pennebaker a Kings, 1999; Pennebaker a Stone, 2003, Pennebaker et al., 2003).

Za účelem ověření těchto hypotéz jsme využívali Pearsonův korelační koeficient, který měří, jak těsně jsou pozorované hodnoty dvou proměnných lineárně propojeny za předpokladu, že nemají normální rozdělení. Koeficient r se pohybuje mezi -1 a $+1$ a je ukazatelem míry a směru závislosti dvou proměnných, míra závislosti se může projevit v kladném nebo záporném směru, kde 0 znamená nulovou, tedy žádnou závislost.

H4a: Jedinci skórující vysoko na škále Extraverze používají méně záporu v písemném projevu.

Hypotéza H4a je formulována k ověření, zda se v našich datech projeví závěry vyplývající z psycholingvistické studie Rude et al. (2004) zaměřující se na výzkum korelací mezi negativní emocionalitou a četností užití záporu a negace ve formálních textech amerických studentů. Z uvedené práce vyplývá, že četnost užití záporu pozitivně koreluje s negativní emocionalitou. Tyto výzkumy nepřímo potvrzují i Pennebaker a King (1999), kteří dokumentují statisticky významné pozitivní korelace mezi depresivitou a četností negativně emočně nabitých slov. Předpokladem formulace hypotézy H4a je očekávání, že Extraverze pojící se spíše s pozitivní emocionalitou (viz kapitola 3.1.4.3) bude se záporům a lingvistickými znaky negace korelovat spíše záporně.

Vycházíme z předpokladu, že jedinci se sklonem k extraverci mají tendenci vyjadřovat svoje stanoviska přímým způsobem a jsou pozitivně nastaveni ke spolupráci se svým okolím, je tudíž pravděpodobné, že se budou vyjadřovat spíše pozitivním způsobem (Pennebaker, et al., 2003; Kučera, 2022).

S tímto ohledem byl způsob vyjadřování negace rozdělen do kategorií podle příslušných slovních druhů, budeme tedy rozlišovat mezi slovesným záporům (*nechceme, nesměl*); záporným tvarem přídavných jmen (*nehezky, nedůležitý*) a příslovci (*nepozorovaně, nedbale*); zápornými determinujícími výrazy (*žádný, žádnému*) a zápornými zájmeny (*nikdo, nic*).

Tabulka 22: Četnost výskytu záporu a negace u jednotlivých slovních druhů v závislosti na míře Extraverze

Celý soubor		Spearmanovy korelační koeficienty			
Slovní druh	Lingvistický jev	Extraverze	Energičnost	Asertivita	Sociabilita
Slovesa	Zápor	0,029	-0,016	0,076	0,009
Příd. jm.	Zápor	-0,012	-0,012	0,036	-0,040
Příslovce	Zápor	-0,005	-0,065	0,014	0,014
Determinující výrazy	Záporné	-0,066	-0,085	-0,042	-0,042
Zájmena	Záporná	0,116*	0,062	0,119*	0,119*

*korelační koeficient je významný na 5% hladině významnosti

Výsledky v Tabulce 22 naznačují negativní směr vztahu mezi mírou a Extraverze a negací, přestože jsou statisticky nevýznamné. V případě záporného tvaru zájmen se četnost jejich užití s mírou Extraverze naopak zvyšuje, což se nejvíce projevuje v subškále Asertivita a Sociabilita. Asertivnější a sociabilnější jedinci mají tendenci užívat více záporných zájmen (*nikdo, nic*) ve svém písemném projevu. Tyto výsledky ještě ověříme v podsouborech mužů a žen.

Tabulka 23: Četnost výskytu záporu a negace u jednotlivých slovních druhů v závislosti na míře Extraverze v podsouboru mužů

Slovní druh	Muži Lingvistický jev	Spearmanovy korelační koeficienty			
		Extraverze	Energičnost	Asertivita	Sociabilita
Slovesa	zápor	-0,055	-0,058	-0,162	0,061
Příd. jm.	zápor	0,028	0,068	-0,089	0,049
Příslovce	zápor	0,085	0,032	0,223	-0,028
Determinující výrazy	záporné	-0,100	0,009	-0,317*	0,054
Zájmena	záporná	-0,044	-0,016	-0,014	-0,062

*korelační koeficient je významný na 5% hladině významnosti

V podsouboru mužů byl potvrzen jeden významný korelační vztah, a to mezi užíváním determinujících negativních výrazů a Asertivity. Tato závislost je negativní, čím více jsou muži asertivní, tím méně používá ve svém písemném projevu negativní determinující výrazy. Ostatní výsledky naznačují negativní směr vztahu mezi mírou Extraverze a četností užití záporu v textech mužských respondentů, avšak neprokázal se také statisticky významný.

Positivní směr korelace se ovšem ukázal v podsouboru žen.

Tabulka 24: Četnost výskytu záporu a negace u jednotlivých slovních druhů v závislosti na míře Extraverze v podsouboru žen

Slovní druh	Ženy Lingvistický jev	Spearmanovy korelační koeficienty			
		Extraverze	Energičnost	Asertivita	Sociabilita
Slovesa	zápor	0,052	-0,008	0,118*	0,004
Příd. jm.	zápor	-0,014	-0,024	0,054	-0,047
Příslovce	zápor	-0,021	-0,080	-0,014	0,018
Determinující výrazy	záporné	-0,058	-0,108	0,010	-0,058
Zájmena	záporná	0,149*	0,079	0,145*	0,082

*korelační koeficient je významný na 5% hladině významnosti

U žen byla potvrzena statisticky významná závislost v pozitivním směru mezi četností užití slovesného záporu (*nemohu, nelze*), negativních zájmen (*nikdo, nic*) a mírou Extraverze na subškále Asertivity. Čím asertivnější tedy ženy jsou, tím více slovesné negace užívají ve svém písemném projevu. Dále se prokázala pozitivní korelace mezi negativními zájmeny a celkovou mírou Extraverze.

Vzhledem k uvedeným výsledkům se naše očekávání nepotvrdila naše očekávání ani výsledky předchozích výzkumů a hypotézu H4a nemůžeme přijmout.

H4b: Jedinci skórující vysoko na škále Extraverze používají komplexnější delší a rozvitější souvětí.

Tato hypotéza byla stanovena na základě předpokladu, že osoby s vyšší mírou Extraverze mají tendenci k vyšší míře Sociability a budou se vyjadřovat pomocí rozvitých

vět a delších souvětí. Jinými slovy budou ve svém písemném projevu košatí a velkorysí (Pennenbaker, Mehl, Niederhoffer, 2003; Holtgraves, 2011; Kučera, 2022).

Větnou skladbu a rozvitost souvětí můžeme sledovat pomocí relativní četnosti spojek podřadicích (u spojení vět vedlejších) a spojek souřadících (u spojení vět hlavních). Pro zachycení dalších typů souvětí byla do analýzy zařazena i interpunkce (závorky, středník a pomlčky) a průměrný počet slov ve větě. Výsledky shrnuje Tabulka 25.

Tabulka 25: Souvislost mezi mírou Extraverze, délkou vět a větnou skladbou

Celý soubor	Spearmanovy korelační koeficienty			
	Lingvistický jev	Extraverze	Energičnost	Asertivita
Průměrná délka věty	-0,052	-0,081	-0,017	-0,031
Souřadící spojky	0,070	0,086	0,077	0,028
Interpunkce	-0,006	-0,030	0,001	-0,013
Podřadící spojky	-0,032	-0,042	-0,010	-0,040

Hypotéza H4b se nepotvrdila ani u jednoho sledovaného lingvistického jevu. Přestože jsou výsledky statisticky nevýznamné, naznačují spíše negativní směr vztahu, s rostoucí mírou Extraverze se tedy snižuje délka vět, četnost výskyt podřadicích spojek a interpunkce. Spojky souřadící se mohou naopak s rostoucí mírou Extraverze objevovat častěji. Tyto tendence však nejsou statisticky významné, a proto hypotézu H4b musíme zamítnout.

H4c: Jedinci skórující vysoko na škále Extraverze používají méně trpného rodu.

Trpný rod je v lingvistice spojován s možností vynechání agenta, přesněji řečeno s možností vynechání plnohodnotného původce děje neboli podmětu. Ve větě *`je řečeno`* neumíme dost dobře určit, kdo nebo co to říká, a proto jde o unikátní možnost vyjádřit se bez toho, aniž bychom označili původce děje jako někoho, kdo je za obsah výroku zodpovědný (Golková, 1983).

V psychologii je Extraverze spojována s vyšší mírou sociální jistoty, Asertivity a impulsivity (Eaves, Eysenck, 1975), lze tedy vycházet z předpokladu, že jedinci skórující výše na škále Extraverze budou mít tendenci vyjadřovat se přímo bez vynechání agenta (podmětu) nebo původce děje. V dalších studiích byl potvrzen zvýšený výskyt trpného rodu v písemném a formálním psaném projevu (Pennenbaker, et al., 2003; Kučera, 2022). Hypotéza H4c tedy předpokládá negativní směr závislosti pasivu na míře Extraverze.

Tabulka 26: Četnost užití trpného rodu u sloves v závislosti na míře Extraverze

Celý soubor	Spearmanovy korelační koeficienty			
	Extraverze	Energičnost	Asertivita	Sociabilita
Činný rod	-0,031	-0,037	-0,029	0,015
Trpný rod	0,030	0,046	0,067	-0,007

Ani jeden korelační koeficient nevyšel v našem souboru statisticky významný. Souvislost mezi četností užití trpného rodu a mírou Extraverze se tedy nepotvrdila. Zajímavé je, že ačkoliv jsou výsledky statisticky nevýznamné, naznačují opačný druh závislosti, než předpokládá stanovená hypotéza. Oproti očekávání, se v našem souboru projevuje statisticky nevýznamný, ale pozitivní vztah mezi užitím trpného rodu a mírou Extraverze. Tyto vztahy jsou však neprůkazné a hypotézu H4c musíme zamítnout.

H4d: Jedinci skórující nízko na škále Extraverze používají více podmiňovacího způsobu

Hypotéza H4d je postavena na psychologickém předpokladu, že jedinci s nízkou mírou Extraverze vykazují nižší míru Asertivita a sociální jistoty (Eaves, Eysenck, 1975), tím pádem budou ve formulaci svých výroků umírněnější a zdrženlivější. Podmiňovací způsob je v lingvistice spojován s vyšší měrou sociální nejistoty a nižší měrou Asertivita na straně mluvčího (Johnson-Laird, Byrne, 2002), dá se tedy předpokládat, že jedinci s nízkou měrou Extraverze se budou častěji ve svých textech uchýlovat k použití podmínkových vět (např. *bylo by dobré* místo *bude dobré*), a to právě z důvodu, že jim chybí sebejistota a sociální zdatnost.

Tabulka 27: Četnost užití podmiňovacího způsobu v závislosti na míře Extraverze

Celý soubor	Spearmanovy korelační koeficienty				
	Lingvistický jev	Extraverze	Energičnost	Asertivita	Sociabilita
Podmiňovací způsob		-0,012	-0,067	-0,019	0,038

Jak naznačují výsledky v Tabulce 27, existuje zde sice negativní vztah mezi užitím podmiňovacího způsobu a mírou Extraverze, není však statisticky signifikantní. Nemůžeme tedy na základě něho potvrdit závěr, že s klesající mírou Extraverze roste četnost užití podmiňovacího způsobu, jak naznačují výsledky předchozích výzkumů. Hypotézu H4d musíme tedy zamítnout.

H4e: Jedinci skórující nízko na škále Extraverze používají více komparativu, tedy 2. stupně přídavných jmen a příslovčí

Stanovení této hypotézy H4e se opírá o psychologický předpoklad, že jedinci s vyšší mírou Extraverze jsou více orientováni na své okolí, mají větší tendenci vytvářet si vztahové vazby ve vnějším světě a více se upínají na děje a procesy kolem sebe (Eaves, Eysenck, 1975). Z lingvistického hlediska je orientace na okolí podmíněna užíváním komparativu (tedy 2. stupně přídavných jmen, např. *lepší, horší*), neboť jde o způsob srovnání dvou jednotek nebo jevů mezi sebou; a pragmaticky se jedná o jazykový důkaz vnitřního postoje mluvčího k danému vztahu nebo situaci ve vnějším světě (Atlas, 1984).

V dřívějším výzkumu byl raportován negativní vztah mezi Extraverzí a užitím diference (méně než, lepší ve srovnání s) v písemném projevu (Pennebaker a King, 1999).

Použití komparativu tak může být právě průvodním jazykovým rysem, který se bude častěji vyskytovat u autorů, jejichž pozornost je primárně směřována na externí vztahy. Potřeba srovnávání však může zároveň dokládat i nižší míru sebejistoty na straně autora a jeho tendenci hledat ujištění v hierarchizaci vnější reality (Kay, 2006).

Tabulka 28: Četnost užití komparativu a superlativu v závislosti na míře Extraverze

Celý soubor		Spearmanovy korelační koeficienty			
Slovní druh	Lingvistický jev	Extraverze	Energičnost	Asertivita	Sociabilita
Příd. jm.	1. stupeň	-0,065	-0,070	-0,046	-0,020
	2. stupeň	-0,028	-0,041	0,010	-0,036
	3. stupeň	-0,023	-0,038	-0,015	0,027
Příslovce	1. stupeň	-0,013	-0,060	0,039	0,006
	2. stupeň	-0,089	-0,117*	-0,050	-0,057
	3. stupeň	0,018	-0,011	-0,013	0,048

*korelační koeficient je významný na 5% hladině významnosti

Souvislost mezi Extraverzí a užíváním komparativu u přídavných jmen se nepotvrdila. Ani jeden korelační koeficient nevyšel statisticky významný. Výsledky Tabulky 28 však naznačují negativní závislost mezi Extraverzí a užitím komparativu, což může nasvědčovat klesající potřebě srovnávání s rostoucí mírou Extraverze. Souvislost mezi užíváním 2. stupně příslovčí se prokázala u Energičnosti. Čím je člověk energičtější, tím méně často užívá 2. stupeň příslovčí. Na základě těchto výsledků však nemůžeme hypotézu H4e přijmout.

H4f: Jedinci skórující vysoko na škále Extraverze používají více osobních zájmen v 1.osobě singuláru.

Hypotéza H4f byla u samotného zrodu myšlenky o sepsání této práce. Je založena na stereotypním předpokladu, že jedinci s vysokou mírou Extraverze jsou orientovaní spíše na své okolí (Eaves, Eysenck, 1975), což se v jejich písemném projevu projeví častějším výskytem 1. osoby množného čísla (*my, náš, předpokládejme*). Oproti tomu jedinci s nízkou měrou Extraverze, kteří jsou orientovaní spíše dovnitř a jejich introvertní zaměření se projevuje tendencí zkoumat spíše vlastní myšlenky než vnější svět (Eaves, Eysenck, 1975), budou ve svém písemném projevu častěji využívat 1. osobu čísla jednotného (*já, mým, myslím si*).

Tyto úvodní předpoklady byly potom vyvráceny díky extenzivnímu studiu psycholingvistických výzkumů studujících souvislosti mezi osobností a jazykem. Bylo totiž prokázáno, že 1. osoba singuláru (*já, myslím, moje*) se opakovaně vyskytuje v pozitivních korelacích s osobnostními rysy jako je sebejistota (Jordan a Pennebaker, 2017), sociální dominance (Jordan et al., 2019), přívětivost (Summer et al., 2011), sociální angažovanost (Cegala, 1989) a loajalita (Kučera, 2020 a 2022), které společně naznačují, že bychom měli ve vztahu k Extraverzi očekávat podobný trend.

Ve výzkumu je však důležité zmínit i protiargumenty, které naše závěry prověří, zde je nutné uvést i studie, které reportují pozitivní korelační vztahy mezi výskytem 1. os. singuláru s negativními emočními stavy a neuroticismem (Pennebaker a King, 1999), depresivitou a emoční labilitou (Rude et al. 2004).

Následující Tabulka 29 shrnuje výsledky korelačních vztahů mezi užitím gramatických kategorií čísla a osoby v souvislosti s mírou Extraverze. Abychom pokryli gramatické a strukturální možnosti českého jazyka, osobu a čas zde sledujeme u několika slovních druhů: (i) pomocného slovesa `být` (*jsem, budu*), (ii) významových sloves (*umím,*

chci), (iii) u zájmen osobních (*já, mně, mnou*), a (iv) determinujících výrazů přivlastňovacích (*můj, moje, mého*).

Tabulka 29: Četnost výskytu 1-3. osoby singuláru a plurálu v písemném projevu ve vztahu k míře Extraverze daného autora

Celý soubor Lingvistický jev	Spearmanovy korelační koeficienty			
	Extraverze	Energičnost	Asertivita	Sociabilita
1.osoba, plural	0,089	0,121*	0,035	0,039
1.osoba, singular	0,147**	0,047	0,129*	0,163**
2.osoba, plural	0,158**	0,137*	0,092	0,152**
2.osoba, singular	-0,039	0,019	-0,086	-0,022
3.osoba, plural	0,016	0,000	-0,024	0,043
3.osoba, singular	0,005	0,035	0,001	-0,008

*korelační koeficient je významný na 5% hladině významnosti

** korelační koeficient je významný na 1% hladině významnosti

Z výsledků korelační analýzy shrnuté v Tabulce 29 jednoznačně vyplývá, že mezi mírou Extraverze a četností výskytu 1. os. jednotného čísla (*já, moje, udělám, budu*) existují statisticky významné vztahy v pozitivním směru $p=0,007$. Nejsilněji se tento vztah potvrdil na subškále Sociability $p=0,003$. H4f tedy na základě této analýzy můžeme přijmout.

Zastavíme se zde krátce ještě u dalšího vztahu, který stojí za povšimnutí, ačkoli se přesně netýká uvedené hypotézy. Jedná se o vztah mezi Extraverzí a četností užití 2. osoby plurálu (*vy, vaše, budete*), i v tomto případě se v našem souboru projevil statisticky pozitivní vztah na 1 % hladině významnosti. S rostoucí mírou Extraverze roste četnost užití 2. os. plurálu $p=0,004$, nejsilněji se tento vztah projevil opět na subškále Sociability $p=0,005$. Čím vyšší míru Extraverze, Energičnosti a Sociability jedinec vykazuje, tím častěji můžeme v jeho písemném projevu očekávat užívání 2. osoby množného čísla.

Oproti původnímu předpokladu, že extravertní jedinci se budou vyjadřovat v 1. osobě čísla množného (*my, jdeme*), se v našem souboru potvrdily výsledky předchozích studií, které indikují pozitivní vztah mezi 1. osobou čísla jednotného (*já, můj, jsem, mohu*) a vyšší mírou Extraverze, zejména na subškále Sociability.

8.6 Výsledky ověření platnosti statistických hypotéz

Zde si krátce shrneme výše uváděné hypotézy a zrekapitulujeme si jejich platnost s ohledem na výsledky kvantitativní analýzy prezentované v předchozích kapitolách 8.2. – 8.5.

Výzkumný záměr 1: ověřit pomocí komparativní analýzy, zda se relativní četnost sledovaných lingvistických jevů mění ve vztahu k pohlaví respondentů. Byly stanoveny dvě obecné hypotézy k plošnému screeningu datového souboru, cílem bylo zjistit, kterých slovních druhů a lingvistických jevů se rozdíl mezi pohlavími bude týkat.

H1a: Existuje rozdíl v četnosti užívání slovních druhů mezi muži a ženami

H1a: Nemůžeme přijmout plošně, protože se potvrdila pouze u některých slovních druhů.

Statisticky významný vztah mezi pohlavími v četnosti užívání jednotlivých slovních druhů se prokázal pouze u předložek (*na, pod, u*), spojek souřadících (*a, ale, nebo*) a zájmen (*jim, oni, který, všem*). Předložky používají významně častěji muži než ženy. Spojky souřadící a zájmena používají významně častěji ženy než muži.

H1b: Existuje rozdíl v četnosti užívání sledovaných lingvistických jevů mezi muži a ženami.

H1b: Nemůžeme přijmout plošně, protože se potvrdila pouze u některých lingvistických jevů.

Statisticky významný vztah mezi pohlavími v četnosti užívání jednotlivých lingvistických jevů se v podsouboru **mužů** prokázal u přídavných jmen v negativním tvaru (*nemilý, nechťěný*), rozkazovacím způsobu u sloves (*určete, uvaž*), ve 3. osobě množného čísla pomocného slovesa být (*jsou, byli*, v 1. stupni u příslovčí (*dobře, rychle*) a v 3. osobě množného čísla u determinujících výrazů (*jejich, jejím*). Tyto lingvistické jevy využívají významně častěji muži než ženy.

Statisticky významný vztah mezi pohlavími v četnosti užívání jednotlivých lingvistických jevů se v podsouboru **žen** prokázal u přítomného času u sloves (*nese, založime*), činného rodu (*jde, založili*), sloves v kladném tvaru (*splňují, směřovali*), oznamovacího způsobu (*hraje, zkoumáme*), činného rodu pomocného slovesa `být` (*budou, je*), oznamovacího způsobu (*je, byl*), zájmen osobních (*ji, vás, jim, si, se*), determinujících výrazů všeobecných (*všechn, každý*) a v 1. osobě množného čísla (naše, našimi). Tyto lingvistické jevy využívají významně častěji ženy než muži.

Výzkumný záměr 2: ověřil pomocí korelačních vztahů, zda se relativní četnost sledovaných lingvistických jevů mění ve vztahu k míře Extraverze jednotlivých autorů textů.

H2a: Existuje statistická závislost mezi četností užívání slovních druhů a mírou Extraverze.

H2a: Nemůžeme přijmout, protože se potvrdila pouze u jediného slovního druhu.

Statisticky významný vztah mezi mírou Extraverze a četností výskytu jednotlivých slovních druhů se prokázal pouze u předložek. Tento kladný korelační vztah je statisticky nejvýznamnější v subškále Sociability. V ostatních subškálách se žádný vztah neprojevil.

H2b: Existuje statistická závislost mezi četností užívání sledovaných lingvistických jevů a mírou Extraverze.

H2b: Nemůžeme přijmout plošně, protože se potvrdila pouze u některých lingvistických jevů.

Statisticky významný vztah mezi mírou Extraverze a četností výskytu jednotlivých lingvistických jevů se prokázal u determinátorů/zájmen přivlastňovacích v 1. osobě singuláru a 2. osobě plurálu, u pomocného slovesa `být` také 1. osobě singuláru a 2. osobě plurálu, dále v kategorii oznamovacího způsobu a přítomného času pom. slovesa `být`, u 2. stupně příslovčí (komparativu), a u zájmen záporných.

Výzkumný záměr 3: ověřit míru účinku pohlaví a Extraverze na četnost výskytu sledovaných jazykových proměnných.

H3: Pohlaví respondentů má ve srovnání s mírou Extraverze vyšší míru účinku na četnost užívání sledovaných lingvistických jevů.

H3: Přijímáme, u pohlaví se potvrdila vyšší míra účinku u všech sledovaných jazykových proměnných.

Ověřovali jsme míru účinku pouze u těch jazykových proměnných, u kterých se prokázal statisticky významný vztah zároveň ve vztahu k pohlaví a k Extraverzi. U lingvistických jevů, u kterých jsme sledovali praktickou významnost (viz kapitola 8.4.), je účinek pohlaví i Extraverze slabý. Avšak dle hodnot měř účinku je o něco silnější efekt pohlaví.

Výzkumný záměr 4: ověřit pomocí korelačních vztahů, zda se v našem souboru potvrdí výsledky předchozích psycholingvistických studií. (Kučera, 2020 a 2022; Pennebaker a Stone, 2003; Pennebaker a King, 1999; Yarkoni, 2010; Lee et al., 2007; Cegala, 1989; Summer et al. 2011; Argamon et al., 2007; Rao et al. 2010; Mukherje a Liu, 2010; Holtgraves, 2011; Johnson-Laird, Byrne, 2002).

H4a: Jedinci skórující vysoko na škále Extraverze používají méně záporu v písemném projevu.

H4a: Nemůžeme přijmout, naměřili jsme statisticky signifikantní vztah v opačném směru.

V našem souboru byl prokázán jediný statisticky významný vztah mezi mírou Extraverze a frekvencí záporu, a to v kladném směru v kategorii záporných zájmen. V podsouboru mužů se jako statisticky významný projevil záporný vztah k frekvenci užití záporných determinačních výrazů, a sice ve formě negativních přivlastňovacích zájmen. V podsouboru žen se objevil kromě zájmen ještě pozitivní vztah k četnosti užití záporu u sloves.

H4b: Jedinci skórující vysoko na škále Extraverze používají komplexnější delší a rozvitější souvětí.

H4b: Nemůžeme přijmout, nepotvrdila se ani u jednoho sledovaného lingvistického jevu.

H4c: Jedinci skórující vysoko na škále Extraverze používají méně trpného rodu.

H4c: Nemůžeme přijmout, nepotvrdila se ani u jednoho sledovaného lingvistického jevu.

H4d: Jedinci skórující nízko na škále Extraverze používají více podmíňovacího způsobu

H4d: Nemůžeme přijmout, nepotvrdila se ani u jednoho sledovaného lingvistického jevu.

H4e: Jedinci skórující nízko na škále Extraverze používají více komparativu, tedy 2. stupně přídavných jmen a příslovčí

H4e: Nemůžeme přijmout, naměřili jsme statisticky signifikantní vztah v opačném směru.

V našem souboru byl prokázán jediný statisticky významný vztah mezi mírou Extraverze a 2. stupněm komparativu, a to u slovního druhu příslovčí v negativním směru.

H4f: Jedinci skórující vysoko na škále Extraverze používají více osobních zájmen v 1. osobě singuláru.

H4f: Přijímáme, potvrdili jsme statisticky signifikantní vztah mezi 1. osobou singuláru a mírou Extraverze v pozitivním směru.

Na našem souboru se potvrdila šetření předchozích psycholingvistických studií v statisticky významném kladném vztahu mezi mírou Extraverze a užitím 1. osoby singuláru. Současně se prokázal statisticky významný kladný vztah i k užití 2. osoby plurálu. Oba vztahy se prokázaly na 0,01 hladině významnosti.

9 DISKUZE

Nyní přistoupíme k diskusi nad způsobem zpracování dat, prezentací statistických výsledků, z nich vyvozovaných závěrů a problémy či výzvami, které jsme během výzkumu museli řešit. Protože získávání psychometrických proměnných z inventáře BFI-2 je v psychologickém výzkumu velmi dobře zdokumentováno (Hřebíčková et al. 2020), budeme se soustředit na popis lingvistických proměnných.

V první řadě musíme vzít v úvahu velikost ($n=336$) a složení našeho statistického souboru (studenti 1. ročníku psychologie), jeho rozsah je omezený a skladba příliš homogenní na to, abychom mohli z této práce vyvozovat jakékoli obecné závěry.

Kučera (2020 a 2022; Kučera et al., 2020) ve svých výzkumech došel k podstatnému zjištění, a sice že žánrová variabilita textů, která měla původně sloužit ke kontrole falešně pozitivních zjištění, zvyšuje variabilitu lingvistických dat, protože textový žánr velice významně souvisí s frekvencí a způsobem užití sledovaných jazykových prostředků (Kučera, 2020, s. 108). V naší studii jsme proto od žánrové variace upustili a omezili se na formální akademický styl psaných esejí.

Hlavním **cílem** této studie je prozkoumat možnosti psycholingvistického výzkumu ve smyslu extrakce jazykových prostředků a jejich výpovědní hodnoty o struktuře osobnosti. V tomto smyslu jsme se snažili navázat na tradici psycholingvistického výzkumu v českém prostředí (Kučera a Haviger, 2014; Havigerová et al., 2019, Kučera a Havigerová, 2015; Kučera et al., 2022).

Mnohá omezení jsme zjistili hned při strojovém zpracování jazyka, kdy výsledné kategorie neodpovídají gramatické stavbě češtiny nebo jí odpovídají jen částečně. Tento deficit jsme se snažili kompenzovat složitou extrakcí takových jazykových proměnných, a jejich kombinací, které by alespoň zčásti splňovaly nároky mluvnického systému češtiny.

Kučera (2020, s. 115) podotýká, že „*operacionalizace jazykových rysů (morfologických či stylistických) v psychologickém výzkumu nemá příliš velké zázemí*“. S tímto problémem jsme se nepotýkali jen my, ale týká se řady výzkumů, někteří lingvistické proměnné stanovují na základě dostupných kategorií ze strojového zpracování jazyka (Kučera, 2020), jiní je stanovují náhodně nebo účelově (Pennebaker, 2013; Pennebaker a King, 1999).

Obecně ovšem platí, že jazyk je natolik komplexním systémem proměnných, že je téměř nemožné izolovat jednotlivé prvky tak, aby je bylo možné zkoumat odděleně a ze statistického pohledu vytvořit jasné operacionalizovatelné definice (Jelínek, 2018; Oberlander a Gill, 2004; Tauszik a Pennebaker, 2010).

S tímto ohledem jsme formulovali i **výzkumné hypotézy**, které byly pojaty obecně a vysloveny s ohledem na holistické pojetí výzkumu jazykových prostředků – tedy ne pouze pomocí izolovaného pozorování jednotlivých, na sobě zdánlivě nezávislých, jevů; ale právě s ohledem na zachycení komplexních fenoménů, které pomocí analýzy přirozeného jazyka můžeme v textech pozorovat a díky pochopení vnitřních jazykových mechanismů i popsat a správně interpretovat ve vztahu k danému rysu osobnosti, v našem případě tedy k míře Extraverze.

Pro ilustraci zde můžeme zmínit kategorii mluvnického záporu (H4a v kapitole 8.5). V rámci strukturálních možností může jazyk vyjadřovat zápor několika způsoby: (i) slovesným (*není pravděpodobné*), (ii) pomocí negativního tvaru přídavných jmen (*je nepravděpodobné*), a (iii) jejich kombinací (*není nepravděpodobné*). Tyto jazykové prostředky jsou v různých jazykových systémech různě omezené a jejich kombinace různě

gramaticky podmíněné, například česká věta (*nikdo nic nechce*) bude v angličtině odpovídat pouze jednomu záporu (*nobody wants anything*).

Ve jménu smysluplnosti výzkumu tedy není možné mnohavrstevnost a mnohotvárnost jazykových prostředků omezovat do striktních kategorií, které potom postrádají smysl (Firbas, 1964; Chamonikolasová, 2005), navíc v kros-lingvistickém pojetí stejně zůstanou nesrovnatelné (Veselovská, 2014; Kučera a Mehl, 2022).

Jedním z výzkumných záměrů bylo ověřit výsledky předchozích studií na našem datovém souboru (H4a-H4f v kapitole 8.5), což nedopadlo příliš úspěšně. Jednou z možných příčin může být právě variabilita jazykového systému:

„Klíčovou podmínkou souměřitelnosti jazykových rysů je jejich variabilita v obou jazycích. Pokud rys variuje, je možné uvažovat o jeho funkci. Z hlediska jednoduché deskripce jazyka mohou mít dva jazyky stejný či podobný repertoár rysů (např. kategorií substantiv, zájmen, rodů, čísla, osoby apod.), avšak mohou ho využívat funkčně odlišně“ (Kučera, 2020, s. 217).

V psycholingvistickém výzkumu obecně dochází opakovaně k situaci, kdy ve výsledcích korelačních studií vychází slabé, ale signifikantní vztahy (Pennebaker a Stone, 2003; Pennebaker a King, 1999; Kučera 2022; Yarkoni, 2010; Lee et al., 2007). Zde je také nutné zdůraznit, že jde ve směr o **frekvenční analýzu jazykových prostředků**.

Kučera (2020, s. 92) ve svém výzkumu poukazuje na fakt, že ačkoliv se slabé statistické vztahy prokazují napříč mnoha studii, může to souviset s prezentací výsledků pomocí p-hodnoty, nejčastěji s hladinou významnosti 0,05, bez uvedení dalších ukazatelů a využití pokročilejších metod zpracování dat.

V naší analýze jsme se drželi pouze základních korelačních postupů, avšak pro prezentaci výsledků z našeho souboru jsme zvolili hodnoty korelačních koeficientů, i když jsme odstoupili od navrhované statistické korekce falešně pozitivních výsledků (Haviger, 2018).

Samotné kvantitativní zpracování dat sestává z několika kroků rozdělených dle formulovaných výzkumných záměrů. Statistickou analýzou jsme se snažili odpovědět na **výzkumnou otázku**, zda existuje souvislost mezi mírou Extraverze a způsobem užití lingvistických prostředků v písemném projevu.

Prvním výzkumným záměrem je pomocí **komparativní analýzy** ověřit, zda se relativní četnost sledovaných jazykových jevů mění v závislosti na **pohlaví** respondentů (H1a, H1b). Vztah mezi pohlavím a četností výskytu jazykových proměnných je častým předmětem psycholingvistických studií (Rao et al. 2010; Pennebaker a Stone, 2003; Mukherje a Liu, 2010).

Z jejich výsledků vyplývá, že **ženy** používají více odkazů na sociální interakce a častěji se vyjadřují v 1. osobě singuláru. Muži se naopak vyjadřují formálněji a častěji používají určité a neurčité členy (Argamon et al., 2007).

V našem souboru se částečně prokázala stejná zjištění. Ženy v našem souboru využívají více intertextových odkazů ve formě nepřímých citací a častěji se vyjadřují v 1. osobě ale plurálu (*naše*). Dále se v našem souboru u žen prokázala vyšší frekvence výskytu spojek souřadících a zájmen ukazovacích (*ten, toto*), což může souviset s popsáním jevem, že ženy se orientují více na okolní sociální děje a užívají proto více ukazovacích zájmen, zdá se tedy, že se v jejich jazykovém projevu se odráží vyšší míra sociální interakce (Argamon et al., 2007).

Muži se v našem souboru naopak častěji vyjadřují ve 3. osobě plurálu (*jejich*), užívají více rozkazovacího způsobu (*zvažte*). V písemném projevu mužů se v našem souboru vyskytuje výrazně větší počet předložek. Užívání předložek je typické pro

akademický styl a značí to tendenci používat v písemném projevu více `předložkových frází` (např. *na rozdíl od*) a idiomatických výrazů (např. *naproti tomu*). Předložkové fráze jsou charakteristickým jazykovým jevem ve formálním psaném projevu a jsou součástí rozsáhlejších funkčních větných částí (Malá, 2017).

Kučera (2020) častější výskyt předložek ve svém výzkumu spojuje s vyšší měrou kontroly a sociálních dovedností a považuje ho za nejspolehlivější prediktor mužského pohlaví ve svých zkoumaných textech. Statisticky významný vztah se v tomto smyslu projevil i v našem datovém souboru.

U jevů, kde se prokázal statisticky významný vztah u pohlaví i Extraverze současně, jsme ověřovali **praktickou významnost**, což souvisí s třetím výzkumným záměrem, a sice ověřením, zda má na četnost výskytu jazykových prostředků větší efekt pohlaví nebo Extraverze (H3). Z našeho měření **míry účinku** sledovaných jevů jsme zjistili slabé, ale jednoznačné efekty potvrzující, že **pohlaví** má na užití jazykových prostředků silnější účinek než Extraverze.

Druhým výzkumným záměrem je pomocí **korelační analýzy** ověřit vztah mezi frekvencí výskytu jazykových jevů a mírou **Extraverze** (H2a, b). Z předchozích studií vyplývá, že Extraverze pozitivně koreluje s emočně pozitivním nábojem slov a nižší četností komparativu (Pennebaker a King, 1999), s větší délkou vět a slov (Holtgraves, 2011), s vyšším počtem záporných zájmen (Cegala, 1999) a čtenějším výskytem 1. osoby (Summer et al. 2011).

Jedním, ze statisticky významných vztahů, který se v našem souboru prokázal v negativním směru, je frekvence výskytu 2. stupně příslovcí, **komparativu**. S vysokou mírou Extraverze klesá četnost výskytu komparativu, což odpovídá předchozím závěrům (Pennebaker a King, 1999). Tento jev můžeme interpretovat jako menší potřebu relativizovat vlastní výroky pomocí srovnání a snižování jejich významu (*významnější, vhodnější*), tato potřeba klesá s rostoucí měrou sociální jistoty a sebejistoty u vyšších skóre Extraverze (Pennebaker a King, 1999).

S rostoucí měrou Extraverze v našem souboru roste i míra užití **záporných zájmen** (*nikdo, nic, ničtí*) a sloves v přítomném čase a v oznamovacím způsobu. Tyto vztahy jsou statisticky významné a příznačné pro gramatické mechanismy, které jazyk používá k vyjádření záporu neslovesnou cestou a přenesením negace na jiné slovní druhy, jako typicky záporné tvary zájmen (Dahl, 2010).

V lingvistice se tento přenosový mechanismus nazývá `funkční gramatika` nebo také `bridging theory` (Matthiessen a Halliday, 2009) a v tomto případě jde o funkční modifikaci větné struktury za účelem přenesení záporného významu mimo sémantické pole autora. V našem výzkumu by to mohlo naznačovat nižší ochotu Extravertních jedinců vyjadřovat negaci přímo ve slovesném tvaru, což se potvrzuje v zahraničních výzkumech (Pennebaker a Stone, 2003; Pennebaker a King, 1999) a souvisí s nastavením jejich osobnosti na pozitivní vztahy a spolupráci.

Posledním bodem, u kterého se zastavíme, bude statisticky signifikantní vztah v kladném směru mezi vysokou mírou Extraverze a frekvencí výskytu **1. osoby čísla jednotého a 2. osoby čísla množného**.

Na samém počátku této studie (dávno před sepsáním této práce) jsme formulovali intuitivní předpoklad, že vysoká míra Extraverze by se mohla vyznačovat vysokou frekvencí výskytu 1. osoby plurálu (*my*), protože by se tím projevila osobnostní orientace na interpersonální a sociální vztahy, kdežto u jedinců orientovaných introvertně by se mohla častěji vyskytovat 1. osoba singuláru (*já*), protože jsou orientováni více na své nitro. Tento náš intuitivní předpoklad byl ale záhy vyvrácen studiem odborné literatury.

Identifikovat žádoucí proměnné ve změní nejednotných kategorií v psycholingvistickém výzkumu bylo výzvou. V některých studiích je rozlišována osoba a

číslo (např. 1. os. singuláru a 1. os. plurálu) (např. Pennebaker a Stone, 2003; Pennebaker a King, 1999), v jiných se tato diferenciace ztrácí a je ponechán výstup ze strojového zpracování dat v původním tvaru, který rozlišuje pouze osobu (1., 2., a 3.) a pouze číslo (singulár, plurál) (např. Kučera, 2020; Cegala, 1999).

Možná i proto se výsledky přechozích studií jeví jako nejednoznačné, někteří 1. osobu spojují se sociální angažovaností a interpersonální orientací (Cegala, 1999), jiní se sociální dominancí (Kučera, 2022), zároveň ale existují výzkumy, ve kterých 1. osoba singuláru pozitivně koreluje s neuroticismem a depresivitou (Kučera, 2021; Rude et al. 2004).

Z těchto zjištění vyplývá značná metodologická nejednotnost v přístupech k analýze lingvistických proměnných. To ztěžuje jednak operacionalizaci výzkumu a jednak možnost mezikulturní a mezijazykové komparace (Kučera, 2020, s. 115).

Na základě předchozích studií jsme tedy přeformulovali svůj původní odhad a postavili výzkumnou hypotézu na pozitivním směru korelace mezi mírou Extraverze a užitím 1. osoby plurálu (*já*). Tento předpoklad se v našem souboru potvrdil na 1% hladině významnosti. To odpovídá poznatkům jiných, zejména anglofonních studií (Pennebaker, 2013; Nikolaeva, 2014).

Vyšší frekvenci výskytu **1. osoby singuláru** u jedinců s vyšší mírou Extraverze interpretujeme ve smyslu vyšší sociální jistoty, lepších interpersonálních dovedností a vyšší mírou orientace na mezilidské vztahy (Chung a Pennebaker, 2007). Tuto interpretaci potvrzuje i fakt, že statisticky nejvýznamnější vztahy se projevují v subškále Sociability.

Interpretaci výskytu 1. osoby singuláru ve smyslu interpersonální orientace podporuje i fakt, že se v našem souboru prokázal na 1 % hladině významnosti ještě další pozitivní vztah mezi Extraverzí a frekvencí výskytu **2. osoby plurálu**, což se také nejvýrazněji projevuje v subškále Sociability. Stejný výskyt v českém výzkumu raportuje i Kučera (2020).

Namísto intuitivního odhadu jsme v této studii na 1 % hladině významnosti prokázali pozitivní korelační vztah mezi mírou Extraverze a užitím 1. osoby singuláru (*já*) a 2. osoby plurálu (*vy*), což souvisí vysokou měrou Sociability u Extravertně orientovaných jedinců.

Dalším záměrem této práce bylo **porovnat**, zda se vztahy nalezené v našem souboru shodují s výsledky předchozích studií¹² (H4a-H4f). Jak již bylo zmíněno výše, statistické vztahy jsou obecně v psycholingvistickém výzkumu slabé (Pennebaker a Stone, 2003; Pennebaker a King, 1999; Kučera 2022; Yarkoni, 2010; Lee et al., 2007).

V našem souboru se u řady lingvistických proměnných vztahy sice projevily v očekávaném směru, ale jsou statisticky nevýznamné, jmenovitě negativní korelace mezi mírou Extraverze a frekvencí výskytu: (i) trpného rodu (Pennebaker et al., 2003; Kučera, 2022), (ii) podmiňovacího způsobu (Johnson-Laird, Byrne, 2002) a délkou slov, vět i textů (Holtgraves, 2011).

V předchozích studiích je trpný rod často spojován se sociální nejistotou (Rao et al., 2010) a s loajalitou (Pennebaker a Stone, 2003); submisivita a nízká míra seberealizace byla zase spojována s užitím podmiňovacího způsobu (Murkherje a Liu, 2010). Vzhledem k tomu, že Extraverze se vyznačuje vysokou měrou sociální jistoty, sebejistoty a dominance, v našem souboru se sice potvrdil negativní směr korelace u těchto proměnných, ačkoliv nebyl statisticky významný. U délky slov a vět se oproti očekávání projevily statisticky nevýznamné korelace v opačném směru.

¹² Psycholingvistické studie, na které se v rámci výzkumného záměru 4 (kapitola 8.5) odkazujeme jsou: (Kučera, 2020 a 2022; Pennebaker a Stone, 2003; Pennebaker a King, 1999; Yarkoni, 2010; Lee et al., 2007; Cegala, 1989; Summer et al. 2011; Argamon et al., 2007; Rao et al. 2010; Mukherje a Liu, 2010; Holtgraves, 2011; Johnson-Laird, Byrne, 2002)

Jak uvedl Kučera (2020, s. 98) v reakci na nízkou úroveň korelací ve svém výzkumu: „*Jelikož identifikované vztahy vysvětlují max. 7 % celkové jazykové variance (což je ostatně výsledek srovnatelný se zahraničními výzkumy), budí úvaha o jejich aplikovatelnosti určité pochybnosti*“. V naší studii se setkáváme s obdobnou problematikou, tento fakt by však neměl být překážkou v pokusech o vytvoření přesnějších nástrojů a operacionalizovatelnějších lingvistických proměnných k zpřesnění výzkumu mezi osobností a jazykem.

Jak jsme již několikrát naznačili, nesnažíme se zde o obecně aplikovatelné závěry, ale spíše o plošné šetření možností psycholingvistického výzkumu. Tím se tedy dostáváme zpět k **teoretickým otázkám** platnosti takového výzkumu a jeho možnostech přinášet nové informace ke studiu osobnosti.

Jedním z **cílů** formulovaných v samém úvodu této práce bylo ověření, zda osobnost vůbec lze zkoumat pomocí frekvenční analýzy přirozeného jazyka. Počet psycholingvistických studií (i těch uvedených pouze v této práci) tuto možnost nepochybně dokládá, musíme se tedy ptát, nakolik je jejich přínos pro výzkum osobnosti relevantní.

Výsledky našeho šetření, stejně jako dostupné výsledky předešlých studií (Kučera 2020 a 2022; Pennebaker a King, 1999; Pennebaker a Stone, 2003), dokumentují opodstatněnost výzkumu **funkčních slov**, tedy jazykových prostředků, které zajišťují gramatičnost jazykového systému (zájmena, předložky, spojky, determinující výrazy a formy negace), právě u nich odhaluje frekvenční analýza největší rozdíly, a právě ty jsou nositeli informací o jinak nesledovatelných procesech uvnitř jazyka (Chung a Pennebaker, 2011).

Frekvenční analýza jazykových znaků nám sice může podat zajímavý vhled do způsobu nejen **užívání jazyka**, ale i **myšlení** ve formě vytváření si koncepčních reprezentací vnější reality; ALE nemůže být samoúčelná, musí úzce nasedat na dokonalou znalost gramatického systému a strukturálních a funkčních možností daného jazykového systému. Pokud budeme jednotlivé jevy zkoumat izolovaně, ztratíme přehled o celku a nebudeme schopni výsledky smysluplně interpretovat (Chung a Pennebaker, 2007 a 2011; Sgall, 1985).

Poslední otázkou formulovanou v této práci je zhodnocení potenciálu frekvenční lingvistické analýzy k **predikci osobnostních rysů** autora. Komputační zpracování jazyka nám přináší informace o jazykovém systému, které jsou bez využití techniky pro běžného člověka nepostřehnutelné (například výpočet četnosti výskytu specifických gramatických jevů). Na samotnou predikci osobnosti autora ale zatím nestačí.

Jako efektivnější způsob predikce se jeví aplikace AI a strojového učení, kde by se na souborech trénovacích dat (podobných naší datové matici) AI sama `naučila` identifikovat jednotlivé jazykové proměnné a začala je dávat do souvislosti s proměnnými psychologickými, například z psychologických skóre autora v inventáři BFI-2, pak by jednoduše stačilo přenést tento postup na data nová (Grimmer et al., 2022). V současné komputační lingvistice takové modely dosahují velice slibné výsledky (Mereu, 2021; Jelínek, 2018;).

Problém ale zůstává s pochopením souvislosti mezi jazykem – a osobností. Způsob, jakým AI totiž zjišťuje potřebné souvislosti, zatím pro člověka zůstává nerozluštitelný, tzv. `black box problem` (Castelbecchi, 2016).

„Proces učení AI je srovnatelný s ději v černé skříňce poskytne nám jen malý náhled na to, jaké souvislosti jsou pro vztahy mezi mluvčím a jeho komunikací podstatné“ (Kučera, 2020, s.117).

Prozatím se tedy budeme muset spokojit s psycholingvistickými nástroji, které jsou pro nás čitelné, a poctivě hledat souvislosti v složité spleti lingvistických proměnných a postupně skládat kousek po kousku konečný jazykový vzor, ve kterém by se přesně odrážely osobnostní rysy daného autora.

10 ZÁVĚR

V této práci jsme pracovali s výzkumnou otázkou, zda existuje souvislost mezi mírou Extraverze a způsobem užití jazykových prostředků v písemném projevu.

Pomocí komparativní analýzy jsme zjišťovali souvislost mezi **pohlavím** a četností výskytu sledovaných jazykových jevů. V našem souboru se jako statisticky významné projevíly následující vztahy:

Ženy používají více intertextových odkazů ve formě nepřímých citací a častěji se vyjadřují v 1. osobě plurálu (*naše*), častěji také užívají spojek souřadících (*a, ale*), vyjadřují se v přítomném čase a způsobu oznamovacím a používají více ukazovacích zájmena a determináčnických výrazů (*ten, toto*).

Muži používají více 3. osoby plurálu (*jejich*), vyjadřují se častěji v rozkazovacím způsobu (*vezměme*), používají více přídavných jmen v negativním tvaru (*nevhodný, nevšední*) a používají více předložek (*bez, nad, naproti*) než ženy, což je podle Kučery (2020) nejspolehlivější prediktor pohlaví autora.

Při ověření **praktické významnosti** jsme u sledovaných jevů zjistili slabé, ale jednoznačné efekty potvrzující, že na užití jazykových prostředků má **vyšší míru účinku pohlaví** než Extraverze.

Extraverze pozitivně koreluje s častějším výskytem **záporných zájmen** (*nikdo, nic, ničí*) a slovesným tvarem v přítomném čase a v oznamovacím způsobu. **1 osobou singuláru** (*já*) a **2. osobou plurálu** (*vy*).

Negativně koreluje s **2. stupněm příslovcí – komparativu** (*hůře, rychleji*).

Ve všech psycholingvistických studiích jsou raportovány slabé až neprokazatelné statistické vztahy (Pennebaker a Stone, 2003; Pennebaker a King, 1999; Kučera 2022; Yarkoni, 2010; Lee et al., 2007).

Výsledky předchozích studií se tedy s korelační analýzou v našem datovém souboru shodují jen v neprokazatelné míře nebo v omezeném rozsahu, a to ze dvou důvodů:

- (i) lingvistické proměnné jsou v počítačném výzkumu generovány strojově formou univerzálních kategorií, které neodpovídají struktuře mluvnického systému češtiny a potom je nelze s anglofonním výzkumem porovnávat;
- (ii) nejednotná metodologie v psycholingvistickém přístupu způsobuje množství odchylek v kategoriích, a tím znesnadňuje operacionalizaci proměnných.

Pomocí prosté frekvenční analýzy jazykových znaků lze sice zjistit řadu statisticky významných vztahů, bez kvalitativní analýzy a detailního pochopení funkčních mechanismů uvnitř konkrétního jazyka však může psycholingvistický výzkum přispět k pochopení principu a způsobu, jakým se osobnostní rysy a jazykové prostředky ovlivňují, jen ve velmi omezené míře (Sgall, 1985; Oberlander a Gill, 2004).

Pro možnost predikce osobnostních rysů z písemného projevu se jeví jako vhodnější využít možností strojového učení a AI, s nimiž se dá dosáhnout překvapivě přesných výsledků, avšak nedozvíme se nic o tom, jaké proměnné jsou použity k predikci jednotlivých rysů osobnosti (Mereu, 2021; Jelínek, 2018; Hui, 2021).

Zbývá si tedy odpovědět na otázku, do jaké míry jazyk jako takový ovlivňuje vnímání osobnosti a nakolik osobnost naopak determinuje způsob užití jazykových prostředků. Tato otázka je myšlena spíše filozoficky, protože kauzalitu vztahu nemůžeme

pomocí aplikovaných statistických nástrojů určit a ani nebylo našim záměrem ji ověřovat (Chen et al., 2014).

Z pohledu lingvistického relativismu na komplementární vztah jazyka a myšlení můžeme jazyk považovat za odraz naší osobnosti, a naopak ve výzkumu zpětně ověřovat, nakolik se osobnost v jazyce odráží.

According to the logic of the lexical strategy, the most important elements of personality variation should be represented, in any human language, by a large number of similar but distinct words (generally adjectives) that are used by lay people in everyday description of their own and others' personalities. Therefore, factor analyses of self- or peer ratings on the familiar personality adjectives of any language would be expected to reveal the fundamental vectors of personality. (Ashton et al., 2014)¹³.

¹³ „Z logické podstaty lexikální strategie vyplývá, že by nejdůležitější variace aspektů osobnosti měly být zastoupeny v každém přirozeném jazyce formou velkého počtu podobných, ale jedinečných slov (obvykle přídavných jmen), která jsou používána obyčejnými lidmi v každodenním popisu vlastní osobnosti a osobnosti druhých. Proto by se dalo očekávat, že faktorová analýza běžných přídavných jmen popisujících osobnost buď formou sebehodnocení nebo hodnocení druhou osobou v jakémkoli jazyce odhalí základní osobnostní vektory“ (Ashton et al., 2014) – vlastní překlad.

11 SOUHRN

Jak samotný název práce napovídá, jedná se psycholingvistickou studii souvislostí mezi jazykovými proměnnými a osobnostními skóry autora v inventáři BFI-2 na škále Extraverze.

Teoretickým předpokladem této práce je myšlenka, že jazyk, myšlení a osobnost tvoří jeden celek, který se vzájemně prolíná a ovlivňuje. V teoretické části práci jsme si nastínili základní princip **myšlení**, tedy rozpoznávat, třídit a kategorizovat konkrétní i abstraktní jevy a vztahy v okolním světě; a utvářet si o nich teoretické představy v podobě *`mentálních reprezentací`* vnější reality (Eco, 1979).

Právě v těchto *`mentálních reprezentacích`* se odráží způsob našeho myšlení, a tedy nepřímo i nastavení naší osobnosti (Eysenck, Keane, 2008). Jak je zřejmé, každá *`mentální reprezentace`* může realitě odpovídat v menší či větší míře, a to v závislosti na přesnosti úsudku, schopnosti pozorování a mentální kapacitě myslícího subjektu. Mentální reprezentaci si můžeme představit jako mentální mapu území (reality), která může být přesná jen do určité míry.

Vztah mezi **osobností** a myšlením se právě hledá v míře přesnosti této mentální mapy. Jestliže má myslící subjekt dobré kognitivní schopnosti, dostatečné IQ a mentální kapacitu dospělého jedince, měla by jeho mentální reprezentace realitu více méně kopírovat. Pokud bude mít nějaké výrazné omezení (např. fyzický handicap, úzkostné nebo depresivní stavy), bude se jeho mentální reprezentace od reality odchylovat (Nakonečný, 1997) a jeho mapa ztratí přesný obraz reality.

`Mentální reprezentace` v paměti uchováváme kromě jiného pomocí jazykových konceptů. V teoretické části práce se tedy zabýváme i vztahem mezi myšlením a **jazykem** (Vygotsky, 2012). Z výrazných lingvistických teorií představujeme Sapir-Whorfovu hypotézu v podobě jazykového determinismu, tedy pojetí, kde je myšlení na jazyce zcela závislé (Whorf, 1954); a v pojetí lingvistického relativismu, kde jazyk myšlení pouze ovlivňuje (Boroditsky, 2011).

Významnou pozornost potom věnujeme vývoji teoretického pojetí osobnosti z pohledu psychoanalýzy, behaviorismu a humanistické psychologie. Osobnost v této práci vnímáme jako mnohodimenzionální dynamický celek (Cakirpaloglu, 2012). Po krátkém představení výzkumných metod a přístupů ke studiu osobnosti se blíže zastavujeme u strukturálních modelů osobnosti založených na analýze hlavních charakterových rysů (Eysenck, 1984).

Z psycholingvistického pohledu je zajímavé připomenout, že vzniku **strukturálních modelů osobnosti**, na kterých dnes stojí většina psychometrických inventářů, předcházela extenzivní jazykový výzkum a hledání popisných znaků (adjektiv) pro deskripci stávajících charakterových rysů osobnosti (Goldberg, 1992). Psychologové při svém výzkumu vycházeli z Galtonovi (1884) lexikální teorie, kterou Allport a Odbert (1936) formulovali jako *`lexikální hypotézu`*, a sice, že základní charakterové dimenze osobnosti se budou odrážet ve struktuře a slovní zásobě přirozeného jazyka. To dodnes platí jako základní jazykové paradigma v současném studiu osobnosti

Na základě tohoto předpokladu byla pomocí lingvistických popisných znaků formulována celá řada psychometrických inventářů, připomeňme si třeba Cattellovu rozsáhlou faktorovou analýzu a inventář 16F (Drápela, 2011), Eysenckův PEN model (Mikšík, 1999), NEO a OCEAN model (Costa a McCrae, 1997) nebo pětifaktorový model

osobnosti Big Five, ze kterého pochází i náš pětifaktorový inventář BFI-2 (Hřebíčková, 2008), který využíváme ke sběru psychometrických dat v praktické části této práce.

Než postoupíme k samotné kvantitativní analýze, zastavíme se u psychologické dimenze **Extraverze** jako jednoho z hlavních rysů pětifaktorového modelu osobnosti. Jedinci s vysokou mírou Extraverze jsou popisováni jako otevření, pozitivně naklonění ke změnám a výzvám, jsou orientovaní na výkon, umí se prosadit a jsou cílevědomí, mají pocit dostatku energie a obvykle mají i dobrou výkonnost, jsou cílevědomí a touží po sociálním kontaktu a ocenění druhých, ve svém projevu jsou spíše dominantní, asertivní, umí vyjádřit svůj názor a jsou otevření ke spolupráci a věří v interpersonálních vztazích (Lucas et al. 2000).

Teoretická část práce končí přehledem relevantních psycholingvistických výzkumů na téma jazykových prostředků ve vztahu k osobnosti. Většina studií vychází primárně z anglického jazyka (Pennebaker a Stone, 2003; Pennebaker a King, 1999; Yarkoni, 2010; Lee et al., 2007; Cegala, 1989; Summer et al. 2011; Argamon et al., 2007; Rao et al. 2010; Mukherje a Liu, 2010; Holtgraves, 2011; Johnson-Laird, Byrne, 2002), v českém prostředí stojí za povšimnutí výzkum Dalibora Kučery (Kučera et al., 2020; Kučera 2020 a 2022), na jehož poznatcích a metodologii budeme stavět v praktické části této práce.

V praktické části práce náš datový soubor (n=336) tvoří eseje studentů prvního ročníku psychologie spárované anonymně přes ID autora s jejich skóry v inventáři BFI-2 na škále Extraverze. **Inventář BFI-2** jsme k extrakci psychologických dat použili ze dvou důvodů: (i) zajistíme tak operacionalizaci a validitu předkládaných hodnot (Hřebíčková, 2022); a zajistíme možnost srovnání s ostatními studiemi v českém prostředí, zejména Kučera et al. (2020 a 2022).

V Kučerově (2022) souboru byly použity žánrově nejednotné texty a korelační výsledky jednoznačně potvrdily, že jazykový žánr (formální nebo neformální) má rozhodující efekt na frekvenci a způsob užití konkrétních jazykových prostředků.

V naší práci jsme proto lingvistická data žánrově omezili a použili jsme pouze jediný styl - eseje. Neklademe si proto nárok, že by předkládaná práce měla mít univerzální platnost. Jsme si vědomi, že je prováděna na příliš homogenním vzorku populace, s nereprezentativním datovým souborem, a nelze z ní vyvozovat všeobecně platné závěry.

Před přístupem k statistickému zpracování dat jsme museli definovat konkrétní jazykové jevy, které chceme pomocí frekvenční analýzy sledovat. Jednotlivé jazykové proměnné jsme získali výstupem z open-source softwaru **UDpipe** sloužícímu ke strojovému zpracování přirozeného jazyka (Straka et al., 2016). Využitelnost tohoto systému hodnotíme jako dostačující, avšak jednotlivé jazykové kategorie jsou v něm formulovány univerzálně a neodpovídají přesně gramatickému systému českého jazyka.

Z automaticky vygenerovaného seznamu proměnných jsme tedy pomocí vytvořených kombinací dále extrahovali takové proměnné, které by smysluplně nasedaly na gramatický systém českého jazyka.

Před provedením výzkumu jsme ke každému ze čtyř výzkumných záměrů formulovali **otevřené hypotézy**, jejichž úkolem bylo sledovat plošně souvislosti a směry naměřených vztahů mezi pohlavím nebo Extraverzí a jednotlivými jazykovými prostředky.

Samotné kvantitativní zpracování dat mělo několik částí. V rámci výzkumného záměru 1 jsme pomocí **komparativní analýzy** ověřili vztah mezi relativní četností sledovaných lingvistických jevů a pohlavím respondentů. Jako statisticky významné se v našem datovém souboru prokázaly vztahy mezi frekvencí výskytu 1. osoby plurálu (*my*), ukazovacích zájmen (*ten, toto*), spojek souřadících (*a, ale*) a počtem nepřímých citací u **žen** (Summer et al., 2011; Pennebaker a Stone, 2003).

U **mužů** jsme naopak zjistili statisticky významné vztahy mezi frekvencí 3. osoby plurálu (*jejich*) a rozkazovacího způsobu (*vezměme*). V podsouboru mužů se prokázala větší frekvence výskytu **předložek**, což je podle Kučery (2020) nejspolehlivější prediktor pohlaví autora a pojí se typicky s formálním stylem a užíváním četných **předložkových frází** v písemném stylu (např. *na rozdíl od*).

U jazykových jevů, kde se prokázal vztah jak s pohlavím autora, tak s jeho mírou Extraverze, jsme zjišťovali **praktickou významnost**, což souvisí s výzkumným záměrem 3. Z našeho měření **míry účinku** u sledovaných jevů jsme zjistili slabé, ale jednoznačné efekty potvrzující, že **pohlaví** má na užití jazykových prostředků silnější účinek než Extraverze.

Druhým výzkumným záměrem je pomocí **korelační analýzy** ověřit vztah mezi frekvencí výskytu jazykových jevů a mírou **Extraverze**.

Jedním, ze statisticky významných vztahů, který se v našem souboru prokázal, je nižší frekvence výskytu 2. stupně příslovcí, **komparativu** u Extravertních jedinců, což odpovídá předchozím závěrům (Pennebaker a King, 1999). Tento klesá s rostoucí měrou sociální jistoty a sebejistoty u vyšších skóre Extraverze (Pennebaker a King, 1999).

S rostoucí měrou Extraverze v našem souboru roste i míra užití **záporných zájmen** (*nikdo, nic, nič*) a sloves v přítomném čase a v oznamovacím způsobu, což se potvrzuje v zahraničních výzkumech a souvisí s potřebou mluvčího přesunovat negaci ze slovesa na jiné větné členy (Pennebaker a Stone, 2003; Pennebaker a King, 1999).

Statisticky signifikantní vztahy v kladném směru se také prokázal mezi vysokou mírou Extraverze a frekvencí výskytu 1. osoby čísla jednotého (*já*) a 2. osoby čísla množného (*vy*). Vyšší frekvenci výskytu **1. osoby singuláru** u jedinců s vyšší mírou Extraverze interpretujeme ve smyslu vyšší sociální jistoty, lepších interpersonálních dovedností a vyšší mírou orientace na mezilidské vztahy ve formě (*já*) – (*vy*) (Chung a Pennebaker, 2007).

Pozitivní vztah mezi Extraverzí a frekvencí výskytu **2. osoby plurálu** je interpretován ve stejném smyslu a potvrzuje se v podobném výzkumu i u Kučery (2020). Interpersonální orientace se s vysokou mírou Extraverze nepochybně pojí, tento předpoklad posiluje i fakt, že se v našem souboru u obou případů tento vztah nejsilněji projevil v subškále Sociability, a to na 1 % hladině významnosti.

Dalším cílem bylo ověřit možnost mezinárodní **srovnatelnosti našich měření**. V českém prostředí obdobné studie potvrzují zahraniční výzkumy jen částečně (Kučera et al., 2020 a 2022), a to ze dvou důvodů – (i) lingvistická data jsou ze zkoumaných textů shromažďována pomocí automatických nástrojů, které odpovídají gramatické stavbě anglického, nikoli českého jazyka, a (ii) v psycholingvistických výzkumech obecně vycházejí statisticky slabé až nevýznamné vztahy (Pennebaker a Stone, 2003; Pennebaker a King, 1999; Kučera 2022; Yarkoni, 2010; Lee et al., 2007; Cegala, 1989).

Kučera (2020, s.117) varuje před falešně pozitivními výsledky a zároveň podotýká, že v psycholingvistickém výzkumu jde vesměs o **frekvenční analýzu jazykových prostředků** a autoři nepoužívají pokročilejší metody statistického zpracování dat.

K replikaci již zaběhnutých postupů s cílem predikce osobnostních rysů z písemného projevu autora se hodí využívat spíše techniky strojového **učení a AI**, které po trénování na připravených datových souborech dokáží dojít k překvapivě přesným výsledkům (Hui, 2021, Grimmer et al., 2022).

Za účelem možnosti komparace mezi různými jazyky a smysluplného objasnění souhry vztahů mezi jazykovými procesy a osobnostními rysy je nutné pomocí **kvalitativní analýzy** zpřesnit zkoumané jazykové kategorie tak, aby odpovídaly funkčnímu, nikoli pouze gramatickému systému daného jazyka - např. systém anglických členů a český

slovosled funkčně nasedá do jedné jazykové kategorie FSP (Firbas, 1964, Sgall, 1985), kterou je třeba zohledňovat (Baayen, 2017), jinak psycholingvistický výzkum generuje sice zajímavé, ale z hlediska systemického vztahu jazyk-osobnost jen málo vypovídající výsledky.

Jak jsme uvedli v teoretickém úvodu, jazykové proměnné úzce nasedají na způsob nejen myšlení, ale i způsob formátování a třídění koncepčních vzorců a tvorbě mentálních reprezentací o struktuře vnější reality (Land, 2014). Jazyková data tedy musíme zkoumat v konkrétním kontextu dané věty, včetně její syntaktické struktury a až naposledy pomocí frekvenční analýzy jednotlivých větných členů a slovních druhů (Oberlander a Gill, 2004; Veselovská, 2014; Sgall, 1985).

LITERATURA

- Allport, G. W. (1937). *Personality: A psychological interpretation*.
- Allport, G. W. (1955). *Becoming: Basic considerations for a psychology of personality* (Vol. 20). Yale University Press.
- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological monographs*, 47(1), i.
- Anisi, J. (2012). Validity and reliability of NEO Five-Factor Inventory (NEO-FFI) on university students. *International Journal of Behavioral Sciences*, 5(4), 351-355.
- Argamon, S., Koppel, M., Pennebaker, J., & Schler, J. (2007). Mining the blogosphere: Age, gender, and the varieties of self-expression. *First Monday*.
- Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, 87(5), 707.
- Atlas, J. D. (1984). Comparative adjectives and adverbials of degree: An introduction to radically radical pragmatics. *Linguistics and Philosophy*, 347-377.
- Baayen, R. H. (2017). Data mining at the intersection of psychology and linguistics. In *Twenty-First Century Psycholinguistics* (pp. 69-84). Routledge.
- Bail, C. A. (2014). The cultural environment: Measuring culture with big data. *Theory and Society*, 43, 465-482.
- Balcar, K., Bedrenová, E., Možný, I., Bútorá, M., Bútorová, Z., Nakonečný, M. (1995). *Psychologie osobnosti*. Praha, Academia.
- Barrett, P. T., Petrides, K. V., Eysenck, S. B., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and individual differences*, 25(5), 805-819.
- Beebe, J. (2012). Psychological types in Freud and Jung. *Jung Journal*, 6(3), 58-71.
- Benešová, M., Faltýnek, D., Kormaníková, L., & Kučera, O. I repeat therefore I am: The parasyntactic perspective. *Journal of Linguistics/Jazykovedný časopis*, 74(2), 477-494.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In *Out of corpora* (pp. 181-190). Brill.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Biggs, M. (2013). Prophecy, self-fulfilling/self-defeating. *Encyclopedia of philosophy and the social sciences*, 765-766.
- Blatný, M. (2006). Metodologie psychologického výzkumu: konsilience v rozmanitosti. *Academia, Praha*.
- Boas, F. (1911). *The mind of primitive man: a course of lectures delivered before the Lowell institute, Boston, Mass., and the National university of Mexico, 1910-1911* (Vol. 41). Macmillan.
- Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences*, 11(5), 515-524.
- Boroditsky, L. (2011) How language shapes thought. *Scientific American* 304(2), 62-65.
- Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current opinion in behavioral sciences*, 18, 63-68.
- Burton, A., & Radford, J. (Eds.). (2022). *Thinking in perspective: critical essays in the study of thought processes*. Taylor & Francis.
- Cakirpaloglu, P. (2012). *Úvod do psychologie osobnosti*. Grada Publishing as.

- Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, 538(7623), 20.
- Cattell, R. B. (1945). The description of personality: Principles and findings in a factor analysis. *The American journal of psychology*, 58(1), 69-90.
- Cegala, D. J. (1989). A study of selected linguistic components of involvement in interaction. *Western Journal of Speech Communication (includes Communication Reports)*, 53(3), 311-326.
- Costa Jr, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and individual differences*, 13(6), 653-665.
- Costa, P. T., & McCrae, R. R. (2014). The NEO Inventories1. In *Personality assessment* (pp. 229-260). Routledge.
- Covey, S. R. (2009). 7 návyků skutečně efektivních lidí. Praha. Management Press.
- Dahl, Ö. (2010). Typology of negation. *The expression of negation*, 9-38.
- Dahlstrom, W. G. (1992). The growth in acceptance of the MMPI. *Professional Psychology: Research and Practice*, 23(5), 345.
- De Marneffe, M. C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal dependencies. *Computational linguistics*, 47(2), 255-308.
- Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and brain sciences*, 22(3), 491-517.
- Dostál, D. (2016). Statistické metody v psychologii. Olomouc. Filozofická fakulta.
- Drapela, V. (1998). *Přehled teorií osobnosti*. Praha. Portál.
- Du, X. (2023). Lexical Features and Psychological States: A Quantitative Linguistic Approach. *Journal of Quantitative Linguistics*, 30(3-4), 257-279.
- Eaves, L., & Eysenck, H. (1975). The nature of extraversion: a genetical analysis. *Journal of personality and social psychology*, 32(1), 102.
- Eco, U. (1979). *A theory of semiotics* (Vol. 217). Indiana University Press.
- Eisenstein, J. (2019). *Introduction to natural language processing*. MIT press.
- Eysenck, H. J. (1984). Cattell and the theory of personality. *Multivariate Behavioral Research*, 19(2-3), 323-336.
- Eysenck, M. W., & Keane, M. T. (2008). *Kognitivní psychologie*. Academia.
- Faltýnek D., Matlach V. *Hapax legomena jako nástroj určování autorství*. Slovo a Slovesnost. 2023.
- Faltýnek, Dan & Matlach, Vladimír & Owsianková, Hana. (2020). Hapax legomena jako indikátor autorského stylu a formální znak koheze textu. 10.13140/RG.2.2.16509.79847.
- Ferjenčík, J. (2008). *Úvod do metodologie psychologického výzkumu: jak zkoumat lidskou duši*. Portal sro.
- Firbas, J. (1964). From comparative word-order studies:(thoughts on V. Mathesius' conception of the word-order system in English compared with that in Czech). *Brno studies in English*, 4(1), 111-128.
- Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, 44, s. 329.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, 32(4), 365-387.
- Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard university press.
- Fruchter, T. G. (2018). *Linguistic relativity and universalism: a Judeo-philosophical re-evaluation of the Sapir-Whorf hypothesis* (Doctoral dissertation, University of the Free State).
- Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychological review*, 102(4), 652.

- Galton, F. (1889). Personal identification and description. *Journal of anthropological institute of Great Britain and Ireland*, 177-191.
- Galton, F. (1949). The measurement of character. *Fortnightly Review*, 36, 179–185.
- Gazzaniga, M.S., Ivry, R.B., & Mangun, G.R. (2002). *Cognitive Neuroscience: The Biology of the Mind* (2nd Ed.). New York: W.W. Norton & Company.
- Gifford, R., & Hine, D. W. (1994). The role of verbal behavior in the encoding and decoding of interpersonal dispositions. *Journal of Research in Personality*, 28(2), 115-132.
- Giles, D. (2013). *Advanced research methods in psychology*. Routledge.
- Gleitman, L., & Papafragou, A. (2005). Language and thought. *Cambridge handbook of thinking and reasoning*, 633-661.
- Goldberg, L. R. (1990). An Alternative Description of Personality: The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, (23), 14.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment*, 4(1), 26.
- Golková, E. (1983). On adverbials of agency in English and Czech. *Brno studies in English*, 15(1), 37-48.
- Grigsby, J., & Stevens, D. (2000). *Neurodynamics of personality*. Guilford Press.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.
- Grossmann, I., Ellsworth, P. C., & Hong, Y. Y. (2012). Culture, attention, and emotion. *Journal of Experimental Psychology: General*, 141(1), 31.
- Hajič, J. (2004). Disambiguation of rich inflection (computation morphology of Czech). Prague: Karolinum.
- Hajičová, E. (2021). Významný pohled na strukturu věty a textu. Jarmila Tárnyiková: Bridging Linguistic Theories and Language Use. *Časopis pro moderní filologii*, 103(1), 125-131.
- Hall, S. C., Lindzey, G., Leohlin, H. C., Manosevitz, M. (1997). *Psychológia*.
- Hamilton, R. V. (1957). A psycholinguistic analysis of some interpretive processes of three basic personality types. *The Journal of Social Psychology*, 46(2), 153-177.
- Hausser, R., & Hausser, R. (2001). *Foundations of computational linguistics*. Springer-Verlag Telos.
- Haviger, J. (2018). Statistická explorace Vztahů mezi osobnostními charakteristikami a textovými parametry. In D. Kučera, J. M. Havigerová, J. Haviger, V. Cvrček, Z. Komrsková, D. Lukeš, T. Jelínek, T. Urbánek, J. Franková (Eds.), *Výzkum CPACT: Komputační psycholingvistická analýza českého textu* (pp. 125–140). České Budějovice: PF JÚ.
- Havigerová, J. M., Haviger, J., Kučera, D., & Hoffmannová, P. (2019). Text-based detection of the risk of depression. *Frontiers in psychology*, 10, 385113.
- Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th Annual meeting of the Association for Computational Linguistics* (pp. 3-10).
- Heider, E.R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1), 10-20.
- Hendl, J. (2016). *Kvalitativní přístup a metody v psychologickém výzkumu*. Praha. Portál
- Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of Research in Personality*, 45(1), 92–99. <https://doi.org/10.1016/j.jrp.2010.11.015>.
- Hornová, L. (2003). Referenční slovník gramatických termínů [Reference dictionary of grammatical terms]. Olomouc: Univerzita Palackého v Olomouci.

- Hřebíčková, M. (1993). Lexikální přístup k osobnosti I: Historický přehled taxonomických výzkumů. *Československá psychologie*, 37(5), 455-468.
- Hřebíčková, M. (1994). Lexikální přístup k osobnosti IV: Seznamy termínů popisujících osobnost. *Československá psychologie*, 38(6), 516-525.
- Hřebíčková, M. (2008). Lexikální a dispoziční přístup k pětifaktorovému modelu osobnosti. *Teze disertace k získání vědeckého titulu ve skupině sociální a humanitní vědy*. Praha: AV ČR.
- Hřebíčková, M. (2011). *Pětifaktorový model v psychologii osobnosti*. Grada Publishing as.
- Hřebíčková, M., & Ostendorf, F. (1995). Lexikální přístup k osobnosti V: Klasifikace přídavných jmen do kategorií osobnostní deskripce. *Československá psychologie*, 39(3), 265-267.
- Hřebíčková, M., & Urbánek, T. (2001). Big five. *NEO pětifaktorový osobnostní inventář*.
- Hřebíčková, M., Čermák, I., & Osecká, L. (1993). Lexikální přístup k osobnosti III. *Československá psychologie*, 491-500.
- Hřebíčková, M., Jelínek, M., Květon, P., Benkovič, A., Botek, M., Sudzina, F., Soto, Ch. J & John, O. P. (2020). Big Five Inventory 2 (BFI-2): Hierarchický model s 15 subškálami. *Československá Psychologie*, 64(4).
- Hui, D. (2021). Understanding the in-depth AI Big Data Analysis application through Diversified Integration and Development of Functional Linguistics. *Journal of Applied Science and Engineering*, 24(5), 777-782.
- Chamonikolasová, J. (2005). Comparing the structures of academic texts written in English and Czech. *Slovak studies in English*, 1, 77-84.
- Chandler, D. (2022). *Semiotics: the basics*. Routledge.
- Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, 36(11), 1514-1528.
- Chen, S. X., Benet-Martínez, V., & Ng, J. C. (2014). Does language affect personality perception? A functional approach to testing the Whorfian hypothesis. *Journal of personality*, 82(2), 130-143.
- Chung, C. K., & Pennebaker, J. W. (2007). The psychological functions of function words. In K. Fiedler (Social communication: Frontiers of social psychology (pp. 343–359). Psychology Press.
- Chung, C., & Pennebaker, J. (2011). The psychological functions of function words. In *Social communication* (pp. 343-359). Psychology Press.
- James, W. (1907). The absolute and the strenuous life. *The Journal of Philosophy, Psychology and Scientific Methods*, 4(20), 546-548.
- Jelínek, T. (2018). Současná východiska komputační lingvistiky a její aplikace. In D. Kučera, J. M. Havigerová, J. Haviger, V. Cvrček, T. Jelínek, Z. Komrsková, D. Lukeš, T. Urbánek, & J. Franková (Eds.), *Výzkum CPACT: Komputační psycholingvistická analýza českého textu* [CPACT Research: Computational psycholinguistic analysis of Czech text]. České Budějovice: Vydavatelství Pedagogické fakulty Jihočeské univerzity v Českých Budějovicích
- John, O. P., & Srivastava, S. (1999). The Big- Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). Guilford Press.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of personality and social psychology*.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, 3(2), 114-158.

- Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological review*, 109(4), 646.
- Jordan, K. N., & Pennebaker, J. W. (2017). The exception or the rule: Using words to assess analytic thinking, Donald Trump, and the American presidency. *Translational Issues in Psychological Science*, 3(3), 312.
- Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, 116(9), 3476-3481.
- Jung, C., & Beebe, J. (2016). *Psychological types*. Routledge.
- Kay, P. (2006). Pragmatic aspects of grammatical constructions. *The handbook of pragmatics*, 675-700.
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis?. *American anthropologist*, 86(1), 65-79.
- Knapp, M. L., Hart, R. P., & Dennis, H. S. (1974). An exploration of deception as a communication construct. *Human communication research*, 1(1), 15-29.
- Kolářová, I., Klímová, K., Hauser, P., & Ondrášková, K. (2012). *Český jazyk*. Praha: Grada.
- Kučera, D. (202). *Osobnostní markery v textu: Aplikace kvantitativní psychologicko-lingvistické analýzy písemného projevu při popisu osobnosti* [Personality markers in text: Application of quantitative psychological-linguistic analysis of written text in personality description]. Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta.
- Kučera, D. (2021). Our Words in a State of Emergency: Psychological–Linguistic Analysis of Utterances on the COVID-19 Situation in the Czech Republic. *Psychological Studies*, 66(3), 239-258.
- Kučera, D., Haviger, J., & Havigerová, J. M. (2020). Personality and text: Quantitative psycholinguistic analysis of a stylistically differentiated Czech Text. *Psychological Studies*, 65, 336-348.
- Kučera, D. (2022). *Osobnostní markery v textu: Aplikace kvantitativní psychologicko-lingvistické analýzy písemného projevu při popisu osobnosti*. Nakladatelství Jihočeské univerzity v Českých Budějovicích.
- Kučera, D. & Haviger, J. (2014). Možnosti využití metod počítačové lingvistiky v psychologii [Possibilities of using computational linguistics methods in psychology]. In M. Blatný et al. (Eds.), *Sborník příspěvků konference Sociální procesy a osobnost 2013: Včera, dnes a zítra* (pp. 175–178). Nové Město: Psychologický ústav AV ČR.
- Kučera, D. & Havigerová, J. M. (2015). Interpersonální charakteristiky komunikátora z pohledu kvantitativní psycholingvistické analýzy (předběžné sdělení z výzkumné studie). [Interpersonal Characteristics of a Communicator from the Perspective of Quantitative Psycholinguistic Analysis (Research Study Preliminary Report).] In M. Bozogánová et al. (Eds.), *Sociálne procesy a osobnosť 2014: Človek a spoločnosť: Zborník príspevkov zo 17. ročníka medzinárodnej konferencie* (pp. 267–274). Spoločenskovedný ústav SAV.
- Kučera, D., & Havigerová, J. M. (2015). Computational psycholinguistic analysis and its application in psychological assessment of college students. *Journal of Pedagogy*, 6(1), 61-72.
- Kučera, D., Haviger, J., & Havigerová, J. M. (2020). Personality and text: Quantitative psycholinguistic analysis of a stylistically differentiated Czech Text. *Psychological Studies*, 65, 336-348.

- Kučera, D., Haviger, J., & Havigerová, J. M. (2022). Personality and word use: Study on Czech language and the big five. *Journal of Psycholinguistic Research*, 51(5), 1165-1196.
- Lee, C. H., Kim, K., Seo, Y. S., & Chung, C. K. (2007). The relations between personality and language use. *The Journal of general psychology*, 134(4), 405–413.
- Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, 5(1), 66-71.
- Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in psychology*, 6, 121301.
- Ľuba, P. (2018). Errors in the use of articles in advanced EFL learners with Czech as L1.
- Lucas, R. E., & Diener, E. (2009). Personality and subjective well-being. *The science of well-being: The collected works of Ed Diener*, 37, 75-102.
- Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of personality and social psychology*, 79(3), 452.
- Lund, N. (2014). *Language and thought*. Routledge.
- Malá, M. (2017). Non-prepositional English correspondences of Czech prepositional phrases. T. Egan and H. Dirdal (eds), 199-217.
- Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, 129(3), 361-368.
- Matthiessen, C. M., & Halliday, M. A. K. (2009). Systemic functional grammar: A first step into the theory.
- Matlach, M. V., & Faltýnek, M. D. Aplikace kvantitativní lingvistiky na analýzu sekvencí.
- Matlach, V., & Faltýnek, D. (2016). Báze nejsou písmena. *Studie z aplikované lingvistiky-Studies in Applied Linguistics*, 7(1), 20-38.
- McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American psychologist*, 52(5), 509.
- McEnery, T. (2019). *Corpus linguistics*. Edinburgh University Press.
- Mereu, A. (2021). Big five personality traits prediction with AI. *European Psychiatry*, 64(S1), S445-S446.
- Mikšík, O. (1999). *Psychologické teorie osobnosti*. Karolinum
- Mikšík, O. (2007) *Psychologická charakteristika osobnosti*. Karolinum
- Mikulová, M., Hlaváčová, J., Hajič, J., Hana, J., Hanová, H., Hladká, B., Štěpánková, B a Zeman, D. (2020). *Manual for morphological annotation, Revision for the Prague Dependency Treebank-Consolidated 1.0*. Technical Report TR-2020-64, Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic.
- Morency, S. (2017). Language Relativism and Determinism. *Psychology of Language*.
- Mukherjee, A., & Liu, B. (2010) Improving gender classification of blog authors. In *Proceedings of the 2010 conference on Empirical Methods in natural Language Processing* (pp. 207–217). Association for Computational Linguistics.
- Nakonečný, M. (1997a). *Encyklopedie obecné psychologie* (2. rozš. vyd.). Academia.
- Nakonečný, M. (1997b). *Motivace lidského chování*. Praha: Academia
- Nakonečný, M. (1995). *Lexikon psychologie*. Vodňář.
- Nettle, D. (2005). An evolutionary approach to the extraversion continuum. *Evolution and Human Behavior*, 26(4), 363-373.
- Nikolaeva, L. (2014). *The secret life of pronouns* (Doctoral dissertation, Massachusetts Institute of Technology).

- Noller, P., Law, H., & Comrey, A. L. (1987). Cattell, Comrey, and Eysenck personality factors compared: More evidence for the five robust factors? *The journal of personality and social Psychology*, 53(4), 775.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes. Replicated factor structure in peer nomination personality ratings. *The journal of abnormal and social Psychology*, 66, s. 574-583
- Oberlander, J., & Gill, A. J. (2004). Individual differences and implicit language: personality, parts-of-speech and pervasiveness. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 26, No. 26).
- Orel, M. (2016). *Psychopatologie: nauka o nemocech duše*. Grada Publishing, as.
- Peabody, D., & De Raad, B. (2002). The substantive nature of psycholexical personality factors: A comparison across languages. *Journal of personality and Social Psychology*, 83(4), 983.
- Pennebaker, J. W. (2013). *The Secret Life of Pronouns: What Our Words Say About Us*. New York: Bloomsburty Press.
- Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, 10(3), 90-93.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296-1312.
- Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: language use over the lifespan. *Journal of personality and social psychology*, 85, 291-301.
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count: Liwc 2007*. Austin, TX: LIWC.
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001), 2001.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1), 547-577.
- Pervin, L. A., & John, O. P. (1999). Handbook of personality. *Theory and research*, 2.
- Peterková, M. (2008). Teorie osobnosti. Personologové: stručný přehled. <https://is.muni.cz/www/mordechai/7159323/teorie-osobnosti.pdf> [cit. 16.3.2024]
- Piaget, J. (2005). *Language and Thought of the Child: Selected Works vol 5*. Routledge.
- Piaget, J. & Inhelderová, B. (2001). *Psychologie dítěte*. Portál.
- Plhánková, A. (2004). *Učebnice obecné psychologie*. Academia. Praha, FÚ AV, s. 489–499.
- Pramodh, K. C., & Vijayalata, Y. (2016). Automatic personality recognition of authors using big five factor model. In *2016 IEEE international conference on advances in computer applications (ICACA)* (pp. 32-37). IEEE.
- Prilleltensky, I., & Walsh-Bowers, R. (1993). Psychology and the moral imperative. *Journal of Theoretical and Philosophical Psychology*, 13(2), 90.
- Progovac, L. (1995). Determiner phrase in a language without determiners. *Working Papers in Linguistics*, 5(2), 81-102.
- Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010) Classifying latent user attributes in twitter. In J. C. Cortizo (Ed.), *Proceedings of the 2nd international workshop on Search and mining user-generated contents* (pp. 37–44). ACM.

- Read, S. J., & Miller, L. C. (2002). Virtual personalities: A neural network model of personality. In *The Dynamic Perspective in Personality and Social Psychology* (pp. 357-369). Psychology Press.
- Robinson, D. L. (2008). Brain function, emotional experience and personality. *Netherlands Journal of Psychology*, *64*, 152-168.
- Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*(8), 1121-1133.
- Řičan, P. (2010). *Psychologie osobnosti*. Grada publishing as.
- Sag, I. A. (1991). Linguistic theory and natural language processing. In *Natural Language and Speech: Symposium Proceedings Brussels, November 26/27, 1991*(pp. 69-83). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Sapir, E. (1921). An introduction to the study of speech. *Language*, *1*, 15.
- Saucier, G., & Goldberg, L. R. (1998). What is beyond the Big Five?. *Journal of personality*, *66*, 495-524.
- Saucier, G., Hampson, S. E., & Goldberg, L. R. (2014). Cross-language studies of lexical personality factors. In *Advances in personality psychology* (pp. 1-36). Psychology Press.
- Sgall, P. (1985). Contributions to functional syntax, semantics and language comprehension. *Contributions to Functional Syntax, Semantics and Language Comprehension*, 1-379.
- Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, *8*(4), 467-487.
- Scherer, K. R. (1979). Social markers in speech. Cambridge: Cambridge University Press.
- Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. (2006) Effects of age and gender on blogging. In N. Nicolov et al. (Eds.), *Proceedings of 2006 AAAI spring symposium on computational approaches for analyzing weblogs* (pp. 199–205).
- Schmiedtová, B. (2003). The use of aspect in Czech L2. *ZAS Papers in Linguistics*, *29*, 177-194.
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, *8*(9), e73791. <https://doi.org/10.1371/journal.pone.0073791>.
- Smékal, V. (1985). Předmět psychologie osobnosti a pojem "osobnost". *Sborník prací Filozofické fakulty brněnské univerzity. I, Řada pedagogicko-psychologická*, *33*(I19), 53-64.
- Smékal, V. (2004). *Pozvání do psychologie osobnosti. Člověk v zrcadle vědomí a jednání*. Barrister & Principal.
- Smolik, F. (2009). Psycholingvistika a čeština: některá slibná témata. *Naše řeč*, (5), 240-251.
- Sobotka, M. (2008) Problém „cogito, ergo sum“. *Filosofický časopis (Philosophical Journal)*, *4*(56), 489-499.
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, *113*(1), 117.
- Soukup, P. (2013). Věcná významnost výsledků a její možnosti měření. *Data & Research: SDA Info/Data a Vyzkum: SDA Info*, *7*(2).
- Sternberg, R.J. (2009) *Kognitivní psychologie*. Praha: Portál.
- Straka, M., & Straková, J. (2017). Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipeline. In *Proceedings of the CoNLL 2017 shared task: Multilingual parsing from raw text to universal dependencies* (pp. 88-99).

- Straka, M., Hajic, J., & Straková, J. (2016). UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 4290-4297).
- Sumner, C., Byers, A., & Shearing, M. (2011). Determining personality traits and privacy concerns from facebook activity. *Black Hat Briefings*, 11(7), 197–221.
- Svoboda, M. (2011). Projektivní významy Rorschachových tabulí. *Československá Psychologie*, 55(2), 188.
- Tackman, A. M., Sbarra, D. A., Carey, A. L., Donnellan, M. B., Horn, A. B., Holtzman, N. S., ... & Mehl, M. R. (2019). Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of personality and social psychology*, 116(5), 817.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24-54.
- Tupes, E. C. & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. USAF Technical Report ASD-TR-61-97.
- Vachek, J. (1976). *Selected writings in English and general linguistics* (Vol. 92). Walter de Gruyter.
- Veltkamp, G. M., Recio, G., Jacobs, A. M., & Conrad, M. (2013). Is personality modulated by language? *International Journal of Bilingualism*, 17(4), 496-504.
- Vernon, M. (1967). Relationship of language to the thinking process. *Archives of General Psychiatry*, 16(3), 325-333.
- Veselovská, L. (2014). Possessives and Their Equivalents in English and Czech: A Comparative Study. *Nominal Structures: All in Complex DPs*, 110-33.
- Vygotskij, L. S. (2004). *Psychologie myšlení a řeči*. Portál.
- Vygotsky, L. S. (2012). *Thought and language*. MIT press.
- Watson, D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In *Handbook of personality psychology* (pp. 767-793). Academic Press.
- Watson, J. B. (1930). *Behaviorism* (Rev. ed.). W W Norton & Co.
- Whorf, B. L. (1952). Language, mind, and reality. *ETC: A review of general semantics*, 167-188.
- Whorf, B. L. (2012). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT press.
- Wierzbicka, A. (2005). There are no "color universals" but there are universals of visual semantics. *Anthropological linguistics*, 217-244.
- Wierzbicka, A. (2013). *Imprisoned in English: The hazards of English as a default language*. Oxford University Press.
- Willig, C. (2019). What can qualitative psychology contribute to psychological knowledge?. *Psychological methods*, 24(6), 796.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the national academy of sciences*, 104(19), 7780-7785.
- Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3), 363-373.
- Yuan, C., Wu, J., Li, H., & Wang, L. (2018, July). Personality recognition based on user generated content. In *2018 15th International Conference on Service Systems and Service Management (ICSSSM)* (pp. 1-6). IEEE.

PŘÍLOHY

Seznam příloh:

1. Ukázka datové matice
2. Příklad položek inventáře BFI-2
3. Seznam tabulek
4. Seznam obrázků
5. Abstrakt v českém jazyce
6. Abstrakt v anglickém jazyce

Příklad položek inventáře BFI-2

Hřebíčková et al., 2020

Následující vlastnosti mohou nebo nemusí vystihovat vaši osobnost. Zapište za každým tvrzením číslo vyjadřující, do jaké míry s ním souhlasíte nebo nesouhlasíte.

- 1 Zcela nesouhlasím
- 2 Spíše nesouhlasím
- 3 Ani nesouhlasím ani souhlasím
- 4 Spíše souhlasím
- 5 Zcela souhlasím

Považuji se za někoho, kdo

1. ...je společenský, družný.
2. ...je soucitný, má dobré srdce.
3. ...má sklon být chaotický.
4. ...je uvolněný, dobře zvládá stres.
5. ...se o umění příliš nezajímá.
6. ...se umí prosadit a projevovat své vlastní názory.
7. ...je uctivý, s ostatními zachází s úctou.
8. ...má sklon být líný.
9. ...zůstává optimistický i po nějakém nezdaru.
10. ...zajímá se o mnoho různých věcí.
11. ...zřídka kdy pocítuje vzrušení a nadšení pro věc.
12. ...má sklon hledat chyby na ostatních.
13. ...je spolehlivý, ve svém chování stálý.
14. ...je náladový, střídá se mu dobrá a špatná nálada.
15. ...je vynalézavý, nachází důmyslné způsoby, jak něco dělat.
16. ...bývá tichý.
17. ...s ostatními příliš nesoucí.
18. ...je systematický, udržuje ve věcech pořádek.
19. ...bývá napjatý.
20. ...je fascinován uměním, hudbou a literaturou.
21. ...je dominantní, zastává roli vůdce.
22. ...s ostatními vyvolává rozepře.
23. ...má obtíže začít s úkoly.
24. ...se cítí sebejistě, je sám se sebou spokojen.
25. ...se vyhýbá intelektuálním a filosofickým debatám.
26. ...je méně činný než ostatní.
27. ...má v povaze odpouštět.
28. ...bývá poněkud ledabylý.
29. ...je emočně vyrovnaný, jen tak něco ho nerozhodí.
30. ...není příliš tvořivý.
31. ...je někdy plachý, introvertní.
32. ...pomáhá ostatním a není sobecký.
33. ...udržuje věci úhledné a uspořádané.
34. ...si hodně dělá starosti.
35. ...oceňuje umění a krásu.
36. ...pokládá za obtížné ovlivňovat druhé.
37. ...je někdy na ostatní hrubý.
38. ...je výkonný, věci dotahuje do konce.
39. ...je často smutný.
40. ...je přemýšlivý, nad věcmi uvažuje do hloubky.
41. ...je plný energie.
42. ...je vůči záměrům ostatních nedůvěřivý.
43. ...je důvěryhodný, vždy se na něj dá spolehnout.
44. ...drží své emoce pod kontrolou.

Seznam tabulek

Tabulka 1: Psychologické aspekty osobnostního rysu Extraverze.....	23
Tabulka 2: Výzkumná otázka.....	37
Tabulka 3: Přehled položek osobnostního rysu Extraverze v inventáři BFI-2.....	41
Tabulka 4: Rozdíl v míře Extraverze (včetně jejích subškál) v podsouboru mužů a žen	45
Tabulka 5: Skladba lingvistického datového souboru.....	45
Tabulka 6: Automatické tagování morfologické struktury přirozeného jazyka pomocí nástroje UDpipe.....	47
Tabulka 7: Seznam jazykových jevů, u kterých byla sledována relativní četnost výskytu v psaných textech jednotlivých autorů.....	47
Tabulka 8: Příklad extrakce trpného rodu z datové matice přes ID autora.....	49
Tabulka 9: Příklad vyhledávacího příkazu v programu Excel k extrakci trpného rodu z datové matice.....	50
Tabulka 10: Příklad kontingenční tabulky pro párování výskytu konkrétních jazykových jevů s textem konkrétního autora.....	51
Tabulka 11: Zátěže z analýzy hlavních komponent (rotace Varimax).....	53
Tabulka 12: Reliabilita položek inventáře BFI-2.....	54
Tabulka 13: Relativní četnost jednotlivých slovních druhů v závislosti na pohlaví.....	55
Tabulka 14: Relativní četnost extrahovaných lingvistických jevů v závislosti na pohlaví.....	57
Tabulka 15: Vztah mezi relativní četností slovních druhů a mírou extraverze ve všech subškálách inventáře BFI-2.....	60
Tabulka 16: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností výskytu slovních druhů v podsouboru žen.....	61
Tabulka 17: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností výskytu slovních druhů v podsouboru mužů.....	62
Tabulka 18: Vztah mezi relativní četností sledovaných lingvistických jevů a mírou Extraverze na všech subškálách BFI-2.....	63
Tabulka 19: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností užití jednotlivých lingvistických jevů v podsouboru žen.....	65
Tabulka 20: Vztah mezi mírou Extraverze (včetně jednotlivých subškál) a četností užití lingvistických jevů v podsouboru mužů.....	65
Tabulka 21: Míra účinku pohlaví a jednotlivých subškál osobnostního rysu Extraverze... ..	67
Tabulka 22: Četnost výskytu záporu a negace u jednotlivých slovních druhů v závislosti na míře Extraverze.....	69
Tabulka 23: Četnost výskytu záporu a negace u jednotlivých slovních druhů v závislosti na míře Extraverze v podsouboru mužů.....	70
Tabulka 24: Četnost výskytu záporu a negace u jednotlivých slovních druhů v závislosti na míře Extraverze v podsouboru žen.....	70
Tabulka 25: Souvislost mezi mírou Extraverze, délkou vět a větnou skladbou.....	71
Tabulka 26: Četnost užití trpného rodu u sloves v závislosti na míře Extraverze.....	71
Tabulka 27: Četnost užití podmiňovacího způsobu v závislosti na míře Extraverze.....	72
Tabulka 28: Četnost užití komparativu a superlativu v závislosti na míře Extraverze.....	73
Tabulka 29: Četnost výskytu 1-3. osoby singuláru a plurálu v písemném projevu ve vztahu k míře Extraverze daného autora.....	74

Seznam obrázků

Obrázek 1: Schéma vzájemné závislosti myšlenkového procesu a subjektu myšlení.....	9
Obrázek 2: Vztah jazyka a myšlení z pohledu lingvistického determinismu.....	13
Obrázek 3: Vztah myšlení, osobnosti a jazyka.....	15

Abstrakt v českém jazyce

V této frekvenční analýze jazykových prostředků zkoumáme vztah mezi mírou Extraverze (měřenou pomocí skóre v sebeposuzovacím inventáři BFI-2) a četností výskytu pozorovaných lingvistických jevů (získaných automatickým zpracováním přirozeného jazyka UDpipe). Cílem této práce je ověřit souvislost mezi osobnostními rysy autora a jeho výrazovými prostředky pomocí korelačních vztahů. V souboru (n=336) studentů 1.roč. psychologie jsme zkoumali jeden textový žánr: psané eseje. Pomocí komparativní analýzy jsme ověřovali rozdíl v užití jazykových prostředků mezi pohlavími. Zatímco ženy užívají více spojek souřadících (*a, ale*), více nepřímých citací a ukazovacích zájmen (*toto, tomu*) a vyjadřují se v 1. os. plurálu (*my, naše*) a oznamovacím způsobu; muži používají více 3. osoby plurálu (*oni, jejich*), více negativních zájmen (*nevhodný, nepodstatný*) a předložek (*naproti, vedle*) a vyjadřují se častěji způsobem rozkazovacím (*vezměme*). Po ověření praktické významnosti můžeme konstatovat, že pohlaví má vyšší míru účinku na užití jazykových prostředků než míra Extraverze. Statisticky významné ($p < 0,05$) vztahy jsme našli mezi mírou Extraverze a frekvencí výskytu záporných zájmen (*nikdo, nic*) v pozitivním směru; a frekvencí výskytu 2. stupně příslovcí komparativu (*lépe, snadněji*) v negativním směru. Obecně se v naší studii projevíly statisticky slabé až nevýznamné vztahy, s výjimkou pozitivní korelační závislosti mezi mírou Extraverze a frekvencí výskytu 1. osoby singuláru (*já*) a 2. osoby plurálu (*vy*) na hladině 1% hladině významnosti. Tyto vztahy se nejsilněji projevíly na subškále Sociability, což potvrzuje náš předpoklad, že se v jazykovém projevu tímto způsobem zrcadlí psychologické rysy osobnosti Extravertních jedinců a jejich orientace na sociální vztahy.

Klíčová slova

NLP – automatické zpracování přirozeného jazyka – psycholingvistika – komparativní studie – lingvistická analýza – kvantitativní lingvistika – textová analýza – statistická analýza – frekvenční analýza – psychologie přirozeného jazyka – jazyk a osobnost

Abstract

This study examines the frequency of occurrence of defined linguistic patterns in relationship to Extraversion levels (measured on a self-report BFI-2 inventory). The linguistic data is obtained through automatic tool for natural language processing (NLP) called UDpipe. The aim of this study is to statistically examine the relationship between personality scores on the scale of Extraversion and linguistic features found in written texts (essays) of (n=336) 1st year students of psychology. The study is limited in scope to on single genre and is not representative in terms of general use. By means of comparative analysis we have first examined the frequency of occurrence of linguistic patterns in relation to gender. Women score higher in the use of conjunctions (and, but), indirect quotations, determiners and demonstrative pronouns (*this, that*), and express themselves in 1st person plural (*we*) in an indicative mood. Men, on the other hand, use more 3rd person plural (*they*), more negative pronouns (*unacceptable, irrelevant*) and prepositions (*next, opposite*) and more often express themselves in the imperative mood (*let's go, fix*). By measuring the effect size on our data, we can conclude that gender has larger effect on the use of observed linguistic features than Extraversion. In relation to Extraversion, we have found statistically significant ($p < 0.05$) correlations in the frequency of use of negative pronouns (*nobody, nothing*) in a positive direction; and the frequency of occurrence of adverbs in 2nd degree comparative form (*better, easier*) in a negative direction. In general, statistically modest to non-significant relationships were observed in our study. The only exception was the positive correlation of the frequency of 1st person singular (*I*) and 2nd person plural (*you*) at $\alpha=0,01$. These lexical features scored highest in the positive direction on the subscale of Sociability, which confirms our assumption that psychological traits are reflected in language. The use of 1st person singular (*I*) and 2nd person plural (*you*) we attribute to the fact that individuals with high score on the scale of Extraversion are positively oriented to social interaction and interpersonal relationships.

Key words

NLP – natural language processing – psycholinguistics – Extraversion – statistical analysis – quantitative linguistics – textual analysis – frequential analysis - psychology of language use – personality and language