

UNIVERZITA KARLOVA

Filozofická fakulta

Katedra psychologie



BAKALÁŘSKÁ PRÁCE

Jiří Štipl

**Osobnost v jazyce vetkaná: lexikální struktura osobnosti
odvozená pomocí strojového učení**

**Personality in language inwoven: lexical personality
structure derived using machine learning**

Praha 2021

Vedoucí práce: Bc. Mgr. Jiří Lukavský, PhD.

Poděkování

Chtěl bych poděkovat svému vedoucímu Bc. Mgr. Jiřímu Lukavskému, PhD za trpělivé vedení a schválení vlastního tématu. Dále bych rád vyjádřil vděčnost všem vědcům, kteří činí vědu otevřenější sdílením svých dat, bez nichž by tato práce nemohla vzniknout. Rád bych také poděkoval Bc. Mgr. Alexandře Elbakyanové za poskytnutí článků, které jsem měl problém získat. V neposlední řadě bych chtěl poděkovat mým rodičům za podporu během studia a mé sestře Kateřině za vytrvalé motivování do dalších pomodoro „sessions“ při psaní této práce.

Prohlášení

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

V Praze dne 10.5.2021

.....

Jiří Štípl

Abstrakt

Práce představuje možnou novou metodu pro studium struktury osobnosti v rámci lexikálního přístupu. V první kapitole přehledu literatury vymezují nejdůležitější teoretické konstrukty studia rysů, lexikální hypotézy a pětifaktorového modelu osobnosti. Poté shrnují současné poznatky v oblasti aplikací informačních technologií do studia struktury osobnosti, konkrétně pro její predikci, predikci jejího dojmu a její simulaci. V praktické části popisují, jak lze odvodit obecnou strukturu osobnosti pomocí studia korelací mezi vektory reprezentujícími význam slov namísto korelací dotazníkového sebehodnocení. Úspěšnost této metody je pak hodnocena zejména pomocí konfirmační faktorové analýzy a shlukové analýzy a ilustrována pomocí několika grafů. Podle mých výsledků navrhovaná metoda poskytuje nový zdroj informací o struktuře osobnostně popisných slov. V diskuzi pak uvádím mnohá úskalí tohoto přístupu a navrhuji agendu pro další výzkum.

Klíčová slova

lexikální hypotéza, velká pětka, word2vec, struktura osobnosti, word embeddings

Abstract

This thesis presents a possible new method for inquiry into the structure of personality within the lexical approach. In the first chapter of the literature review, I define the most important theoretical constructs for trait research, lexical hypothesis and five factor model. Then I summarize the current findings in the area of information technology application to the study of personality structure, concretely to its prediction, prediction of its impression and its simulation. In the empirical part I describe how to devise general personality structure through the study of correlation between vectors representing meanings of words instead of correlation of self-report questionnaires. Successfulness of this method is then evaluated using confirmatory factor analysis, cluster analysis and illustrated with the aid of several graphs. According to my results, the proposed method provides new source of information about the structure of personality descriptive words. In the discussion section I then list several possible shortcomings of this approach and propose agenda for further research.

Keywords

lexical hypothesis, big five, word2vec, personality structure, word embeddings

Obsah

Obsah	6
Úvod	1
Literárně přehledová část.....	3
1 Lexikální přístup ke studiu struktury osobnosti	3
1.1 Prvotní studium rysů	3
1.2 Allport & Odbert (1936)	3
1.3 Podstata rysů osobnosti	4
1.4 Kritika rysového přístupu k osobnosti	6
1.4.1 Nejasná definice rysů	6
1.4.2 Vysvětlující nebo popisné.....	6
1.4.3 Stabilita v čase	7
1.4.4 Situacionismus	7
1.4.5 Psychologie cizince	7
1.5 Lexikální hypotéza	8
1.6 Lexikální projekty	10
1.7 Cesta k velké pěťce	10
1.8 Dimenze Velké pěťky	11
1.8.1 Extraverze	12
1.8.2 Svědomitost	12
1.8.3 Přívětivost	12
1.8.4 Emoční stabilita	13
1.8.5 Otevřenost	13
1.8.6 Upřímnost-Pokora	13
2 Dosavadní studie struktury osobnosti využívající informační technologie.....	14
2.1 Inference osobnosti	15
2.1.1 Klasické výpočetní metody pro inferenci osobnosti	15

2.1.2	<i>Využití neuronových sítí pro inferenci osobnosti</i>	16
2.2	Odhad dojmu osobnosti.....	17
2.3	Emulace osobnosti.....	18
	Výzkumná část	20
1	Výzkumný problém, výzkumné otázky a hypotézy	20
1.1	Výzkumné otázky a hypotézy	20
2	Design výzkumného projektu.....	21
2.1	Typ výzkumu.....	21
2.2	Metody získávání dat	21
2.3	Metody zpracování a analýzy dat.....	23
2.3.1	<i>Příprava dat pro analýz</i>	23
2.4	Etika výzkumu.....	26
3	Výzkumný soubor	27
4	Výsledky.....	28
4.1	První výzkumná otázka	28
4.2	Druhá výzkumná otázka.....	30
5	Diskuse	36
5.1	Reflexe výsledků	36
5.2	Limity výzkumného projektu	36
5.3	Uplatnění výsledků a doporučení pro další výzkum	37
6	Závěr.....	40
	Seznam použité literatury	41
	Seznam obrázků.....	48
	Seznam tabulek.....	49
	Seznam zkratk.....	50

Úvod

Struktura lidské osobnosti ve smyslu vyčerpávající hierarchické taxonomie psychologických charakteristik je jedním ze základních východisek psychologie výzkumné i aplikované. Kromě samotného poznání lidské osobnosti nám její znalost umožňuje vyvíjet lepší metody pro výběr zaměstnanců, diagnostiku psychických problémů, predikci školní úspěšnosti, ale třeba i stability partnerského vztahu. Podoby modelů struktury osobnosti tedy dopadají na širokou škálu nejpodstatnějších oblastí lidského života, a proto je třeba jim věnovat adekvátní výzkumné úsilí. Když Allport a Odbert (1936) sestavili první vyčerpávající seznam osobnostně relevantních slov, Cattell (1947) a další jej zpracovávali tehdy nejmodernějšími metodami až došli k tzv. Velké pěťce jako téměř univerzálnímu modelu. Nicméně po tomto úspěchu metodologická inovace poněkud ustala a krom drobných vylepšení výzkum pokračoval stejnými prostředky jako na konci první poloviny 20. století. V poslední době velmi narůstá relevance IT, datového inženýrství a datové vědy, což vede k tvorbě nových metod použitelných i pro společenskovední výzkum. O tento vývoj se zajímám, a tak jsem si uvědomil, že algoritmus word2vec, vyvinutý Čechem Tomášem Mikolovem v laboratořích Google (Mikolov et al., 2013) pro zlepšení strojového překladu, lze aplikovat na tento letitý problém struktury osobnosti jako nové metodologické východisko. V práci se zaměřím na prezentaci relevantních teoretických východisek jak z oblasti teorie rysů, lexikální hypotézy a velké pěťky, tak z oblasti výpočetní psychologie osobnosti (personality computing). V praktické části ukážu, jak lze tyto poněkud vzdálené oblasti propojit pomocí nového metodologického přístupu.

V první části přehledu literatury se zabývám rysy, lexikální hypotézou, lexikálními projekty a rysovými modely, které těmito přístupy vzešly. V prvních dvou podkapitolách nastíním vznik myšlenky lexikálního studia struktury rysů, ve třetí podkapitole uvedu hlubší teoretické ukotvení pojmu rys, neboť se jedná o centrální pojem této práce a v následující podkapitole shrnu kritiku rysového přístupu k osobnosti. V následujících třech podkapitolách popíšu východiska použití slov jako osobnostních deskriptorů, lexikální hypotézu a tzv. lexikální projekty. V poslední řadě v závěru první části popíši nejprominentnější model struktury rysů tzv. Velkou pěťku, neboť z ní budu vycházet v praktické části práce.

Ve druhé části přehledu literatury shrnuji výsledky vědní oblasti v angličtině nazývané „personality computing“ (Vinciarelli & Mohammadi, 2014). Tomuto tématu se věnuji ve třech podkapitolách, které jsou vymezeny podle zvyku v této literatuře na 1.) odhad osobnosti. Tato kapitola je nejpodrobnější, neboť je nejrelevantnější pro praktickou část práce, zejména její části o odhadu osobnosti pomocí analýzy textu strojovým učením 2.) odhad dojmu osobnosti 3.) emulaci osobnosti umělými bytostmi, například roboty nebo virtuálními agenty.

V práci vycházím zejména z článků, které jsem našel vyhledáváním v databázi Scopus a z článků na které odkazovaly nebo jimi byly citovány. Dále jsem využil službu Connected Papers (Tarnavsky Eitan et al., 2020) pro lepší orientaci v literatuře a nalezení souvisejících článků. Literatura první části obsahuje i poněkud starší články, ale zahrnul jsem je vzhledem k jejich zásadnímu formujícímu významu na další výzkum. Některé podstatné diskuze proběhly v literatuře zejména ve vzdálenější minulosti, proto není možné dané kapitoly postavit pouze na člancích z posledních deseti let. Naopak literatura druhé části vychází vzhledem k novosti tématu zpravidla z velmi recentních článků. V práci je citováno podle citační normy sedmého vydání publikačního manuálu Americké psychologické asociace (APA, 2020).

Literárně přehledová část

1 Lexikální přístup ke studiu struktury osobnosti

1.1 Prvotní studium rysů

Strukturu osobnosti ve smyslu osobnostních rysů a z nich vyplývajících interindividuálních rozdílů se poprvé v kontextu jazykového lexikonu zabýval Galton (1884), jehož zajímaly zejména morálně relevantní vlastnosti osobnosti. V rejstříku tehdy široce využívaného slovníku Rogetova Thesauru (Roget, 1852) prošel jeho značnou část a odhadl, že celý slovník obsahuje zhruba 1000 slov relevantních pro charakter. Galton zde činí zásadní poznámku „...it contained fully one thousand words expressive of character, each of which has separate shade of meaning and , while shares a large part of its meaning with some of the rest“. Jedná se pravděpodobně o první vyjádření zásadního předpokladu lexikálního přístupu ke studiu struktury osobnosti, a sice, že osobnostně relevantní slova sdílí určitou část významu, v čemž je implicitně obsaženo, že existují nějaké latentní struktury, které by tyto podobnosti mohly vysvětlit. Bohužel tato velmi podnětná myšlenka byla několik desetiletí opomíjena, do diskurzu ji vrátili až Klages a Johnston (1933), kteří odhadli, že se v němčině nachází zhruba 4000 slov užitečných pro popis osobnosti (zaměřili se ale spíše na morální aspekt) čímž stimulovali již nepřerušenu vlnu zájmu o lexikální hypotézu. Baumgartenová (Allport & Odbert, 1936) zaujatá Klagesovými výsledky se rozhodla jeho závěry systematicky ověřit, po prostudování několika německých slovníků došla k závěru, že němčina obsahuje 941 přídavných jmen a 688 podstatných jmen vhodných k popisu osobnosti. Tyto dva seznamy obsahovaly ale mnoho příbuzných slov, celkový počet unikátních významů (unikátnějších než přátelský vs. přátelskost) byl pouze 1 093.

1.2 Allport & Odbert (1936)

Revolučním posunem v lexikálním přístupu je slavná studie Allporta a jeho spolupracovníka Odberta (1936), která přišla s prvním opravdu vyčerpávajícím seznamem osobnostně relevantních slov v určitém jazyce (angličtině). Allport a Odbert identifikovali veškerá osobnostně relevantní slova ve velkém reprezentativním slovníku Webster's unabridged New International Dictionary (Webster et al., 1903), čímž dali vzniknout seznamu 17 953 slov, která zorganizovali do čtyř kategorií, jež podle tehdejšího tiskového vydání článku nazývají sloupce.

První sloupec s nadpisem „*Neutral Terms Designating Possible Personal Traits*“ sdružoval osobnostně popisná adjektiva bez hodnotící dimenze, takových identifikovali 4 504 (%). Druhý sloupec nazvaný „*Terms Primarily Descriptive of Temporary Moods or Activities*“ obsahoval 4541 slov označujících dočasné stavy osobnosti, tedy slova nevhodná pro popis stabilních vlastností. Allport a Odbert navrhuji, že by tento sloupec mohl být užitečný psychologům zabývajících se emocemi, prožíváním, vůlí nebo postoji. Třetí sloupec nadepsaný „*Weighted Terms Conveying Social or Characterial Judgements of Personal Conduct, or Designating Influence on Others*“ obsahoval 5 226 hodnotících atributů. Allport a Odbert poznamenávají, že s tímto sloupcem by neměli psychologové pracovat, vyjma potřeby sociálních psychologů studovat subjektivní pojmání morálky. Čtvrtý sloupec pak obsahuje zbylá slova různých významů, která jsou nevhodná pro objektivní popis osobnosti, kvůli jejich metaforickému významu, neostrosti, zaměřením na fyzický vzhled a podobně.

Autoři zastávají pozici, že osobnostní rysy ve striktním smyslu lze považovat pouze ty obsažené v prvním sloupci, nicméně se rozhodli zahrnout i některé další kategorie s ohledem na výzkumníky, kteří by s nimi nemuseli souhlasit (Allport & Odbert, 1936). Vydání tohoto vyčerpávajícího seznamu osobnostních deskriptorů zásadním způsobem ovlivnilo další výzkum rysů, jednalo se téměř až o paradigmatickou změnu, protože do této doby nebyl přístup typu „bottom up“ téměř využíván. Význam výsledků Allporta a Odberta (1936) je i v tom, že ukázali, jak podhodnocený počet osobnostně popisných slov byl v předešlých studiích (Klages & Johnston, 1933).

1.3 Podstata rysů osobnosti

V článku „What is a trait of personality“ Allport (1931) podává jednu z prvních explicitních analýz pojmu rys (trait) v osmi bodech. Zejména se jej snaží odlišit od příbuzných pojmů, které mu přijdou příliš sémanticky úzké, špatně definované či jinak neodpovídající, především jde o reflexy, zvyky, postoje a tendence. 1) a 2) rys není jen popisný abstraktní pojem, odkazuje se k opravdovému chování, podobně jako vznikají komplexní zvyky integrací těch specifitějších, tedy Allport vnímá rysy jako další zobecnění těchto komplexních zvyků. Popisuje, že u mnoha zvyků si lze všimnout vzájemné statistické asociace a že různé rysy pod sebe shrnují různě širokou škálu zvyků, předjímá tak svůj pojem kardinálních a periferních rysů (Zuroff, 1986). 3.) Rysy jsou dynamické a determinující, což Allport dokládá tím, že chování podle něj není tolik odpovědí na stimulus, ale je spíše dané tím, jak je odpověď na daný stimulus ovlivněna rysy osobnosti. Dnes bychom řekli, že rysy

jsou moderující proměnnou vztahu stimulů a chování. 4.) rysy mohou být demonstrovány statisticky nebo empiricky. Z dnešního pohledu je poněkud zvláštní stavět proti sobě statistické výsledky a empirii, neboť prvé by mělo být podmnožinou druhého, nicméně podle kontextu tím Allport myslí neformální nesystematickou zkušenost, například říká že můžeme mluvit o zvyku okusování nehtů, aniž bychom měřili četnost tohoto chování. 5.) rysy jsou pouze relativně nezávislé jeden na druhém, to že jsou součástí komplexní adaptace na prostředí způsobuje jejich interkorelace. 6.) Rys není to stejné jako morální kvalita. Allport varuje, že rysy osobnosti, které objevíme mohou, ale nemusí, odpovídat již existujícímu sociálnímu konstrukt. Ideálně bychom se měli snažit pojmenovávat rysy neutrálně, bez naznačené morální valence, což přiznává že může být složité u rysů typu „upřímnost“, které jsou zároveň společensky významné. 7.) Přítomnost rysu neznamena dokonale konzistentní chování s daným rysem, záleží na centralitě rysu v osobnosti člověka, jde o přetrvávající tendenci, ne výlučnost chování. 8.) Rysy lze nahlížet buď jako variabilitu vlastností napříč populací nebo se lze zaměřit na konkrétního člověka a identifikovat jeho rysy. Allport odhaduje že prvního typu, který je smysluplné zkoumat jako určitou míru u každého člověka existuje několik stovek, kdežto když se zaměříme na jednotlivce, dokážeme u něj identifikovat tisíce různých rysů. Po 35 letech se k tomuto článku Allport (1966) vrátil v „*Traits revisited*“ a i když říká, že většina jeho předchozích tezí je stále obhajitelná a pro něj rozumná, prezentuje na rysy méně vyhraněný názor. Předně již netrvá na rozlišování rysů od zvyků, sklonů, reflexů a ani na tom, že rysy musí mít motivační složku. Přiznává jak rysy konativní/motivační, tak emoční a kognitivní. Také již netrvá na svém rozlišení rysů obecných a rysů individuální, neboť to nebylo přijato většinou ostatních výzkumníků.

Hampshire (1953) podobně jako Allport vymezuje obecnější pojem dispozice v osmi bodech. 1) Dispozice je tvrzení o tom, co má tendenci se stávat, nikoliv o tom, co se děje. 2) Dispozice se může časem změnit, ale nesmí se měnit příliš často. 3) Dispozice se musí manifestovat v reálném chování člověka. 4) Pokud je stanovení nějaké dispozice sporné, je třeba vážit poměr manifestací dispozice a alternativních dispozic. 5) Osoba vynášející soud o dispozici měla dlouhodobou příležitost pozorovat chování a uvažování hodnoceného člověka a u každého hodnocení může jmenovat mnoho příkladů vyjádření dané dispozice a téměř žádné vyjádření opačné dispozice. 6) Mezi dispozicí a chováním je pouze pravděpodobnostní vztah, chování v nesouladu s dispozicí není vyloučeno. Dispozice se většinou ukazují rovným dílem ve stylu myšlení, v emočních reakcích a v behaviorálních projevech. 7) Dispozice nejsou většinou polární konstrukty, absence nějaké dispozice nutně

neznamená přítomnost opačné tendence, může se jednat o pouhou absenci popsateľné pravidelnosti. Hampshire vychází z kategorického pojetí rysů, v moderním pojetí stupňovitých rysů je tato situace ekvivalentní oblasti kolem 50. percentilu. 8) Ani dobře popsaná přítomnost rysu nezaručuje chování vždy v duchu tohoto rysu.

1.4 Kritika rysového přístupu k osobnosti

V roce 1994 bylo celé druhé číslo *Psychological Inquiry* věnováno právě reflexi rysového přístupu k osobnosti (Pervin, 1994). Cílovým (target) článkem byl Pervinův „*A critical analysis of current trait theory*“. Kritizuje zejména pojetí rysů v kontextu pětifaktorového modelu a lexikální hypotézy. Například se strefuje do snahy zastánců pětifaktorového modelu stále opakovat, že se tento model velmi dobře replikuje, čehož ale dosahují tak, že považují i nalezení jiného počtu faktorů za, alespoň částečný, důkaz platnosti, což vystihuje vtipnou parafrází Millerova magického čísla (Miller, 1956): „*Magic Number Five, Plus or Minus Two*“. Obecněji v tom vidí trend ve výzkumu rysů, existuje zde přílišná snaha potvrdit apriorní teorie na úkor lepšího vystihnutí dat. Jde o odklon od původního smyslu lexikálních projektů.

1.4.1 Nejasná definice rysů

Pervin si všimá, že rysoví teoretici téměř vždy definují rys jako „stabilní vzorce v chování“, ale dotazníky používané k měření těchto rysů obsahují kromě chování i vysoký poměr otázek na postoje, emoce, hodnoty nebo touhy. Také odvozuje, že rysy se nejeví samy o sobě, ale vždy v nějakém chování, které je reakcí na situaci a že je základním předpokladem psychologie existence nějakých pravidelností v tomto chování, jinak by nemělo smyslu studovat osobnost, ba možná i celou psychologii, čímž zpochybňuje přínos rysů jako teoretického konstruktů a nevidí jasně odlišující znaky od jiných psychologických kategorií.

1.4.2 Vysvětlující nebo popisné

Další nejistotu ohledně rysů přináší jejich status jako konceptů. Jde o pramenné vysvětlující zdroje chování nebo jen o deskriptivní sumarizace chování? Nejzásadnější kritikou rysů v tomto ohledu byla zřejmě ta z pozice pozitivismu a behaviorismu. Například Skinner (Jones & Skinner, 1939) tvrdí, že se rysy běžně pojmají jako kauzální fenomény, ale když vysvětlujeme např. fakt že někdo hraje dobře na piano jeho hudebním nadáním, tak pouze přidáváme zbytečný popis, který nevysvětluje nic navíc. Podobně jako Pervin (1994), Wiggins (1977) považuje rysy spíše za popisné než kauzální, což jadrně vystihuje větu

"Traits are 'lost causes'; their existence requires, rather than provides, a scientific explanation" (Wiggins, 1997), stejný pohled sdílí například i Funder (1994).

1.4.3 Stabilita v čase

Další přednost, kterou Pervin (1994) napadá je stabilita rysů v čase. Jejich zastánci sice předkládají mnoho studií, které demonstrují korelace dotazníkových měření v čase, ale interpretují tyto stability jako výlučný důsledek rysů, čímž ignorují často existující stabilitu prostředí, sociálních rolí, kulturních vlivů atd. v čase.

1.4.4 Situacionismus

Dalším často zmiňovaným argumentem proti rysům bývá determinace chování situací, např. známý Lapierův paradox (Lapierre, 1934), popisuje, jak rasové předsudky vyjadřujeme více vůči nepřítomným lidem a mnohem méně fyzicky přítomným. Po druhé světové válce vyšla v USA rozsáhlá studie „The American Soldier“ (Stouffer et al., 1949), která popisovala různé psychologické aspekty vojenské služby. Mimo jiné také ukázala, že postoje a vlastnosti vojáků významně souvisí s jejich hodností, specializací, délkou služby, jednotkou a dalším sociálními rolemi. V obou těchto případech se tedy chování zdá závislé na situaci/rolí, a ne na vlastnostech. Ale Allport (1966) nesouhlasí a obhajuje rysově pojetí. V případě Lapierova experimentu nelze tvrdit, že jde o absenci konzistentního rysu, rozdíl v chování může být vysvětlen tím, že v různých situacích jsou relevantní různé rysy. Když čínský pár přišel fyzicky do restaurace, mohla být tendence reagovat agresivně v důsledku „rasistického rysu“ vyvážena tendencí chovat se přátelsky před ostatními hosty v důsledku rysu přívětivosti. Další slabinou situačních studií je, že se většinou nezabývají variabilitou v rámci jedné podmínky, ale pouze rozdíly mezi podmínkami, tedy zákonitě nemohou ověřit význam rysů.

1.4.5 Psychologie cizince

McAdams (1994) zavedl dnes již známý pojem „psychologie cizince“, kterým kritizuje přílišnou abstrakci rysového přístupu. Odkazuje se na Goldbergovu větu „*English lexicon as it has evolved over time wished to know the answers to at least five types of questions about a stranger they were soon to meet;*“ (tučné písmo přidáno) (Goldberg, 1981), ve které, podle podle McAddamse Goldberg nevědomky přiznává, že rysově-lexikální přístup zachycuje pouze rudimentární informace o osobnosti člověka, takové, jaké bychom si asi přáli v rychlosti se dozvědět před setkáním s úplně cizím člověkem, ale nedokáží plasticky popsat osobnost. Tento problém ilustruje i tzv. podmíněnými vzorci chování „conditional patterns“,

např. *I talk the most when I am nervous*“, tedy nelze vysvětlit chování pouze rysy nebo situací, ale je třeba zahrnout i specifické reakce na konkrétní situace, interakční efekty, což rysový přístup nečiní (McAddams, 1994). Nicméně v poslední době se objevují již přístupy, které se tyto podmíněnosti snaží začlenit, např. Whole trait theory (Jayawickreme et al., 2019). McAddams (1994) ale považuje rysový přístup za velký posun v psychologii osobnosti a uznává jeho úspěchy, pouze vyjadřuje obavu, že díky těmto úspěchům jej neoprávněně jeho zastánci povyšují na komplexní teorii osobnosti, kterou nemůže být.

1.5 Lexikální hypotéza

Slovní zásoba obecně a slova převzatá z běžné komunikace při lexikálních projektech obzvláště, bývají kritizovány pro jejich nedostatečně definovatelný a proměnlivý význam. Vědci zabývající se lexikálním přístupem si byli dobře vědomi těchto limitů, například Allport (1961) na adresu tohoto problému poznamenává „*But dissatisfied (with the natural language) as we may be, verbal bags are all we have to work with. ... Our fate is to analyze traits in words, doing the best we can along the way to define our terms*“

Je vlastně s podivem, že ač se tzv. „lexikální hypotéza“ nazývá hypotézou, její přímé testy nebyly v minulosti prováděny. Prvním takovým přímým teste je studie Wooda (2015), V předchozích výzkumech zabývajících lexikální hypotézou je s touto hypotézou nakládáno spíše jako s axiomem, nejde o ověření platnosti této hypotézy ale o důsledky plynoucí z její platnosti (Eliasz et al., 2005). Wood pro účel své studie definuje tři varianty lexikální hypotézy: rysy, které jsou více důležité budou 1) použity častěji pro popis osob, 2) mít více synonym, 3) mít větší počet vysoce korelujících slov. Toto pojetí do značné míry vystihuje, jak se lexikální hypotéza běžně pojímala (Wood, 2015) až na její aspekt, že podstatné vzorce chování pro danou společnost mají vyšší pravděpodobnost, že budou zakódovány jako jedno slovo, nicméně tento předpoklad je alespoň nepřímě obsažen v druhé hypotéze. Dalším krokem této studie bylo definování důležitosti rysu, která též nebyla dříve operacionálně definována. Většina studií se spokojila s nejasnými pracovními definicemi jako např. Goldberg (1982) „*most significant in the daily transactions of persons with each other*“. Wood se zabývá zejména relevancí rysu měřenou „vztahovým dopadem“ (relational impact), který definuje jako moderátor rozhodnutí účastnit se určitého vztahu s hypotetickou osobou. Konkrétně agreguje tato hodnocení z pěti oblastí: (1) interagovat nebo strávit nějaký čas (2) mít romantický vztah s někým (3) pracovat s někým na projektu nebo v týmu (4) být přáteli s někým nebo (5) podpořit někoho, aby se stal členem týmu. Taková volba jistě dobře zachytí varianci interpersonálních vlastností, ale například doména neuroticismu se primárně

projevuje v intrapsychickém životě a až druhotně v interpersonální komunikaci (Widiger, 2017), což neznamená, že nemá dopad na společnost, jen se nebude tolik manifestovat v interpersonálních vztazích. Dále se Wood ptá, jak důležité pro respondety je, aby daná vlastnost byla známá jejich okolí, a naopak jak moc by se rádi o dané vlastnosti dozvěděli u ostatních lidí. Wood (2015) také nechal participanty ohodnotit 504 nejpoužívanějších osobnostních adjektiv v angličtině v jejich pozorovatelnosti, sémantické šíři, sebehodnocení a hodnocení blízkou osobou, včetně hodnocení toho, jak je hodnotitel má rád.

Nejpodstatnějším zjištěním studie je, že vztahový dopad velmi dobře koreluje ($r=0.91$) s faktorovými náložemi rysu v sedmifaktorovém modelu použitých adjektiv, ale i s dalšími mírami lexikální hustoty a důležitosti. S důležitostí, aby vlastnost znali ostatní je ale síla vztahu nižší ($r=0.31$), ještě nižší je síla vztahu s důležitostí znalosti vlastnosti pro ostatní ($r=0.13$), nicméně všechny tyto korelace jsou signifikantní. Vztahový dopad dále koreloval signifikantně s frekvencí použití v korpu anglického jazyka ($r=0.23$), udávanou frekvencí používání participanty ($r=0.21$), počtem ostatních slov s korelací nad 0.3 ($r=0.23$) až na hrubý počet synonym podle slovníku ($r=0.02$) s kterýmž korelace nebyla signifikantní. Wood interpretuje, že znalost vlastnosti s vyšší faktorovou náloží je důležitější, protože odráží širší škálu chování, oproti tomu znát vlastnost se specifitější významem, která má mnoho synonym není tak užitečné.

1.5.1 Lexikální hypotéza v době pandemie

Lexikální hypotéza předpokládá, že osobnostně relevantní slova v jazyce vznikají jako reakce na potřebu popsat relevantní vlastnosti ostatních lidí pro přežití v širším smyslu (Wood, 2015). Vývojový přístup nebyl v této oblasti uplatňován, protože životní styl celé společnosti se vyvíjí relativně pomalu vzhledem k vzniku nových slov kódujících „nové“ vlastnosti. Určitou výjimkou mohou být dlouhodobé krize, jako je např. nynější pandemie SARS-CoV-2. Podle projektu Čeština 2.0 (Kavka, 2021), který shromažďuje nové výrazy v češtině, vzniklo několik osobnostně popisných neologismů. Například „covidiot“ – člověk iracionálně odmítající opatření, „rouškař“ – člověk nosící roušku nad rámec povinnosti, „koronograf“ – člověk extenzivně sledující vývoj pandemických statistik. Bez definice ještě uvedu „virozvěst“, „roušmírák“, „rouštébák“ nebo „pandemagog“. Zda se tato slova v lexikonu udrží trvale není nyní jasné, ale tento jev je určitou podporou pro lexikální hypotézu.

1.6 Lexikální projekty

Pojmem „lexikální projekt“ bývá označován výzkumný záměr induktivního zmapování osobnostní struktury v nějakém jazyce, který vychází z lexikální hypotézy. Lexikální projekty časem získaly standardní podobu, která se příliš nelišila napříč různými výzkumnými týmy v rozdílných kulturních prostředích. John, Goldberg a Angleitner (1984) shrnuli pět hlavních problémů, které je nutné vyřešit při provedení lexikální taxonomické studie.

Prvními dvěma problémy jsou samotná definice zkoumaného fenoménu a vymezení kritérií pro zahrnutí jednotlivých reprezentací daného fenoménu, který odpovídá definici. Například biologické taxonomie tvořené na základě reálných atributů organismů lze dobře vymežit, ale taxonomie psychologických fenoménů jako je například osobnost jsou neostré významové prostory. Má patřit do pojmu osobnosti pouze temperament? I morálka? A co například fyzická atraktivita? Tento krok je pro lexikální projekty kruciólní, neboť používaná definice osobnosti přímo determinuje výběr slov pro další části výzkumu, a tak částečně i celkový výsledek. Například se lze zabývat všemi rysy osobnosti (Allport & Odbert, 1936), pouze stabilními rysy osobnosti (Goldberg, 1992) nebo pouze interpersonálními rysy osobnosti (Wiggins, 1980).

Dalším krokem je rozhodnutí o tom, jaká kritéria budou uplatněna při objevování nebo konstruování taxonomie na základě nasbíraných dat. John, Goldberg a Angleitner (1988) činí rozdíl mezi objevováním a konstruováním taxonomie podle epistemologického názoru badatele, zda je následovníkem spíše behavioristického nebo sociálně-konstruktivistického přístupu.

1.7 Cesta k velké pěťce

Raymond Cattell (1947) navázal na již popisovanou studii Allporta a Odberta (1936). Využil první sloupec jejich seznamu osobnostně popisných slov, k němu přidal ještě několik set slov z druhého sloupce, čímž se ukázala užitečnost předpokladu Allporta a Odberta (1936), že s nimi nemusí budoucí výzkumníci souhlasit ohledně klasifikace, a proto raději zahrnují i podle nich nevhodná kandidátní slova na rysy osobnosti. Ze vzniklého vzorku 2200 slov vytvořil dle vlastní úvahy 171 synonymních trsů. Dále prostudoval vědeckou i klinickou literaturu, aby doplnil rysy, jež nejsou adekvátně reprezentovány v běžné komunikaci, ač kódují významné pravidelnosti chování. Tento finální set 171 vlastností Cattell administroval několika vzorkům. Na základě korelační analýzy nakonec určil finální seznam

35 vlastností vystihujících osobnost, který pak aplikoval na několika vzorcích a podrobil faktorové analýze. Sám Cattell z těchto 35 vlastností vyvodil 11-12 faktorů, nicméně ostatní autoři jich za použití identických slov nalézali méně (Fiske, 1942; Tupes Christal 1961; Norman 1963). Nejdůležitějším navazujícím výzkumem byl ten Normanův (1963), ve kterém se pokusil o vytvoření ještě lepšího vyčerpávajícího seznamu osobnostně popisných slov, nicméně původní seznam Allporta a Odberta (1936) rozšířil pouze o dalších 171 slov. Výsledný soubor 18 125 slov nejdříve apriorními metodami zredukoval na 7 000 a poté rozdělil do tří sloupců, „*biophysical traits*“, „*temporary states*“ a „*activities*“, což nápadně připomíná první tři sloupce Allporta a Odberta (1936). Do prvního sloupce „skutečných vlastností osobnosti“ bylo zařazeno 2797 slov, která byla dále rozdělena do 75 synonymních trsů. Pro tyto téměř tři tisíce slov získal, i za tehdejších technicky omezených možností, hodnocení jejich sociální žádoucnosti, známosti, funkčnosti, sebehodnocení i hodnocení blízkými lidmi. Norman v datech, které nasbíral pomocí těchto 2797 položek identifikoval 5 faktorů, které byly pojmenovány již ve stejném duchu jako moderní modely Velké pětky, „*Extraversion*“, „*Agreeableness*“, „*Conscientiousness*“, „*Emotional stability*“ a „*Culture*“, které tak vešly ve známost jako „Norman’s five“ (Eliasz et al., 2005).

Později se prominentním výzkumníkem v lexikálních modelech stal Goldberg, přímo navazující na Normana. Jeho seznam přídatných jmen zkrátil na 1710 a dále agregoval nasbíraná data do Normanových 75 synonymních trsů. Hodnocení získaná předtím Normanem na těchto agregovaných 75 klastrech analyzoval pomocí různých metod extrakce a rotace faktorů a pokaždé došel ke stejnému počtu a přibližně stejnému obsahu vzniklých faktorů jako Norman. Goldberg (1990) po několika dalších replikacích pětifaktorového řešení kovarianční struktury odpovědí na dále zredukováný seznam adjektiv Normana získal důvěru v pětifaktorovou strukturu osobnosti, což komentoval slovy: „*Consequently, it now seems reasonable to conclude that analyses of any reasonable large sample of English trait adjectives in either self- or peer-descriptions will elicit a variant of the Big Five factor structure and therefore that virtually all such terms can be represented within this model.*“ Postupně se díky Goldbergově vlivu nahradilo označení „Norman’s Five“ značkou „Big Five“ neboli v češtině „Velká pětka“ (Goldberg, 1990).

1.8 Dimenze Velké pětky

V této podkapitole bych rád definoval nejčastěji identifikované dimenze v lexikálních výzkumech osobnostní struktury a dále popíši, se kterými podstatnými životními proměnnými (life outcomes) souvisí, a to zejména podle rozsáhlého replikačního projektu

(Soto, 2019) který ověřil 78 udávaných vztahů těchto proměnných s faktory Velké pětky na více než adekvátních vzorcích populace (mediánové N =1504).

1.8.1 Extraverze

Extraverze v pojetí lexikálního přístupu bývá statisticky nejvýraznějším rysem. Její kladný pól je v češtině vymezen např. adjektivy „*společenský*“, „*hovorný*“, „*družný*“, „*aktivní*“, „*průbojný*“ a záporný pól např. „*uzavřený*“, „*málomluvný*“, „*plachý*“ nebo „*rezervovaný*“ (Hřebíčková, 2011).

Extraverze na individuální úrovni predikuje subjektivní well-being, existenciální well-being, pocit vděčnosti, pocit inspirace, resilienci, míru depresivity a některé poruchy osobnosti. Extraverze v interpersonální sféře pozitivně souvisí s přijetím vrstevníky, variabilitou partnerů, atraktivitou, statusem mezi vrstevníky a spokojeností v romantických vztazích. V pracovní sféře extraverze predikuje pracovní spokojenost a pocit závazku k organizaci, sklon k vedení lidí, tendenci zapojovat se do dobrovolnických aktivit, míru zapojení do profesních činností, podnikatelskou a sociální pracovní orientaci (Soto, 2019).

1.8.2 Svědomitost

V angličtině „*conscientiousness*“. Vysokou míru toho rysu v češtině značí např. adjektiva „*svědomitý*“, „*pracovitý*“, „*pečlivý*“, „*přesný*“ a nízkou míru „*bezcílný*“, „*chaotický*“, „*líný*“ nebo „*nedbalý*“ (Hřebíčková, 2011).

Svědomitost na individuální úrovni predikuje religiózní přesvědčení a chování, negativně rizikové chování, zneužívání návykových látek a negativně i pozitivně poruchy osobnosti u minorit identifikaci s vlastní kulturou (oproti asimilaci s většinou kulturou). Svědomitost na interpersonální úrovni predikuje spokojenost v romantických vztazích. Na úrovni společensko-organizačních svědomitost koreluje s pracovní výkonností, vnitřními měřítky úspěchu a negativně s antisociálním a kriminálním chováním (Soto, 2019).

1.8.3 Přívětivost

V angličtině „*agreeableness*“, souvisí v češtině např. s adjektivy „*laskavý*“, „*dobrosrdečný*“, „*smířlivý*“ a naopak je v nepřímé úměře k adjektivům jako je např. „*bezcitný*“, „*útočný*“, „*pomstychtivý*“ nebo „*panovačný*“ (Hřebíčková, 2011).

Na individuální úrovni koreluje míra přívětivosti s religiózním chováním a přesvědčením, pocitem vděčnosti, pocitem odpuštění, smyslem pro humor a negativně s

některými poruchami osobnosti. V oblasti interpersonálních vztahů koreluje míra přívětivosti se spokojeností v romantických vztazích. V pracovní oblasti predikuje přívětivost podnikatelské profesní zájmy, ochotu zapojit se do dobrovolnictví, schopnost vést ostatní a negativně koreluje s externími měřítky úspěchu a kriminálním chováním (Soto, 2019).

1.8.4 Emoční stabilita

V angličtině „*emotional stability*“. V dispozičním přístupu je tento rys nazván podle opačné polaritě jako neuroticismus. V českém lexikonu jej vymezují například adjektiva „*napjatý*“, „*popudlivý*“, „*zranitelný*“ naopak jedinci s vyšší mírou stability jsou spíše „*klidní*“, „*vyrovnaní*“, „*sebejistí*“ nebo „*nezdolní*“ (Hřebíčková, 2011).

Na individuální úrovni emoční stabilita predikuje subjektivní well-being, existenciální well-being, smysl pro humor, zvládnání stresu, integraci identity, negativně i pozitivně s různými poruchami osobnosti a negativně s depresivitou a úzkostností. V oblasti interpersonálních vztahů predikuje emoční stabilita rodinnou spokojenost, status mezi vrstevníky (u mužů), spokojenost v romantických vztazích a negativně se zneužíváním v partnerských vztazích a rozchodovostí. V pracovní oblasti emoční stabilita koreluje s profesní spokojeností, pocitem závazku k organizaci, vnímaných úspěchem podle vnitřních i vnějších měřítek, finančním zabezpečením a negativně s antisociálním chováním (Soto, 2019).

1.8.5 Otevřenost

Pátý faktor bývá pojmenován různě, v lexikálních přístupech se nejčastěji používá označení „*intelligence*“, „*kultura*“, „*kreativita*“, ale v dispozičním přístupu „*otevřenost ke zkušenosti*“ (Costa & McCrae, 2008). V češtině je tato dimenze zakotvena na kladném pólu adjektivy např. „*otevřený*“, „*zvědavý*“, „*inteligentní*“, „*originální*“, a naopak záporný pól obsahuje slova jako „*konvenční*“, „*realistický*“ nebo „*nevzdělaný*“ (Hřebíčková, 2011).

Otevřenost v individuální oblasti koreluje s pocitem inspirace, s integrací identity a negativně s rigidností identity. V oblasti interpersonálních vztahů koreluje s výzkumnou a uměleckou profesní orientací a negativně s pravicovým autoritářstvím a konzervatismem.

1.8.6 Upřímnost-Pokora

V angličtině „*honesty-humility*“. DeRaad a Szirmaková (1994) v maďarském lexikálním projektu poprvé objevili jako samostatný faktor dimenzi vztahující se

k dodržování společenských pravidel, slušnosti a soucitu, která bývá v pětifaktorovém modelu obsažena v dimenzi přívětivosti a pojmenovali ji „integrita“. Původně byl tento faktor považován za artefakt tehdejší politické situace v Maďarsku (Hřebíčková, 2011) ale později byl identifikován jako důvěryhodnost v italštině (di Blas & Forzi, 1999) a jako pravdivost v korejštině (Hahn et al., 1999). Saucier přichází s hypotézou, že tato dimenzi bývá salientnější v méně rozvinutých společnostech (Wood, 2015), což by mohlo být způsobené tím, že v rozvinutějších společnostech je dodržování společenských pravidel a zákonů lépe vymáháno, tedy je naprostá většina obyvatel dodržuje, čímž klesá variabilita v tomto rysu, a tedy se snižují unikátní kovariance adjektiv, která by ho vymezila.

Ve zmiňovaných jazycích a později například i v němčině, holandštině a francouzštině byl tento faktor identifikován, ale například v angličtině, češtině a tagalštině zachycen lexikálním projektem nebyl. V řadě dalších jazyků byl tento faktor potvrzen top-down přístupem ověřením přeložených dotazníků obsahujících tuto dimenzi (Thielman et al., 2019). V češtině sice nebyla objevena poctivost-pokora v lexikálním projektu, ale převod do češtiny dotazníku HEXACO-PI-R, který tuto dimenzi obsahuje, se těší dobrým psychometrickým vlastnostem (Zášková, Dostál, 2016). Ač v českých dotaznicích měřících pětifaktorový model NEO-FFI, NEO-PI-R a NEO-PI-3 tato dimenze jako samostatný faktor nevystupuje, můžeme její kladný pól definovat adjektivy „*poctivý*“, „*spravedlivý*“, „*důvěryhodný*“ a protipól například adjektivy „*pokrytecký*“, „*domýšlivý*“ nebo „*nenasytný*“. Esencí tohoto rysu je tendence využívat druhé v situacích absence penalizace za takové chování (Hřebíčková, 2011).

2 Dosavadní studie struktury osobnosti využívající informační technologie

První a dosud nejcitovanější přehledový článek (Vinciarelli & Mohammadi, 2014) zaměřený na užití výpočetních metod v psychologii osobnosti si všímá výrazného nárůstu výzkumného zájmu mezi rokem 2004 a 2006, který nadále pokračuje (Remaida et al., 2020). Autoři článku za tím vidí tři hlavní důvody, 1) zvyšující se množství digitálních osobních údajů, 2.) zlepšující se možnost kontinuálního sběru těchto dat v online prostředí 3) stále častější interakce člověka s počítači/programy, které se snaží působit lidsky, k čemuž by jim mělo pomoci i porozumění osobnosti konverzačního partnera – člověka. Dosavadní články rozdělují na tři hlavní oblasti. První z nich je „*Personality recognition*“, tedy rozpoznání osobnosti, jedná se o inferenci sebehodnocení z externích vodítek v projevech jedince,

například odhad osobnosti pisatele z jeho textů. Další oblastí je „*Automatic Personality Perception*“, ve které jde o predikci hodnocení osobnosti externími hodnotiteli, tedy predikuje se dojem, jaký daná osobnost vyvolává a ne „skutečná“ osobnost. Třetí uvedenou oblastí je „*Automatic Personality Synthesis*“, která se týká programů a přístrojů, které mají navozovat iluzi živého komunikačního partnera. Cílem je, aby si lidský protějšek utvořil požadovaný dojem o „osobnosti“ programu. V dalším textu představím konkrétní studie z těchto tří oblastí s důrazem na inferenci osobnosti a osobnostního dojmu pomocí analýzy textů, protože tyto podoblasti nejvíce souvisí s praktickou částí práce.

2.1 Inference osobnosti

2.1.1 Klasické výpočetní metody pro inferenci osobnosti

Jednou z prvních studií zabývajících se výpočetním odhadem osobnosti z textu byla Argamona (2005). V této studii bylo 1200 studentů rozděleno do kategorií podle jejich percentilů v Extraverzi a Neuroticismu měřeném pomocí NEO-FFI. Cílem studie bylo klasifikovat studenty do horní nebo dolní třetiny v těchto dvou vlastnostech na základě lingvistických ukazatelů. Těmito ukazateli byly relativní frekvence čtyř kategorií slov funkční slova, hodnotící slova, slova udržující kohezi, slova ukazující postoje. Klasifikací pomocí Support Vector Machines bylo dosaženo 58 % přesnosti klasifikace (50 % úroveň náhody). Stejná data byla dále v navazující studii analyzována pomocí Linguistic inquiry and word count (LIWC), což je podrobnější metoda kategorizace slov, která dělí slova do 88 kategorií (Mairesse et al., 2007). Tentokrát bylo cílem studie klasifikovat studenty do horní nebo dolní poloviny v daném rysu. Přesnost pro jednotlivé domény byla v rozmezí 50 % (Svědomitost) až 62 % (Otevřenost ke zkušenosti). V podobné studii (Mehl et al., 2006) byla v nahrávkách běžné konverzace také identifikována Otevřenost ke zkušenosti jako nejlépe odhadnutelný rys (cca 62% přesnost, 50 % hranice náhody).

Otevřenost ke zkušenosti se ukázala jako nejlépe identifikovatelná z blogů v další studii, která přišla s řazením identifikovatelnosti vlastností: otevřenost ke zkušenosti > extraverte > neuroticismu > přívětivost > svědomitost (Vazire & Gosling, 2004). Několik dalších studií odhadovalo osobnost blogerů na základě jejich publikovaných textů, například studie Gill a kolegů (2009) predikovala tři stupně zastoupení rysu ordinální logistickou regresí z LIWC kategorií, naměřené celkové korelace modelů byly pod 0.1. Jiná studie (Minamikawa & Yokoyama, 2011) dosáhla lepší maximální klasifikační úspěšnosti 85 %, což bylo zřejmě dáno tím, že místo klasicky používané Big5 použila model EgoGram, který

je zaměřený na komunikaci a tedy by jeho rysy mělo být snazší identifikovat než například svědomitost z Velké pětky, která nemá pro komunikaci takovou relevanci (Widiger, 2017). Kromě individuálních slov použitých v přechozích studiích lze využít také specifické kombinace několika slov, tzv ngramy. Tímto přístupem dosáhl Nowson a Oberlander (2007) nejvyšší klasifikační přesnosti 65 % ve vzorku 1672 blogerů.

Kromě textu lze osobnost odhadovat také z gest nebo fyziologických ukazatelů. Již zmiňovaná studie (Mairesse et al., 2007) se pokusila kategorizovat participanty podle neverbálních charakteristik jejich promluv zejména podle průměrů a variability hlasitosti, tempa a výšky ale nedosáhla statisticky významných výsledků. Podobnou metodologií se podařilo (Pianesi et al., 2008) dosáhnout až 90 % přesnosti v predikci extraverze a místa kontroly z krátkých úryvků skupinových diskuzí. Překvapivě dobrých výsledků v mediánové klasifikaci se podařilo dosáhnout studii (Zen et al., 2010), která odhadovala míru extraverze a neuroticismu s 66 %, respektive 75 % přesností z rychlosti chůze a vzdálenosti podle kamerových záznamů skupinové interakce. Velmi zajímavý je komplexní přístup k přístrojové diagnostice osobnosti ASCERTAIN, participantům jsou prezentovány krátké emočně zabarvené klipy a zároveň jsou snímány různé fyziologické signály, zejména EEG, EKG, kožní odpor a výrazy obličeje. Rozdílné fyziologické odpovědi na dané stimuly jsou použity pro predikci osobnostních rysů Velké pětky (Subramanian et al., 2018).

Sociální média jsou skvělým zdrojem pro studium osobnosti, neboť facilitují a zaznamenávají online komunikaci a sebeodhalování. Kromě četných studií zabývajících se analýzami fotografií z obrazově orientovaných sítí (Cooper et al., 2020; Kim & Kim, 2018) bylo publikováno mnohé o využití analýzy textů zejména na sociální síti Facebook a Twitter.

2.1.2 Využití neuronových sítí pro inferenci osobnosti

Remaida a kolegové (2020) se snažili úsporně shrnout, či spíše zachytit obecnou strukturu, vědeckých článků využívajících neuronové sítě k odhadu osobnostních vlastností. Zajímavé je, že použili pro tuto klasifikaci výstupy modelu GENSIM (kóduje význam slov do číselných vektorů), což je model, který se zároveň používá v některých člancích pro odhad osobnosti. Nakonec se rozhodli pro řešení se třemi tématy: 1) Kulturní a socialně-kognitivní asociace s osobnostními vlastnosti (39.1% článků), 2) Koreláty mozkových struktur a osobnostních rysů (30.7% článků), 3.) Interakce lidského posuzování a rozhodování s osobností (30.3% článků).

Algoritmus word2vec umožňuje převést slova na číselné vektory reprezentující jejich význam (Mikolov et al., 2013), viz podrobně v kapitole 2.2.1 v empirické části práce.. Většinou je využíván pro modelování latentních témat (využívá jej Google překladač) nebo pro analýzu sentimentu online příspěvků (v poslední době ale začíná být využíván i pro účely sociálních věd). Článek Majumdera a kolegů (2017) byl zřejmě první studií, která word2vec použila v kontextu psychologie osobnosti. Pokusili se odhadnout z volného psaní („*stream of consciousness writing*“) participanta, zda se nachází nad nebo pod mediánem v rysech Velké pětky. Za tímto účelem vytvořili sedmivrstvou konvoluční neuronovou síť jejímž vstupem byly sémantické vektory pro slova v jednotlivých větách, vektory slov byly dále sdružovány do bigramů, trigramů, vět a nakonec byl spočítán vektor kódující celý dokument pro dané participanta. Tento vektor byl zkombinován s vypočítanými charakteristikami celého dokumentu podle Mairesse protokolu (Mairesse et al., 2007) obsahujícího např. standardní LIWC parametry. Takto vzniklý vektor reprezentující každé volné psaní byl poté použit v klasifikaci pomocí vícevrstvé perceptronové sítě a support vector machines. Nejlepší výsledky různými kombinacemi metod byly 62.68 % pro otevřenost, 59.38 % pro neuroticismus, 58.09 % pro extraverci, 57.30 % pro svědomitost a 56.71 % pro přívětivost. Výsledky pro samotný Mairesse protokol bez neuronových sítí byl většinou pouze o méně než dva procentní body nižší a ve srovnání se sítěmi se významně nelišil, ač je jeho použití mnohem snadnější než u neuronových sítí. Je tedy otázkou, zda oproti interpretaci výsledků autory není toto spíše podnětem k rozšiřování apriorních technik.

2.2 Odhad dojmu osobnosti

Dojem, jakým lidé působí v krátkých úryvcích řeči, ve smyslu lokace nad či pod mediánem v doménách Velké pětky lze predikovat charakteristikami samotných nahrávek, zejména výškou, hlasitostí a poměrem pomlk. Tímto přístupem bylo dosaženo přesnosti klasifikace mezi 60 a 73.5 % (Mohammadi & Vinciarelli, 2012; Polzehl et al., 2010).

Velmi zajímavým projektem byla The Interspeech Speaker Trait Challenge (Schuller et al., 2012), soutěž, která se snažila objevit nejlepší přístup k analýze osobnostního dojmu. Účastníkům soutěže byl poskytnut rozsáhlý korpus záznamů řeči a jejich úkolem v soutěži bylo předpovědět, zda jsou participanti vnímáni jako nad nebo pod mediánem v dimenzích Velké pětky. Účastníci soutěže zvolili mnoho různých statistických metod, např. Gaussian-Mixture Models, Sequential Floating Forward Search, Support Vector Machines a zaměřili se na různé charakteristiky nahrávek LIWC, prozodii, Modulation Spectrum Analysis a další. Nejvíce se dařilo predikovat svědomitost (téměř 80 % úspěšnost v průměru) a

přívětivost (kolem 75 % úspěšnost v průměru). Nicméně žádný analytický postup se svou úspěšností významně nelišil od ostatních.

Zajímavý přístup zvolily studie (Evans et al., 2008; Steele Jr et al., 2009), které porovnávaly skutečnou osobnost uživatelů a osobnost přisouzenou na základě jimi prezentovaného obsahu. Profilové obrázky 440 uživatelů hodnotilo 736 participantů. Kromě osobnosti byly obrázky hodnoceny i z hlediska obsahové analýzy, například přítomnost zvířat, částí těla ale i výrazů tváře, úhlu pohledu a podobně. Na rozdíl od předchozích studií nebylo cílem predikovat samotné vlastnosti, ale shodu vlastností skutečných a přisuzovaných, tedy odhalit jaké profilové fotky nevíce odhalují osobnost. Tato shoda byla nejvyšší, když obrázek ukazoval pouze jedince (vlastníka profilu), s úsměvem a bez pokrývky hlavy. První ze studií (Evans et al., 2008) se také zabývala korelací shody skutečných a vnímaných vlastností s online prezentací. Tentokrát byly zahrnuty všechny běžné atributy online profilu. Ukázalo se, že nejvyšší shoda nastává, když uživatelé sdílí své hodnotové pozice, názory na víru, své zájmy a zábavu

2.3 Emulace osobnosti

Výzkumy v oblasti emulace osobnosti se zaměřují většinou pouze na manipulaci s dimenzí introverze-extroverze. Vytvoří se dvě varianty systému, introvertní a extravertní, a následně se porovnávají reakce uživatelů.

Prvním typem studií v této oblasti jsou studie hlasových asistentů bez virtuálního těla. Například Nass a Lee (2001) zjistili, že úpravou prozodických vlastností lze vytvořit dojem extravertního nebo introvertního syntetického řečníka. Počet slov za minutu pro extravertní dojem by měl být 216, pro introvertní 184, výška hlasu extraverta by měla být 140 Hz, 84 Hz u introverta a hlasitost introverta o 15 % nižší než ta u extraverta. Velmi zajímavým výsledkem této studie je, že se lidé nechají více ovlivnit při výběru knih hlasovým asistentem se stejnou polaritou na ose introverze-extraverze.

Pokud má virtuální řečník i virtuální tělo, hovoříme o tzv. umělých agentech. Již bylo například demonstrováno, že lze vytvořit umělé agenty, kteří jak svým zjevem, tak chováním vykazují požadované psychologické vlastnosti. Konkrétně se podařilo u čtyř virtuálních agentů emulovat vysoké nebo nízké míry rysů v Eysenckových dimenzích extraverze, neuroticismu a psychoticismu (Mcrorie et al., 2012). Podobně se podařilo (Neff et al., 2010) manipulovat s vnímanou extraverzí robota úpravou expresivity jazyka, prozodie a doprovodných gest. Dokonce i tak hrubá charakteristika jako je pozice hlavy ovlivňuje

dojem o osobnostních vlastnostech. Konkrétně 133 participantů hodnotilo 54 pohledů na hlavu virtuálního agenta v dimenzích extraverze, přívětivosti a emoční stability. Signifikantní efekty úhlu pohledu byly nalezeny pro všechny dimenze, například nejvíce přívětivě agenta působil, když hleděl doprostřed a mírně dolů (Arellano et al., 2011).

Pokud je virtuálním agentům dáno i fyzické tělo nazýváme je roboty. Lidé, kteří strávili 25 minut s robotickým psem AIBO, byli schopni konzistentně rozpoznávat, zda jde o introvertní nebo extravertní verzi robota (Lee et al., 2006). Na rozdíl od ostatních studií zde bylo nalezeno několik komplementárních preferencí, tedy extraverti raději trávili čas s introvertním robotem a introverti s extravertním. Robot s opačnou mírou extraverze byl také hodnocen jako více inteligentní, atraktivní a sociálně angažovaný v daném momentu. Autoři předpokládají, že nalezení efektu komplementarity u robotů a konzistentní nalézání efektů stejnosti u virtuálních agentů je důsledek lidského vnímání komunikace, protože stejný rozdíl nalézáme ve studiích popisu cizího člověka vs. skutečná interakce s cizím člověkem (Lee et al., 2006; Nass & Moon, 2000). Praktickým uplatněním studia „osobnosti“ u robotů může být například zlepšení vztahu pacientů po cévní mozkové příhodě s terapeutickými roboty (Tapus et al., 2008). Zde se ukázalo, že pacienti preferovali roboty se stejnou mírou extraverze. Vývojáře sociálních robotů často zajímá, jaký zájem a zapojení (engagement) robot vzbuzuje u ostatních lidí. Salam se spolupracovníky (2017) zjistili, že zapojení lidí do interakce s robotem záleží do jisté míry i na jejich osobnosti a že je těžší odhadnout míru zapojení u introvertů oproti extravertům, což znamená, že robot správně odhadující osobnost dokáže i lépe odhadnout míru zapojení, a tím i poskytnout lepší zpětnou vazbu vývojářům.

Výzkumná část

1 Výzkumný problém, výzkumné otázky a hypotézy

Lexikální přístup ke studiu struktury osobnosti se spoléhá na analýzu osobnostně popisných slov z přirozeného jazyka, pro která jsou obvykle získávána sebehodnocení anebo externí hodnocení a dále se zpracovávají korelační vztahy těchto (sebe)hodnocení. Osobnostně relevantních slov je ohromné množství, například známá Allportova a Odbertova studie (1936) uvádí celkově 17 953 osobnostně relevantních slov v anglickém lexikonu a 4 504 specificky osobnostních rysů vyjádřených převážně adjektivy. Získat participanty ochotné odpovědět na takto vysoká množství položek by bylo velmi náročné, lexikální výzkumné projekty vždy využívají apriorní vyřazování slov malým počtem hodnotitelů. Hlavním kritériem bývá subjektivní dojem hodnotitele, zda dané slovo dobře odráží nějakou vlastnost. Tímto postupem je snižována validita lexikálních projektů, hrozí ovlivnění výsledků předpoklady hodnotitelů, jimiž jsou často sami výzkumníci.

Ve druhé kapitole literárně-přehledové části jsem ukázal, že metody výpočetní lingvistiky jsou již nějakou dobu využívány pro studium korelátů osobnostní struktury nebo pro odhad osobnosti konkrétních lidí. Při rešerši literatury jsem ale nenarazil na nikoho, kdo by tyto metody využil pro studium samotné struktury osobnosti. Metody typu word2vec umožňují z přirozeného jazyka odvodit významy slov zakódované do číselných vektorů, pokud získáme tyto definující vektory pro osobnostně relevantní slova, můžeme spočítat jejich vzájemné vztahy a tyto vztahy použít jako měřítko sémantické podobnosti, ekvivalentně jako je tomu u korelací (sebe)hodnocení participanty.

Hlavním cílem empirické části této práce je tudíž ověřit, zda studium přirozeného jazyka pomocí metod výpočetní lingvistiky může být užitečné pro studium struktury osobnosti. Kombinace lexikálního přístupu a výpočetní lingvistiky může přinést řešení pro analýzu vztahů velkého množství slov, neboť nevyužívá participantů, ale spoléhá se pouze na výsledky lidské činnosti v podobě textů, které mohou být převzaty z rozsáhlých korpusů, nemusí být vytvořeny specificky pro účel takové studie.

1.1 Výzkumné otázky a hypotézy

Výzkumná otázka 1 slouží k prvotnímu ověření navrhované metody na menším vzorku slov, výzkumná otázka 2 pak přímo porovnává struktury vzniklé navrhovanou metodou a běžně používanou metodou (sebe)hodnocení.

Výzkumná otázka 1: Do jaké míry budou osobnostní adjektiva s dobře ověřenou strukturou v sebehodnoceních vykazovat podobnost se strukturou odvozenou z vektorových reprezentací osobnostně relevantních slov?

Hypotéza 1: Konfirmační faktorová analýza korelací vektorů reprezentujících adjektiva s pozitivní valencí z Goldbergova Big Five Indicators potvrdí apriorní strukturu.

Výzkumná otázka 2: Do jaké míry odpovídá struktura osobnosti odvozená ze sebehodnocení osobnostní struktury odvozené z vektorových reprezentací osobnostně relevantních slov?

Hypotéza 2: Příslušnost adjektiva do shluku podle klastrové analýzy vztahů odvozených ze sebehodnocení nebude nezávislá na příslušnosti do shluku podle word2vec reprezentací.

Hypotéza 3: Korelace osobnostních adjektiv vypočítané podle sebehodnocení a podle vektorových reprezentací spolu korelují.

2 Design výzkumného projektu

2.1 Typ výzkumu

Předkládané otázky budou v podobě hypotéz ověřeny v rámci kvantitativního výzkumného plánu. Podle klasické metodologické klasifikace výzkumných projektů se jedná o observační studii stavějící na analýze korelačních vztahů, které jsou odvozeny z dotazníkově získaných dat a z dat získaných analýzou stop a výsledků činnosti v podobě počítačové analýzy přirozeného textu. Některé naměřené hodnoty budou interpretovány i z kvalitativního hlediska, například podobnosti nalezených faktorů s faktory z jiných studií a jejich přepokládaný význam.

2.2 Metody získávání dat

Vzhledem k vysoké náročnosti získání původních dat jak pro lexikální výzkum, tak pro stanovení vektorových reprezentací slov, jsou v této práci použita data převzatá z otevřených zdrojů dostupných v repozitáři vědeckých dat Open Science Forum a dalších volně dostupných zdrojů.

2.2.1 *Word2vec*

Především byly využity vektorové reprezentace slov z anglického jazyka vytvořené pomocí algoritmu word2vec zveřejněné společností Google (2013). Na konceptuální úrovni lze říct, že vstupem tohoto algoritmu je velký balík různých textů a jeho výstupem je N čísel pro každé slovo v těchto textech, které kódují význam daných slov. Word2vec je relativně jednoduchý algoritmus spadající do neuronových sítí umožňující tzv. strojové učení. Při tréninku tohoto modelu je sekvenčně procházen vstupní text po určitých „oknech“. Například u varianty (architektury) algoritmu Continuous bag of words (CBOW) je při šířce okna 5 vstupem vždy 5 slov, tak jak jdou za sebou ve větě. Prostřední slovo slouží jako fokální a jej obklopující 4 slova jako jeho prediktory. Algoritmus se tak snaží ze znalosti 4 slov jimž chybí prostřední slovo předpovědět, jaké slovo se do prázdného místa nejlépe hodí. Učiní predikci, vyhodnotí, zda předpověděl správné slovo a pokud ne, tak upraví parametry modelu tak, aby příště byla jeho predikce blíže fokálnímu slovu. Tento postup se opakuje v mnoha iteracích. Právě tímto opakovaným zpětným opravováním predikčních chyb se model „naučí“ význam slov v textu. I když cílem tohoto modelu z implementačního hlediska je predikce fokálního slova, jeho nejpoužívanějším výstupem jsou právě parametry definující význam slov. U CBOW nezáleží na pořadí slov kolem fokálního slova, proto je v názvu ono slovo „bag“. Alternativou je architektura skip-gram, která obecně funguje stejně akorát jako vstup přijímá jedno slovo a snaží se predikovat jaká slova by se měla nacházet v jeho okolí a zohledňuje i vzdálenost slov od vstupního slova. Mikolov s kolegy uvádí, že CBOW je rychlejší pro výpočet ale skip-gram vede k lepším reprezentacím, zejména pro vzácnější slova.

Jakmile je model „natrénován“ (jsou jím vytěženy informace dostupné ve vstupním textu), můžeme začít používat jeho parametry jako slova definující vektory. Pro nějaké slovo je v našem modelu např. 300 seřazených parametrů (čísel), které zachycují jeho význam. A protože je tento význam reprezentovaný matematicky, lze s ním provádět výpočty. Slavnou ilustrací je slovo „king“ od kterého můžeme odečíst slovo „men“ a přičíst slovo „women“, tím nám vznikne vektor, který by měl teoreticky znamenat to co král, když se jeho mužská část vymění za ženskou. A skutečně, nejbližším vektorem pro takto vypočítaný vektor byl vektor slova „queen“ – královna (Mikolov et al., 2013).

Ve své výzkumné části využívám word2vec model s architekturou skip-gram obsahující 2 883 863 slov, z nichž každé je reprezentované 300dimenzionálním vektorem. Reprezentace těchto slov byly získány na vzorku 100 miliard slov ve větách z korpu

anglofonních zpravodajských textů Google news dataset (Google, 2013). Mikolov s kolegy, (2013) uvádí, že podobný datový soubor (6 miliard slov korpus, milion měřených slov, 1000 dimenzí) trval touto metodou připravit dva a půl dne výpočetního času na 125 procesorových jádrech, tedy na běžném osobním počítači by to bylo řádově více. Toto je jeden z důvodů, proč jsem se rozhodl použít již existující model. Dalším důvodem pro výběr tohoto konkrétního modelu je jeho rozsah. Jedná se o nejobsáhlejší model pro anglický jazyk z modelů zveřejněných Nordic Language Processing Laboratory (*NLPL Word Embeddings Repository*, n.d.), což zvyšuje pravděpodobnost, že bude obsahovat i infrekventní osobnostní adjektiva a zároveň je tento model odvozen z nejrozsáhlejšího korpu z dostupných modelů, což zvyšuje reliabilitu a validitu jeho odhadů.

2.2.2 *Sebehodnocení*

Sebehodnocení osobnostními adjektivy bylo získáno z OSF Gerarda Sauciera, jež tato data využívá ve svých studiích (Saucier, 2019). Jedná se o sebehodnocení 310 lidmi na 1710 osobnostních adjektivech, z těchto adjektiv byly v průběhu studie vybírány relevantní subsety podle dostupnosti jejich vektorových reprezentací. Saucier bohužel uvádí u těchto sebehodnocení pouze zkratky pro delší slova, což je nedostatečné pro nalezení daných slov v natrénovaném modelu a opravování 1710 slov je pracné a náchylné k chybě. Naštěstí jsem na OSF nelezl dostupný seznam stejných adjektiv z jiné studie (Fischer et al., 2020). Po kontrole kompatibility zkratk Sauciera a celých znění od Fischera a kolegů jsem zjistil, že slovo „inharmonious“, které používá Saucier bylo zaměněno za „unharmonious“ a zařazeno na konec datasetu. Tento problém byl vyřešen spárováním slov, protože jejich předpokládaný sémantický rozdíl je zanedbatelný.

2.3 **Metody zpracování a analýzy dat**

2.3.1 *Příprava dat pro analýzy*

2.3.1.1 Příprava dat pro ověření první výzkumné otázky

Z word2vec modelu byly extrahovány vektory pro osobnostní adjektiva použitá Goldbergem (1992) v jeho Big five markers. Pouze pro tři slova se nepodařilo nalézt odpovídající vektor, jednalo se o bigramy „guilt free“, „at ease“ a „not envious“ (Goldberg chtěl mít u každého markeru pozitivní i negativní pól, proto byl nucen zahrnout i trochu krkolomnější formulace).

2.3.1.2 Příprava dat pro ověření druhé výzkumné otázky

Jak bylo popsáno v kapitole 2.2.1, sebehodnocení byla získána z volně dostupných souborů Sauciera (2019). Zkratky slov byly nahrazeny plnými zněními, tím vznikl soubor 1710 osobnostních adjektiv, ke kterým je dostupné sebehodnocení 310 lidmi. Původní záměr byl analyzovat co největší počet osobnostně relevantních slov, nicméně u druhé hypotézy jsem pro snadnější prezentaci výsledků, z 1710 dostupných slov vybral 310, které Saucier označuje jako „common denominator adjectives“. Jedná se o adjektiva, která se nejčastěji vyskytují v běžném jazyce, což je užitečné i z důvodu, že toto zvyšuje pravděpodobnost, že budou daná slova obsažena v natrénovaném modelu vektorových reprezentací. Z těchto 310 adjektiv v modelu bylo obsaženo 302, vyřazena byla „bull_headed“, „down_to_earth“, „happy_go_lucky“, „high_strung“, „nosey“ „open_minded“ a „self_disciplined“ s jednou výjimkou se jedná o samé bigramy nebo trigramy, které se liší svým zápisem více než běžná slova (varianta dohromady, zvlášť, využití různých znaků pro oddělení slov), proto bývá těžší je tímto „bottom up“ přístupem identifikovat, ale celá řada dalších n-gramů je modelem zachycena. Jako přílohu přikládám datový soubor pdas1710.self.invec.csv obsahující sebehodnocení pro všech 1223 zahrnutých slov (včetně oné podmnožiny 302 adjektiv) a soubor pdas1710.vec.csv obsahující word2vec reprezentace pro tato stejná slova.

Vektorové reprezentace slov byly uloženy na disk počítače a dále zpracovávány pomocí skriptů v jazyce Python (verze 2.7) zejména pomocí knihovny Gensim (Rehurek & Sojka, 2011), která umožňuje snadnou manipulaci s existujícími modely i trénink nových modelů. Pro manipulaci s maticemi a soubory byla využita knihovna NumPy (Harris et al., 2020). Po nahrání modelu do operační paměti (alokace 1.8 GB RAM) byly z modelu extrahovány reprezentace osobnostně popisných slov. Tyto reprezentace byly uloženy na disku v souboru formátu comma separated value (CSV).

2.3.2 Analýza dat

Analýzy jsem provedl pro první hypotézu v programu JASP (JASP team, 2020) a pro zbylé dvě v programovacím jazyce R (R Core Team, 2019) v prostředí RStudio (Rstudio, 2020) s využitím několika balíčků, které budou citovány v relevantních kapitolách.

2.3.2.1 Analýza dat první výzkumné otázky

Pro prvotní exploraci vektorů kódující 47 slov zachycujících dimenze Velké pětky jsem si vytvořil síťový graf, který shrnuje podobnou informaci jako korelační matice, ale je intuitivně srozumitelný. Uzly v této síti odpovídají slovům a hrany mezi nimi jejich vztahům,

čím silnější čára, tím vyšší hodnota příslušné korelace. Relativní pozice dvou slov by měla vycházet z jejich sémantické příbuznosti dané vztahy se všemi ostatními slovy. Dále jsem obarvil pozadí jednotlivých slov podle příslušné dimenze, pro přehlednost uvádím konkrétně ve výsledkové části. Použitím projekce do dvoudimenzionálního prostoru dochází ke zkreslení, bohužel v textu lze dostatečně přehledně v tomto případě zobrazit pouze dvě dimenze.

Klasickou metodou psychologické vědy pro testování shody apriorní struktury s naměřenými vztahy v datech je konfirmační faktorová analýza. Vzhledem k tomu, že word2vec neměří pouze podobnost relevantní pro osobnost, ale i např. podobnost na základě stejného slovního druhu, rodu, čísla a podobně (Mikolov et al., 2013), zvolil jsem strukturu modelu s obecným faktorem na první úrovni, který by měl zachytit tyto podobnosti, a na druhé úrovni faktory Velké pětky, které sytí příslušná adjektiva. Jako metodu odhadu parametru jsem zvolil metodu maximální věrohodnosti, zvolil jsem si standardizaci všech parametrů a pro škálování zafixování variance latentních faktorů k 1.

2.3.2.2 Srovnání struktury sebehodnocení a vektorových reprezentací

2.3.2.3 Analýza shluků

Pro prvotní ilustraci struktury souboru 302 slov byla provedena shluková analýza. Nejdříve byly spočítány eukleidovské vzdálenosti mezi sebehodnoceními a vektorovými reprezentacemi, dále byly tyto vzdálenosti zobecněny metodou multidimenzionálního škálování (Mardia, 1978). Souřadnice vzniklé multidimenzionálním škálováním byly využity ve shlukové analýze provedené metodou k-means clustering (Hartigan & Wong, 1979) za využití funkcí z balíčku stats (R Core Team, 2019). Tato metoda se snaží iterativním procesem nalézt předem definovaný počet shluků tak, aby vzdálenost slov navzájem od sebe uvnitř daného shluku byla minimální a zároveň maximalizovala vzdálenost od slov v jiných shlucích. Výsledky této shlukové analýzy slouží primárně pro exploraci dat. Pro intuitivní prezentaci struktury těchto dat jsem vytvořil graf prezentující vzniklé shluky jak pro sebehodnocení, tak tak pro word2vec vektory pomocí balíčku ggpubr (Kassambara, 2020).

Dále mě zajímalo, jak jsou si shluky vzniklé z obou typů dat podobné. Za tímto účelem jsem vytvořil kontingenční tabulku, která obsahuje zejména informaci o tom, kolik slov se překrývá mezi shluky napříč druhy dat. Tedy například se dozvím, že 25 slov by bylo

shodně klasifikováno do shluku 1 oběma metodami. Číslování shluků je zde arbitrární, jejich význam posoudím kvalitativně na základě překryvů s ostatními shluky a obsahu.

Kvantitativním testem validity klasifikace z vektorových reprezentací bude chí-kvadrát test nezávislosti, který provedu na zmiňované kontingenční tabulce. Pokud je alespoň triviálně nenáhodný vztah mezi oběma klasifikacemi, měl by tento test vyjít signifikantní. V průběhu práce používám hladinu statistické významnosti 0.05. Jako míru velikosti toho vztahu použiji Cramerovo V.

2.3.2.4 Porovnání korelací

Základní možností, jak porovnat strukturu dat ze sebehodnocení a vektorových reprezentací je porovnání přímo jejich korelačních matic. Zajímalo mě, jak podobné korelace přisuzují jednotlivé metody vztahům mezi slovy. Pro ilustraci hodnot korelací pro soubor 1223 slov jsem vytvořil histogram. Dále byla vypočítána Spearmanova korelace pořadí velikosti korelací podle jednotlivých metod. Konkrétně, vzniklé korelační matice vztahů mezi adjektivy byly převedeny po řádcích každá na jeden vektor obsahující všechny dané korelace. Dále bylo v rámci každého vektoru stanoveno pořadí velikosti korelací, poté byly spočítány korelace těchto pořadových hodnot pro obě dvojice vektorů.

Pro získání lepší ilustrace těsnosti vztahu zavádím ještě speciálně navrženou metriku. Vypočítal jsem rozdíl pořadí velikosti korelací pro každý vztah mezi adjektivy a vydělil jej celkovým počtem korelací. Například když by mezi slovy „friendly“ a „talkative“ byla korelace podle word2vec 0.33 a tato korelace by měla pořadí velikosti 234 a podle sebehodnocení by korelace mezi těmito slovy byla 0.55 s pořadím velikosti v rámci korelací pro sebehodnocení 357, tak by diference pořadí byla 123 a pokud by celkový počet korelací v tomto hypotetickém případě byl 400, tak relativní (k celkovému počtu) diference by byla 0.3075. Tedy bychom mohli říct, že rozdíl mezi typy dat pro velikost korelací je takový, že se v přisuzovaném pořadí velikosti tohoto vztahu od sebe liší o 30.75 %. K tomuto srovnání jsem přistoupil, protože velikosti korelace mezi vektorovými reprezentacemi jsou obecně vyšší, aniž by toto nutně znamenalo vyšší míru vztahu, který nás zajímá a proto by srovnání na základě hodnot pouhého rozdílu korelací nebylo smysluplné.

2.4 Etika výzkumu

Výzkumný projekt využívá statistické agregace veřejných textů a veřejně dostupné anonymizované datasety, tedy není nutné řešit otázky ochrany osobních údajů nebo anonymizace a dále ani není relevantní posuzovat splnění ostatních etických principů

psychologického výzkumu, neboť v žádné fázi se neúčastnili lidé ani zvířata a ani nepřímo se jich realizace výzkumného projektu nedotkla. Autorské právo též není dotčeno, neboť relevantní natrénovaný word2vec model byl publikován pod licenci Apache License 2.0 (Sinclair, 2010) a data obsahující sebehodnocení v osobnostních adjektivech byla zveřejněna volně k dispozici autorem na Open Science Framework v již anonymizované podobě (Saucier, 2019).

3 Výzkumný soubor

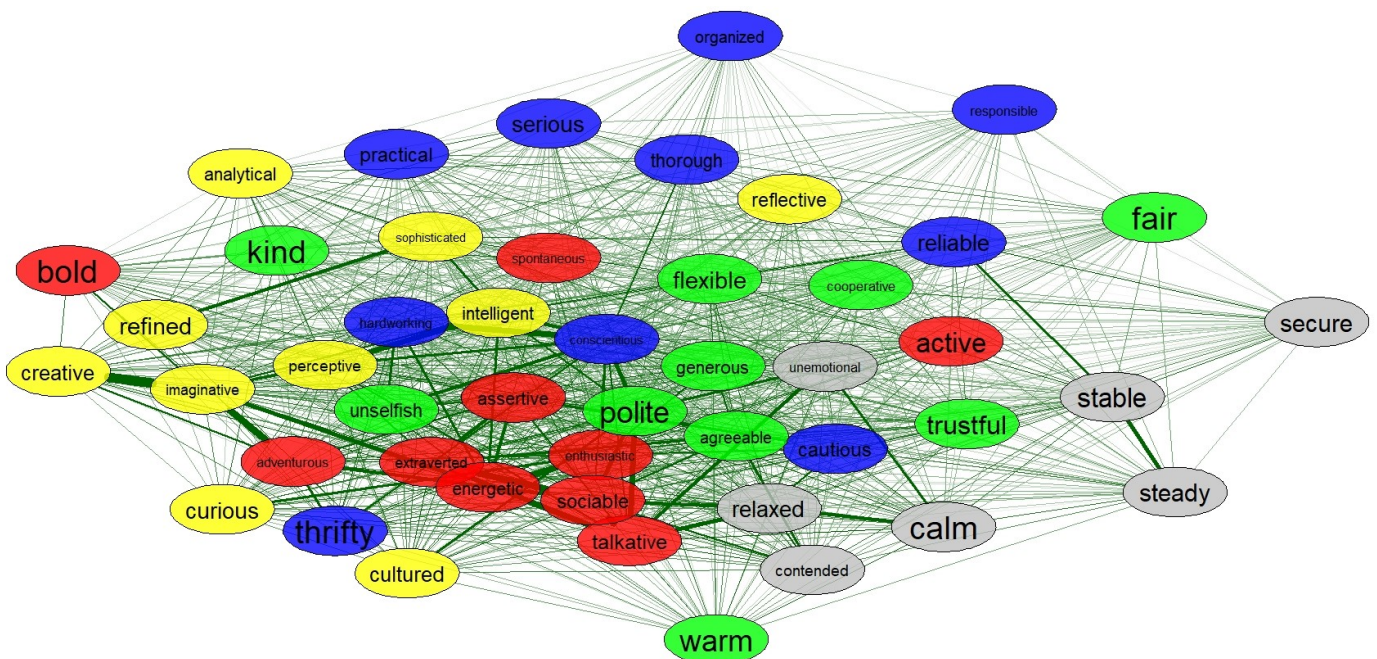
V této studii nebyli využiti participanti přímo, jediným lidským vzorkem, jehož data jsou využita jsou sebehodnocení převzatá od Sauciera (2019). Jedná se o vzorek 310 respondentů, z nichž je 187 z USA a 123 z Austrálie. Počet žen ve vzorku je 200 (64.5 %). Saucier neuvádí další demografické údaje. Povaha vzorku textů, který sloužil pro stanovení vektorových reprezentací je popsána v kapitole 2.2.2 empirické části.

4 Výsledky

4.1 První výzkumná otázka

4.1.1 Explorace dat

Síťový graf uvedený jako Obrázek 1 na první pohled ukazuje, že se v těchto datech nachází relativně konzistentní shluky, nepůsobí jako náhodná distribuce slov. Abych ilustroval některá úskalí metody word2vec vyplývající z analýzy přirozeného jazyka, tak popíši některé fenomény v grafu podrobněji. Modrá slova spadají do dimenze svědomitosti, zaujímají zejména horní část obrázku, nicméně 4 slova se nachází od toho clusteru dále. U slova „hardworking“ tomu zřejmě bude silnou vazbou na slovo „intelligent“ neboť tato dvě slova se v amerických textech častěji vyskytují u sebe při popisu osob. Osoby jsou popisovány jako „intelligent and hardworking“ když se mluví o výjimečném pracovním výkonu, protože obě tyto charakteristiky s ním korelují. Tedy může být problém, že vztahy mezi adjektivy mohou být zkresleny jejich často zmiňovaným vztahem v přirozeném jazyce s třetí proměnnou. Podobně tomu může být se slovem „conscientious“, ale u „thrifty“ mě nenapadá plausibilní vysvětlení pro jeho polohu mimo apriorní dimenzi. Poslední mimoklastrové slovo „cautious“ se nachází v blízkosti faktoru emoční stability, což by mohlo být tím, že opatrnost je jak znakem uměřenosti, tak znakem obav, právě tento druhý význam, nebo spíše častý kontext, zřejmě vychýlil polohu tohoto slova. Nejkoherentnějším faktorem je zde extraverteze označená červeně. Jedná se o tradičně nejvýznamnější faktor v lexikálních faktorech co do vysvětlené variance a zároveň je pravděpodobně



Obrázek 1 Síťová analýza vztahů word2vec vektorů pro 47 osobnostně popisných adjektiv

nejdiskutovanější lidskou charakteristikou (a z poslední kapitoly přehledu literatury plyne, že se jedná i o nejdiskutovanější „psychologickou“ charakteristiku robotů). U extraverte zde vidíme tři mimoklastrová slova, „active“ a „bold“ mohou trpět tím, že jsou mnohoznačná a použitelná ve více kontextech. Významové vektory totiž kódují každé unikátní slovo jako jeden vektor nehledě na význam, takže vektor pro „active“ bude smíšen z kontextů jak popisu osobnosti, tak ale i významů jako aktivní sopka nebo aktivní opatření. Podobně „bold“ krom odvahy či troufalosti znamená i tučné písmo. Naopak většina červeně podbarvených slov tvořících klastr jsou spíše jednoznačná, použitelná zejména pro popis osobnosti.

4.1.2 Konfirmační faktorová analýza

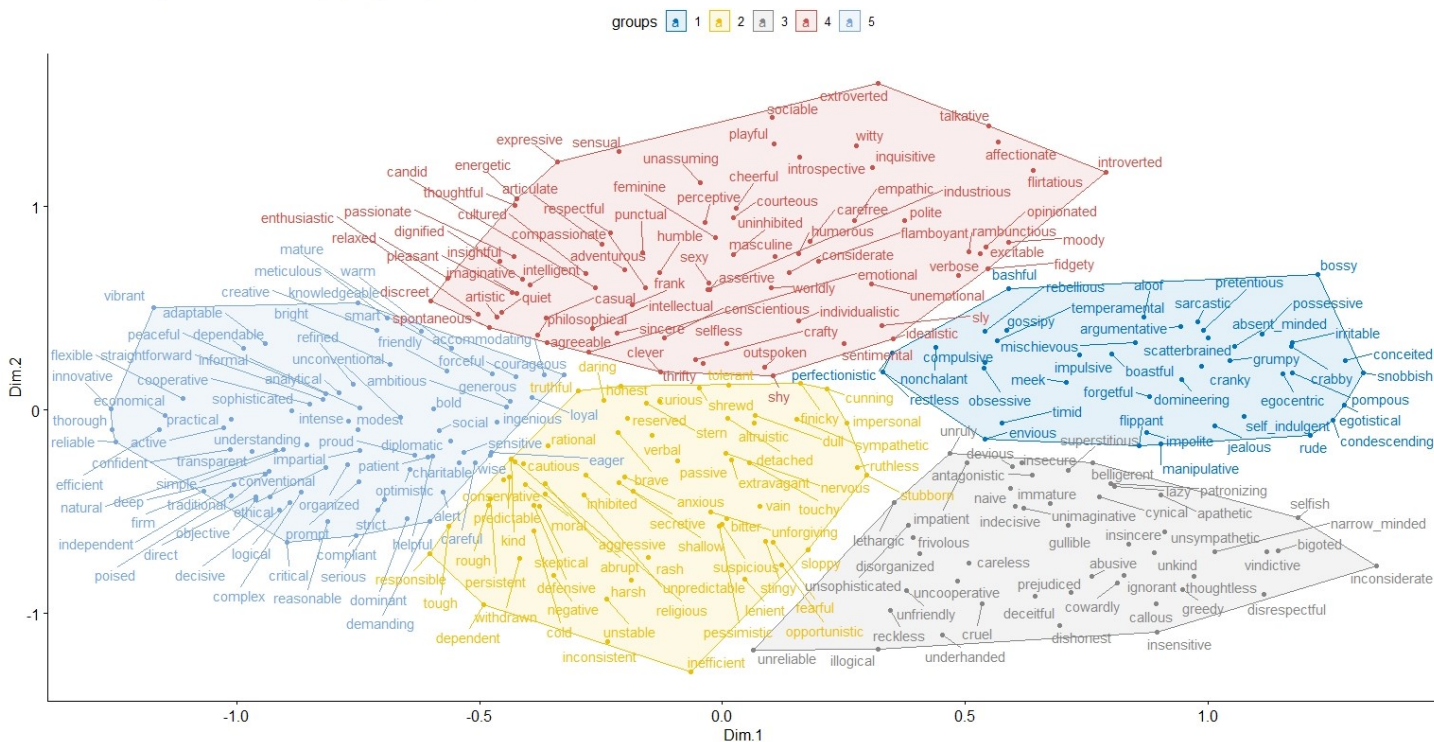
Tabulka 1 Konfirmační faktorová analýza – word2vec vektory seskupené podle Velké pětky– std. faktorové nálože

Extraverte		Přívětivost		Svědomitost		Emoční stabilita		Otevřenost	
adjektivum	Std. Odhad.	adjektivum	Std. Odhad.	adjektivum	Std. Odhad.	adjektivum	Std. Odhad.	adjektivum	Std. Odhad.
extraverted	0.7	warm	0.416	organized	0.225	calm	0.624	intelligent	0.684
energetic	0.669	kind	0.4	responsible	0.318	relaxed	0.713	perceptive	0.701
talkative	0.701	cooperative	0.464	reliable	0.423	stable	0.492	analytical	0.452
enthusiastic	0.543	unselfish	0.533	conscientious	0.758	contended	0.586	reflective	0.373
bold	0.283	polite	0.655	practical	0.433	secure	0.308	curious	0.451
active	0.461	agreeable	0.648	thorough	0.519	unemotional	0.513	imaginative	0.613
spontaneous	0.47	trustful	0.478	hardworking	0.475	steady	0.397	creative	0.456
assertive	0.665	generous	0.552	thrifty	0.548			cultured	0.509
adventurous	0.592	flexible	0.495	cautious	0.52			refined	0.435
sociable	0.756	fair	0.315	serious	0.41			sophisticated	0.563

Model úspěšně zkonvergoval, nicméně jeho fit byl podle běžně uplatňovaných hranic nedostatečný, a to i po přidání residuálních kovariancí u položek, kde to bylo teoreticky obhajitelné, např. mezi „stable“ a „steady“. Většinou jsem tak činil v rámci stejné dimenze a pouze tam, kde byl modifikační index takové změny vyšší než 10. Klasický chí-kvadrát test pro rozdíl modelové a residuální kovarianční matice vyšel signifikantní, ($X^2(1008)=2351.77, p<.001$), což je známkou modelu nedostatečně odpovídajícího datům. Stejně tak Comparative fit index byl výrazně pod hranicí 0.9 (0.738) a RMSEA nad 0.05 - 0.067 (90 %CI= (0.63,0.70)), SRMR vyšel relativně dobrý, 0.07. V Tabulce 1 uvádím hodnoty faktorových náloží pro příslušné dimenze. Většina těchto náloží má vyšší hodnotu než 0.4, která se běžně používá jako hranice pro přijatelnou položku v psychometrii, ale je otázkou, zda je možné stejné kritérium uplatňovat v tomto speciálním případě. Tato hranice

dimenzi dosahují slova s negativní valencí, např. „unkind“, „cruel“, „unstable“, „dull“. Druhá dimenze (osa Y) je o něco méně jasná. V opozici proti sobě stojí slova „masculine“ (nízká hodnota) a „feminine“ (vysoká hodnota). Ale jedná se spíše o odraz stereotypů reprodukováných v přirozeném jazyce, než že by se jednalo o dimenzi genderu. Vysoké hodnoty napříč shluky zachycují např. slova „withdrawn“, „timid“, „patient“, „peacefull“ a nízké hodnoty naopak „rough“, „agressive“, „confident“ a „independent“. Jedná se tedy zřejmě o dimenzi sociální dominance nebo možná přímo extraverze, což by bylo v souladu i s tím, že extraverze bývá nejvýraznější dimenzí v lexikálních přístupech (valence se standardně nezkoumá jako samostatný faktor).

První dimenze reprezentací vypočítaných z word2vec vektorů zřejmě stejně jako u sebeuposuzování zachycuje zejména valenci nebo sociální desirabilitu. Velmi nízké hodnoty jsou charakterizovány např. slovy „reliable“, „vibrant“, „innovative“ a naopak velmi vysoké např. „inconsiderate“, „bossy“, „rude“. Negativní slova jsou obsažena dominantně v druhém (žlutém) shluku a částečně v pravé části pátého (světle modrého) a pravých dvou třetin třetího (šedého) faktoru, tedy ideální osa pro zachycení valence by vznikla rotací souřadnicového prostoru zhruba kolem bodu -0.5 na ose X o 60 stupňů proti směru běhu hodinových ručiček. Interpretace druhé dimenze je ale i zde složitější, nejvýše položený první shluk obsahuje nejvýše slovo „extraverted“ a i v okolí se nachází slova pro vysoké



Obrázek 3 Shluková analýza 302 osobnostně popisných adjektiv podle word2vec vektorů

hodnoty extraverze, tedy se nabízí tato dimenze, ale je třeba zvážit další hlediska. Předně slovo „introverted“ je velmi blízko slovu „extraverted“, což ukazuje na fundamentální rozdíl mezi word2vec a sebehodnoceními. Zde jsou jako velmi podobná hodnocena i slova s opačnou valencí, protože valence je jen jeden z odstínů významu slova, kdežto u sebehodnocení je valence přímo obsažená ve skóru, který definuje význam slova.

Vzniklé shluky bychom mohli rozdělit podle valence. Shluky 1 a 3 z vektorových reprezentací a 1 a 5 ze sebehodnocení obsahují dominantně slova s negativní valencí a shluky 5 a 4 z vektorových reprezentací a shluky 3 a 4 ze sebehodnocení slova s valencí pozitivní. Klastř 2 z vektorových reprezentací i klastř 2 ze sebehodnocení je pak složen vyrovnaněji ze slov s oběma valencemi. Překryvy mezi klastřy podle typu dat shrnuje tabulka 3. Číslice v nadpisu sloupců a řádků se odkazují ke konkrétním shlukům, řádky jsou pro word2vec vektory a sloupce pro sebehodnocení. Dvojkřížek označuje počet slov v daném překryvu, X% označuje kolik procent daný počet tvoří ze součtu v řádku (pro vektory), a Y% značí kolik procent tvoří daný počet ze součtu sloupce (pro sebehodnocení). Tedy například když by mě zajímalo kolik % slov ze shluku 5 podle vektorů je obsaženo ve shluku 3 podle sebehodnocení, tak naleznu řádek tabulky označený „5“, podívám se do sloupce nadepsaného „3“ a odpovědí je číslo z podsloupečku X%, tedy 35.5%. Číslo z podsloupečku Y% naopak označuje kolik % slov ze třetího shluku podle sebehodnocení je obsaženo v pátém shluku podle vektorů. Zde bych rád znovu připomněl, že číslování shluků je arbitrární a nezávislé napříč daty, tedy není důvod očekávat, že např. shluky označené stejným číslem by měly sdílet více slov.

Na této kontingenční tabulce jsem provedl chí-kvadrát test nezávislosti, jež vyšel signifikantní, $X^2(16, N = 302) = 177.22, p < .001$. Velikost efektu měřená Cramerovým V vyšla 0.383, což odpovídá střední velikosti efektu.

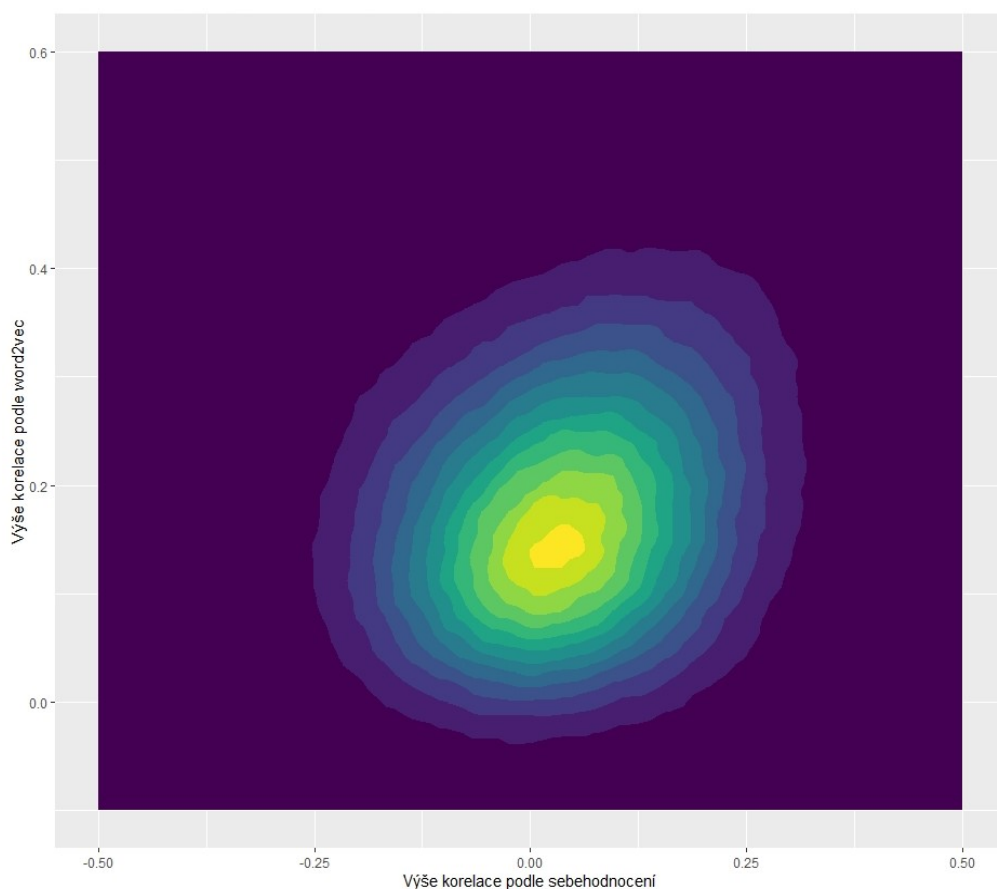
Tabulka 2 Kontingenční tabulka shlukové analýzy pro word2vec vektory (řádky) a sebehodnocení (sloupce)

	1			2			3			4			5			
	#	X%	Y%	#	X%	Y%	#	X%	Y%	#	X%	Y%	#	X%	Y%	suma
1	14	34.1%	24.1%	13	31.7%	26.5%	0	0.0%	0.0%	3	7.3%	3.9%	11	26.8%	17.5%	41
2	19	30.6%	32.8%	10	16.1%	20.4%	8	12.9%	14.3%	10	16.1%	13.2%	15	24.2%	23.8%	62
3	10	21.3%	17.2%	1	2.1%	2.0%	0	0.0%	0.0%	2	4.3%	2.6%	34	72.3%	54.0%	47
4	9	11.8%	15.5%	14	18.4%	28.6%	21	27.6%	37.5%	29	38.2%	38.2%	3	3.9%	4.8%	76
5	6	7.9%	10.3%	11	14.5%	22.4%	27	35.5%	48.2%	32	42.1%	42.1%	0	0.0%	0.0%	76
suma	58			49			56			76			63			302

4.2.2 Souvislost vztahů mezi adjektivy podle sebehodnocení a word2vec

Obrázek 4 zobrazuje 2D hustotu mezi výši korelací podle sebehodnocení a word2vec, jedná se vlastně o společný histogram, který zobrazuje distribuci korelací podle těchto dvou typů dat napříč úrovněmi obou proměnných současně. Z grafu lze jednak odhadovat, že mezi oběma proměnnými existuje korelace zhruba na přelomu nižší a střední výše. Velmi zajímavý je fakt, že tam kde sebehodnocení přisuzuje nulový vztah, word2vec vidí nějakou souvislost, je zde dobře patrná tendence word2vec systematicky přisuzovat vyšší velikost asociace. Za tímto fenoménem je fakt, že word2vec kóduje komplexní význam slov, takže například všechny osobnostní slova nebo třeba být jen slova relevantní k živým věcem, nebo všechna adjektiva spolu budou korelovat nehledě na jejich vztah na úrovni reprezentovaných osobnostních rysů.

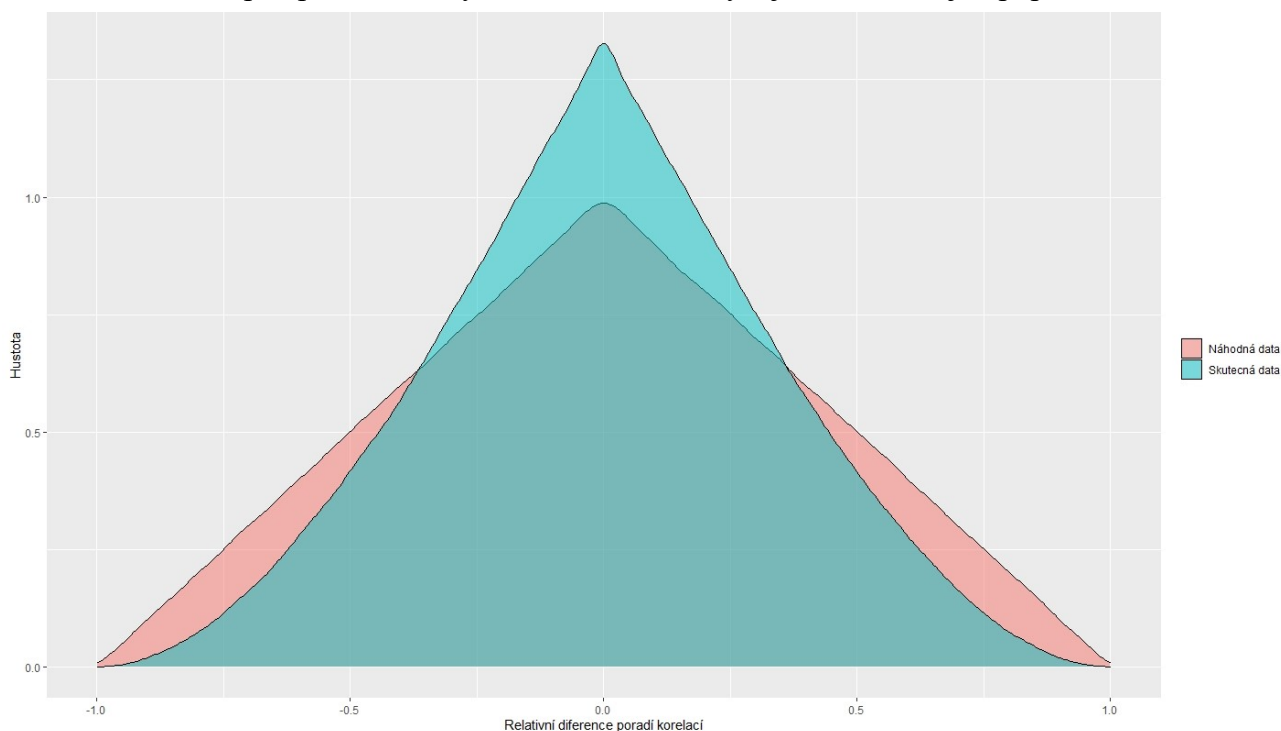
Pro ověření nakolik obecně souvisí korelace mezi vektory a mezi sebehodnoceními jsem využil shodná data jako v předchozí podkapitole, ale tentokrát pro 1223 osobnostně popisných adjektiv. Každá z použitých korelačních matic byla tedy tvořena 1 495 729 korelacemi, když pomínu symetrii korelační matice podle uhlopříčky. Spearmanova



Obrázek 4 2D graf hustoty pro výši korelací podle sebehodnocení a word2vec

korelace mezi velikostmi korelací podle obou metod vyšla díky obrovskému vzorku samozřejmě signifikantní ($p < .001$) ale i s netriviální hodnotou 0.334. Tato hodnota sice odpovídá spíše slabšímu efektu v tradiční interpretaci korelací ale je třeba zdůraznit, že toto je vztah mezi všemi myslitelnými asociacemi pro daný set slov, u většiny z nich bychom nepředpokládali, že by měl nějaký vztah existovat a správnost pořadí těchto korelací v oblasti kolem 0 není tak významná jako v absolutně vyšších hodnotách.

Pro lepší prezentaci výsledků těchto analýz jsem zavedl již popsanou metriku

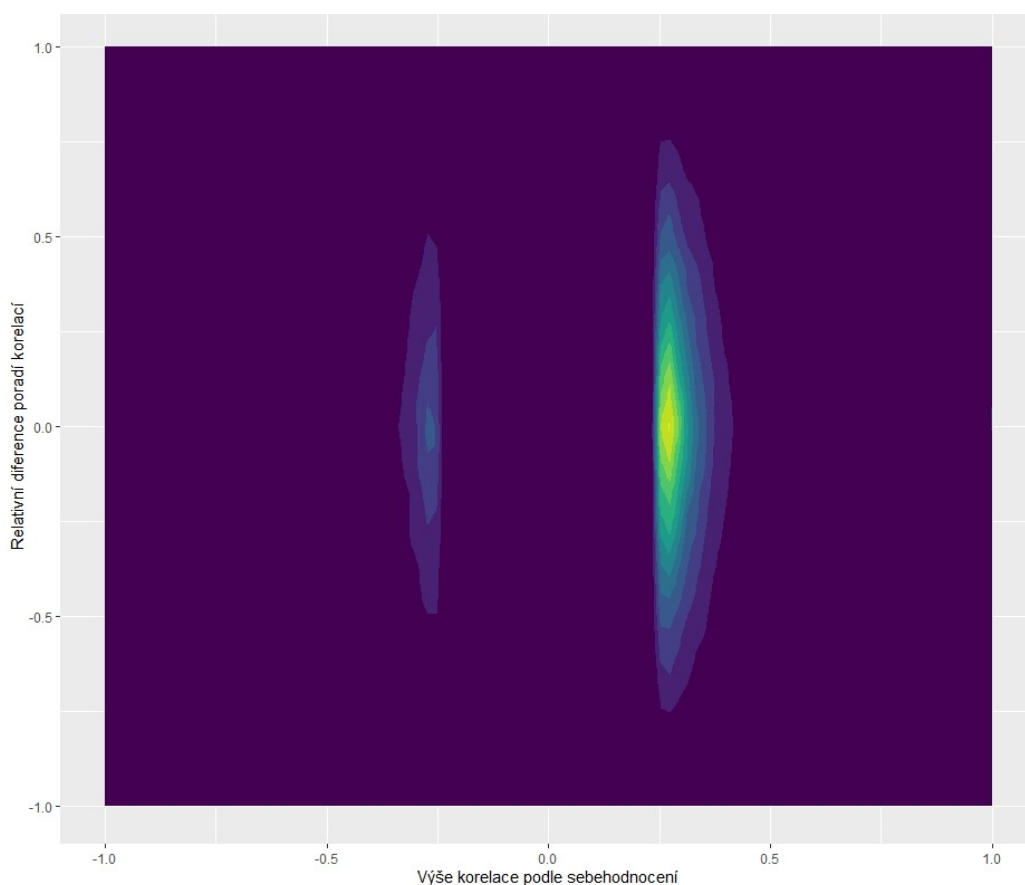


Obrázek 5 Hustota relativních diferencí pořadí velikosti korelací mezi korelacemi adjektiv podle sebehodnocení a word2vec na skutečných a náhodných datech

relativní diference pořadí korelací. K názorné ilustraci výsledků této metriky jsem se rozhodl využít graf, viz Obrázek 5 popisující hustotu těchto jednotlivých diferencí. Jedná se vlastně o zobecněnou formu histogramu s arbitrární „četností“ na ose Y, na ose X je již zmiňovaná metrika. Plocha se zelenou barvou (včetně překryvu s červenou plochou) vyznačuje hustotu pro skutečná data popsaná výše. Vzhledem k tomu, že tato metrika nemá známá interpretační pásma a jedná se o ad hoc zavedený ukazatel, rozhodl jsem se ji porovnat s náhodnými daty. Nasimuloval jsem náhodné pořadí pro dva vektory stejné velikosti jako měly vektory spočítané na skutečných datech, tedy $2 \cdot 1223 \cdot 1223$ čísel. Na takto vzniklých pořadích jsem pak aplikoval stejný postup výpočtu popsané metriky. Výsledek pro simulovaná data je na grafu vyznačen červeně (včetně překryvu). Vzhledem k obrovské velikosti „vzorku“ a jasnému efektu neprovádím test statistické hypotézy. Pokud bychom se snažili podle

velikosti seřadit vztahy mezi osobnostními adjektivy v sebehodnoceních, tak by se nám dařilo lépe pokud bychom jako nápovědu využili jejich vztahy dané word2vec vektory, protože hustota náhodného řešení se přesouvá díky této nápovědě směrem doprostřed (k 0 na ose X), tedy směrem k nižšímu rozdílu a větší shodě. Z grafu je patrné, že jednak ubývá vztahů, které mají velmi vysoké hodnoty diferencí, např. 0.8 = chyba v odhadu pořadí o 80 % a zároveň výrazně přibylo těch s diferencí velmi blízko 0, tedy těch na jejichž pořadí se obě metody shodnou.

Poslední výsledek vzbuzuje další otázku. Nejsou slova s nízkou diferencí pouze ta, která označují pro klasifikace ne tak důležité nulové vztahy? Tedy zda se nejedná disproporčně o korelace s nízkou absolutní hodnotou. Abych mohl toto ukázat vytvořil jsem 2D graf hustoty, viz Obrázek 6, který zobrazuje hustoty pro velikost korelací $\langle -1, -0.25 \rangle$ a $\langle 0.25, 1 \rangle$, tedy pro pásmo vyšších hodnot korelací, kterých bylo minimum (kdybych zahrnul všechny korelace, v grafu by nebyla opticky patrná situace v oblasti vyšších korelací). Z grafu vyplývá, že většina těchto netriviálních korelací se nachází v oblasti s nízkou diferencí, a tedy lze říci, že pozorovaný predikční úspěch modelu není dán pouze vztahy, který by byly na okraji zájmu.



Obrázek 6 2D graf hustoty pro výši korelací podle sebehodnocení a relativní difference pořadí korelací

5 Diskuse

5.1 Reflexe výsledků

Tato práce je prvním mně dosud známým představením tohoto alternativního pojetí studia struktury osobnosti lexikálním přístupem, a proto v literatuře dosud nejsou srovnatelné výzkumy s jejichž výsledky by bylo možné práci přímo porovnat. I proto jsem se snažil použít několik různých analytických přístupů pro hodnocení předkládané metody, abych měl co nejkomplexnější srovnání výsledků v rámci práce.

Za nejdůležitější výsledek považuji vztah mezi klasifikací podle sebehodnocení a word2vec. Ukázalo se, že alespoň částečně lze predikovat strukturu osobnosti podle užívání slov v přirozeném jazyce bez použití dotazování jakýchkoliv konkrétní lidí ale i bez apriorního teoretizování výzkumníka. Velikost tohoto vztahu sice není výrazná, ale považuji tento výsledek za „proof of concept“ – důkaz, že něco takového lze provést. Vzniklé klastry slov neodpovídaly výrazně klasickým dimenzím Velké pětky, ale vzhledem k relativně nízké závislosti existuje riziko, že část tohoto nesouladu je dána spíše náhodou chybou měření než systematickou odlišností struktury osobnosti.

Za důležité též považuji, že se podařila demonstrovat možnost s téměř nulovými náklady zpracovávat data o 1223 osobnostně popisných adjektivech. Analýza tohoto výběru slov vyvrátila, že by word2vec byl dobrý v predikci pouze nulových vztahů. Pomocí simulace náhodných dat jsem také ukázal, že je tento vztah výrazný a jak se mění kvalita predikce oproti náhodě napříč spektrem chyb odhadů.

V neposlední řadě jsem využil konfirmační faktorovou analýzu pro test dobře známé apriorní struktury adjektiv na datech z word2vec. Tyto výsledky vzbudily ale spíše více otázek než odpovědí. Podle klasických kritérií pro interpretaci CFA sice model neodpovídal datům, je ale otázkou, zda je tento model vhodný pro test takových dat. Bohužel v literatuře o word2vec jde vždy o induktivní postupy a tento problém není tedy zřejmě statisticky ještě vyřešen. Lepší indikátor podobnosti vztahů než korelační koeficient by mohla být kosinová vzdálenost, tu ale zase nelze využít jako vstup CFA.

5.2 Limity výzkumného projektu

Prvním limitem tohoto projektu byla samotná vstupní data. Vzhledem k vysoké náročnosti sběru reprezentativní korpů, ať již co do velikosti nebo struktury, a jejich dalšího výpočetního zpracování, jsem využil již natrénovaný word2vec model, a tedy jsem si nemohl

volně stanovit jeho parametry. Kdybych tuto možnost měl, pravděpodobně bych ale zvolil parametry velmi podobné. Tento dataset byl tvořen převážně články amerických novinářů a tedy může v modelu reprodukovat různé jejich kulturní stereotypy nebo i obecné kognitivní biasy jako např. haló efekt, ten by mohl být zodpovědný např. za mnou pozorovaný vztah mezi některými adjektivy inteligence a adjektivy svědomitosti, protože pokud je někdo úspěšný, máme tendenci mu přisoudit pozitivní atributy, které mu ve skutečnosti nenáleží (Lachman & Bass, 1985). Další nevýhodou word2vec vektorů je nemožnost odlišit různé významy stejného slova, což bylo zřejmě také důvodem některých výsledků, zejména patrných v síťové analýze u první výzkumné otázky. V neposlední řadě to, že se na rozdíl od klasických lexikálních projektů, kde jsou na začátku zahrnuta všechna adjektiva nehledě na zdroj jejich použití, tento word2vec model byl natrénován pouze na textech určených k veřejné prezentaci. Rysy se mohou lišit v tom, jak dobře se o nich komunikuje nebo i tím, jak je společensky vhodné o nich komunikovat, či jen jiným pojmáním, např. laikové do inteligence často zařazují i sociální dovednosti, a proto bychom zde zřejmě našli word2vec přístupem vztah, ale klasickým studiemi pomocí testů a dotazníků zřejmě nikoliv (Hoskens & de Boeck, 1991). Limitující mohlo být i využití dat Gerarda Sauciera, nicméně jedná se o jedinečný dataset, který bylo velmi náročné vytvořit, a tedy jsem neměl jinou možnost než využít již dataset existující. Zároveň stejně jako v předchozím případě bych při sestavování takového datasetu zřejmě postupoval velmi podobně.

Samotná metoda nahrazení dat ze sebehodnocení daty z word2vec reprezentací má mnoho potenciálních úskalí. Za nejzásadnější považuji to, že nelze snadno sbírat nové, a hlavně nezávislé, vzorky populace jako u dotazníkových metod. Velmi problematickou vlastností těchto modelů je fakt, že vlastně nevíme, podobně jako u ostatních metod strojového učení, proč spolu daná slova vlastně souvisí. Kromě souvislosti na úrovni příbuzného osobnostního rysu se do naměřeného vztahu mohou promítat další vlastnosti jako například slovní druh (pokud bychom se nezaměřili pouze na adjektiva), časté nerelevantní kontexty slov nebo mnohoznačnost. Další nevýhodou je, že takto lze studovat pouze slova, která se vyskytují v jazykových korpech s určitou četností, kdežto do dotazníku můžeme umístit jakékoliv srozumitelné slovo.

5.3 Uplatnění výsledků a doporučení pro další výzkum

Celkové výsledky ukazují, že vektorové reprezentace významu slov zachycují minimálně část relevantní variance v osobnostních dimenzích, a tedy má smysl nad touto metodou uvažovat jako nad dalším možným úhlem pohledu na lidskou osobnost. Tyto

výsledky mohou sloužit jako opodstatnění navazujícího výzkumu ať již specificky v oblasti lexikální hypotézy, obecněji pro překryv výpočetní lingvistiky a psychologie osobnosti nebo zcela obecně mohou stimulovat zájem o uplatnění nových IT technologií v psychologickém výzkumu. Podařilo se demonstrovat, že analýzy i s tímto komplexním datovým typem lze provádět pomocí relativně jednoduché statistiky, a tedy by se mohly stát pravidelnější součástí psychologických studií. Jak jsem ukázal v přehledu literatury, již teď se používají word2vec modely pro odhad osobnosti konkrétního člověka, ale zcela ateoreticky, zřejmě by zde šlo docílit lepších výsledků zapojením apriorní a explicitní struktury osobnostně relevantních slov.

Navazující výzkum by se mohl například zaměřit na možnosti získávání co nejčistších dat. Bylo by užitečné vědět, jak se naměřené vztahy mezi slovy mění s velikostí základního korpu, jeho reprezentativitou, jazykem a podobně. Možným způsobem, jak řešit rozdíly mezi různými word2vec modely v rámci jednoho jazyka (např. v angličtině je těchto modelů veřejně dostupných několik), by bylo zprůměrování naměřených vztahů napříč různými korpusy. Další možností, jak zvýšit kvalitu dat by bylo identifikovat, které konkrétní parametry v daných vektorech reprezentují irelevantní variabilitu mezi slovy, např. zda jde o adjektivum nebo adverbium. Tohoto lze docílit například tak, že se sestrojí vektor zachycující tento rozdíl a od všech použitých vektorů slov se odečte. Velmi zajímavou možností této metody je také objevení „nových“ osobnostně popisných slov. Je pravdou, že word2vec model pravděpodobně neobsahuje všechna slova dohledatelná ve slovníku, zejména archaická. Ale na druhou stranu, pokud je trénován na aktuálních datech, např. internetových článcích a příspěvcích na sociálních sítích lze zachytit i slova, která zatím nejsou ve standardních slovnících uvedena. Tohoto by se dalo technicky docílit tak, že se vyberou slova reprezentující určitou dimenzi, jejich vektory se zprůměrují, tím vznikne vektor kódující danou dimenzi (za předpokladu absence interakčních efektů v rámci vektoru), spočítají se vzdálenosti tohoto vektoru od ostatních slov v modelu a seřadí podle velikosti. Pokud daná slova nejsou již zahrnuta v našem kódujícím vektoru, možná jsme právě narazili na nová slova relevantní pro výzkum osobnosti.

Až se podaří více ověřit validitu této metody, objasnit kvalitu dat a případně implementovat metody jejich purifikace, bylo by ultimátní aplikací této metody zopakování lexikálních projektů tímto novým přístupem. Mohli bychom objevit nové osobnostní dimenze, nebo alespoň alternativní strukturu těch stávajících. Minimálně by nám tyto lexikální projekty bez lidské účasti (alespoň na straně participantů) přinesly dosud nevyužitý

pohled na rozdíl reprezentace osobnosti v přirozeném jazyce, tedy jakési implicitní folkové teorie osobnosti, a strukturou sebeposouzení předloženými slovy.

6 Závěr

V literárně přehledové části jsem definoval relevantní konstrukty pro studium struktury osobnosti lexikálním přístupem a vymezil dimenze Velké pětky, ke kterým se často odkazují v druhé části přehledu literatury a výzkumné části práce. V druhé kapitole přehledu literatury jsem popsal, jak rozvoj IT technologií, čím dál tím více proniká i do psychologického výzkumu. Zejména jsem ukázal stupňující se pokusy o predikci osobnosti z online chování pomocí metod výpočetní lingvistiky, ale i některých dalších metod.

V praktické části práce jsem představil nový možný přístup ke studiu struktury osobnosti lexikálním přístupem využívající „vetkanost“ konstruktů osobnosti do přirozeného jazyka. Podrobně jsem vysvětlil metodu word2vec, který tento přístup umožňuje a pak ji aplikoval na osobnostní adjektiva k nimž jsem pro srovnávání získal zároveň data z klasického dotazníkového přístupu. Celkově lze výsledky shrnout tak, že tato metoda dává nové odpovědi na staré otázky, ve kterých je sice velká míra šumu, ale v několika analýzách bylo ukázáno, že tato metoda má i jasně nenáhodný přínos.

Předkládané výsledky jsou pouze prvotním pilotním ověřením tohoto přístupu, a proto v diskuzi navrhuji řadu dalších možných námětů pro navazující výzkumy. Věřím, že jednoho dne bude možné zopakovat lexikální projekty v mnoha světových jazycích právě tímto přístupem, a i když vzhledem k limitům metody které popisují pravděpodobně nenahradí dotazníkové paradigma, přinesou cenný pohled na osobnost v lidském jazyce a jazyk v lidské osobnosti.

Seznam použité literatury

- Arellano, D., Varona, J., Perales, F. J., Bee, N., Janowski, K., & André, E. (2011). Influence of Head Orientation in Perception of Personality Traits in Virtual Agents (Extended Abstract). *N Proc. Int. Conf. Auton. Agents Multiagent Syst*, 1093–1094. www.ifaamas.org
- Cooper, A. B., Blake, A. B., Pauletti, R. E., Cooper, P. J., Sherman, R. A., & Lee, D. I. (2020). Personality Assessment through the Situational and Behavioral Features of Instagram Photos. *European Journal of Psychological Assessment*, 36(6), 959–972. <https://doi.org/10.1027/1015-5759/a000596>
- di Blas, L., & Forzi, M. (1999). Refining a descriptive structure of personality attributes in the Italian language: The abridged Big Three circumplex structure. *Journal of Personality and Social Psychology*, 76(3), 451–481. <https://doi.org/10.1037/0022-3514.76.3.451>
- Eliasz, A., Hampson, S. E., & de Raad, B. (2005). Advances in personality psychology. In *Advances in Personality Psychology* (Vol. 2). <https://doi.org/10.4324/9780203000953>
- Evans, D. C., Gosling, S. D., & Carroll, A. (2008). What elements of an online social networking profile predict target-rater agreement in personality impressions? *ICWSM 2008 - Proceedings of the 2nd International Conference on Weblogs and Social Media*.
- Fischer, R., Karl, J. A., Luczak-Roesch, M., Fetvadjev, V. H., & Grener, A. (2020). Tracing Personality Structure in Narratives: A Computational Bottom-Up Approach to Unpack Writers, Characters, and Personality in Historical Context. *European Journal of Personality*, 34(5). <https://doi.org/10.1002/per.2270>
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In *Review of personality and social psychology: Vol.2* (Vol. 2).
- Goldberg, L. R. (1990). An Alternative “Description of Personality”: The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, 59(6), 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216>

- Goldberg, L. R. (1992). The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, 4(1), 26–42. <https://doi.org/10.1037/1040-3590.4.1.26>
- Google. (2013). *Google Code Archive - Long-term storage for Google Code Project Hosting*. <https://code.google.com/archive/p/word2vec/>
- Hahn, D. W., Lee, K., & Ashton, M. C. (1999). A factor analysis of the most frequently used Korean personality trait adjectives. *European Journal of Personality*, 13(4). [https://doi.org/10.1002/\(SICI\)1099-0984\(199907/08\)13:4<261::AID-PER340>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1099-0984(199907/08)13:4<261::AID-PER340>3.0.CO;2-B)
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. In *Nature* (Vol. 585, Issue 7825). <https://doi.org/10.1038/s41586-020-2649-2>
- Hartigan, A., & Wong, M. A. (1979). A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society*, 28(1).
- Hoskens, M., & de Boeck, P. (1991). An Implicit Theory of Intelligence-Related Mental Activities. *Journal of Personality*, 59(4), 793–814. <https://doi.org/10.1111/j.1467-6494.1991.tb00932.x>
- JASP Team (2020). JASP (Version 0.14.1)[Computer software].
- Jayawickreme, E., Zachry, C. E., & Fleeson, W. (2019). Whole Trait Theory: An integrative approach to examining personality structure and process. *Personality and Individual Differences*, 136, 2–11. <https://doi.org/10.1016/j.paid.2018.06.045>
- Jones, F. N., & Skinner, B. F. (1939). The Behavior of Organisms: An Experimental Analysis. *The American Journal of Psychology*, 52(4). <https://doi.org/10.2307/1416495>
- Kassambara, A. (2020). Package ‘ggpubr’: “ggplot2” Based Publication Ready Plots. *R Package Version 0.4.0*.
- Kavka, M. (2021). *Slovník, který tvoříte* vy. Čeština 2.0. www.cestina20.cz.

- Kim, Y., & Kim, J. H. (2018). Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study. *Information Processing and Management*, 54(6), 1101–1114.
<https://doi.org/10.1016/j.ipm.2018.07.005>
- Klages, L., & Johnston, W. H. (1933). The Science of Character. *Journal of Philosophy*, 30(20), 557–559. <https://doi.org/10.2307/2016365>
- Lachman, S. J., & Bass, A. R. (1985). A direct study of halo effect. *Journal of Psychology: Interdisciplinary and Applied*, 119(6).
<https://doi.org/10.1080/00223980.1985.9915460>
- Lapierre, R. T. (1934). Attitudes vs. Actions. *Social Forces*, 13(2), 230–237.
<https://doi.org/10.2307/2570339>
- Lee, K. M., Peng, W., Jin, S. A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication*, 56(4), 754–772.
<https://doi.org/10.1111/j.1460-2466.2006.00318.x>
- Mairesse, F., Uk, M. A., Mehl, M. R., & Moore, R. K. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial Intelligence Research* (Vol. 30).
- Mardia, K. v. (1978). Some Properties of Classical Multi-Dimensional Scaling. *Communications in Statistics - Theory and Methods*, 7(13).
<https://doi.org/10.1080/03610927808827707>
- Mcrorie, M., Sneddon, I., Mckeown, G., Bevacqua, E., de Sevin, E., & Pelachaud, C. (2012). Evaluation of four designed virtual agent personalities. *IEEE Transactions on Affective Computing*, 3(3), 311–322. <https://doi.org/10.1109/T-AFFC.2011.38>
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5), 862–877. <https://doi.org/10.1037/0022-3514.90.5.862>

- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. <https://doi.org/10.1037/h0043158>
- Minamikawa, A., & Yokoyama, H. (2011). Blog tells what kind of personality you have : Egogram Estimation from Japanese Weblog. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 217–220. <https://doi.org/10.1145/1958824.1958856>
- Mohammadi, G., & Vinciarelli, A. (2012). Automatic personality perception: Prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing*, *3*(3), 273–284. <https://doi.org/10.1109/T-AFFC.2012.5>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Neff, M., Wang, Y., Abbott, R., & Walker, M. (2010). Evaluating the effect of gesture and language on personality perception in conversational agents. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6356 LNAI*, 222–235. https://doi.org/10.1007/978-3-642-15892-6_24
- NLPL word embeddings repository*. (n.d.). Retrieved May 8, 2021, from <http://vectors.nlpl.eu/repository/>
- Pervin, L. A. (1994). A Critical Analysis of Current Trait Theory. *Psychological Inquiry*, *5*(2), 103–113. https://doi.org/10.1207/s15327965pli0502_1
- Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., & Zancanaro, M. (2008). Multimodal Recognition of Personality Traits in Social Interactions. *ICMI'08: Proceedings of the 10th International Conference on Multimodal Interfaces*, 53–60. <https://doi.org/10.1145/1452392.1452404>

- Polzehl, T., Möller, S., & Metze, F. (2010). Automatically assessing personality from speech. *Proceedings - 2010 IEEE 4th International Conference on Semantic Computing, ICSC 2010*, 134–140. <https://doi.org/10.1109/ICSC.2010.41>
- R Core Team. (2019). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*.
- Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).
- Remaida, A., Abdellaoui, B., Moumen, A., & Idrissi, Y. E. B. el. (2020, April 1). Personality traits analysis using Artificial Neural Networks: A Literature Survey. *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2020*. <https://doi.org/10.1109/IRASET48871.2020.9092076>
- Roget, P. M. (1852). *Thesaurus of English Words and Phrases Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition* (3rd ed.). Longman, Brown, Green and Longmans.
- Rstudio, T. (2020). RStudio: Integrated Development for R. In *Rstudio Team, PBC, Boston, MA URL <http://www.rstudio.com/>*. <https://doi.org/10.1145/3132847.3132886>
- Schuller, B., Steidl, S., Batliner, A., Nöth, E., Vinciarelli, A., Burkhardt, F., van Son, R., Weninger, F., Eyben, F., Bocklet, T., Mohammadi, G., & Weiss, B. (2012). The INTERSPEECH 2012 Speaker Trait Challenge. *Proc. Annu. Conf. Int. Speech Commun. Assoc.* <http://www.isca-speech.org/archive>
- Sinclair, A. (2010). Licence Profile: Apache License, Version 2.0. *International Free and Open Source Software Law Review*, 2(2). <https://doi.org/10.5033/ifosslr.v2i2.42>
- Soto, C. J. (2019). How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, 30(5), 711–727. <https://doi.org/10.1177/0956797619831612>
- Steele Jr, F., Evans, D. C., & Green, R. K. (2009). Is your profile picture worth 1000 words? Photo characteristics associated with personality impression agreement. *Third International AAAI Conference on Weblogs and Social Media*.

- Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., & Williams Jr., R. M. (1949). The American soldier: Adjustment during army life. (Studies in social psychology in World War II), Vol. 1. In *The American soldier: Adjustment during army life. (Studies in social psychology in World War II), Vol. 1*. Princeton Univ. Press.
- Subramanian, R., Wache, J., Abadi, M. K., Vieriu, R. L., Winkler, S., & Sebe, N. (2018). Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing*, 9(2), 147–160. <https://doi.org/10.1109/TAFFC.2016.2625250>
- Tapus, A., Țăpuș, C., & Matarić, M. J. (2008). User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, 1(2), 169–183. <https://doi.org/10.1007/s11370-008-0017-4>
- Tarnavsky Eitan, A., Smolyansky, E., & Knaan Harpaz, I. (2020). Connected Papers. *Connected Papers. About*.
- Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. In *Journal of Personality and Social Psychology* (Vol. 87, Issue 1, pp. 123–132). <https://doi.org/10.1037/0022-3514.87.1.123>
- Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. In *IEEE Transactions on Affective Computing* (Vol. 5, Issue 3, pp. 273–291). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/TAFFC.2014.2330816>
- Webster, N., Porter, N., Harris, W. T., Wharton, F., & Brewer, D. J. (1903). Webster's International Dictionary of the English Language. *Michigan Law Review*, 1(8). <https://doi.org/10.2307/1273778>
- Widiger, T. A. (2017). The Oxford handbook of the Five Factor Model. In *The Oxford handbook of the Five Factor Model*.
- Wiggins, J. S. (1997). In Defense of Traits. In *Handbook of Personality Psychology* (pp. 95–115). Elsevier. <https://doi.org/10.1016/b978-012134645-4/50005-6>

- Wood, D. (2015). Testing the lexical hypothesis: Are socially important traits more densely reflected in the English lexicon? *Journal of Personality and Social Psychology*, *108*(2), 317–335. <https://doi.org/10.1037/a0038343>
- Zen, G., Lepri, B., Ricci, E., & Lanz, O. (2010). Space speaks - Towards socially and personality aware visual surveillance. *MPVA '10 - Proceedings of the 2010 ACM Workshop on Multimodal Pervasive Video Analysis, Co-Located with ACM Multimedia 2010*, 37–42. <https://doi.org/10.1145/1878039.1878048>
- Zuroff, D. C. (1986). Was Gordon Allport a Trait Theorist? *Journal of Personality and Social Psychology*, *51*(5), 993–1000. <https://doi.org/10.1037/0022-3514.51.5.993>

Seznam obrázků

Obrázek 1 Síťová analýza vztahů word2vec vektorů pro 47 osobnostně popisných adjektiv	28
Obrázek 2 Shluková analýza 302 osobnostně popisných adjektiv podle sebehodnocení	30
Obrázek 3 Shluková analýza 302 osobnostně popisných adjektiv podle word2vec vektorů.....	31
Obrázek 4 2D graf hustoty pro výši korelací podle sebehodnocení a word2vec.....	33
Obrázek 5 Hustota relativních diferencí pořadí velikosti korelací mezi korelacemi adjektiv podle sebehodnocení a word2vec na skutečných a náhodných datech.....	34
Obrázek 6 2D graf hustoty pro výši korelací podle sebehodnocení a relativní difference pořadí korelací	35

Seznam tabulek

Tabulka 1 Konfirmační faktorová analýza – word2vec vektory seskupené podle Velké pětky– std. faktorové nálože	29
Tabulka 2 Kontingenční tabulka shlukové analýzy pro word2vec vektory (řádky) a sebehodnocení (sloupce).....	32

Seznam zkratek

2D	Dvoudimenzionální
ASCERTAIN	dataBASE for impliCit pERsonaliTy and Affect recognitIoN
CBOW	Continious bag of words
CFA	Confirmatory Factor Analysis
CI	Confidence interval
EEG	Elektroencefalografie
EKG	Elektrokardiografie
HEXACO-PI-R	HEXACO Personality Inventory-Revised
IT	Informační technologie
JASP	Jeffreys's Amazing Statistics Program
LIWC	Linguistic Inquiry and Word Count
NEO-FFI	NEO Five Factor Inventory
NEO-PI-3	NEO Personality Inventory 3rd version
NEO-PI-R	NEO Personality Inventory Revised
OSF	Open Science Forum
RMSEA	Root Mean Square Error of Approximation
SRMR	Standardized Root Mean Squared Residual
USA	United States of America