

Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví / Studia nových médií

Diplomová práce

Bc. Ľubica Stančíková

Genderový pohľad na prezentáciu žien na Wikipédii

Presentation of Women on Wikipedia from Gender Perspective

Vedúca práce: Mgr. Michaela Slussareff, Ph.D.

Konzultant: Mgr. Josef Šlerka, Ph.D.

Praha 2019

Pod'akovanie:

Týmto by som chcela poďakovať za pomoc pri písaní diplomovej práce predovšetkým Mgr. Michaele Slussareff, Ph.D. za dobré rady a usmernenia v rámci teoretickej časti práce a metodológie výskumu, a takisto aj Mgr. Josefovi Šlerkovi, Ph.D. za odborné vedenie počas spracovávania výskumných dát a celkového zhodnotenia.

Prehlásenie:

Prehlasujem, že som diplomovú prácu vypracovala samostatne, že som náležite citovala všetky použité pramene a literatúru, a že práca nebola použitá v rámci iného vysokoškolského štúdia či k získaniu iného alebo rovnakého titulu.

V Prahe, dňa 23. 7. 2019

Bc. Ľubica Stančíková

Identifikačný záznam

STANČÍKOVÁ, Ľubica. Genderový pohľad na prezentáciu žien na Wikipédii. 112 s. Diplomová práca (Mgr.). Univerzita Karlova, Filozofická fakulta, Ústav informačných štúdií a knihovníctví. Vedúca práce Mgr. Michaela Slussareff, PhD.

Kľúčové slová

Wikipédia, kvantitatívna analýza, dátová analýza, text mining, gender, genderové role, stereotypizácia, genderová nerovnosť, gender bias, diverzita

Abstrakt

Predkladaná diplomová práca skúma, charakterizuje a kvantifikuje spôsob prezentácie žien a mužov na českej a slovenskej Wikipédii a ich prípadných rozdielov. Teoretická časť práce popisuje aktuálny stav v rámci Wikipédie a všeobecne v médiách a upozorňuje na genderovú nerovnosť v zastúpení pohlaví v rámci editorskej základne, ktorú dáva do súvislosti s všeobecným trendom vo free culture komunitách, kde je takisto badateľná určitá nerovnosť. Ďalej sa teoretická časť práce v skratke venuje aj otázke genderu všeobecne a podrobne rozpracováva aktuálny stav poznatkov, štúdií a postupov skúmajúcich otázku genderu na Wikipédii. V rámci teoretickej časti sa práca špecificky venuje aj česko-slovenskému kontextu. Praktická časť nadväzuje a čiastočne replikuje štúdiu *First Gender, Second Sex: Gender bias on Wikipedia* (Graells-Garrido, Lalmas a Menczer, 2015), ktorá bola realizovaná metódou text miningu v softvéri RStudio a pozostávala z kvantitatívnej analýzy biografických textov o mužoch a o ženách v oboch jazykoch a sledovaním frekvencie výskytu vybraných kategórií slov týkajúcich sa genderu, referencie na opačné pohlavie, rodiny a kariéry. Pomocou štatistického softvéru RStudio sme spolu zanalyzovali 24 510 slovenských biografických článkov a 110 866 českých biografických článkov, pričom analýza získaných dát potvrdila predpokladanú nerovnosť a stereotypizáciu v prezentácii žien na Wikipédii v oboch jazykoch.

Keywords

Wikipedia, quantitative analysis, data analysis, text mining, gender, gender roles, stereotyping, gender inequality, gender bias, diversity

Abstract

This thesis examines, characterizes and quantifies how women and men are being presented on Czech and Slovak Wikipedia and further elaborates on their possible differences. The theoretical part of the thesis describes the current state of Wikipedia and media in general and draws attention to gender inequality within its editorial base, which is linked to the general trend in free culture communities, where there is also a visible inequality. Furthermore, the theoretical part of the thesis also deals with the issue of gender in general and summarizes current state of knowledge and methods of studying the issue of gender on Wikipedia. The practical part of this thesis partly replicates the study *First Gender, Second Sex: Gender Bias on Wikipedia* (Graells-Garrido, Lalmas and Menczer, 2015) using RStudio programme to do basic text mining and quantitative analysis of biographical texts of men and women in both languages, tracking word frequencies from selected word categories, namely Gender, Reference to the opposite sex, Family and family status and Career. Overall, we have analysed 24 510 Slovak biographical articles and 110 866 Czech biographical articles, and our findings have confirmed an imbalance and stereotyping in the presentation of women on Wikipedia in both languages.

OBSAH

ÚVOD.....	8
Štruktúra práce.....	8
Cieľ práce.....	9
1 Wikipédia a jej popularita.....	10
1.1 Kto sú jej používatelia a ako k nej pristupujú.....	10
1.2 Ako používatelia pristupujú k Wikipédii.....	11
2 Je Wikipédia mužská záležitosť?.....	13
2.1 Kto číta Wikipédiu.....	13
2.2 Kto edituje Wikipédiu.....	15
3 Špecifiká free culture komunity.....	19
3.1 Vzostup geekov a internetovej kontrakultúry.....	19
3.2 Nastavenie kultúry participácie.....	20
3.3 Chýbajúca inklúzia.....	22
3.4 Aké je prostredie českej a slovenskej Wikipédie a FLOSS komunity.....	25
4 Ako sa hovorí o ženách.....	27
4.1 Ako sú prezentované ženy na Wikipédii.....	27
4.2 Aké sú tendencie v profesionálnej sfére online médií a spravodajstva?.....	31
4.3 Špecifiká českého a slovenského prostredia.....	32
4.3.1 Mediálne prostredie.....	32
4.3.2 Prostredie Wikipédie.....	37
5 Stručný exkurz do genderovej lingvistiky.....	39
5.1 Ženská otázka.....	39
5.2 Populárno-vedecké genderové testy.....	41
5.3 Otázka genderu.....	43
6 Metodológia výskumu.....	45
6.1 Na úvod.....	45
6.2 Formulácia hypotéz.....	47
6.3 Metodológia.....	48
6.4 Zber dát a príprava datasetu.....	49
6.4.1 Zber dát.....	49
6.4.2 Príprava dát na analýzu.....	50
6.4.3 UDPipe.....	51
6.5 Použitie metódy analýzy.....	52
7 Čiastkové výsledky výskumu.....	54

7.1	Počet článkov a počet slov v datasete	54
7.1.1	Česká verzia Wikipédie	54
7.1.2	Slovenská verzia Wikipédie.....	57
7.1.3	Vyhodnotenie hypotéz H1 a H2 pre českú aj slovenskú Wikipédiu.....	59
7.2	Frekvencie výskytu slov.....	61
7.2.1	Česká verzia Wikipédie	61
7.2.2	Frekvencie výskytu vybraných kategórii slov	63
7.2.3	Slovenská verzia Wikipédie.....	75
7.2.4	Frekvencie výskytu vybraných kategórii slov	77
7.2.5	Vyhodnotenie hypotéz H3 a H4 pre českú aj slovenskú Wikipédiu.....	87
7.3	Porovnanie českej a slovenskej Wikipédie	90
7.4	Sumarizácia výsledkov hypotéz	92
ZÁVER	93
Sumár výsledkov výskumu	94
Hlavný prínos práce a odporúčania pre prax		96
Limity výskumu		97
Zoznam použitej literatúry		98
Zoznam použitých tabuliek.....		109
Príloha 1		111
Príloha 2		115

ÚVOD

Wikipédia je jedným z najznámejších a najvyužívanejších informačných zdrojov, z ktorého čerpajú milióny ľudí po celom svete. Dlhé roky bola Wikipédia terčom kritiky a takisto aj predmetom skúmania predovšetkým v súvislosti s kvalitou príspevkov a dôveryhodnosťou poskytovaných informácií. Tieto obavy však už boli do značnej miery prekonané a v súčasnosti sa čoraz častejšie pozeráme na Wikipédiu a na kvalitu jej článkov aj z iných hľadísk; z hľadiska wikipedickej komunity a editorov, z pohľadu jednotlivých regiónov, jazykov, kultúr, a takisto aj z hľadiska genderu a menšín. Práve problematika genderu dnes veľmi rezonuje naprieč všetkými odvetvami spoločnosti, a to predovšetkým v súvislosti so stereotypizáciou a nerovnosťou medzi pohlaviami.

Podnikajú sa rôzne kroky - od edukácie, odporúčaní a návodov až po normatívne vyhlásenia, ktoré majú prispieť k vyrovnanejším podmienkam pre obe pohlavia. Najnovšie sa opatrenia týkajú aj oblasti reklamy, ktorá bola už dlhšie terčom kritiky. Až priveľmi známe spojenie „mama varí, otec číta“, ktoré pozná snáď každý priamo z televíznej obrazovky prípadne školskej učebnice, je od decembra 2018 považované *The U.K.'s Advertising Standards Authority* za sexistické a stereotypné, a preto sa v britských reklamných vodách viac nebude vyskytovať, podobne ako aj iné príklady neslávne známe z reklamy na pracie prášky či autá.¹ Táto problematika sa značne dotýka aj zbytku populárnej kultúry, ktorá je tiež vo veľkom kritizovaná za nerovné zobrazovanie žien a mužov v audiovizuálnej produkcii (Bechdel test) a rovnako aj v žurnalistike, kde sa takisto ukazuje tendencia stereotypizácie žien a marginalizácia ich úspechov (Finkbeiner test). Keďže sa tieto tendencie ukazujú aj v tradičnej offline/online žurnalistike, Wikipédia sa pre svoju popularitu prirodzene nemohla vyhnúť podobnému skúmaniu. V súčasnosti sa genderovej problematike na Wikipédii venuje predovšetkým Claudia Wagnerová, Jahna Otterbacherová, prípadne Eduardo Graells-Garrida, ktorých výskumy ukazujú a potvrdzujú existenciu nerovnosti v zobrazovaní pohlaví aj na tomto používateľsky vytváranom a spravovanom portáli.

Štruktúra práce

V rámci predkladanej práce si postupne predstavíme všeobecne Wikipédiu s fokusom na jej českú a slovenskú verziu (kapitola 1 *Wikipédia a jej popularita*), pričom sa špeciálne v kapitole 2 *Je Wikipédia mužská záležitosť?* zameriame na zastúpenie mužov a žien

¹ <https://www.nytimes.com/2019/06/14/style/uk-gender-stereotype-ads-ban.html>

v čitateľskej a editorskej základni. Následne sa bližšie pozrieme na wikipedickú komunitu editorov ako takú a tiež na (ne)rovnú kultúru, ktorá v nej panuje (kapitola 3 *Špecifická free culture komunity*). Postupne sa náš fokus bude zužovať len na problematiku genderu všeobecne a samozrejme na genderovú problematiku na Wikipédii a v online a offline médiách vo svete a v našej geografickej oblasti (kapitola 4 *Ako sa hovorí o ženách* a kapitola 5 *Stručný exkurz do genderovej lingvistiky*). Posledná časť práce je časť praktická, ktorá pozostáva z metodológie, samotného výskumu a zhodnotenia jeho výsledkov.

Cieľ práce

Cieľom predkladanej práce je preskúmať (kvantifikovať a charakterizovať) spôsob, akým sú na českej a slovenskej Wikipédii prezentované ženy a muži a poukázať na možné rozdiely, predovšetkým stereotypné zobrazovanie a nerovnosť, čo do počtu článkov a ich kvality. Praktická časť nadväzuje a čiastočne replikuje štúdiu *First Gender, Second Sex: Gender bias on Wikipedia* (Graells-Garrido, Lalmas a Menczer, 2015), ktorá bola realizovaná metódou text miningu. V rámci práce sme posťahovali a spracovali dokopy 24 510 slovenských biografických článkov a 110 866 českých biografických článkov. Ako metódu spracovania výskumu sme zvolili kvantitatívnu dátovú analýzu s použitím voľne dostupného štatistického softvéru RStudio. Zamerali sme sa špecificky na frekvenčnú analýzu, resp. frekvencie výskytu vybraných skupín slov v rámci ženského a mužského datasetu, ktoré sme následne porovnali. Konkrétne sa jednalo o slová z kategórie Gender, Referencia na opačné pohlavie, Rodina a sociálny status a Kariéra. V rámci práce sme uvideli aj viaceré čiastkové výsledky, ktoré slúžili ako kontext k zhodnoteniu hypotéz.

V rámci práce sme si vybrali za predmet nášho výskumu českú a slovenskú Wikipédiu a to predovšetkým pre dobrú znalosť kultúrneho prostredia a jazyka. Takisto sme však chceli poukázať na fakt, že tejto problematike sa v našom geografickom prostredí nevenuje dostatočná pozornosť a stále stojí na okraji záujmu. Preto tento výskum chápeme aj ako pilotný, resp. ako prvú genderovú sondu do českého a slovenského wikipedického prostredia a dúfame, že výsledky nášho výskumu poslúžia ako základ pre ďalšie skúmania, prípadne ako podklad pre praktické opatrenia.

1 Wikipédia a jej popularita

V januári v roku 2001 Jimmy Wales a Larry Sanger spustili po prvý krát portál Wikipédia, online encyklopédiu, ktorá má v súčasnosti cez 50 miliónov článkov vo viac ako 300 jazykových mutáciách a celkovo až niečo cez 82 miliónov registrovaných používateľov². Bez väčších pochybností teda môžeme tvrdiť, že Wikipédia má dnes významné postavenie medzi voľne dostupnými informačnými zdrojmi na internete ale rovnako aj medzi etablovanými encyklopédiami, akými sú napríklad Encyclopædia Britannica. Podľa stránky Alexa Rank je celosvetovo 5. najnavštevovanejšou stránkou na webe³. Podobne populárna je aj v českom a slovenskom prostredí; v Česku jej patrí 6. a na Slovensku 10. priečka návštevnosti.⁴ Len za mesiac marec 2019⁵ používatelia navštívili slovenskú Wikipédiu 16,85 miliónov-krát a v Česku sa toto číslo vyšplhalo až na necelých 74 miliónov zhliadnutí. Okrem voľnej dostupnosti, čím predchádza tradičné encyklopédie, k jej popularite do veľkej miery prispieva aj fakt, že až 70 % návštevníkov prichádza na Wikipédiu z populárnych vyhľadávačov, pričom až 50 % prichádza z vyhľadávača Google⁶ (Prescott, 2007), ktorý je momentálne najrozšírenejším vyhľadávačom; v slovenskom prostredí bezkonkurenčne najvyužívanejším⁷ a v českom prostredí⁸ sa o používateľov delí s vyhľadávačom Seznam.cz.

1.1 Kto sú jej používatelia a ako k nej pristupujú

Lepšiu predstavu o naozaj masovom využívaní Wikipédie si môžeme vyvodit' z počtu používateľov pripojených na internet, keďže je Wikipédia voľne dostupným zdrojom, všetci s internetovým pripojením či s prístupom na internet patria medzi jej potenciálnych používateľov. V súčasnosti je v USA k internetu pripojených celkovo 65% domácností (Pew Internet Research Center, 2018). V Česku a na Slovensku je situácia ešte výrazne lepšia; podľa Eurostatu (2018) je k internetu pripojených 83 % českých a 81 % slovenských domácností.

² Štatistiky sú aktualizované na dennej báze, údaje boli aktuálne ku dňu 18. 6. 2019 a pochádzajú z portálu https://meta.wikimedia.org/wiki/List_of_Wikipedias#Grand_Total

³ Podľa štatistiky Alexa Rank Top Pages by Country aktuálne ku dňu 15. 4. 2019: <https://www.alexa.com/topsites/countries>

⁴ Podľa štatistiky Alexa Rank Top Pages by Country aktuálne ku dňu 15. 4. 2019: <https://www.alexa.com/topsites/countries>

⁵ Údaje pochádzajú zo stránky Wikimedia Statistics a boli aktuálne ku dňu 15. 4. 2019:

<https://stats.wikimedia.org/v2/#/sk.wikipedia.org>, <https://stats.wikimedia.org/v2/#/cs.wikipedia.org>

⁶ I keď sa jedná o starší výskum, štatistiky z Alexa Rank do značnej miery stále potvrdzujú jeho aktuálnosť; ku dňu 15. 4. 2019 uvádzajú približne 54% organickú návštevnosť: <https://www.alexa.com/siteinfo/wikipedia.org>

⁷ Ku dňu 16. 4. 2019 bol vyhľadávač Google.com na 1. priečke v rebríčku návštevnosti Alexa Rank a Google.sk na 3. priečke návštevnosti: <https://www.alexa.com/topsites/countries/SK>

⁸ Ku dňu 16. 4. 2019 bol vyhľadávač Google.com na 1. priečke v rebríčku návštevnosti Alexa Rank a Seznam.cz na 3. priečke návštevnosti: <https://www.alexa.com/topsites/countries/CZ>

Neprekvapivo, najvyššie percentá dosahuje prieskum predovšetkým u mladej generácie. Najpočetnejšia skupina americkej populácie, ktorá má prístup na internet, sú 18-29-roční (98 %) a s miernym poklesom ľudia vo veku 30 - 49 rokov (97 %) (Pew Internet Research Center, 2018). Podobné čísla síce nie sú k dispozícii pre Česko a Slovensko, no môžeme predpokladať, že sa od tých amerických nebudú výraznejšie líšiť. Ak sa zameriame konkrétne na Wikipédiu zistíme, že ju pravidelne využíva až 36 % respondentov v rámci vekových skupín 18 - 29 a 30 - 49 rokov, ktoré takisto dosahujú najvyššie percento, čo sa týka pripojenia na internet (Pew Internet Research Center, 2018).

1.2 Ako používatelia pristupujú k Wikipédii

Podľa amerických výskumov sa teda Wikipédia teší veľkej popularite predovšetkým u mladších používateľov; žiakov a študentov. Až jedna tretina respondentov pravidelne využíva Wikipédiu pri plnení si školských povinností predovšetkým na rýchle overenie faktov a všeobecný prehľad o problematike (Lim, 2009). Neskoršie výskumy ukazujú ešte vyššie čísla. V niektorých prípadoch sa jedná až o 87,5 % (Selwyn a Gorard, 2016) či pri menšej vzorke opýtaných dokonca až 100% (Blikstad-Balas, 2016). Po prečítaní a získaní základného prehľadu, študenti ďalej siahajú po bibliografických odkazoch pri článkoch, ktoré ich navedú k ďalším relevantným informáciám. Teda ak sa zameriame na využívanie informácií z Wikipédie, je Wikipédia považovaná a zároveň aj využívaná ako počiatočný zdroj informácií, ktorý môže viesť k objaveniu ďalších súvisiacich zdrojov (Shaw 2008; Lim 2009; Garrison 2015), na rýchle a jednoduché overenie si faktov a získanie prehľadu o problematike (Blikstad-Balas, 2016; Garrison 2015). V súlade s utilitárskym princípom, nad kvalitou informácií vyhráva ich jednoduchá dostupnosť (Lim, 2009; Blikstad-Balas, 2016).

Wikipédia je encyklopédiou, ktorá sa však nevyužíva striktne len popri štúdiu, ikeď to môže evokovať pomerne veľké percento výskumov zameraných práve na Wikipédiu v školskom prostredí. Wikipédia však dosiahla medzi ľuďmi takú mieru dôvery, že ju bežne nájdeme citovanú aj v súdnych spisoch (Gerken, 2010), vedeckých časopisoch (Nielsen, 2007) či v spravodajstve (Lih, 2004). Okrem štúdia a pracovných povinností sa ľudia obracajú na Wikipédiu aj v bežnom živote, a to predovšetkým, keď majú chuť sa ďalej neformálne vzdelávať (37 %), keď sa chcú dozvedieť viac o téme, ktorú zachytili v médiách (25 %) alebo počas konverzácie (24 %) (Lemmerich 2018). Všestrannosť Wikipédie a jej presah do každodenného života, predovšetkým ako voľnočasová zábavka, potvrdzuje aj Spoerri (2007), podľa ktorého najvyššie percento návštevnosti na Wikipédii dosahujú články z kategórie Zábava; až 43 %.

Okrem faktorov, ktoré motivujú používateľov siahnuť po Wikipédii, sa primerane tomu líši aj ich informačné správanie. Keď používatelia zavítajú na Wikipédiu, vo väčšine prípadov sú už s témou oboznámení (55 %) a hľadajú len doplňujúce informácie, 35 % opýtaných vyhľadáva na Wikipédii konkrétne fakty, 33 % chce získať všeobecný prehľad o téme a 32 % študuje tému do hĺbky (Lemmerich, 2018).

Širokému využívaniu Wikipédie, nielen v akademickom kontexte, nahrávajú aj ďalšie faktory. Podľa štúdie Connaway et al., (2011) najvýraznejšie ovplyvňuje využívanie informačných zdrojov pohodlnosť (*convenience*), ktorá je platným argumentom pri výbere informácií, tak v akademickom ako aj v bežnom živote, kedy sa prikláňame k zdrojom, ktoré nemusia byť najlepšie, ale ich kvalita je momentálne dostačujúca. Jednoduchá dostupnosť predstavuje ďalší zo smerodajných faktorov. *„Za najjednoduchšie dostupné zdroje informácií môžu byť považované internetové vyhľadávače, elektronické databázy, virtuálne služby knižnice, online elektronické repozitáre, e-knihy a online knihkupectvá; zistenia tiež ukazujú, že obzvlášť dôležitý je pre mladú generáciu vyhľadávač Google.“* (Connaway, et al., 2011), ktorý je zodpovedný aj za veľké percento návštevnosti Wikipédie. Posledným rozhodujúcim faktorom je časová náročnosť vyhľadávania informácií, ktorá predeterminuje výber informačného zdroja. S týmito zisteniami korešpondujú aj výskumy zamerané na informačné správanie tzv. Google generácie, teda práve generácie mladých používateľov, ktorí si pri hľadaní zdrojov a informácií žiadajú *„nepretržitý prístup 24/7, okamžité uspokojenie informačnej potreby jediným kliknutím a stále väčším príklonom skôr k odpovedi na otázku ako celému zdroju.“* (Rowlands, et al., 2008, s. 293).

2 Je Wikipédia mužská záležitost?

Wikipédia je podľa vlastných slov „*encyklopédia, ktorú môže editovať každý*“⁹. Napriek tomu sa v súčasnosti čoraz častejšie ozývajú hlasy, ktoré toto známe wikipedické motto do veľkej miery spochybňujú, a to práve v súvislosti s otázkou skutočnej otvorenosti wikipedickej komunity, diverzity a v neposlednom rade, genderu. Súčasný pohľad na Wikipédiu trefne zhrnula v názve svojej štúdie Heather Ford a Judy Wajcman; Wikipédia je, podľa autoriek, encyklopédia, ktorú „*môže editovať každý, ale nie každý ju aj skutočne edituje*“¹⁰ (2017).

V nasledujúcej časti sa fókus práce pomaly začne zužovať práve na vyššie spomínanú otázku genderu, ktorú sa budeme snažiť predstaviť a zodpovedať z viacerých uhlov pohľadu. Predmetom tejto kapitoly budú v prvom rade čitatelia a editori Wikipédie vo všeobecnosti, teda naprieč jej jazykovým verziami, do značnej miery oprostí od kultúrnych rozdielov. Z väčšej časti čerpáme z výskumov staršieho dáta, ktoré sú podporené novšími zisteniami. Tieto výsledky kopírujú trend v skúmaní Wikipédie a jej obsahu, ktorý smeruje od využívania wikipedického obsahu a čitateľov k tvorcom obsahu, editorskej praxi a komplexnejšiemu hodnoteniu kvality obsahu.

2.1 Kto číta Wikipédiu

Podľa aktuálneho výskumu americkej organizácie Pew Research Center (2018) používa internet celkovo 89 % mužov a 88 % žien, zdá sa teda, že vo vyspelých krajinách sa pomaly stierajú rozdiely v prístupe k internetu. Rozdiely sa však predsa len nájdu, a to aj v prostredí Wikipédie napríklad v konzumácii obsahu.

Výskum Sook Limovej a Nahyun Kwonovej (2010) pri skúmaní vzťahu študentov k tejto online platforme zistili, že muži konzumujú obsah na Wikipédii v priemere častejšie ako ženy (Lim a Kwon, 2010). Podobné čísla ukazujú aj výsledky prieskumu Pew Research Center (2011); 56 % amerických mužov a 50 % amerických žien využíva Wikipédiu k vyhľadávaniu informácií. I keď sa nejedná o signifikantný rozdiel, novšie výskumy zamerané na jednotlivé jazykové verzie Wikipédie ďalej potvrdzujú existenciu rozdielu medzi mužskými a ženskými čitateľmi v rámci jej jazykových verzií (Protonotarios et al., 2016, vid' Tabuľka 1). Tieto výsledky sú však tiež skôr orientačného charakteru, jedná sa o výskumy staršieho dáta, ktoré neboli ďalej opakované, prípadne sa značné rozdiely ukázali len na menších vzorkách, resp. v

⁹ https://en.wikipedia.org/wiki/Wikipedia:The_Free_Encyclopedia

¹⁰ Voľný preklad názvu štúdie 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap (Ford a Wajcman, 2017).

jednotlivých jazykových verziách Wikipédie. Výsledky však napriek tomu naznačujú určitý nepomer medzi čitateľmi a čitateľkami, ktorý sa ale neukazuje byť výrazne smerodajný a všeobecne platný.

Zaujímavejšie výsledky sa však týkajú využívania Wikipédie. Okrem nerovnováhy medzi čitateľmi a čitateľkami, poukazuje výskum Limovej a Kwonovej (2010) aj na rozdielne využívanie Wikipédie medzi mužmi a ženami, predovšetkým v rámci voľného času. Podľa autoriek (Lim a Kwon, 2010) si muži túto online encyklopédiu čítajú pre zábavu a vo voľných chvíľach častejšie ako ženy, čo môže byť signifikantnejší indikátor, keďže, ako uvádza Spoerri (2007), najvyššie percento návštevnosti na Wikipédií dosahujú články z kategórie Zábava, a to v priemere 43 %. Táto situácia môže mať viacero vysvetlení, často sa stretávame s pretrvávajúcim a pomerne stereotypným poukazovaním na rozdielne množstvo voľného času (Eckert a Steiner, 2013), no tento názor bol z veľkej časti prekonaný už v roku 2007 výskumom mapujúcim zmenu využívania voľného času medzi americkými mužmi a ženami počas 4 desaťročí (Aguiar a Hurst, 2007), i keď sa stále miestami vynorí ako pádny argument. Na jedno z pravdepodobnejších vysvetlení poukázal novinár The New York Times Noah Cohen (2011), ktorý ilustračne porovnal články na anglickej Wikipédii a pozrel sa na ne z pohľadu čitateľa, lepšie povedané, z pohľadu čitateľky. Články, u ktorých sa predpokladá, že budú viac zaujímať ženské publikum ako to mužské, boli oveľa menej prepracované, zväčša sa obmedzili len na niekoľko odsekov písaných strohým štýlom s malým počtom detailov, a to aj v prípade populárneho seriálu Sex v meste, slávnych návrhárov a dizajnérov, akými sú Manolo Blahnik či Jimmy Choo. Na porovnanie uvádza Cohen (2011) seriál Sopranovci, ktorého každá epizóda je dopodrobna rozpracovaná čo do dĺžky textov, ako aj do počtu uvedených detailov.

V súčasnosti tento nepomer do veľkej miery pretrváva, i keď sa transformoval do podoby vymazávania článkov o ženách, prípadne článkov, pri ktorých nad mužskými prevažuje množstvo ženských čitateľiek. Pre ilustráciu, boli medializované kauzy vymazania článku o vedkyňi Donne Strickland¹¹, ktorá podľa editorov Wikipédie nebola dostatočne známa ani významná v odbore, a to len pár týždňov pred získaním Nobelovej ceny za fyziku (Bazely, 2018). Podobné špekulácie o významnosti témy nastali aj v prípade svadobných šiat Kate Middleton¹², ktoré mali vlastnú stránku na anglickej Wikipédii a ktorá bola objektom značných špekulácií, no nakoniec zmazaná nebola (Bosch, 2012). Ako uviedol Jimmy Wales na každoročnej konferencii Wikimania, príklad šiat Kate Middleton pekne ilustruje genderovú

¹¹ <https://www.washingtonpost.com/outlook/2018/10/08/why-nobel-winner-donna-strickland-didnt-have-wikipedia-page/>

¹² <https://slate.com/technology/2012/07/kate-middleton-s-wedding-gown-and-wikipedia-s-gender-gap.html>

priepasť a možnú tematický bias v rámci Wikipédie (Bosch, 2012). Wikipedická stránka svadobných šiat Kate Middleton viacerých poburovala, avšak len málo ľudí si všimne, ako uviedol Wales, že „*máme viac ako 100 článkov popisujúcich rôzne verzie Linuxu, niektoré z nich dosť obskúrne... a ktoré nemajú doslova žiadny vplyv na mainstreamovú kultúru, ale to sa zdá byť úplne v poriadku*” (Bosch, 2012).

Tento fakt môže ďalej zohrať u žien úlohu aj pri celkovom vnímaní Wikipédie ako projektu a jej kvality. Podľa výskumu Limovej a Kwonovej (2010) pri vyhľadávaní informácií na Wikipédii, muži Wikipédiu hodnotia všeobecne kladnejšie ako dámy, a to vrátane dôvery v samotný projekt, všeobecnej kvality výsledkov, doterajších skúseností s kvalitou informácií, emocionálnych stavov pri využívaní alebo očakávaných dopadov informácií. Dámy naopak viackrát zmieňujú riziká spojené s jej využívaním, predovšetkým pri vyhľadávaní informácii (Lim a Kwon, 2010).

2.2 Kto edituje Wikipédiu

Rozdiely nepretrvávajú iba na strane konzumácie wikipedického obsahu, ale aj na strane editorskej. Rôzne štúdie uvádzajú, že z celkového počtu iba 9 - 11 % prispievateľov tvoria ženy (Glott a Ghosh, 2010; Wikimedia, 2011). Tieto zistenia ďalej potvrdzuje aj najnovší prieskum z roku 2018 realizovaný samotnou Wikipédiou, podľa ktorého je len 9 % celosvetových prispievateľov ženského rodu (Wikimedia, 2018). Slovenská a česká Wikipédia sú v tejto oblasti značne nezmapované, no minimálne pre slovenskú Wikipédiu uviedol približné percentá žien editoriek jej regionálny správca Michal Matúšov, podľa ktorého „*podiel žien je na úrovni 9 až 15 percent. Väčšina [editorov] sú mladí ľudia, prevažne muži 15 až 30 rokov*” (Tvardzík, 2015).

Väčší nepomer medzi editormi Wikipédie môžeme postrehnúť aj pri jej ďalších jazykových verziách. Je to do veľkej miery kultúrnou otázkou, no štúdie itak vychádzajú v neprospech žien – čitateľiek a žien - editoriek. Prehľadný sumár výsledkov prináša grécka štúdia, ktorej autormi sú Ioannis Protonotarios, Vasiliki Sarimpei a Jahna Otterbacher (2015), ktoré uvádzame v Tabuľke č.1 nižšie. Prvé štyri uvedené výskumy sa týkajú Wikipédie naprieč jazykovými verziami, posledné dva výskumy sa sústreďujú konkrétne len na francúzsku a grécku Wikipédiu. Protonotarios a kolektív (2016) vo svojej štúdii ďalej uvádzajú, že práve zistená *gender gap*, teda genderová priepasť medzi mužmi a ženami v kategórii čitateľa a editori, je jedným z mála zistení, ktoré sú konzistentné naprieč kultúrami a jazykmi Wikipédie.

	Čitatelia		Editori	
	Muži	Ženy	Muži	Ženy
UNU-MERIT Survey (Glott, Schmidt a Ghosh, 2010)	68,9 %	30,5 %	86,7 %	12,7 %
Wikimedia Foundation Survey (2011)	56,0 %	44,0 %	90,0 %	9,0 %
Wikipedia Gender Gap Revisited (Hill a Shaw, 2013)			84,0 %	16,0 %
Massa a Zelenkauskaitė (2014)			82,6 %	17,4 %
Bourdelloise a Vincente (2014) (francúzská Wikipédia)	61,5 %	38,5 %	81,4 %	18,6 %
Greek Wiki survey (2015) (grécka Wikipédia)	62,5 %	37,5 %	85,2 %	14,7 %

Tabuľka č. 1 Percentuálne rozdelenie používateľov Wikipédie podľa genderu (Protonotarios, et al., 2016, s. 81)

Zdá sa teda, že nielen čitateľská základňa je čiastočne nevyvážená, ale už samotnému prostrediu Wikipédie dominujú muži, a teda i mužský pohľad na jednotlivé témy, a to nielen na celosvetovej úrovni, ale aj v našich geografických šírkach. Pôvodom tohto fenoménu sa venujú napríklad Julia Bear a Benjamin Collier (2016), ktorí pri definícii 3 hlavných faktorov nadväzujú na ich predchádzajúcu štúdiu z roku 2012. Vysvetľujú, že nižšie percento žien editoriek je dané tromi hlavnými psychologickými procesmi, ktoré okrajovo súvisia s vyššou koncentráciou mužov - prispievateľov.

Špecificky zmieňujú: (1.) sebavedomie a dôveru vo vlastnú expertízu, čo koreluje so zisteniami z predošlej štúdie autorov (Bear a Collier, 2012), kedy sa ženy – editorky necítili komfortne pri editácii cudzích článkov, a zároveň nemali pocit, že majú dosť znalostí na to, aby na Wikipédii prispievali. Pre úplnosť, nedôvera vo vlastnú expertízu sa ukazuje aj pri počítačových zručnostiach, ktoré môžu mať vplyv na rozhodnutie neparticipovať na vytváraní

obsahu. Nepreukázalo sa, že by ženy boli menej kompetentné, čo sa týka práce s nástrojmi v online prostredí, viac ako zručnosti im chýba dôvera vo vlastné schopnosti a expertízu, na rozdiel od mužských participantov (Hargittai a Shafer, 2007); (2.) vysporiadanie sa s konfliktami, pričom až 31 % uviedlo strach z konfliktu a negatívnych emócií ako dôvod prestať na Wikipédiiu prispievať (Bear a Collier, 2016) a (3.) schopnosť vyrovnáť sa s negatívnou spätnou väzbou, pričom až 22 % editoriek sa už stretlo s negatívnymi správami od iných editorov prípadne s inou formou nepriateľskej komunikácie (Bear a Collier 2016).

Problém s negatívnou spätnou väzbou sa tak ďalej odráža v nižšej motivácii prispievať, ako ukázal najnovší experiment Shane-Simpsonovej a Gillespie-Lynchovej (2017). V prípade, že sa editorky v rámci celého procesu nestretnú so žiadnou ďalšou ženou – editorkou a ani s pozitívnou afirmáciou alebo feedbackom, ženy prispievajú menej (Shane-Simpson a Gillespie-Lynch 2017). Naopak mužom neutrálne prostredie bez pozitívneho feedbacku vyhovuje – v týchto podmienkach prispievajú podstatne viac.

Bear a Collier (2016) nakoniec upozorňujú na fakt, že aj keď je proces vylepšovania Wikipédie označovaný ako prispievanie, častejšie táto aktivita zahŕňa úpravy a mazanie práce iných editorov než vytváranie úplne nového obsahu, dennodennou praxou je teda súperenie s ostatnými (mužmi) o presadenie vlastného názoru. Pričom viaceré výskumy so špecifickým zameraním na prostredie Wikipédie dokazujú, že ženám viac vyhovuje priateľské prostredie, ktoré je naklonené vzájomnej kolaborácii ako konfrontácii (Bear a Collier 2012; Laniado et al., 2012; Shane-Simpson a Gillespie-Lynch 2017).

Aby sme pochopili, ako veľmi je tento nepomer naozaj spôsobený genderom, je potrebné zohľadniť, do akej miery sa téma týka všeobecného rozdielu prístupu mužov a žien k Wikipédii s ohľadom na rýdzo subjektívne hodnoty, akými sú napríklad inklinácie k typicky ženským a mužským témam a ich disproporcionálne obsadenie, čomu sme sa už čiastočne venovali v predchádzajúcej kapitole. Viaceré zistenia totiž naznačujú, že ženy a muži majú tendencie prispievať do odlišných tematických celkov (Laniado, et al., 2012; Hinnosaar, 2019). Podľa zistení Davida Laniada a kolegov (2012), ženy najčastejšie prispievajú do tém Umenie, Zdravie, Matematika a Počítačová veda (o 15 % viac ako muži), pričom v rámci uvedených tém sú interné diskusie o zverejnení článku citeľne pozitívnejšie ladené ako tie zvyšné, čo zodpovedá zisteniam zo súčasných výskumov. Najmenej zaujímavé sú pre ne podľa Laniada (2012) História a Geografia, kde naopak dominujú ich mužskí kolegovia. Podobné výsledky prináša aj štúdia Shyonga K. Lama a kolegov (2011), podľa ktorej ženy viac prispievajú do kategórie Umenie a Osobnosti, muži naopak preferujú Geografiu a Vedu.

Výsledky potvrdzuje aj novší výskum Paula Schrijvera (2016) zameraný na zastúpenie mužských a ženských editorov naprieč témami v anglickej Wikipédii. Čerpajúc dáta z viac ako 3,7 milióna článkov v 470 kategóriách uvádza na editorskom profile svoje pohlavie 176 tisíc editorov. Z týchto čísel je približne 85 % editorov mužského a 15 % ženského rodu, čo viac menej zodpovedá predchádzajúcim odhadom, vzhľadom na fakt, že množstvo editorov tieto údaje neuvádza, prípadne nie sú registrovaný. Zaujímavejšia je však distribúcia editácií naprieč témou a pohlavím - práca uvádza 5 kategórií s dominantne mužskými editáciami a 5 kategórií s prevažne ženskými editáciami. Do mužského zoznamu sa dostali kategórie „YearInSpaceflight” (rok v kozmickom lietaní), „Asteroid” (asteroid), „BaseballSeason” (basebalová sezóna), „MotorsportSeason” (pretekárska sezóna) a „FormulaOneTeam” (tím Formula 1). Ďalej autor uvádza, že práve šport je najčastejšou oblasťou, kde prevažujú muži, často sa tiež objavuje aj doprava a politika. Ženská prevaha sa naopak ukázala v menej očividných témach, resp. „FigureSkater” (krasokorčuliar/krasokorčuliarka), „Skater” (korčuliar/korčuliarka), „Garden” (záhrada), „GaelicGamesPlayer” (hráč/hráčka galských počítačových hier) a „Mollusca” (slimáky). Okrem uvedených tém však vo všeobecnosti ženské editácie prevažujú skôr v kultúrnych témach. Uvedených 5 príkladov dominancie naznačuje, ktoré témy sú pre danú komunitu aktuálne najzaujímavejšie.

Okrem exkurzu do najaktuálnejších tém, ktorými komunita žije, sú špecifické aj preferencie biografických článkov na Wikipédii. Až 31 % žien si na editáciu vyberie radšej článok o žene ako o mužovi, spravidla z oblasti kultúry (Hinnosaar, 2019). Význam tohto čísla je o to väčší, že z celkového počtu biografických článkov majú biografie o ženách zastúpenie len 15,2 % (Hinnosaar, 2019).

Pokiaľ by sme chceli tejto kapitole dať všeobecné zhrnutie, tak dochádzame k záveru, že vzhľadom k špecifickým preferenciám a nižšiemu zastúpeniu žensky atraktívnych tém na Wikipédii sa znižuje aj celkový počet žien, ktoré Wikipédiu využíva, a tým prirodzene klesá aj motivácia do nej aktívne prispievať. Preto je v porovnaní s počtom mužov, počet žien – prispievateľiek a v menšom merítku aj počet čitateľiek, nízky. A samozrejme to môže platiť aj opačne a príčinou nižšieho záujmu a zastúpenia žensky atraktívnych tém je nízky počet editoriek, a teda obrazne povedané, hovoríme o začarovanom kruhu.

3 Špecifiká free culture komunity

Ženy ako editorky v neposlednom rade čelia ešte jednej výzve, ktorú naznačujú predchádzajúce výskumy a tou je celková atmosféra, ktorá vládne v rámci *open source* a *open access* komunity, vrátane Wikipédie (Eckert a Steiner, 2013). V porovnaní s mužskými editormi, má Wikipédia stále málo nových žien editoriek; v roku 2011 to bolo na anglickej Wikipédii len 16 % z celkového počtu (Shyong K. Lam, et al., 2011) celosvetový prieskum ukazuje len cca 9 % (Wikimedia, 2018). Častejšie sú ženy ako editorky odmietnuté, teda neschválené, a zároveň, ak už sa im podarí do Wikipédie dostať, majú vyššiu šancu, že ich úpravy článkov nebudú schválené; všeobecne sa ženy podpísali len pod 9% editácii všetkých článkov (Shyong K. Lam, et al., 2011), takže genderová priepasť, okrem celkových počtov editorov, má značné presahy aj do počtu editácii. Koniec koncov ženy Wikipédiu opúšťajú skôr ako ich mužskí kolegovia (Shyong K. Lam, et al., 2011).

Viaceré výskumy potvrdili (Shyong K. Lam, et al., 2011; Bear a Collier, 2012; Shane-Simpson a Gillespie-Lynch, 2017; Laniado, et al., 2012), že ženám nevyhovuje konkurenčné prostredie editorov Wikipédie. Treba však zdôrazniť, že reakcie začínajúcich žien editoriek a mužov editorov na kritiku a neschválenie úprav sa nelíšia, teda neplatí, že sa ženy nechajú odradiť už pri prvotnom neúspechu, čo dokazuje výskum Shyong K. Lama a kolegov z roku 2011. Súčasné výskumy naznačujú skôr postupné upadávanie počiatočného odhodlania prispievať, a teda prakticky bojovať s veternými mlynmi. Podobný pocit prevláda aj u laickej verejnosti. Internetová verejnosť za najväčšiu prekážku (36 %) vyššieho počtu žien prispievateľiek pokladá práve editorské prostredie Wikipédie (Eckert a Steiner, 2013).

Zaujímavý argument do diskusie o nerovnosti v zastúpení pohlaví prináša výskum Davida Lanianda (2012), podľa ktorého „*by sme sa na problém nerovnosti v zastúpení pohlaví na Wikipédii nemali pozerať len ako na problém menšieho počtu editoriek. Tento stav naznačuje, že ženy – editorky, ktoré sa rozhodnú prispievať nepredstavujú reprezentatívnu vzorku, a teda čím viac skúseností ženy naberajú, tým viac majú tendenciu sa prispôbovať a prevziať (ne)emocionálny spôsob jednania ostatných editorov*“ (Laniand, et al., 2012, s. 5).

3.1 Vzostup geekov a internetovej kontrakultúry

K lepšiemu pochopeniu nastavenej kultúry prispievateľov nielen na Wikipédiu sa vrátíme trochu do histórie. Nové počítačové technológie sa začali už okolo roku 1970 spájať s kultúrou geekov, neskôr s myšlienkami osobnej slobody a celkovo kontrakultúry. Nástup internetu túto asociáciu len ďalej posilnil. Spoločne so vznikom WWW konzorcia a spoločnou

globálnou internetovou infraštruktúrou sa nahrnuli nadšenci ochotní sami a bez nároku na honorár vyvíjať webové aplikácie prístupné širokej verejnosti. Túto periódu vývoja do roku 2000 charakterizuje José van Dijk vo svojej knihe *Culture of Connectivity* (2013) ako *dot.communism*. Po splasnutí *dot.com* bubliny sa tieto myšlienky postupne vrátili s ešte väčšou silou spoločne s nástupom Webu 2.0 a jeho nových možností participácie, ktoré sľubovali skutočnú demokratizáciu online priestoru. Nástup veľkých firiem a korporácií však zmenili *dot.communism* na *dot.commercialism* (van Dijk, 2013) prakticky cez noc. Časom sa aj menšie nezávislé iniciatívy priklonili ku korporátnemu modelu fungovania a financovania, ktoré sľubovalo väčšiu udržateľnosť projektu a nové možnosti monetizácie. I keď sa stále hlásajú počiatočné heslá oslavujúce slobodu na internete, nekonečné možnosti kooperácie, participácie a liberalizujúci potenciál internetu, svojim vnútorným fungovaním dnes tieto heslá v zásade popierajú, platformy kupujú väčší hráči, prípadne sa z nich samých stávajú významní hráči na trhu, postupne sa uzatvárajú, menia a monetizujú. Odkláňajú sa od počiatočných myšlienok geek kultúry a kontrakultúry na internete, ako uvádza José van Dijk, „už len zbežný pohľad na 100 najpopulárnejších sociálnych médií potvrdzuje fakt, že 98 % z nich patria pod veľké korporácie, ktoré sa na internet pozerajú v prvom rade ako na trh a nie ako na verejné fórum“ (2013, s.16). Jedinou významnou výnimkou je podľa van Dijk (2013) práve Wikipédia, ktorá si dokázala niečo z tohto odkazu a spôsobu fungovania ponechať.

3.2 Nastavenie kultúry participácie

Wikipédia a jej podobné projekty s otvorenou participáciou, *open source* a *open access*, ktoré toto obdobie ustáli bez výrazných transformatívnych zmien, môžeme označiť širšou definíciou *free culture*, ako ju v rovnomennej knihe z roku 2004 zdefinoval Lawrence Lessig. Atmosféra fungovania, ktorá sa v rámci tejto komunity zachovala, si stále nesie počiatočné rysy geekovstva a kontrakultúry spojené s výrazným akcentom na slobodu voľby a slobodu jednotlivca. Že aj v rámci týchto komunit, ktoré stavajú na liberálnych hodnotách slobody a otvorenosti, prevláda do značnej miery genderová nevyváženosť prispievateľov, resp. participantov, potvrdili viaceré výskumy (Lin, 2005; Nafus, 2006; Reagle, 2013; GitHub, 2017).

Joseph Reagle je jedným z popredných vedcov, ktorý sa v rámci svojej akademickej dráhy zameriava prevažne na kultúru *open source*, *open access* a FLOSS komunity. Reagle (2013) v rámci svojho 6-ročného výskumu fungovania *free culture* komunit zistil 3 najproblematickejšie znaky fungovania týchto komunit, ktoré, ako sám tvrdí, prispievajú k genderovej nevyváženosti zastúpenia pohlaví a sú aspoň z časti výsledkom kultúry, dynamiky

a hodnôt, ktoré vyznávajú. V prvom rade (1.) free culture môže byť pre niektorých ľudí nepríťažlivá, hlavne v súvislosti so stereotypným imageom online geeka, napríklad aj spájanie vlastnej identity s nekritickým a hlbokým záujmom o danú problematiku a s tým spojené agresívne obhajovanie a vystupovanie; (2.) otvorenosť prostredia, resp. samoregulácia, umožňuje niektorým minoritným skupinám komplikovaných členov disproporcionálne meniť dynamiku a tón komunikácie, a nakoniec (3.) ideológia a rétorika spojená s otvorenosťou a slobodou môže koniec koncov spôsobiť, že sa niektoré aj pretrvávajúce sťažnosti na fungovanie zametú pod koberec s odkazom na slobodné fungovanie, hrozbu cenzúry a celkovo slobodnú voľbu jednotlivca sa zapájať či nezapájať (Reagle, 2013). Posledný bod je často aplikovaný špecificky na problematiku genderu, resp. žien. Racionalizujú sa tak ich nízke počty v rámci týchto komunit.

Podobné výsledky prinášajú aj štúdie realizované na participantoch FLOSS komunity¹³, teda komunity, ktorá spája iniciatívy zamerané na vývoj *open source* a *open access* aplikácií a softvérových riešení (Firefox, Linux a podobne). FLOSS komunity majú síce explicitne technickejší charakter ako Wikipédia, avšak svojimi stanovmi sú si veľmi podobné, postavené na rovnakých princípoch meritokracie, egalitarizmu a slobody. Podobne ako Wikipédia, aj tieto komunity trpia nízkym zastúpením žien vo svojich radoch; najnovší prieskum z oblasti *open source* uvádza len 3 % ženských prispievateľiek (GitHub, 2017), pričom podobne, ako aj ženy na Wikipédii sa častejšie stretávajú s nepriateľskými vyjadreniami či obsahom (25 % vs. 15 %), častejšie sú obeťou stereotypizácie (12 % vs. 2 %) či nežiadúcich sexuálnych návrhov (6 % vs. 3 %) (Bear a Collier, 2012). Rovnako je do určitej miery prekvapivá vysoká a striktná úroveň hierarchizácie v tomto prostredí, tak v prípade FLOSS komunity (Leach, et al., 2006) ako aj Wikipédie (van Dijk, 2013, Ford a Wajcman, 2017), pričom v oboch prípadoch boli zistené značné komunikačné nedostatky, predovšetkým agresivita a ponižovanie členov (GitHub, 2017; Leach, et al., 2006), kritika a všeobecne nepriateľské vystupovanie (Bear a Collier 2012). A to napriek prevládajúcemu názoru, že väčší počet žien vo FLOSS komunite by bol vítaný a vnímaný pozitívne u 66 % mužských a 98 % ženských súčasných členov (Leach, et al., 2006).

Reagle (2013) hovorí o nedostatkoch FLOSS a wikipedickej komunity v širšom kontexte, naopak Fordová a Wajcmanová (2017) sa pozerajú na štruktúru a fungovanie komunity Wikipédie špecificky z genderového a feministického hľadiska. Fordová a Wajcmanová chápu Wikipédiu ako jeden z ďalších príkladov STS, resp. socio-technologického systému, ktorý je rovnakou mierou definovaný technologickými nástrojmi,

¹³ FLOSS je skratkou pre združenie Free/Libre and Open Source Software.

manažérskou hierarchiou a samotnými pracovníkmi. Wikipédia sa navonok stále prezentuje ako viac menej projekt nadšencov amatérov, no pod povrchom sa vytvára hierarchická organizácia disponujúca novým typom expertízy, a teda sily, ktorá je, vzhľadom na nerovnomerné zastúpenie žien, konštituovaná ako mužská, a v princípe replikuje staré systémy moci. Infraštruktúra fungovania Wikipédie, jej platforma (ktorá je často terčom kritiky kvôli sťaženej orientácii a práci s ňou) a celkovo predpokladaný špecializovaný typ poznatkov, dopredu limitujú kto sa môže stať wikipedistom a koniec koncov zvyhodňujú tých, ktorí už sú na Wikipédii nejaký čas aktívni a zavedení. Takýto členovia, ktorí disponujú tak znalosťou platformy ako aj určitým typom znalostí, sa stávajú tzv. *power brokers*. Na druhej strane tí, ktorým sa z nejakého dôvodu, a tým majú autorky na mysli predovšetkým disponovanie iným typom znalostí, ako je norma, prípadne kvôli negatívnym sociálnym interakciám v rámci Wikipédie, nepodariť osvojiť si tento systém, zostanú na jeho okraji, bez možnosti prispievať a s veľkou pravdepodobnosťou aj bez toho, aby bol ich špecifický typ znalostí na Wikipédii zastúpený. Celý text Fordová a Wajcmanová (2017) koncipovali s ohľadom na ženy, teda, keď hovoríme o „nich“, sú tým myslené predovšetkým ženy.

3.3 Chýbajúca inklúzia

V súčasnosti sa o týchto problémoch v rámci komunit vie, vedú sa o nich diskusie a dost sa tlačí na princíp otvorenosti a inklúzie, ktoré podľa Kata Walshy (Ross, 2011), dlhoročného prispievateľa a člena správnej rady Wikimedia Foundation, prirodzene prinesie do prostredia viac žien. Výsledky týchto snažení sú však v zásade neuspokojivé.¹⁴ Počet žien na Wikipédii a v princípe aj v ostatných podobne ladených komunitách, je stále malý a ani sa v priebehu času zásadne nezvyšuje (Shyong K. Lam, 2011) napriek aktivitám a iniciatívam. Wikipédia si v roku 2011 dala za cieľ do roka 2015 zvýšiť celkový počet editoriek na 25 % (Cohen, 2011). V roku 2018 však štúdia Wikimedia uvádza približne 9 % žien editoriek z celkového počtu editorov. Podobná situácia panuje aj vo FLOSS komunite. Po zverejnení extenzívnej štúdie FLOSSPOL (Leach, et al., 2006), ktorá sa venuje genderovej problematike a participácii v rámci FLOSS projektov, sa štúdia takisto snažila upozorniť na problém genderovej nevyváženosti editorov,

¹⁴ V rámci Wikipédie sú známe hlavne dve iniciatívy a to: editačné maratóny, nazývané aj edit-a-thon a iniciatíva Women in Red. Editačné maratóny sú organizované nezávisle po celom svete s cieľom jednorazovo prispieť čo najviac do vybranej témy, zväčša v súvislosti s chýbajúcou genderovou diverzitou. Podobne zameraná je aj iniciatíva Women in Red, ktorá sa sústreďuje na zvýšenie počtu článkov o významných ženách, ktoré na Wikipédii chýbajú. V rámci FLOSS komunity sú známe iniciatívy podporujúce aktivitu žien v rámci open source a open access projektov, obzvlášť známe sú Arch Women, Women in Drupal, LinuxChix, The Debian Women Project, Fedora Women, GnomeWomen, KDE Women, PHPWomen, Ubuntu Women a Women & Mozilla (WoMoz).

avšak posledná revízná štúdia z roku 2016 (Robles, et al.) len potvrdila stálu platnosť stavu z roku 2006, len s malými skôr nevýznamnými zmenami. V tejto súvislosti označujú Robles spolu s kolegami (2016) uplynulú dekádu dokonca za „stratenú dekádu“ v súvislosti s riešením problematiky genderu a inklúzie.

Podobné výsledky genderovej nevyváženosti, ikeď z trochu iného súdka, hlásia aj najnovšie výskumy komunity GitHub, len 3 % žien (2017) a Stack Overflow, len 6,3 % žien (Stack Overflow 2018 survey: age, gender, sexuality, 2018). Aj v rámci týchto komunit sa podobne ako v rámci Wikipédie participantí často explicitne nehlásili k svojmu genderu, počet používateľov, ktorí sa prezentovali ako ženy bol značne nevyvážený oproti otvorene mužským používateľom, pričom v prípade Stack Overflow majú mužskí prispievatelia aj viac reputačných bodov a viac sa zapájajú do komunitných aktivít (Vasilescu et al., 2014). Podľa Vasilescu (2014) sa ako výsledok formuje „*relatívne nezdravá komunita, v rámci ktorej ženy strácajú o ňu záujem skôr ako muži, ikeď úrovne ich aktivity sú si podobné*“ (Vasilescu, et al., 2014, s. 19).

Keďže zdanlivo vo *free culture* komunite neprebehla výraznejšia zmena, myšlienka aktívneho, ajkeď možno nevedomého vynechávania žien prostredníctvom nastavených sociálnych procesov Jamesa Leacha a kolegov (2006) je stále aktuálna. Ženy nie sú pasívne ani neprejavujú nezáujem, ich vylúčenie je maskované za slobodné rozhodnutie jednotlivca a v princípe popiera akúkoľvek existenciu problému genderovej nerovnosti (Leach, et al., 2006). Podobný názor prevláda aj u časti laickej internetovej verejnosti, ktorá hádže vinu na ženy, pretože sa podľa nich nechcú zapájať, keďže ide o ich vlastné rozhodnutie a majú iné, viac dievčenské záujmy (Eckert a Steiner, 2013).

Súčasťou problému môže byť aj skreslená, resp. naivná predstava o otvorenosti a meritokracii v rámci komunity samotnej. V eseji, ktorou Joseph Reagle (2017) nadväzuje na predošlý výskum FLOSS komunity, rozoberá mýtické rovnostárstvo, resp. egalitarizmus a meritokraciu, ktoré sú jedným zo základných princípov nielen v *open source* a *open access* komunite, ale všeobecne v technologickom svete. Joseph Reagle (2017) uvádza, že čo sa týka meritokracie, vo všeobecnosti poznáme dva druhy naivity, ktoré sa miestami môžu prekrývať, a to (1.) naivita, ktorá si nepripúšťa chyby (*ignorant naiveté*) a (2.) subjektívna naivita (*subjective naiveté*), ktorá si je vedomá kritiky meritokracie, ale skúsenosti ju presvedčili, že napriek nedostatkom je to systém, ktorý funguje dostatočne dobre a nie je potrebné ho meniť. Reagle (2017) zdôrazňuje, že nie je možné, aby technologické firmy a IT prostredie stelesňovalo ideál otvorenosti prístupu a šancí do takej miery, ako sa o ňom traduje, predsa len

je načase zaujať po technooptimizme realistickejší pohľad na vec a pripustiť si, že meritokracia, rovnosť prístupu a šanci je ideál, ku ktorému sa chceme priblížiť a nie aktuálny stav. Neochota pripustiť si tento stav znemožňuje snahy sa k nemu bližšie priblížiť.

Postoj členov a zodpovedných predstaviteľov uvedených komunit, ako stručne sumarizuje Reagle, prirovnávajú MacAuleyová a Visserová (2016) k postojom, ktoré zaujali predstavitelia akadémie voči nedostatku rasovej diverzity, ktoré sú detailne popísané v knihe Sara Ahmedovej *On Being Included* (2012), ktorá čerpá z extenzívneho prieskumu situácie vo vysokom školstve v Austrálii a vo Veľkej Británii.

On Being Included popisuje problém diverzity zamestnancov univerzít ako *inštitucionálne divadlo*, v rámci ktorého sa univerzity prezentujú, že „robia niečo“, pričom adekvátne neadresujú, v niektorých prípadoch ani explicitne nepomenujú, problém (MacAuley a Visser, 2016). Autorky veria, že Wikipédia a iné projekty s otvorenou participáciou, trpia podobnými problémami ako univerzity, keďže sa rovnako snažia o progresívnosť a v určitom zmysle aj o sociálnu zmenu, aj v prípade, že sa im nedarí zmeniť zaužívané status quo.

V kultúre, ktorá sa sama zdráha primerane adresovať svoje nedostatky a prispôbiť sa, i keď navonok môže takéto signály vysielat', je veľmi ťažké dosiahnuť požadovanej zmeny, resp. väčšej diverzity. Za všetko hovorí posledný výskum Terrella a kolegov (2017); napriek snahám a iniciatívam, ešte aj v roku 2017 majú ženy väčšiu šancu sa dostať do *open source* komunity v prípade, že sa neprezentujú ako ženy a ani ich tak nie je možné identifikovať inými spôsobmi. Skrytý sexizmus, stereotypy a dvojité meter prevládajú aj neskôr v procese spolupráce FLOSS komunity a sú súčasťou tzv. *brogramming kultúry*¹⁵ (Lin a Besten, 2018). Tieto praktiky odrážajú situáciu, ktorá panuje aj v iných (pracovných) oblastiach, kde ženy musia čeliť genderovej nerovnosti, je od nich vyžadovaná väčšia aktivita a vyšší štandard, čelia neobjektívnemu výberu a musia sa prispôbiť, aby prežili (Terrell, et al., 2017).

Okrem iného sa, obzvlášť v prípade Wikipédie, sa „*prispievatelia musia naučiť ovládať softvérové nástroje ale zároveň si musia osvojiť aj zodpovedajúce presvedčenie, hodnoty, všeobecné ustanovenia a prevládajúce praktiky*“ (Dijk, 2013, s. 718), čo je práve jedna z vecí, ktoré ženám vyhovovať nemusí a vzhľadom na súčasné výskumy ani nevyhovuje. Ako tvrdí Reagle (2013), Wikipédia je asociovaná s geek komunitou, ktorá často v rámci princípu

¹⁵ slangový výraz pre stereotypne vnímaného programátora - muža a maskulínnu kultúru okolo počítačového programovania, často vnímaný ako pejoratívne označenie kultúry, ktorá nerešpektuje ženy v IT odvetví. Zdroj: <https://en.wikipedia.org/wiki/Brogrammer>

otvorenosti prenecháva prílišnú slobodu konfliktnejším minoritám a neprikladá dostatočný dôraz sťažnostiam argumentujúc osobnými preferenciami a slobodným výberom. Situácii takisto nenapomáha ani fakt, že sa zároveň ženy musia vyrovnat' aj inými stereotypnými názormi na ich online aktivitu (Eckert a Steiner, 2013), akým je napríklad názor, že ženy a muži majú iné biologické predispozície, a preto neprispievajú (29 % opýtaných), prípadne majú málo voľného času a iné priority, naznačujúc rodinné povinnosti (27 % percent opýtaných). Preto sa môžeme opäť vrátiť k predpokladu Davida Lanianda a kolegov (2012), že ani to percento žien, ktoré sú v súčasnosti na Wikipédii aktívne, nemusí predstavovať reprezentatívnu vzorku. Namiesto boja s veternými mlynmi sa rozhodli do istej miery prispôbiť nastavenej kultúre a jej očakávaniam. Prípadne ak nedôjde k výraznejšej zmene nastavenia, väčšina z tých, ktoré budú čeliť zabehnutej infraštruktúre wikipedického STS zostanú, podľa Fordovej a Wajcmanovej (2017), na jej okraji.

Jedným z riešení tejto situácie je podľa Josepha Reagla (2013) kontinuálne posúvanie hraníc toho, čo by sme mohli nazvať geekovstvom, aby táto identita asociovaná s prostredím *free culture* bola prístupnejšia, a zároveň je potrebné prestať sa schovávať za otvorenosť, meritokraciu a slobodu jednotlivca, naopak sa tomu postaviť čelom a adresovať problematické aspekty systému. Koniec koncov obzvlášť Wikipédia a jej podobné projekty majú potenciál vyťažiť čo najviac z tzv. kolektívnej inteligencie¹⁶, ktorá sa, zhodou okolností, podľa najnovších výskumov neviaže len na výšku individuálnej inteligencie jednotlivých členov, naopak súvisí so senzitivnosťou, empatiou jednotlivých členov, rovnomerným striedaním sa pri vyslovení vlastného názoru počas diskusie a v neposlednom rade, koreluje aj s počtom žien v skupine (Woolley, 2010).

3.4 Aké je prostredie českej a slovenskej Wikipédie a FLOSS komunity

V našich zemepisných šírkach sú výskumy zaoberajúce sa wiki komunitou, či všeobecnejšie FLOSS komunitou takmer neexistujúce, preto je ťažké hovoriť o česko-slovenskom prostredí. Ako jeden z mála akademikov sa k téme FLOSS komunity na Slovensku vyjadruje Martina Kuchynářová (2010) v staršom texte *Prečo ľudia na Slovensku robia FLOSS*. Na Slovensku pôsobia viaceré organizácie venujúce sa podpore FLOSS projektov. Medzi najznámejšie patrí Slovak Open Source Initiative, skratka SKOSI, ktorá sa zameriava na projekty venované verejnej správe, školstvu a výskumu. Ďalšou známejšou skupinou je Slovenské združenie používateľov Linuxu (skratka SKLUG). A poslednou významnejšou

¹⁶ Ako ju zafinoval Pierre Lévi vo svojej rovnomennej knihe z roku 1997.

organizáciou je Občianske združenie Platon. Ako uvádza Kuchynářová (2010), slovenskú FLOSS komunitu môžeme rozdeliť zhruba do troch skupín: (1.) jednotlivci a malé skupinky pracujúce na menších projektoch; (2.) komunity pracujúce na jazykových verziách väčších medzinárodných projektov; (3.) a nakoniec komunity zapájajúce sa priamo do veľkých medzinárodných projektov (Kuchynářová, 2010). Autorka takisto nahliadla aj do motivácie jednotlivých skupín komunit. „*Motiváciou prvej skupiny býva najčastejšie potreba danej aplikácie, alebo aj nadšenie pre vytvorenie nástroja na uľahčenie si každodennej práce s počítačom*” (Kuchynářová, 2010, s. 2). V prípade druhej skupiny FLOSS komunity býva podľa Kuchynářovej (2010) motivácia pre zapojenie sa a obetovanie času takémuto projektu rôzna. Jednou z najčastejších býva snaha ovplyvniť a pomôcť svojim kúskom kódu a zároveň aj zdokonaliť si svoje programátorské schopnosti. Ďalším z dôvodov pre prácu na FOSS projekte môže byť podľa Kuchynářovej (2010) aj nekompromisná politika softvérových gigantov, ktorí sa snažia zbaviť nežiadúcej konkurencie nevýhodnými licenčnými zmluvami - FLOSS komunita vyvíja produkty, ktoré im s prehľadom konkurujú. Posledným motivačným faktorom je aj snaha byť vo víťaznom tíme vývojárov Open Source na rôznych súťažiach. Motivácia poslednej skupiny participantov na FLOSS produktoch vo veľkej miere kopíruje motiváciu druhej skupiny a autorka ju preto v texte ďalej nerozvíja.

Text Mariany Kuchynářovej je jediný dohľadateľným zdrojom, ktorý sa v súčasnosti venuje FLOSS komunitu na Slovensku, resp. ktorý je špecificky zameraný na participantov a na komunitu. V českom prostredí sme nenarazili na podobný typ textu v prípade FLOSS komunity ani v prípade Wikipédie. Naopak o slovenskej Wikipédii vieme aspoň približné počty žien editoriek, z ktorého si vieme vyvodiť približný obrázok situácie v komunite. Ako sme už uviedli v kapitole 2 *Je Wikipédia mužská záležitosť?*, Michal Matúšov v rozhovore pre časopis Trend z roku 2015 uviedol, že „*podiel žien je na úrovni 9 až 15 percent. Väčšina [prispievateľov] sú mladí ľudia, prevažne muži 15 až 30 rokov*” (Tvardzík, 2015). Inak absentujú ďalšie údaje a analytické texty týkajúce sa či už českej alebo slovenskej wikipedickej komunity. Preto sa môžeme len domnievať, že vďaka globalizácii a globálnej spolupráce na projektoch nielen v rámci FLOSS ale aj Wikipédie, v týchto komunitách všeobecne panuje podobná atmosféra ako inde vo svete a že čelí podobným problémom.

4 Ako sa hovorí o ženách

V nadväznosti na predchádzajúce kapitoly nie je prekvapivý záver výskumu Cohenovej (2011) a Reagle a Rhueho (2011), že predovšetkým kvôli rodovej priepasti v prostredí Wikipédie sa na tejto platforme častejšie objavujú témy, ktoré primárne zaujímajú mužské publikum. Wikipédia je v posledných rokoch v hľadáči výskumníkov nielen pre disproporčné obsadenie editorských pozícií mužmi a ženami, no čoraz viac sa výskumníci zameriavajú na dopady, aké takáto situácia môže mať na kvalitu textov, obzvlášť čo sa týka predpojatého a stereotypného zobrazovania žien (napr. Wagner, 2015; Wagner, et al., 2016; Zagovora, et al., 2017; Otterbacher, et al., 2019), prípadne iných komunit a menšín (napr. Raval, 2014). Podobné výskumy sa realizujú aj na ďalších obdobných platformách ako IMDb (napr. Otterbacher, 2011), ktoré v sebe inkorporujú *user generated content*¹⁷, a takisto sú terčom výskumu aj profesionálne výstupy, akými sú napríklad online spravodajstvo a online magazíny (napr. Yun, et al., 2007; GMMP, 2015; D'Heer, et al., 2019).

4.1 Ako sú prezentované ženy na Wikipédii

Viacero výskumov sa momentálne zaoberá predovšetkým spôsobom písania textov či použitej lexiky. Konkrétne wikipedickému obsahu sa v súčasnosti najintenzívnejšie venuje nemecká výskumníčka Claudia Wagner. Wagnerovej štúdia *It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia* (Wagner, et al, 2015) sa na otázku rodovej rovnosti na Wikipédii snaží odpovedať za pomoci kvantitatívnej analýzy sledujúcej hlavné 3 faktory: (1.) dôraz na pokrytie žien a mužov v článkoch, (2.) dôraz na štruktúru a prepojenie článkov, (3.) dôraz na použitú lexiku a na viditeľnosť, teda porovnanie, koľko článkov o ženách a koľko o mužoch sa dostalo na hlavnú stránku anglickej verzie Wikipédie. V otázke pokrytia sa ukazuje nepomer medzi mužmi a ženami, pričom článkov o ženách je výrazne menej ako tých o mužoch, čo rovnako platí aj na priemernú dĺžku článku, resp. medián, ktorý je v mužskom datase o niečo vyšší, ikeď nie signifikantne. Čo sa týka štruktúry, oveľa častejšie bývajú články o ženách prepojené na tie mužské, než je tomu naopak (napr. keď sa článok týka ženy, editori odkazujú na článok o jej manželovi, naopak to však nefunguje). Líši sa aj použitie slovnej zásoby, anglické články o ženách často obsahujú označenie „woman“ (žena), „female“ (ženský) alebo „lady“, zatiaľ čo výrazy pre mužov „male“ (mužský), „man“ (muž) a „gentleman“ nie sú zďaleka tak časté. U žien sa tiež častejšie v textoch objavujú slová, ktoré sa

¹⁷ Obsahy tvorené samotnými používateľmi.

viažu k rodinnému stavu, ako napríklad vydatá, rozvedená, deti alebo rodina. Ako riešenie teda Wagnerová a kolegovia (2015) navrhujú využívať viac rodovo vyváženú slovnú zásobu.

Podobnú metódu ako Wagnerová zvolila aj výskumná trojica Graells-Garrido, Lalmas a Menczer pre štúdiu *First Woman, Second Sex: Gender Bias on Wikipedia* (2015). Takisto skúmali v prvom rade metadáta, resp. počet článkov, priemerný počet slov na článok, ako druhý faktor si vzali lexiku a frekvenciu výskytu slov v mužskom a ženskom datasete, a nakoniec sa zamerali na prelinkovanie medzi mužskými a ženskými biografiami. Výskum bol realizovaný na anglickej verzii Wikipédie, pracovali priamo s plnými textami článkov a zároveň aj osobitne s anotáciami. Výsledky výskumu Eduarda Graells-Garridu a kolegov ukázali, že na anglickej Wikipédii bolo v roku 2015 z celkového počtu biografických článkov len niečo okolo 15,5 % o ženách, pričom identifikovali aj menší priemerný počet slov na článok, čím potvrdzujú zistenia z predchádzajúcej štúdie C. Wagnerovej a kolektívu (2015). V rámci lexiky sa v ženskom datasete najčastejšie vyskytovali nasledujúce slová: „actress“ (herečka, 15,9 %), „women’s“ (ženin, 8,8 %), female (ženský, 5,6 %), „her husband“ (jej manžel, 4,1 %), „women“ (ženy, 5,3 %), „first woman“ (prvá žena, 1,9 %) and „film actress“ (filmová herečka, 1,6 %) (Graells-Garrido, et al., 2015). Celkovo sa ženy najčastejšie dávali do kontextu s umením, s úspechom, ktorý bol často uvádzaný formulkou „first woman“ a s genderom a rodinnými väzbami. Na druhej strane muži, sú najčastejšie spájaní so slovami „played“ (hral, 14,2 %), „footballer who“ (futbalista, ktorý, 3,0 %), „football“ (4,5 %), „league“ (liga, 5,9 %), „John“ (7,9 %), „major league“ (hlavná liga, 1,8 %), a „football league“ (futbalová liga, 1,6 %), takže od pohľadu je v mužskom datasete dominantný šport (Graells-Garrido, et al., 2015). A nakoniec, čo sa týka prelinkovania medzi článkami, tak dáta ukázali častejšie linkovanie na mužské biografie ako na ženské (Graells-Garrido, et al., 2015).

Na pretrvávajúce rozdiely v prezentácii žien a mužov na Wikipédii ďalej upozorňuje Wagnerová spoločne s kolegami aj vo svojej štúdii z roku 2016, kde tvrdí, že rozdiel medzi zastúpením žien a mužov na Wikipédii je vyšší u tzv. „lokálnych hrdinov“ než u „superstars“, konkrétne teda nižšie zastúpenie žien ako mužov v prípade lokálnych hrdinov. Vysvetľujú si to aj tým, že čím sú daní ľudia menej známi, tým viac je ich zaradenie do Wikipédie subjektívne a špekulatívne. Zároveň tým potvrdzujú existenciu skleneného stropu, teda hoc neviditeľnú ale reálnu bariéru. Skutočne je teda ťažšie pre ženy než pre mužov dostať sa do Wikipédie, pokiaľ nie je žena natoľko slávna, že jej nezaradenie by bolo nezmyselné (Wagner, et al., 2016). Toto tvrdenie je v súlade so starším výskumom Reagla a Rhucho (2011), podľa ktorých majú ženy v porovnaní s mužmi 2,6x väčšiu šancu, že sa na Wikipédiu ich biografia nedostane,

v porovnaní s klasickou autoritatívnou encyklopédiou - Encyclopædia Britannica, ktorú zostavujú profesionáli.

Okrem už naznačených rozdielov v opise žien a mužov, štúdie ukazujú aj ďalšie rozdiely v používaní lexiky. Z historického hľadiska je z veľkej časti pochopiteľné, že v rámci encyklopédií budú aspoň čiastočne dominovať muži, vzhľadom na odlišné príležitosti, ktorým sa historicky dostalo mužom a ženám. Avšak spôsob, akým sa o ženách v rámci týchto článkov píše, by v súčasnosti nemal byť odlišný. Opak je však pravdou. Wagner (2016) ďalej pri porovnávaní datasetu biografii osobností pôsobiacich do konca 18. storočia a od začiatku 19. storočia po súčasnosť ukazuje rozdiely nielen v celkovej počte biografii žien a mužov, ale i v spôsobe písania. V oboch dátových setoch, pred 19. aj po 19. st. v prípade žien boli najčastejšie používané slová evokujúce feminitu, konkrétne „her husband“, (jej manžel), „woman’s“ (ženský) v zmysle patriaci žene a „actress“ (herečka). V prípade novšieho setu sa najčastejšie vyskytovali slová „actress“ (herečka), „woman’s“ (ženský) v zmysle patriaci žene a vzťahové prídavné meno „female“ („ženský“). Muži boli naopak v oboch datasetoch spájaní s vojenskými ťaženiami, politikou („served“, „elected“, „politician“) a neskôr so športom („played“, „league“ a „football“). Wagnerovej štúdia (2016) ukazuje, že i v dnešnej dobe je v prípade žien stereotypne vyzdvihovaná ich feminita, sú stále spájané s tradičnými rolami, a teda kategóriami ako rodina, gender a vzťahy. Naďalej sú vykresľované pasívne, na rozdiel od mužov, ktorí sú asociovaní s činnosťami (slovesá), a teda dostojia stereotypnému zobrazovaniu mužov ako aktívnych a pribojných. Naznačuje to vnímanie muža ako subjektu a ženy ako objektu.

Okrem nemeckej verzie Wikipédie je možné dohľadať aj výskumy pracujúce s jej gréckou verziou, ktoré sa podobne ako prvotné štúdie Claudie Wagnerovej sústreďujú na biografie a špecificky na použitú lexiku. Jahna Otterbacher sa spoločne s kolegami vo výskume z roku 2019 zamerala na tzv. *linguistic bias*¹⁸, teda na jazykový bias v textoch biografii gréckej a anglickej verzie Wikipédie. Aj jej výskum potvrdzuje existenciu výrazných rozdielov v použitej lexike, a to fakt, že ženy sú v textoch vo všeobecnosti viac spájané s rodinou a ich sociálnym životom, resp. vzťahmi a naopak muži sú častejšie opisovaní s využitím adjektív

¹⁸ Lingvistický bias je charakterizovaný ako systematická (nie nevyhnutne uvedomelá) nevyváženosť výberu slov a pomenovaní, ktoré slúžia ako prostriedok priradenia objektu záujmu do určitej sociálnej kategórie. Použité výrazy a slovné spojenia vychádzajú z konkrétnej predstavy o objekte, reprezentujú tieto predstavy a slúžia ako médiá ich ďalšieho prenosu. Často sú tieto jazykové prostriedky nositeľom stereotypov, predsudkov či iných foriem znevýhodnenia (napr. podstatné mená ženy vs. ženské) (Beukeboom, 2017).

a pozitívnych slov, a teda vykresľovaní menej objektívne s príklonom k pozitívnemu zobrazovaniu. Táto tendencia je viditeľnejšia v gréckej jazykovej verzii ako anglickej.

Ďalšou autorkou, ktorá sa aktívne venuje lingvistickým otázkam je okrem Claudii Wagner aj Olga Zagovora. Stereotypnému zobrazovaniu žien a mužov na Wikipédii, s explicitným zameraním na profesionálnu oblasť, sa venovala vo výskume z roku 2017. Výskum z roku 2017 (Zagovora, et al., 2017) bol realizovaný na nemeckej verzii Wikipédie a ukazuje disproporcie medzi mužským a ženským zastúpením v prezentácii jednotlivých profesií. Špecifikom nemeckého jazyka, na rozdiel od angličtiny, ktorá všeobecne dominuje aj ako predmet výskumu, má viac možností ako explicitne vyjadriť rod (napr. príponou -in) napríklad v názvoch profesií; doktor – doktorka (Arzt – Ärztin), čím je nemčina bližšia českému a slovenskému jazyku. Vo všeobecnosti na nemeckej Wikipédii nájdeme názvy profesií uvedené pod mužskými predstaviteľmi jednotlivých profesií; 885 článkov venovaných jednotlivým profesiám uvádza vykonávateľov profesie v mužskom rode, v ženskom rode len 25 (Zagovora, et al., 2017). Neutrálne názvy profesií sa vyskytovali len v 36 prípadoch. Takisto tieto stránky v zásade neodkazujú na ženský ekvivalent uvedenej profesie (len 5 stránok), opačne však ženské názvy profesií odkazujú na tie mužské až v 214 prípadoch (Zagovora, et al., 2017).

Jedným z možných argumentov pre tento stav je jednoducho také, že Wikipédia len kopíruje aktuálny prevládajúci stav v spoločnosti a to spôsobom reflektujúcim stav na pracovnom trhu. Z časti tento predpoklad potvrdzuje Zagorovej štúdia (2017); heslá profesií na Wikipédii uvedené len v ženskom rode zodpovedajú vysokému zastúpeniu žien v tejto profesii na nemeckom pracovnom trhu (80 – 100 % zastúpenie), naopak heslá s výhradne mužskými protagonistami majú na pracovnom trhu zastúpenie žien od 10 – 60 %. Toto vysoké percento vypovedá o čiastočnom nerealistickom zobrazení a o stereotypizácii.

Podobne ako Zagorová interpretuje nemecké výsledky s odkazom na reálnu situáciu na pracovnom trhu aj Wagnerová (2015), ktorá poukazuje na možnú súvislosť lexikálnych rozdielov a stereotypizácie s celkovým postavením žien v spoločnosti meraným nestranné v rámci celosvetových výskumov *Gender Inequality Index*¹⁹.

¹⁹ <http://hdr.undp.org/en/composite/GII>

4.2 Aké sú tendencie v profesionálnej sfére online médií a spravodajstva?

Situácia v profesionálnych online médiách a v médiách všeobecne však nie je natoľko odlišná tej na Wikipédii, ikeď je vidieť pozitívne tendencie. Pri skúmaní (Yun, et al., 2007) prominentných online magazínov v USA, Veľkej Británii, Kórei a Mexiku, sa ukázala prevládajúca tendencia zobrazovať ženy nestereotypne, a to celkovo v 49,1% online článkov. Napriek tomu sú stále viditeľné rozdiely v spôsobe opisu žien a mužov; podobne ako je tomu aj na Wikipédii. V článkoch o ženách sa častejšie spomínali charakterové vlastnosti žien, predovšetkým priateľská, pokojná a atraktívna, taktiež sa viac venovali ich súkromiu. Muži boli zase častejšie prezentovaní ako lídri (92,2 % vs. 68 %) a opisovaní v kontexte pracovného prostredia (Yun, et al., 2007). Tieto výsledky potvrdzuje aj najnovší výskum zameraný na internetovú spravodajskú platformu Vice (D'Heer, et al., 2019). Ženy sa častejšie vyskytujú v spravodajstve v tzv. *soft news*, resp. zábava a osobné témy (71,8 % vs. 41,9 %) ako v tzv. *hard news*, ktoré sa týkajú závažných tém so signifikantným dopadom na spoločnosť, a to len v 28,2 % vs. 58,1 % (D'Heer, et al., 2019).

Rovnako D'Heer a kolegovia (2019) potvrdili výsledky štúdie z roku 2007 a to, že ženy sa výrazne častejšie vyskytujú v kontexte domácej, súkromnej sféry (48 % žien vs. 24,6 % mužov) ako vo verejnom a profesionálnom kontexte (51,5 % žien vs. 75,4 % mužov). Takisto výsledky predchádzajúcich štúdií potvrdzuje aj posledná celosvetová štúdia Global Media Monitoring Project (GMMP, 2015), všeobecne sa v médiách (TV, rádio, print a online spravodajstvo) odkazovalo na sociálny status a rodinu v prípade žien až v 21 % prípadov, u mužov len v 7 %. Ženy sú podľa štúdie GMMP (2015) predmetom len zhruba 25 % celkového mediálneho obsahu, sú tiež častejšie v médiách spomínané v súvislosti s kategóriami ako veda a zdravie (41 %), sociálne a právne témy (28 %), naopak len zriedka sú oproti mužom spájané s ekonomickými (23 %) a politickými záležitosťami (19 %). Podobne dopadli výsledky prieskumu GMMP (2015) aj v online prostredí. „Fakt, že ženy sú v tradičných médiách skoro neviditeľné platí rovnako aj pre platformy online spravodajstva; len 26 % osôb v internetových článkoch a spravodajstve sú ženy“. Okrem toho ženy oproti mužom častejšie v médiách zastávajú rolu očitého svedka, zdieľajú osobnú skúsenosť alebo zastupujú hlas ľudu, len zriedka sa vyskytnú v roli experta alebo hovorca, čo so zanedbateľnou odchýlkou niekoľkých percent (1-2 %) platí podľa štúdie aj v online priestore (GMMP, 2015).

Podobné výsledky ukazujú aj novšie štúdie (Ross, et al., 2018). Spájanie mužov s profesionálnym životom a žien s menej prestížnymi sektormi, súkromným, poprípade domácim, prevláda aj naďalej v médiách všeobecne, ilustračne len 20 % článkov a príspevkov

v britských médiách cituje ženu ako experta, či už je to TV, rádio alebo internetové spravodajstvo (Ross, et al., 2018).

Stereotypné zobrazenie žien a mužov v online médiách nachádzame aj v prípade použitej lexiky, podobne, ako je tomu na Wikipédií (Wagner, 2015; Wagner, et al., 2016; Otterbacher et al., 2019). Výsledky z výskumu korpusu článkov domáceho spravodajstva z 18 prominentných austrálskych novín ukázali značnú disproporciu v používaní niektorých slov a slovných spojení, ktoré naznačujú genderovú nevyváženosť a stereotypizáciu. V austrálskych novinách sa častejšie vyskytuje zámeno „on” na rozdiel od zámena „ona” a to v pomere 3,4 : 1. Slovné spojenie „on povedal” vs. „ona povedala” sú podobne nevyvážené, dosahujú pomer 3,6 : 1, čo, ako tvrdí prieskum, nám ukazuje koho hlas je v médiách dominantný (Lukin, 2016). Pri kalkulácií frekvencie podobne vychádzajú aj iné štáty (Nový Zéland, USA, GB, Kanada, India, Pakistan); zámeno „on” prevyšuje svojim počtom zámeno „ona” vo všetkých skúmaných krajinách viac ako raz, v niektorých, ako India a Pakistan, viac ako 4-krát (Lukin, 2016).

Ak sa pozrieme na stereotypizácia v austrálskom spravodajstve, ukazuje sa napríklad v najčastejších frázach spojených s podstatným menom muž a žena; pre podstatné meno muž je typické spojenie „muži a ženy”, pre ženu naopak spojenie „ženy a deti” a „ženy a dievčatá”. Celkovo počet slov odkazujúcich na maskulinitu (konkrétne: on, jeho, muž, muži, mužský, pán) prevažuje 3 :1 nad slovami odkazujúcimi na feminitu (ona, jej, žena, ženy, ženský, pani, slečna) (Lukin, 2016).

4.3 Špecifiká českého a slovenského prostredia

Podobné výskumy, ako sme uviedli v predchádzajúcej podkapitole, nájdeme aj v česko-slovenskom prostredí, týkajú sa však skôr tradičných médií ako online priestoru, ktorý zachytávajú okrajovo, no itak poskytujú náhľad do aktuálnej situácie v oboch krajinách a na ich základe je možné vytvoriť paralelu medzi situáciou vládnuou v offline s tou v online prostredí. České a slovenské prostredie sú si vo veľkej miere kultúrne a historicky najbližšie z ostatných európskych krajín, a ako si ukážeme v nasledujúcej podkapitole, v českom a slovenskom mediálnom priestore vládnu aj veľmi podobné pomery, čo sa týka genderovej otázky.

4.3.1 Mediálne prostredie

V prvom rade sa pozrieme na genderové zastúpenie žien a mužov v novinárskej profesii. Prieskum European Journalism Observatory z roku 2018, ktorý sa uskutočnil v 11 krajinách Európy, vrátane Českej republiky, bol realizovaný ako obsahová analýza dvoch vybraných

tradičných printových a dvoch online médií z každej skúmanej krajiny, pričom jeden zdroj vždy reprezentoval kvalitu, druhý zase popularitu. Pre Česko boli vybrané printové noviny MF Dnes a Právo, za online spravodajstvo portál Seznamzpravy.cz a Aktualne.cz. Celkovo pri pohľade na pomer mužov a žien v aktívnych v novinárskej profesii jednoznačne prevládajú muži s 57 % vs. ženy 43 % (European Journalism Observatory, 2018). Pri bližšom pohľade, konkrétne na počet článkov, pod ktorými sa podpísali muži a ženy, tento rozdiel ešte narastá. European Journalism Observatory (2018) zistilo, že v čase prieskumu bolo v médiách, printových aj online spoločne, 41 % článkov písaných mužskými novinármi a 19 % ženami novinárkami (u zvyšných percent bolo pohlavie nezistené). Táto priepasť sa ešte viac roztvára v prípade biznysovo ladených článkov (80 % mužov) a komentárov (90 % muži) (European Journalism Observatory, 2018). Jediným médiom, ktoré počas štúdie vykazovalo opačný trend bol spravodajský portál Seznamzpravy.cz, ktorý mal viac článkov od novinárk ako od novinárov (53 % ženských vs. 40 % mužských) (European Journalism Observatory, 2018).

Tento prieskum v slovenskom prostredí replikovala slovenská pobočka českej firmy Newton Media, ktorá sa zameriava na analýzu a archiváciu českého a slovenského mediálneho prostredia. Newton Media (Feriková, 2018) tento prieskum realizovala takisto v roku 2018 na vzorke údajov z prvého polroka 2018, v prípade denníkov a televízneho spravodajstva. U spravodajských webov bolo sledované len 3-mesačné obdobie, keďže články na weboch vychádzajú vo výrazne väčšom objeme. Z printových médií boli vybrané 4 celoštátne denníky SME, Pravda, Hospodárske noviny a Denník N. Zo spravodajských portálov boli predmetom analýzy online mutácie denníka Sme.sk, Pravda.sk, Hnonline.sk a Dennikn.sk a spravodajské relácie (okrem športových) v televíziách Markíza, JOJ, TA3 a Jednotka (Feriková, 2018). Výsledky analýzy (Feriková, 2018) ukázali, že celkovo sa v mediálnom prostredí viac presadzovali muži (55 %) ako ženy (45 %). Najväčšie rozdiely môžeme vidieť v printoch, kde až v 67% bol ako autor článkov uvedený muž (oproti 33 % žien). Pri online spravodajských portáloch sa nepomer mierne znižuje, ikeď stále zostáva signifikantný. Muži boli autormi 62% všetkých online článkov, ženy len 38%, pričom najviac genderovo vyváženým online portálom boli Hospodárske noviny. Prekvapivým zistením analýzy bol fakt, že situácia je úplne opačná v TV spravodajstve, kde viac ako polovica reportáží (53%) patrí ženám (Feriková, 2018).

Ako uvádza autorka analýzy Emília Feriková (2018), z výsledkov vidíme, že výsledok štúdie o genderových pomeroch v európskom žurnalizme zo správy European Journalism Observatory (2018) platí aj na poli slovenskej žurnalistiky.

Pri porovnaní výsledkov z českého (European Journalism Observatory, 2018) a slovenského mediálneho prostredia (Feriková, 2018) vidíme, že v oboch prípadoch v produkcii obsahu printových a online spravodajstiev vo veľkej miere dominujú muži. Na Slovensku ďalej platí, že ženy dostávajú viac priestoru vo webovom spravodajstve ako v tlači (Feriková, 2018), čo zodpovedá aj všeobecne európskym výsledkom správy European Journalism Observatory (2018) a do určitej miery to platí aj pre Česko, ktoré bolo zaradené do analýzy, no správa neuvádza presné percentá, len konštatuje že výsledky printu a online médií sú si podobné. Lepšiu predstavu o česko-slovenskom mediálnom online aj offline priestore získame pri pohľade na celkové čísla z oboch výskumov. V oboch krajinách tvoria väčšinu offline aj online článkov mužskí autori (ČR: 61 %; SR: 65 %), a teda rozdiely v týchto susedných krajinách sú minimálne (Feriková, 2018). V určitých oblastiach je slovenská novinárska scéna, čo sa týka genderu na tom trochu lepšie ako Česko; pod komentáre a hodnotiace články sa na Slovensku podpísalo 80 % mužov (Feriková, 2018), v Česku muži tvoria až 90 % takéhoto typu obsahu (European Journalism Observatory, 2018).

Celkovo môžeme povedať, ženy majú celkovo v českom aj slovenskom mediálnom priestore menší hlas ako ich mužský kolegovia, čo by sa takisto mohlo prejaviť aj na spôsobe, ako sú v médiách prezentované. K tejto tematike však v našom geografickom prostredí nevzniká veľa štúdií. Jana Valdrová, česká germanistka a špecialistka na genderovú lingvistiku, sa tejto téme venovala vo svojej staršej štúdii *Stereotypy a kliše v mediální projekci genderu* z roku 2001. V rámci tejto štúdie autorka kontinuálne počas 4 rokov sledovala český mediálny priestor a jej štúdia tvorí rozbor výberu výpovedí z českých médií, pričom ťažisko tvorí rok 2000. Ako hlavné zdroje uvádza autorka českú televíziu, prvý a druhý program, komerčné televízie Prima a Nova a z periodík Lidové noviny, Mladá fronta Dnes, časopisy pre mladých a pre rodiny (napr. Reflex a iné) a ojedinele aj ženské časopisy. Jednotlivé mediálne citáty Valdrová (2001) analyzuje skrz optiku genderu a lingvistiky. Záver štúdie tvorí súhrnný mediálny obraz ženy a muža v súdobých českých médiách. Autorka uvádza, že „česká „mediálna žena“ v podstate postráda vlastnú identitu. Dbá síce o zovňajšok a erotickú príťažlivosť, ale úspešnosť jej snahy posudzuje muž. Nekonkuruje mužovi profesne, svoje vzdelanie skôr skrýva. Vtipy o ženskej inteligencii znáša dobre, vie totiž, že by inak v spoločnosti neobstála. Na vzdelávanie jej zostávajú len večery – počas dňa sa venuje predovšetkým výchove detí, kde je nezastupiteľná. Nemôže za to čakať uznanie, pretože rodina je jej prirodzená sféra, nenáročná doména jej sebarealizácie. Ak má problémy, tak si ich zavinila sama vlastnými intrigami (Valdrová, 2001, s. 203). Na rozdiel od mediálneho obrazu ženy, „český „mediálny muž“ má ťažší život ako žena, pretože na rozdiel od nej podáva vysoký výkon v práci. Od

používajú výrazy stará (14 %) a mladá (13 %), ktoré však nie sú prídavné mená, ale podstatné mená (spojené s frázou „moja stará”, pre mladšie ročníky aj „moja mladá”), ďalej „holka“ s 12 % (baba), kráska (12 %) a značnú skupinu tvoria aj vulgarizmy (12 %). Nad 5 % získalo len označenie matka, ostatné oslovenia („dívka“, „ženština“, „naivka“, „milienka“, „šéfka“ a „dáma“), dosahujú len nízke percentá od 4 po 1 % (Mahdalová, 2015).

Ďalšia časť štúdie *Obraz žien ve verejnom diskurzu* (Mahdalová, 2015) sa zameriava na vnímanie rolí žien a mužov v spoločnosti a porovnáva online médiá a všeobecnú internetovú verejnú mienku, ktorá je voči ženám kritickejšia ako tá mediálna. Najvýraznejší nepomer nájdeme u vnímania rolí manžela a manželky; médiá z 81 % spájajú s manželstvom ženy a len 19 % mužov, naopak internetová verejnosť spája manželstvo so ženou až v 79 % a muža s rolou manžela len v 21 %. Zaujímavý je aj otvorene nepriateľský postoj internetovej verejnosti voči ženám, ktoré nezapadajú do tradičného genderového rodinného stereotypu, kedy žena nie je považovaná za slušnú: „Do slušnej a pracovitej ženy má strašne ďaleko. Tie sa starajú o svoje deti, manžela a o poriadok v byte.“ (Mahdalová, 2015). Štúdia Kateřiny Mahdalovej (2015) prináša aktuálny pohľad na vnímanie žien na internete. V štúdiu sa však autorka venuje výlučne len ženám, takže súčasný obraz o tom, ako sú na internete vnímaní muži, nám chýba.

V porovnaní s množstvom mediálnych analýz v Česku, Slovensko značne zaostáva a viac menej by sa dalo povedať, že vo veľkej miere vzťahuje české výsledky aj na slovenskú spoločnosť, i keď sa už skoro tri desaťročia vyvíjame samostatne. Jednou z akademičiek, ktorá sumarizuje mediálny obraz slovenských žien a mužov je Božena Markovič Balúchová (2010) vo svojej staršej štúdiu *Rodová (ne)citlivosť na Slovensku (a úloha médií v tomto probléme)*. Médiá podľa Balúchovej (2010) „predkladajú príliš zjednodušené obrazy mužskosti a ženskosti, zaťažené rodovými konotáciami, zároveň svojim vplyvom spolupôsobia ako sociálna kontrola pri určovaní prípustného správania sa žien a mužov. Najvýraznejšie sa to prejavuje v pretrvávajúcej dichotómii verejnej a privatej sféry, kedy muži sú zobrazovaní na titulkách ako kľúčoví aktéri verejného života, kým ženy spravidla ako konzumentky, matky alebo idoly mýtu krásy“ (Balúchová, 2010, s. 2). Tieto zistenia sú v súlade s novšími európskymi mediálnymi prieskumami (GMMP, 2015; Ross, et al., 2018; D’Heer, et al., 2019).

4.3.2 Prostredie Wikipédie

České a slovenské wikipedické prostredie je z hľadiska genderu značne neprebádaná oblasť. V našich zemepisných šírkach sa viac priestoru venovalo skôr využívaniu Wikipédie v procese štúdia či plnenia si študijných povinností (Filová, 2015; Gazdíková, 2014) a celkovo jej kvalite a dôveryhodnosti (Dobšovičová, 2011; Sakastrová, 2011; Žitná, 2013).

Dôvodom, prečo táto problematika aktuálne absentuje, môže byť okrem iného aj relatívne menšia prepracovanosť a celkovo menej hesiel na českej (420 000 článkov a 27. priečka v celkovom rebríčku jazykových verzií podľa počtu článkov)²¹ a slovenskej (229 943 článkov a 37. priečka)²² Wikipédii v porovnaní napr. s jej anglickou verziou. Rovnako návštevnosť slovenskej a českej Wikipédie je s anglickou neporovnateľne nižšia. Počty editorov sa značne líšia aj medzi českou a slovenskou Wikipédiou. Ako nakoniec uvádza sama slovenská Wikipédia²³ na vlastných stránkach, ich redaktorská základňa je pomerne malá ale stabilná, aktuálne s mierne klesajúcim trendom. V roku 2018 na tvorbe spolupracovalo približne 128 aktívnych (5 a viac úprav mesačne) a 15 veľmi aktívnych (100 a viac úprav mesačne) redaktorov. Česká Wikipédia uvádza na svojich stránkach²⁴ aktuálne 32 správcov, v porovnaní so slovenskou stránkou mala česká Wikipédia 575 aktívnych redaktorov a 91 veľmi aktívnych redaktorov²⁵. Porovnanie s anglickou verziou Wikipédie v číslach za rok 2018 sa nachádza v tabuľke nižšie:

Wikipédia	Počet článkov	Aktívni redaktori	Veľmi aktívni redaktori	Návštevnosť
		(5+ úprav mesačne)	(100+ úprav mesačne)	
Anglická	5 781 000	30 639	3 521	92,04 bil.
Česká	419 000	575	91	789 mil.
Slovenská	228 000	128	15	12,7 mil.

Tabuľka č. 2 Základné porovnanie štatistík pre českú, slovenskú a anglickú Wikipédiu²⁶

²¹ Údaje z wikipedickej stránky, aktuálne k 16. 1. 2019:
https://cs.wikipedia.org/wiki/%C4%8Cesk%C3%A1_Wikipedie

²² Údaje z wikipedickej stránky, aktuálne k 4. 6. 2019:
https://sk.wikipedia.org/wiki/Slovensk%C3%A1_Wikip%C3%A9dia

²³ https://sk.wikipedia.org/wiki/Slovensk%C3%A1_Wikip%C3%A9dia

²⁴ https://cs.wikipedia.org/wiki/%C4%8Cesk%C3%A1_Wikipedie

²⁵ https://sk.wikipedia.org/wiki/Slovensk%C3%A1_Wikip%C3%A9dia

²⁶ Údaje boli použité zo stránky slovenskej Wikipédie:
https://sk.wikipedia.org/wiki/Slovensk%C3%A1_Wikip%C3%A9dia a zo stránky Wikimedia Statistics:
<https://stats.wikimedia.org/v2/#/en.wikipedia.org/reading/total-page-views/normal|bar|2018-01-01~2018-12-31|-total|monthly>

Bližšie porovnávanie pomerov panujúcich v rámci českej a slovenskej Wikipédie však zatiaľ neudáva žiadny zdroj, či už z prostredia akadémie alebo priamo z Wikipédie samotnej. Táto podkapitola má byť venovaná výskumom, ktoré nám bližšie priblížia spôsob, akým sa v rámci či už mediálneho alebo konkrétne wikipedického prostredia komunikuje o ženách. Takýto typ výskumov však v našom regióne chýba. Jediné, čo naznačuje, že minimálnej v českej Wikipédii je táto téma genderu skloňovaná, je druhý ročník iniciatívy WikiGap 2019 s fókusom na Oslavu MDŽ, kde počas editačného maratónu bolo doplnených 26 článkov o významných českých ženách, keďže počet biografii o mužoch a o ženách je značne nevyvážený (Wikigap, 2019). Rovnako sa v rámci českej Wikipédie diskutuje aj o téme prechýľovania ženských priezvisk²⁷, domácich alebo zahraničných, čo je takisto jedna z tém, ktorú rozoberajú akademičky v Čechách aj na Slovensku (Komanická, 2016; Valdrová, 2018). Podobné genderovo orientované projekty sme však na slovenskej Wikipédii dosiaľ nezaznamenali²⁸.

Môžeme teda predpokladať, že minimálne česká Wikipédia do určitej miery bojuje s určitou formou biasu voči ženám. V prípade slovenskej Wikipédie takéto indície nemáme, ale vzhľadom na našu kultúrnu spriaznenosť je dosť možné, že slovenská Wikipédia môže mať rovnaký problém. Tomuto faktu nahráva aj štatistika návštevnosti českej Wikipédie, ktorá ukazuje, že až 12 % dotazov zo Slovenska smeruje do českej Wikipédie, ktorá je treťou najpoužívanějšíou jazykovou verziou hneď po slovenskej (44 %) a anglickej (34,5 %) (Česká Wikipédie, 2019). Na záver môžeme ešte dodať Rosenzweigove (2006) konštatovanie o Wikipédii všeobecne a to, že „*najväčší bias je itak uprednostňovanie západnej kultúry, pred napr. grécky či iným regionálnym biasom alebo populárnou kultúrou*” (Rosenzweig, 2006, s. 128). A teda aj celkový spôsob vykresľovania žien v českej a slovenskej Wikipédii bude pravdepodobne reflektovať všeobecný postoj západnej kultúry na ženy a ich postavenie v spoločnosti.

²⁷ https://cs.wikipedia.org/wiki/%C4%8Cesk%C3%A1_Wikip%C3%A9die

²⁸

https://sk.wikipedia.org/wiki/Kateg%C3%B3ria:Wikip%C3%A9dia:Projekty_Wikim%C3%A9die_Slovensk%C3%A1_republika

5 Stručný exkurz do genderovej lingvistiky

Ikeď sme zvýšený záujem o prezentáciu žien v online a offline mediálnych obsahoch zaznamenali len v posledných rokoch, téma zobrazovania žien vo vede, umení a produktoch pop kultúry, však v spoločnosti rezonuje už podstatne dlhšie. Na neadekvátne zobrazovanie žien v literatúre sa sťažuje už aj známa americká feministická autorka Virginia Woolfová vo svojej eseji *A Room of One's Own* (v slovenskom preklade *Vlastná izba*) z roku 1929: „*Všetky tieto vzťahy medzi ženami, ako si narýchlo vybavujem tú bohatú výkladnú skriňu ženských postáv, sú myslím až príliš jednoduché. ... A skúšala som si spomenúť na aspoň jeden príklad z mojej čitateľskej minulosti, kedy boli dve ženy zobrazené ako priateľky. ... Občas sú stvárnené ako matky a dcéry. Ale skoro bez výnimky sú zobrazené na základe ich vzťahu k mužom. Bolo zvláštne si myslieť, že až do príchodu Jane Austenovej, všetky tie skvelé ženy nielenže boli zobrazované len z pohľadu muža, ale zároveň iba v súvislosti s mužským pohlavím. Akú malú časť ženského života to predstavuje*“ (Woolf, 1929).

5.1 Ženská otázka

K profesionalizácii záujmu o túto otázku došlo však približne až o 50 rokov neskôr. Ženskú otázku do lingvistiky priniesla až vedkyňa, dnes profesorka na Univerzite v Berkeley, Robin Lakoffová v roku 1973 s prelomovou štúdiou *Language and Woman's Place*, ktorá odštartovala spoločensko-vedný diskurz o využívaní jazyka v súvislosti s postavením žien. Lakoffová sa vo svojej knihe zaoberá tak spôsobom vyjadrovania sa žien a používanou lexikou, ako aj spôsobom, ako sa o ženách hovorí. Uvádza množstvo príkladov, v ktorých poukazuje na rozdiely medzi vyjadrovaním sa mužov a žien, ktoré sú však podmienené kultúrne, teda podliehajú určitým stereotypom a očakávaniam čo do spôsobu vyjadrenia, tak i v prípade použitých výrazov, ktoré sú vyhradené ženám a mužom.

Lakoffová pre ilustráciu uvádza príklad s malým dievčatkom, ktoré sa len objavuje svoju identitu a postupne sa učí vyjadrovať svoje názory. V tomto procese učenia ju usmerňujú tak v škole ako aj v rodinnom kruhu, je napomínaná ak rozpráva „drsne ako chlapci“. Tento proces socializácie a teda učenia sa sociálnych a kultúrnych noriem je samozrejme prirodzený, no problém podľa Lakoffovej nastáva v prípade, že sa dievča plne podvolí a naučí sa „rozprávať ako dáma“. „*Potláča to vlastnú identitu ženy tým, že jej na jednej strane znemožňuje vyjadriť sa jasne a ostro, tým že je dotlačená k využívaniu slovných zvrátov, ktorými sama naznačuje trivialitu diskutovanej témy alebo vlastnú neistotu, na druhej strane v prípade, že sa diskutuje*

o žene samotnej, hovorí sa o nej ako o objekte – aj v sexuálnom význame, no nikdy nevystupuje ako seriózna osoba s vlastným názorom“ (Lakoff, 1973, s. 48).

Od publikácie prelomového textu Robin Lakoffovej skúmanie otázky genderu a jazyka v rokoch 1980 - 1990 prinieslo štyri odlišné pohľady na problematiku, ktoré vo svojej publikácii *Woman, Men and Language* (2016) približuje Jennifer Coates. Historicky prvý je princíp (1.) deficitu, neskôr vznikol princíp (2.) dominancie, (3.) rozdielu a v súčasnosti prevláda princíp (4.) dynamiky, resp. sociálneho konštraktu.

Princíp deficitu, najstaršiu formu nazerania na genderový problém, nájdeme v prvej štúdií Robin Lakoffovej z roku 1973, v ktorej zadefinovala tzv. *woman language*, teda ženský jazyk, ktorý je vo svojej podstate slabý, neasertívny, inými slovami deficitný vo vzťahu k normatívnemu mužskému spôsobu vyjadrovania. „*Tento prístup bol neskôr zavrhnutý, predovšetkým preto, lebo naznačoval fundamentálne chyby v ženskom jazyku, a teda že by sa ženy mali naučiť vyjadrovať ako muži na to, aby boli brané vážne*“ (Coates, 2016).

Druhý prístup sa zameriava na princíp dominancie, ktorý považuje ženy za utlačované, čomu prispôsobuje aj interpretáciu ich spôsobu reči, a to v zmysle mužskej dominancie a ženskej subordinácie. Jazyk je v tomto prípade prostriedkom nastoľovania mužskej dominancie – *doing power* je v istom zmysle *doing gender* (Zimmerman, 1987). Tento prístup nájdeme predovšetkým v prácach Dale Spendera, Don Zimmermana a Candace West.

Tretí v poradí, princíp rozdielu, je akoby postavený na populárnom „muži sú z Marsu a ženy z Venuše“, a teda zdôrazňuje rozdiely s jazyku mužov a jazyku žien. Vznikol ako pokus o vymanenie sa zo starých jazykových mocenských štruktúr, hlásajúc vlastnú ženskú kultúru, teda inú, ako je v spoločnosti dominantná. Vychádza z predpokladu existencie odlišných ženských a mužských subkultúr, ktoré vznikli ako dôsledok odlišnej socializácie, a ktorej výsledkom sú rôzne komunikačné štýly mužov a žien. Najznámejšou predstaviteľkou tohto princípu je Deborah Tannen, ktorá v knihe *You just don't understand* (1990) na základe zdanlivej príslušnosti do odlišných subkultúr popisuje aj jej odlišnú manifestáciu v jazyku; muži preferujú strohý štýl komunikácie (*report style*), naopak ženy sa uchýľujú k obsiahlejším a detailnejším konverzáciám, ktoré majú za cieľ nielen predať fakty, ale aj budovať a posilňovať vzťahy medzi aktérmi (*rappor style*) (Tannen, 1990).

Posledný prístup vyzdvihuje dynamický aspekt jazyka a komunikácie. Gender je v tomto prípade vnímaný ako výsledok sociálnej interakcie, a teda dynamicky sa meniaci

sociálny konštrukt. Podľa tejto definície je teda skôr pravdivé slovné spojenie *doing gender* (Zimmerman, 1987), spomínané už v predchádzajúcich odstavcoch. Tu je na mieste spomenúť, že všetky 4 princípy vznikali viac menej v rovnakom období a ich hranice sú priepustné. V súčasnosti sa však predchádzajúce tri koncepty považujú už za prekonané a vo vedeckom diskurze dominuje posledný princíp dynamiky, resp. sociálnej konštrukcie. Gender je teda niečo, čo vzniká počas interakcie, nie je statický, skôr svojimi vlastnosťami pripomína sloveso ako podstatné meno. Tento prístup je zjavný aj pri súčasných definíciách genderu, ako si neskôr ukážeme v ďalšej časti práce.

V súčasnosti sa genderovej lingvistiky vo svete venuje predovšetkým Deborah Cameron, ktorá sa sústreďuje na poukazovanie možného negatívneho dopadu princípu deficitu prípadne princípu rozdielu na súčasnú spoločnosť a kultúru. Ako sme už naznačili, prvotné tri princípy genderovej lingvistiky vedci dnes považujú za prekonané, avšak stále sú populárne u laickej verejnosti a v kruhoch tzv. *pop-science* (Coates, 2016). *The Myth of Mars and Venus* je najnovšou knihou Deborah Cameron (2008), ktorá si sugestívne kladie otázku či muži a ženy skutočne rozprávajú inými jazykmi. Vznikla ako priama reakcia na populárno-vedeckú knihu vzťahového poradcu Johna Graya (1992) *Man are from Mars and Women are from Venus*. Cameronová poukazuje na škodlivý vplyv zjednodušovania a prílišnej kategorizácie nielen jazykového prejavu mužov a žien, ktoré vedie k prehlbovaniu stereotypizácie a stereotypného nazerania na mužov a ženy.

V českom a slovenskom prostredí sa v súčasnosti genderovej otázke jazyka venujú predovšetkým Jana Valdrová a Jana Cviková. Obe nadväzujú na princíp vnímania genderu ako dynamicky sa meniaceho sociálneho konštraktu. Česká lingvistka Jana Valdrová poukazuje na špecifiká slovanských jazykov v genderovej otázke jazykového vyjadrenia, a to predovšetkým v otázke prechýľovania ženských priezvisk (Valdrová, 2018). Podobná tematika rezonuje aj na Slovensku (Komanická, 2016), skúmajú sa aj špecifiká menšín a ich jazyka (Molnár Satinská, 2014), výrazný dôraz sa kladie na vzdelávanie a implementáciu genderovo-vyváženého jazyka (Ako používať rodovo vyvážený jazyk, 2014).

5.2 Populárno-vedecké genderové testy

Indikátorom, že téma nerovného zobrazovania mužov a žien všeobecne v spoločnosti rezonuje, je fakt, že v posledných rokoch aj vďaka sociálnym médiám naberajú na popularite testy viac menej populárno-vedeckého charakteru, podľa ktorých je možné si vytvoriť predstavu o zastúpení a postavení žien napríklad v rámci filmu. Pre účely tejto práce je

najznámejším nástrojom je tzv. Finkbeiner test z roku 2013, ktorý patrí medzi desiatku najvyhľadávanejších testov zameraných na genderovú analýzu výstupov populárnej kultúry (Rughinis, et al., 2016). Tento test sa na rozdiel od väčšiny populárnych testov (Rughinis, et al., 2016) zameriava explicitne na textové výstupy. Jeho autorkou je novinárka Christie Aschwandenová, ktorá vytvorila krátky zoznam pravidiel ako písať o ženách vo vede. Na to, aby článok týmto testom prešiel, nesmie obsahovať nasledujúce: explicitne uvádza, že sa jedná o ženu, povolanie jej manžela, prípadne partnera, spôsob, akým je postarané o deti, ako sa správa k svojim podriadeným, ako ju prekvapila konkurencia v jej odbore, že predstavuje vzor pre ďalšie ženy a nakoniec obligátnu frázu „je prvou ženou, ktorá...“ (Curtis, 2013). Tento test vznikol ako odpoveď na odmietnutie zdôrazňovania genderového aspektu v článku o americkej astronómke, ktorý odmietla napísať novinárka Ann Finkbeinerová (Finkbeiner, 2013). Tento test je okrem iného často využívaný ako jedno z kritérií genderového výskumu napríklad aj v prostredí Wikipédie (napr. Wagner, et al., 2016 a iné).

Okrem Finkbeinerovej testu sa v posledných rokoch objavilo množstvo ďalších populárnych testov so zameraním na reprezentáciu žien a mužov v produktoch populárnej kultúry, ktoré sa postupne dostali aj do výskumnej praxe. Na prvej priečke obľúbenosti sa stále drží tzv. Bechdelovej test, resp. Bechdel-Wallace test, z roku 1985 (Rughinis, et al., 2016).

I keď sa v tomto prípade nejedná explicitne o skúmanie jazyka, a teda verbálneho jazykového prejavu a použitej lexiky, ale naopak o audio-vizuálne umenie, Bechdelovej test si v súčasnosti našiel nové uplatnenie ako východisko pri analýze textových výstupov sociálnych médií (napríklad Garcia, et al., 2014 a iné).

Bechdelovej test, inak známy aj ako Bechdel-Wallace test vychádza z priekopníckeho amerického komixu *Dykes to Watch Out For* (v českom preklade *Lesby k pohľadání*). V komixe *Dykes to Watch Out For*, konkrétne v časti príhodne nazvanej *The Rule* (Pravidlo), dve hlavné postavy rozoberajú 3 kritériá výberu film, ktorými sa postava riadi, keď chce ísť do kina. Práve tieto tri základné kritériá sa spopularizovali a dostali označenie Bechdelovej test, podľa autorky komixu Alison Bechdelovej. Test sa skladá z nasledujúcich troch kritérií: (1.) Film musí mať minimálne dve ženské postavy, (2.) ktoré sa spolu rozprávajú (3.) a vedú rozhovor o niečom inom ako o mužovi (Waletzko, 2017).

5.3 Otázka genderu

Genderová otázka je úzko spätá s rolami a postavením v spoločnosti, ako už bolo naznačené v predchádzajúcom texte. Gender ako sociálny konštrukt, je poslednou z vývojových štádií definície genderu.

Prvé koncepcie hľadali odpoveď na otázku, či individuálne rozdiely medzi pohlaviami majú biologický alebo sociálny základ²⁹.

V druhej vlne záujmu o genderovú otázku sa od biologického pôvodu upustilo a pozornosť sa obrátila na spoločenskú štruktúru ako tvorca rozdielu v správaní medzi pohlaviami. Konkrétne sa spomína predovšetkým rodina, v ktorej sa deti tieto koncepty učia a je od nich následne vyžadované ich dodržiavanie. Teda gender ako identita a rola, ktorú si človek osvojil počas detstva.

Tretia vlna prízvukuje princíp sociálnej interakcie, ktorej základom je vôľa prispôbovať sa očakávaniam spoločnosti či skupiny, ktoré posilňuje stereotypizáciu. Posledný princíp chápe gender ako sociálny konštrukt, v zmysle inštitucionalizovaného systému sociálnych praktík, podporujúci systém sociálnej stratifikácie, teda rozvrstvenie spoločnosti na sociálne vrstvy, skupiny, spoločensvá a podobne, v tomto prípade konkrétne vytváranie dvoch odlišných kategórií, muža a ženy, mužského a ženského, na základe čoho im je im priradený status v spoločnosti, ktorý v súčasnosti vnímame ako nerovný. (Risman, 2004; Ridgeway, 2004).

Ruka v ruke so spoločnosťou ide aj kultúra. „*Gender je teda sociálny konštrukt, avšak je kultúrne podmienený. Každá kultúra si vytvára svoje vlastné významy, ktoré chápe pod pojmom mužský a ženský a má vlastné očakávania ohľadom typu správania, ktoré sa od nich očakáva. Ak by sme zašli do extrému, tak tieto očakávania vytvárajú rodové stereotypy.*“ (Eksi, 2009, s.38)

Jazyk je takisto silným prostriedkom vytvárania a udržiavania genderových stereotypov, kdeže je jedným z kľúčových faktoroch pri formovaní identity jednotlivca, tak ženskej ako i mužskej. Pande (2004) konkrétne uvádza, že: „*Výskum na tému gender a jazykové štruktúry rovnako ako štúdie zaoberajúce sa jazykom a vytváraním sociálnych identít jednoznačne ukázali, že jazyk nemôže byť neutrálnym prostriedkom vyjadrenia vlastného ja... Jazyk je úzko*

²⁹ V anglickom jazyku je rozdiel medzi biologickým pohlavím a pohlavím ako sociálnou rolou jednotlivca v spoločnosti zjavný už podľa pomenovania. Ak hovoríme o pohlaví z biologického hľadiska, nazýva sa „sex“. Ak naopak hovoríme o pohlaví v inom ako biologickom zmysle slova, používa sa „gender“.

prepojený so spoločenským, kultúrnym a historickým dedičstvom a je rovnako plný ďalších systémov symbolov. Je to len v rámci konkrétneho diskurzu, kedy jednotlivci interpretujú významy týchto symbolov v konkrétnom kontexte debaty.“ (Pande, 2004, s. 47 - 48).

Podľa zistení Ridgewayovej a Corellovej (2004) v západnej kultúre napriek celospoločenskej snahe o odbúranie stále existuje značne stereotypné vnímanie muža a ženy. Vo všeobecnosti sa ženy považujú za viac spoločenské a muži zas za akčnejších a vecných. Na dôvažok majú tieto rozdiely aj charakter nerovnováhy postavenia v spoločnosti a celkovej moci. Muži sú vnímaní ako kompetentnejší a to predovšetkým v oblastiach, ktoré si západná kultúra cení najviac, akými sú práca či kariéra, racionalita, vecnosť, konštruktívny prístup. Naopak ženy sú považované za menej kompetentné, ale milšie a lepšie v spoločensko-komunikačných situáciách, ikeď tie sú spoločnosťou považované za menej hodnotné.

Tieto a iné stereotypné predstavy je možné vypozerovať aj na jazyku a formulácií vyjadrení o mužoch a o ženách, v slovnej alebo písanej forme. Okrem toho „*v rozpore s domnienkou, že jazyk len čisto reflektuje sociálne štruktúry, akými sú aj rodové stereotypy, výskum v oblasti lingvistiky a sociálnej psychológie dokázal, že tieto role sú v skutočnosti, naopak, podporované a posilňované jazykom*“ (Hellinger, 1984, s. 361). Záleží teda na tom, ako sa vyjadrujeme o mužoch a o ženách, a to nielen slovne, ale aj písomne. Inak sa vzorce a stereotypy ďalej ukotvujú a prehľbujú. Zmena v jazyku je aj zmenou v myslení a opačne.

6 Metodológia výskumu

V nasledujúcej časti si postupne predstavíme základné princípy, z ktorých sme vychádzali pri formulácii výskumu a hypotéz, detailne popíšeme výber dát, prípravu datasetu spolu s metódami kvantitatívnej analýzy, ktoré boli v rámci výskumu použité. V neposlednom rade sa budeme venovať konečným výsledkom, ich interpretácii a možným implikáciám.

6.1 Na úvod

Wikipédia si veľmi zakladá na kvalite svojich článkov. Na jej stránkach nájdeme návody³⁰, podľa ktorých sa majú editori riadiť a na základe ktorých, budú ich príspevky spätne posudzované. Rovnaké stránky s pravidlami nájdeme aj v českej a slovenskej verzii Wikipédie, najviac sa svojimi odporúčaniami približujú tým genderovým v oblasti nezaujatého pohľadu³¹, špecificky sa však ženskej problematike nevenujú. Naopak špeciálnu kategóriu v anglickej verzii Wikipédie tvorí sekcia *Writing about Women*, kde nájdeme niekoľko užitočných bodov, ktorých dodržiavaním máme vyššiu šancu vyhnúť sa (neúmyselnému) znevýhodňovaniu žien v texte a ich stereotypnému zobrazovaniu. Ako sa uvádza priamo na wiki stránke s pravidlami, jej cieľom je „*pomôcť identifikovať subtílne ale aj očividnejšie spôsoby, ktorými oslovovania, jazyk, ilustračné obrázky a prelinkovania na anglickej Wikipédii môžu diskriminovať ženy*” (Wikipedia, 2019). Medzi odporúčaniami (Wikipedia, 2019) nájdeme napríklad používanie priezviska na adresovanie žien v texte namiesto mena, ktoré môže ženu infantilizovať, zbytočne nevyzdvihovať fakt, že sa jedná o ženu, vyhýbať sa slovnému spojeniu „prvá žena” (s odkazom na Finkbeinerovej test), dôraz na genderovo neutrálne slová a dokonca aj upozornenie na znevýhodňovanie žien v rámci prelinkovania v infoboxoch. Veľká pozornosť je ďalej venovaná aj spôsobu popisu vzťahov, manželstva a rodiny v ženských biografiách, ktoré by nemali byť excesívne a príliš ženu definovať (Wikipedia, 2019).

Ako už bolo povedané na začiatku, Wikipédia si zakladá na kvalite a kontrole kvality svojho obsahu, čo koniec koncov ilustrujú aj vyššie uvedené ukážky z pravidiel. Fakt, že viaceré výskumy na takúto skutočnosť poukazujú však skôr ukazuje jednak hĺbku zakorenenia tohto problému a zároveň aj fakt, že postrehnúť takýto jazykový bias (*lingvistic bias*) je často zložité, ako upozorňuje Otterbacher (2019), keďže nemá explicitný ale skôr subtílly charakter.

³⁰ Ako sa uvádza v: https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style

³¹ https://cs.wikipedia.org/wiki/Wikipedie:Nezaujat%C3%BD_%C3%BAhel_pohledu a https://sk.wikipedia.org/wiki/Wikip%C3%A9dia:Nestrann%C3%BD_uhol_poh%C4%BEadu

V rámci prípravy hypotéz predkladanej práce sme sa inšpirovali predovšetkým štúdiou *First Women, Second Sex: Gender Bias in Wikipedia* (2015), ktorá sa na otázku rodovej rovnosti na Wikipédii snaží odpovedať za pomoci kvantitatívnej analýzy sledujúcej nasledovné hlavné faktory: (1.) dôraz na pokrytie žien a mužov v článkoch, (2.) dôraz na štruktúru a prepojenie článkov a (3.) dôraz na použitú lexiku. V rámci práce budeme čiastočne replikovať predovšetkým prvú a tretiu časť, ktorá na základe frekvencie výskytu slov sleduje použitú lexiku v mužských a ženských biografiách. A teda prvé dve stanovené hypotézy budú po vzore štúdie Graells-Garrida a kolegov (2015) sledovať počet biografii o mužoch a o ženách a priemerný počet slov na článok, resp. medián.

Táto časť výskumu je ďalej doplnená o výskum *Size Matters: Word Count as a Measure of Quality on Wikipedia* o zjednodušení určovania kvality wikipedických článkov od Joshua E. Blumenstocka (2008). Blumenstock (2008) považuje za validný indikátor kvality wikipedického obsahu práve počet slov v článku, keďže je to ľahko dostupná metrika, na rozdiel od iných metód si nevyžaduje komplexný prístup a je jednoducho zrozumiteľná aj pre laickú verejnosť (čím viac slov, tým kvalitnejší článok). Pri pohľade na počet slov v datasete, na priemerný počet slov na článok a hodnotu mediánu si vieme spraviť predstavu o kvalite spracovania článkov.

Pri výbere kategórií slov, ktoré budeme sledovať pri frekvenčnej analýze, sme vychádzali zo zistení Claudie Wagnerovej a kolektívu (2015) a Graells-Garrida a kolegov (2015), ktoré oba ukázali častejší výskyt slov odkazujúcich na gender v ženskom datasete, rovnako ako sa u žien častejšie objavujú slová, ktoré sa viažu k rodinnému stavu, ako napríklad vydatá, rozvedená, deti alebo rodina. Preto sme sa rozhodli do hypotéz, konkrétne do frekvenčnej analýzy, zahrnúť frekvencie slov odkazujúcich na gender (aj na opačný gender), sociálny status a rodinu, a v neposlednom rade, slová viažuce sa na kariéru. Kompletný zoznam kategórií a slov, ktoré sme v rámci nich uvádzame v Prílohe 2.

Posledným bodom výskumu je porovnanie českých a slovenských výsledkov. V našich geografických šírkach zatiaľ výskumy, ktoré sa pozerajú na Wikipédiu genderovým pohľadom, absentujú. Menšie množstvo štúdií a prieskumov (Mahdalová, 2015; European Journalism Observatory, 2018; Feriková, 2018) sa zameriava na české a slovenské mediálne prostredie všeobecne a tie do veľkej miery vykazujú podobné výsledky týkajúce sa nerovnosti v prezentovaní žien a mužov ako svetové výskumy. Posledná, piata hypotéza vychádza z výsledkov existujúcich českých a slovenských mediálnych výskumov.

6.2 Formulácia hypotéz

V praktickej časti tejto práce si budeme všímať:

1. POČTY ČLÁNKOV A SLOV:

Výsledky výskumu E. Graells-Garrida a kolegov (2015) ukázali, že v rámci anglickej Wikipédie sú zásadné rozdiely v počte biografí o mužoch a o ženách a menej významné, ikeď stále existujúce rozdiely medzi dĺžkou ich textu, resp. ženský dataset obsahoval menej článkov o ženách a tie boli aj menej početné na slová. Tento fakt sa však môže líšiť regionálne, vzhľadom na očakávanú menšiu prepracovanosť iných jazykových mutácií Wikipédie a celkovo inej národnej kultúry. Z tohto predpokladu vychádza Hypotéza H1.

H1: v počte článkov bude v obidvoch jazykoch prevládať mužský dataset.

Druhá hypotéza nadväzuje na výskum E. Graells-Garrida a kolegov (2015) a ďalej stavia na výskume J. E. Blumenstocka (2008), ktorý predstavil jednoduchú metriku merania kvality wikipedického obsahu, ktorou je počet slov v článku; čím viac slov, tým je vyššia šanca, že je článok kvalitnejšie spracovaný. Pomocnou metriku bude v tomto prípade medián.

H2: články v mužskom datase budú kvalitnejšie spracované, ako články v ženskom datase vzhľadom na signifikantne vyšší priemerný počet slov na článok a vyššiu hodnotu mediánu, a to v obidvoch jazykoch.

2. FREKVENCIA SLOV:

Naprieč rôznymi výskumami, ktoré sa uvádzajú v tejto práci, sa vyskytuje zásadný rozdiel v prezentácii mužov a žien v online aj offline mediálnom prostredí (Yun, et al., 2007; GMMP, 2015; D'Heer et al., 2019), vrátane špecificky Wikipédie (Wagner et al., 2015; Graells-Garrida, et al., 2015; Wagner et al., 2016; Otterbacher, et al., 2019). Pri formulácii tretej hypotézy vychádzame zo všeobecných tendencií prezentácie žien v mediálnom prostredí a špeciálne zo zistení výskumu *First Women, Second Sex: Gender Bias in Wikipedia* (Graells-Garrida, et al., 2015) a výskumu *It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia* (Wagner, et al., 2015), ako už boli prezentované.

H3: v slovenskom a českom ženskom datasete bude častejší výskyt slov odkazujúcich na ich gender, sociálny status a rodinu a slov odkazujúcich na opačné pohlavie, ako v mužskom datasete.

H4: v ženskom datasete bude menej častý výskyt slov odkazujúcich na kariéru. V prípade mužského datasetu budú tieto zistenia opačné, a to v prípade oboch jazykov.

3. POROVNANIE:

Prieskumy českého a slovenského offline aj online mediálneho prostredia ukazujú značne podobné výsledky, čo sa týka vnímania a prezentácie žien v mediálnych obsahoch (Datová žurnalistika, 2015; European Journalism Observatory, 2018; Feriková, 2018). Napriek veľmi príbuznej kultúre a jazyku však očakávame určité rozdiely medzi českou a slovenskou Wikipédiou v zobrazovaní mužov a žien.

H5: pri porovnávaní výsledkov analýzy dosiahne česká Wikipédia celkovo lepšie výsledky, a teda preukáže menšiu mieru stereotypizácie žien ako slovenská.

Pod lepšími výsledkami rozumieme predovšetkým percentuálne vyšší podiel článkov o ženách na Wikipédii v porovnaní so slovenskou verziou, celkovo vyšší priemerný počet slov na článok v ženskom datasete, resp. kvalitnejšie články o ženách ako na slovenskej Wikipédii a menej častý výskyt jednotlivých kategórií slov, v porovnaní so slovenskou verziou Wikipédie, ktoré sa vzťahujú na kategórie gender, slov odkazujúcich na opačné pohlavie, sociálny status a rodinu a naopak vyšší výskyt slov z kategórie kariéra.

6.3 Metodológia

Ako metóda výskumu bola zvolená kvantitatívna dátová analýza korpusu českých a slovenských biografíí o mužoch a o ženách z českej a slovenskej verzie Wikipédie. Predkladaný výskum čerpá a čiastočne replikuje výskum Eduarda Graells-Garrida *First Women, Second Sex: Gender Bias in Wikipedia* (2015). Kvôli špecifickosti českého a slovenského jazyka a práce s prirodzeným jazykom bol zvolený ako nástroj analýzy softvér RStudio a bližšie knižnica UDPipe³², teda nástroje iné, ako boli použité vo výskume C. Wagnerovej, ktorej výskum prebiehal na anglickej Wikipédii.

V prípade češtiny a slovenčiny bolo potrebné nájsť a použiť nástroj (model), ktorý dokáže text a slová previesť na základný tvar. Knižnica UDPipe takéto nástroje, resp. modely,

³² Viac informácií o knižnici vrátane dokumentácie a použitých príkazov: <https://bnosac.github.io/udpipe/en/>

poskytuje a pri základnom teste sa kvalita výstupu ukázala byť dostatočná³³. Špeciálne pre slovenský jazyk existuje len minimum takýchto nástrojov.

Po výbere vhodného nástroja analýzy sme jednotlivé datasety postáhovali za pomoci skriptu a ďalej spracovávali vo voľne dostupnom štatistickom programe RStudio za pomoci predtrénovaného českého a slovenského jazykového modelu na prácu s textom v prirodzenom jazyku, ktorý existuje v rámci knižnice UDPipe. V rámci analýzy sme pracovali so 4 separé dátovými korpusmi; české biografie o mužoch, české biografie o ženách, slovenské biografie o mužoch a slovenské biografie o ženách.

6.4 Zber dát a príprava datasetu

6.4.1 Zber dát

Dáta sme v oboch prípadoch stiahli z oficiálneho a voľne dostupného balíka dát, Wikimedia dumps³⁴. Spolu so sťahovaním dát sme následne využili skript Wikiextractor na čistenie, resp. odstránenie nepotrebných syntaxe a metadát špeciálne vyvinutý pre Wikimedia dump³⁵, pomocou ktorého vznikol dataset v plain texte (.txt formáte). Ešte však pred spustením Wikiextractora³⁶ sme označili biografické články o mužoch a o ženách, ktoré sme určili na základe kategórií v metadátach. Česká Wikipédia má v tomto prípade pre naše účely lepšie prispôsobenú štruktúru a kategorizáciu, biografické články sme vybrali a roztriedili podľa označenia Žena³⁷ a Muž³⁸, uvedených ako kategóriu v rámci metadát jednotlivých stránok. Následne sme spustili Wikiextractor, ktorý dáta očistil od nepotrebných syntaxe. Po vytriedení biografických článkov nám spoločne teda vznikol z českej Wikipédie dataset s 92 517 biografiami o mužoch a dataset s 18 349 biografiami o ženách.

Podobný priebeh mala aj príprava dvoch slovenských datasetov. Stiahnutie obsahu ako Wikimedia dump³⁹ a očistenie pomocou skriptu Wikiextractor. Avšak slovenský dataset si vyžadoval ešte určité úpravy navyše, keďže na slovenskej Wikipédii nie sú súhrnné kategórie pre všetky články o ženách a o mužoch, čím sa sťažila identifikácia článkov. Z wikipedického

³³ Jedná sa o knižnicu predtrénovaných modelov, kvalita modelov sa teda môže líšiť od jazyka k jazyku.

³⁴ <https://dumps.wikimedia.org/cswiki/latest/cswiki-latest-pages-articles.xml.bz2>

³⁵ <https://github.com/attardi/wikiextractor>

³⁶ Wikiextractor očistí dáta od nepotrebných syntaxe a porozdeľuje dáta do menších súborov - články skriptu prechádza rad za radom (po poradí identifikačných čísel) a na základe toho, aké sú jednotlivé články veľké, ich ukladá do súborov tak, aby súbor nepresiahol určitú danú veľkosť.

³⁷ <https://cs.wikipedia.org/wiki/Kategorie:%C5%BDeny>

³⁸ <https://cs.wikipedia.org/wiki/Kategorie:Mu%C5%BEi>

³⁹ <https://dumps.wikimedia.org/skwiki/latest/skwiki-latest-pages-articles.xml.bz2>

dumpu⁴⁰ bolo potrebné vybrať len biografické články, ktoré sme vybrali na základe toho, že v rámci metadát biografických článkov sa vždy nachádza pod skratkou * „Dátum narodenia“. Z takto vzniknutého súboru sme ďalej roztriedili články do ženského a mužského datasetu na základe najčastejších povolání (politik, tenistka, herečka a pod., keďže názvy povolání v slovanských jazykoch indikujú gender), ktorých názov sme nechali vyhľadať v texte. Následne sme zostatok článkov (približne 3 000) roztriedili na mužské a ženské ručne, podľa mena. Tieto úpravy sme urobili ručne cez príkazový riadok, bez špeciálneho softvéru. Po očistení dát cez Wikiextractor nám spoločne vznikol dataset ženských biografíí pozostávajúci z textu z 3 803 článkov a dataset mužských článkov skladajúci sa z 20 707 biografíí.

Kým z českej Wikipédie máme istotu, že náš dataset obsahuje presne toľko článkov o mužoch a o ženách, koľko ich bolo v tom čase dostupných (pri porovnaní so zoznamom na webe), pri slovenskej Wikipédii takúto istotu nemáme, keďže si nemáme s čím spätne overiť konečný počet článkov. Mohlo v procese prípravy dôjsť k určitej strate, prípadne k zlému roztriedeniu určitého počtu článkov.

Ani v jednom jazyku neboli dáta smplované, datasety predstavujú balík všetkých v tom čase dostupných biografických článkov. Dáta boli v oboch prípadoch stiahnuté a uložené vo formáte .txt s jazykovým kódovaním UTF-8⁴¹. V rámci prípravy oboch datasetov boli nakoniec takisto odstránené prázdne riadky a špeciálne znaky ako napr. znaky zahraničnej abecedy, úvodzovky a pod., keďže sa ukázalo, že tento typ znakov nie je zadaný a podporovaný v UTF-8 znakovom formáte, s ktorým pracuje UDPipe.

6.4.2 Príprava dát na analýzu

Pripravené dáta sme následne nahrali do softvéru RStudio. V prípade českej Wikipédie sme mužský dataset nahrávali do RStudia postupne po častiach a s použitím knižnice UDPipe a českého jazykového modelu (czech-pdt-ud-2.3-181115.udpipe) bol postupne anotovaný celý text. Jednotlivé anotované časti, teraz už vo formáte .csv⁴², sme následne načítali do premenných ako dataframe a tie spojili do jedného anotovaného datasetu pre mužov s použitím príkazu rbind⁴³, ktorý sme následne vyexportovali a uložili. Celkový dataset českých mužských biografíí bol príliš veľký na spracovanie naraz (viac ako 2 GB), preto sme súbor anotovali

⁴¹ Knižnica UDPipe si špecificky vyžaduje súbory len vo formáte .txt a s jazykovým kódovaním UTF-8.

⁴² Pri anotácii sa k jednotlivým slovám dopĺňujú kategórie. Takýto súbor (tabuľka) sa bežne ukladá vo formáte .csv (coma separated values).

⁴³ Funkcia rbind() spája vektory, matice alebo data framy po riadkoch na základe spoločného stĺpca, teda súbory, ktoré majú rovnakú štruktúru a sú bezomeny (bez straty) napojené na seba.

v RStudiu postupne po menších častiach a následne spojili⁴⁴. Dataset ženských biografí bol spracovaný naraz, keďže bol podstatne menší ako mužský, len 31,8 MB. Rovnakým procesom anotácie prešiel aj slovenský dataset, v tomto prípade bol však použitý jazykový model predtrénovaný na slovenčinu (slovak-snk-ud-2.3-181115.udpipe) a oba datasety boli anotované celé naraz.

6.4.3 UDPipe

K spracovaniu a následne k ďalšej práci s dátami sme v RStudiu použili knižnicu UDPipe, ktorá je voľne prístupná na stiahnutie na internete na portáli GitHub⁴⁵. UDPipe pôvodne vznikol na akademickej pôde na Katedre matematiky a fyziky na Karlovej Univerzite v Prahe. Pomocou tejto knižnice je možné natrénovať si vlastný jazykový model, alebo využiť už existujúce modely na prácu s textom v prirodzenom jazyku. Takto môže dataset v rámci anotácie prejsť nasledovnými procesmi: tokenizácia – vytvorenie tokenov, tagovanie – priradenie tagov, resp. označení, lematizácia – prevod slova do jeho základného tvaru, a nakoniec dataset prejde syntaktickou analýzou.⁴⁶

Vyššie spomíname anotovanie obsahu datasetu. Keď hovoríme o anotácii textu znamená to vytvorenie kategórií pre každé jedno slovo, vrátane interpunkčných znamienok, ktoré umožňujú ďalšiu prácu s textom. Pre každé jedno slovo vo vete sú určené kategórie, ktoré ho jasne identifikujú a zároveň určia, aký je to slovný druh (tagovanie, tokenizácia), prevedú na základný slovníkový tvar (lematizácia) a určia jeho postavenie vo vete v rámci ostatných vetných členov (*Dependency Parsing* ako syntaktická analýza, ktorej výsledkom sú *Parts of Speech*⁴⁷ označenia zodpovedajúce úlohe slova vo vete). Takto spracovaný text má už podobu tabuľky (.csv) s určenými vlastnosťami každého slova vo vete a v celom texte. Ukážku výsledného anotovaného datasetu uvádzame nižšie.

⁴⁴ Veľkosť sa dataset blížil hornej hranici 2 GB, ktorú dokáže bežný počítač spracovať naraz, hlavne čo sa týka veľkosti operačnej pamäte, s ktorou pracuje softvér R. Pri použití knižnice UDPipe tento súbor narastie ešte niekoľkonásobne kvôli pridaniu kategórií, preto bolo potrebné dataset spracovať po častiach, keďže nato už operačná pamäť kapacitne nestačila. Ak by sme chceli dáta spracovať naraz, bola by potrebná operačná pamäť väčšia ako je v bežnom počítači (8GB operačná pamäť).

⁴⁵ <https://github.com/bnosac/udpipe>

⁴⁶ Podrobnejšie informácie a celkovo dokumentáciu je možné nájsť na oficiálnej stránke: <http://ufal.mff.cuni.cz/udpipe>

⁴⁷ Presný zoznam Parts of Speech kategórií: <https://universaldependencies.org/u/pos/index.html>

X	doc_id	paragraph_id	sentence_id	sentence	token_id	token	lemma	upos	xpos	feats	head_token_id	dep_rel	deps	misc
1	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	1	ci	ci	PUNCT	Z	N/A	3	punct	N/A	SpaceAfter=No
2	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	2	"	"	PUNCT	Z	N/A	3	punct	N/A	SpaceAfter=No
3	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	3	Zacberal	zacberal	VERB	Vldscm=	Animacy=Anim Aspect=Perf Gender=Masc Number=Sing P...	0	root	N/A	N/A
4	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	4	sa	sa	PRON	R	PronType=Prs Reflex=Yes	3	explprv	N/A	N/A
5	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	5	jadrovou	jadrovou	ADJ	AAfstx	Case=ins Degree=Pos Gender=Fem Number=Sing	6	amod	N/A	N/A
6	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	6	fyzikou	fyzika	NOUN	SSf7	Case=ins Gender=Fem Number=Sing	3	obl	N/A	N/A
7	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	7	a	a	CCONJ	O	N/A	9	cc	N/A	N/A
8	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	8	kvantovou	kvantovuj	ADJ	AAfstx	Case=ins Degree=Pos Gender=Fem Number=Sing	9	amod	N/A	N/A
9	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	9	mechanikou	mechanika	NOUN	SSf7	Case=ins Gender=Fem Number=Sing	3	conj	N/A	SpaceAfter=No
10	1	1	1	"Zacberal sa jadrovou fyzikou a kvantov...	10	.	.	PUNCT	Z	N/A	3	punct	N/A	N/A
11	1	2	2	Spolu s B. R. Mottelsonom vyvinuli teóriu...	1	Spolu	spolu	ADV	Dx	Degree=Pos	7	case	N/A	N/A
12	1	2	2	Spolu s B. R. Mottelsonom vyvinuli teóriu...	2	s	s	ADP	Eu7	AdpType=Prep Case=ins	1	fixed	N/A	N/A
13	1	2	2	Spolu s B. R. Mottelsonom vyvinuli teóriu...	3	B	b	X	W	Abbr=Yes	7	rmod	N/A	SpaceAfter=No
14	1	2	2	Spolu s B. R. Mottelsonom vyvinuli teóriu...	4	.	.	PUNCT	Z	N/A	3	punct	N/A	N/A
15	1	2	2	Spolu s B. R. Mottelsonom vyvinuli teóriu...	5	R	r	X	W	Abbr=Yes	7	rmod	N/A	SpaceAfter=No

Obrázok č. 1 Náhľad výslednej podoby spracovaného datasetu pomocou jazykového modelu knižnice UDPipe

6.5 Použité metódy analýzy

Celá analýza prebehla za použitia voľne dostupného softvéru RStudio, knižnice UDPipe a doplnkových knižníc na vizualizáciu výsledkov lattice, igrph, ggraph a ggplot2, ktoré sú všetky voľne dostupné.

V rámci výskumu sme využili nasledujúce metódy, ktoré boli súčasťou balíka UDPipe⁴⁸:

- **Frekvenčná analýza:** skúma počty, resp. frekvencie výskytu jednotlivých slov alebo typov slov v datasete. S pomocou UDPipe je možné vypočítať frekvencie pre jednotlivé slovné druhy osve (podstatné mená, prídavné mená, slovesá a pod.), čo prináša zrozumiteľnejšie výsledky ako výpočet frekvencií pre všetky slová v datasete, keďže je možné si už priamo vybrať, ktorý z typov slov je pre nás relevantný (najčastejšie podstatné mená, prídavné mená, prípadne slovesá). Nie je potrebná zdĺhavá úprava výsledkov (Basic Analytical Use Cases I, 2019). V rámci výsledkov frekvenčnej analýzy je možné filtrovať medzi slovami, prípadne vizualizovať výsledky na grafe.

Obrázok č. 2 Ukážka filtrovania v rámci frekvenčnej analýzy

- **Co-occurrence alebo spoluvýskyt slov:** nám umožňuje vidieť, ako sú slová použité buď celkovo vo vete alebo jedno za druhým. Napríklad nám zodpovedá otázku, ktoré

⁴⁸ Presný postup, resp. príkazy pre jednotlivé úkony frekvenčnej analýzy, spoluvýskytu slov a ich vizualizácie: <https://bnosac.github.io/udpipe/docs/doc5.html> a <https://bnosac.github.io/udpipe/docs/doc7.html>

slová za sebou najčastejšie nasledujú, prípadne sa len nachádzajú v tej istej vete s rozdielom niekoľkých slov medzi nimi. UDPipe je schopný určiť postavenie slova vo vete na základe syntaktickej analýzy (*Dependency Parsing*), a preto umožňuje vypočítať a vizualizovať na grafe tieto výskyty slov (pomocou *Parts of Speech* označení) (Basic Analytical Use Cases I, 2019).

7 Čiastkové výsledky výskumu

Vzhľadom na to, že česká a slovenská Wikipédia zostávajú zatiaľ do značnej miery neprebádaným teritóriom, na začiatku si ukážeme základné ukazovatele pre českú aj slovenskú verziu, ktoré nám poskytnú určitý prehľad o aktuálnej situácii.

7.1 Počet článkov a počet slov v datasete

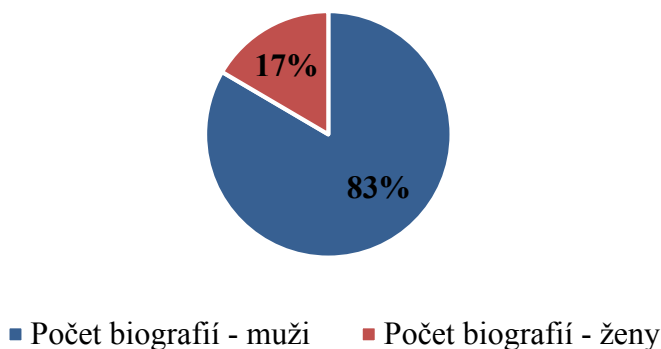
7.1.1 Česká verzia Wikipédie

Z českej Wikipédie sme stiahli 110 866 biografických článkov o mužoch a o ženách spoločne. Biografické články tvoria približne 12 % z celkového počtu článkov na českej Wikipédii⁴⁹. Pri bližšom pohľade na články o mužoch a o ženách začínajú byť viditeľné prvé rozdiely (viď Tabuľka č. 2 nižšie).

Počet biografíí - muži	92 517
Počet biografíí - ženy	18 349
Spolu	110 866

Tabuľka č. 3 Počet biografických článkov na českej Wikipédii

Počet článkov v ženskom datasete je značne menší ako v mužskom, celkovo články o ženách tvoria len 17 % z celkového počtu biografických článkov na českej Wikipédii (viď Graf č. 1 nižšie).



Graf č. 2 Percentuálne zastúpenie mužských a ženských biografíí na českej Wikipédii

Ďalej sa pozrieme na celkové počty slov v jednotlivých datasetoch, ktoré naznačujú prepracovanosť (kvalitu) článkov. Aj v tomto prípade sa ukazujú rozdiely. Mužský dataset je,

⁴⁹ Vychádzajúc z približných celkových počtov článkov stiahnutých v rámci Wikimedia dump za mesiac apríl a rovnako aj z čísel z Wikimedia statistics za mesiac apríl, a to celkovo 909 912 článkov: <https://stats.wikimedia.org/v2/#/cs.wikipedia.org/content/pages-to-date/normal|2019-02-01~2019-05-01|~total>

prirôdzené, početnejší na slová. Ak sa však pozrieme na priemerný počet slov na článok, tak pri porovnaní so ženským datasetom je tento počet približne rovnaký, teda sa neukazujú priepastné rozdiely. Rovnako to platí aj pre kontrolnú hodnotu mediánu. Výsledky mužského datasetu sú zhrnuté v tabuľke nižšie.

Biografie muži cs	
Celkový počet článkov	92 517
Celkový počet slov	23 643 593
Priemerný počet slov na článok	255,56
Hodnota mediánu (počet slov)	1 146

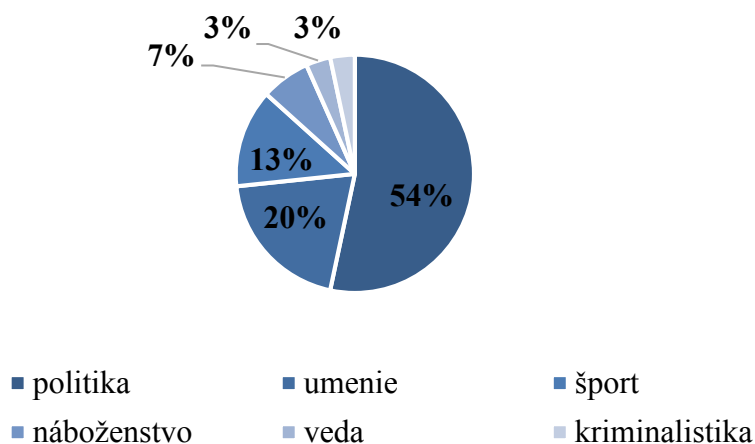
Tabuľka č. 4 Počet slov v českom mužskom datasete

Ženský dataset je na rozdiel od mužského menej početný na slová, no priemerný počet slov na článok je porovnateľný s mužským datasetom, a teda môžeme povedať, že síce sú články o ženách zastúpené na českej Wikipédii v menšom počte, kvalitou prepracovanosti, resp. priemerným počtom slov na článok a mediánom, sa mužskému datasetu značne približujú, a teda nevidíme priepastné rozdiely (255 slov vs. 246 slov). Výsledky pre ženský dataset sú zhrnuté v Tabuľke č. 4 nižšie.

Biografie ženy cs	
Celkový počet článkov	18 349
Celkový počet slov	4 514 615
Priemerný počet slov na článok	246,04
Hodnota mediánu (počet slov)	1 097

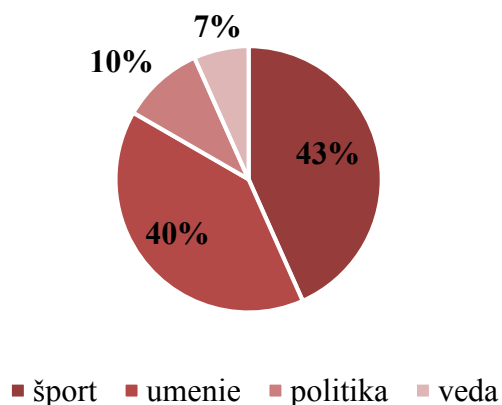
Tabuľka č. 5 Počet slov v českom ženskom datasete

Zaujímavé výsledky sa ukazujú aj pri pohľade na prvých 30 článkov najpočetnejších na slová, resp. ktoré môžeme označiť za najkvalitnejšie a najprepracovanejšie. V prípade českého mužského datasetu vidíme značnú prevahu osobností z oblasti politiky (16 článkov, 54 %), nasledované osobnosťami z oblasti umeleckej tvorby (6 článkov, 20 %), športu (4 články, 13 %), náboženstva (2 články, 7 %), vedy (1 článok, 3 %) a kriminalistiky (1 článok, 3 %). Toto rozdelenie ilustruje Graf č. 2 nižšie. Uvedené percentuálne zastúpenie článkov ukazuje určité preferencie tradične braných mužských tém, podobne ako kopírujú do značnej miery historický vývoj, ikeď sa skôr jedná o ilustráciu, keďže sa pozeráme iba na prvých 30 najpočetnejších článkov. Podrobný zoznam článkov uvádzame v tabuľke v Prílohe 1.



Graf č. 3 Tematické kategórie 30 najpočetnejších článkov - český mužský dataset

V českom ženskom datasete naopak prevládajú medzi 30 najpočetnejšími článkami, články o ženách v športe (13 článkov, 43 %) nasledované osobnosťami z oblasti umenia (12 článkov, 40 %) a malé zastúpenie majú aj osobnosti z oblasti politiky (3 články, 10 %) a vedkyne (2 články, 7 %). Kategórie náboženstvo a kriminalistika sa v rámci týchto článkov nevyskytli. Toto rozdelenie kategórií ilustruje Graf č. 3 nižšie.



Graf č. 4 Tematické kategórie 30 najpočetnejších článkov - český ženský dataset

Uvedené percentuálne zastúpenie článkov opäť ukazuje určité preferencie tradične braných mužských tém, resp. mužského pohľadu na ženy, ktorý sa ukazuje v dominancii tém ako šport, ikeď stále berieme do úvahy len ilustračný charakter týchto výsledkov. Práve prevaha športovkýň naznačuje príklon k mužsky zaujímavejším kategóriám. Zastúpenie ostatných kategórií do veľkej miery kopíruje historický vývoj, kedy sa ženy preslávili najviac v kategórií umenie, politika, resp. ako panovníčky, prípadne ako vedkyne. Podrobný zoznam článkov uvádzame v tabuľke v Prílohe 1.

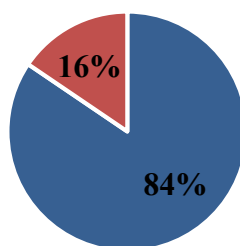
7.1.2 Slovenská verzia Wikipédie

Slovenská Wikipédia má oproti českej verzii celkovo menší počet článkov, v čase zberu dát (apríl 2019) sa na slovenskej Wikipédii nachádzalo spolu približne 434 259, biografické články z toho tvoria len 5,6 % z celkového počtu.⁵⁰ Z toho sú počty mužských a ženských biografíí nasledovné:

Počet biografíí - muži	20 707
Počet biografíí - ženy	3 803
Spolu	24 510

Tabuľka č. 6 Počet biografických článkov na slovenskej Wikipédii

Počet ženských biografíí je teda značne menší ako počet mužských biografíí, pričom ženské biografie tvoria z celkového počtu len 16 % a mužské zvyšných 84 %.



■ Počet biografíí - muži ■ Počet biografíí - ženy

Graf č. 5 Percentuálne zastúpenie mužských a ženských biografíí na slovenskej Wikipédii

Ďalej sa pozrieme na celkové počty slov v jednotlivých datasetoch, ktoré berieme ako jeden z indikátorov kvality článkov. Aj v tomto prípade sa ukazujú rozdiely medzi mužmi a ženami. Nielen, že mužský dataset obsahuje viac článkov, takisto je aj početnejší na slová ako ten ženský. I keď sa môže zdať, že ženské biografie nie sú tak kvalitne spracované ako mužské, v priemernom počte slov a na článok a v hodnote mediánu sú oba datasety porovnateľné. A teda síce na slovenskej Wikipédii prevažujú mužské biografie, vzhľadom na kvalitu spracovania (priemerný počet slov na článok) sú si značne podobné. Súhrn výsledkov mužského datasetu uvádzame v tabuľke nižšie.

⁵⁰ Vychádzali sme z čísel z Wikimedia dumpu a zo štatistických údajov za apríl 2019 na stránke: <https://stats.wikimedia.org/v2/#/sk.wikipedia.org/content/pages-to-date/normal|2019-02-01~2019-05-01|-total>

Biografie mužů sk	
Celkový počet biografii	20 707
Celkový počet slov	5 113 678
Priemerný počet slov na článok	246,954
Hodnota mediánu (počet slov)	894

Tabuľka č. 7 Počet slov v slovenskom mužskom datasete

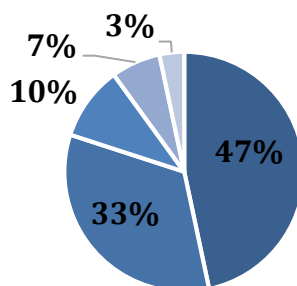
Výsledky ženského datasetu sú nasledovné:

Biografie ženy sk	
Celkový počet článkov	3 803
Celkový počet slov	936 300
Priemerný počet slov na článok	246,200
Hodnota mediánu (počet slov)	858

Tabuľka č. 8 Počet slov v slovenskom ženskom datasete

Výsledky ukazujú, že počet ženských a mužských biografii je nevyrovnaný, no pri lepšom pohľade zistíme, že čo do priemerného počtu slov a hodnoty mediánu, mužské aj ženské články sú si kvalitou spracovania značne podobné.

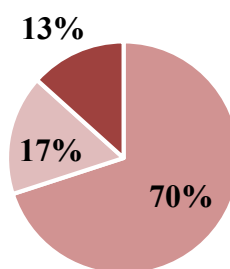
Ďalšie zaujímavé zistenia prináša pohľad na prvých 30 článkov najpočetnejších na slová. V slovenskom mužskom datasete dominujú osobnosti z oblasti politiky (14 článkov, 47 %), hneď za nimi nasledujú osobnosti z umeleckej oblasti (10 článkov, 33 %), športovci (3 články, 10 %), vedci (2 články, 7 %) a osobnosti z kategórie náboženstvo (1 článok, 3 %). Výsledky ilustruje graf nižšie. Podobne ako v českom, aj v slovenskom mužskom datasete dominujú osobnosti z oblasti politiky, v tesnom závese s umelcami. Percentuálne rozdelenie kategórii môže naznačovať prevahu tém typicky zaujímavejších pre mužské publikum – politika (47 %), šport (10 %), veda (7 %), náboženstvo (3 %), v porovnaní s umením (33 %). Tieto výsledky sa však dajú interpretovať aj s ohľadom na históriu, kedy najznámejšie a najvýznamnejšie osobnosti boli práve vladári a politici, cirkevní hodnostári, prípadne umelci, neskoršie aj vedci a športovci.



■ politika ■ umenie ■ šport ■ veda ■ náboženstvo

Graf č. 6 Tematické kategórie 30 najpočetnejších článkov - slovenský mužský dataset

Na druhej strane, ženský dataset už tak české výsledky nekopíruje. V slovenskom ženskom datasete, pri pohľade na prvých 30 článkov, dominujú so 70 % umelkyne (21 článkov), nasledované ženami z oblasti politiky (5 článkov, 17 %) a nakoniec športovkyne (4 články, 13 %). Výsledky dopĺňa graf nižšie. Podrobné výsledky pre oba datasety uvádzame v tabuľke v Prílohe 1.



■ umenie ■ politika ■ šport

Graf č. 7 Tematické kategórie 30 najpočetnejších článkov - slovenský ženský dataset

Uvedené percentuálne zastúpenie článkov neudáva taký jasný signál preferencie mužských tém, resp. mužského pohľadu na ženy, najviac článkov je z kategórie umenie, čo je tradične podľa výskumov v centre pozornosti žien - editoriek, ostatné dve kategórie sú zastúpené len minimálne (5 a 4 články). Tento stav takisto môže reflektovať aj historické prerozdelenie síl v spoločnosti, kedy sa ženy tradične presadili skôr ako umelkyne, známe panovníčky a neskôr aj športovkyne. Aj pri týchto výsledkoch stále berieme do úvahy ich ilustračný charakter, keďže sa pozeráme len na prvých 30 výsledkov.

7.1.3 Vyhodnotenie hypotéz H1 a H2 pre českú aj slovenskú Wikipédiu

V tejto časti zhrnieme výsledky za predchádzajúcu časť vzhľadom na stanovené hypotézy H1 a H2.

H1: v počte článkov bude v oboch jazykoch prevládať mužský dataset.

Dáta ukazujú, že počet biografických článkov o mužoch a o ženách na českej Wikipédii je rozdielny a vychádza v neprospech žien. Výsledky takisto ukazujú, že z celkového počtu biografíí len 17 % (18 349) sú biografie o ženách a zvyšných 83 % (92 517) tvoria biografie o mužoch. Slovenské dáta takisto ukazujú nižší počet biografických článkov o ženách, a to len 3 808 článkov (16 % celkovo) v porovnaní s mužskými biografickými článkami, ktorých počet je 20 707, čo predstavuje až 84 % z celkového počtu biografických článkov na slovenskej Wikipédii. **Tieto výsledky potvrdzujú pravdivosť hypotézy H1 pre obe jazykové verzie.**

H2: články v mužskom datasete budú kvalitnejšie spracované, ako články v ženskom datasete vzhľadom na signifikantne vyšší priemerný počet slov na článok a vyššiu hodnotu mediánu, a to v oboch jazykoch.

Dáta ďalej ukazujú rozdiely v počtoch slov v jednotlivých datasetoch. Český mužský dataset sa ukazuje byť početnejší na slová (23 643 593) ako ženský (4 514 615), čo v zásade primerane odpovedá veľkému rozdielu v celkovom počte článkov v jednotlivých datasetoch. Naopak, ak sa zameriame priamo na priemerný počet slov na článok a hodnotu mediánu, oba datasety, mužský aj ženský, ukazujú porovnateľné hodnoty. A teda priemerný počet slov na 1 článok v českom ženskom datasete je 246,04 slova, v mužskom 255,56 (rozdiel približne 10 slov, resp. 4 %). Hodnota mediánu v českom ženskom datasete je 1 097 slov a v mužskom 1 146 (rozdiel 49 slov, resp. 4,49 %).

Slovenské dáta ukazujú podobné výsledky ako tie české. Slovenský ženský dataset je výrazne menej početný na slová (936 300) ako mužský (5 113 678), čo opäť zodpovedá veľkému rozdielu v počte článkov. Avšak priemerný počet slov na článok (246,2 slov ženy vs. 246,954 slov muži) a medián (858 slov ženy vs. 894 slov muži) ukazujú len malé rozdiely medzi datasetmi. Rozdiely v počte slov na článok nepredstavuje ani jedno slovo a rozdiely v hodnote mediánu sú len 36 slov, resp. 4,19 %. Odhliadnuc od celkových počtov slov v jednotlivých datasetoch, vidíme, že rozdiely v ženskom a mužskom datasete v prípade češtiny aj slovenčiny sú len malé, a preto presvedčivo nedokazujú nižšiu kvalitu spracovania biografických článkov o ženách.

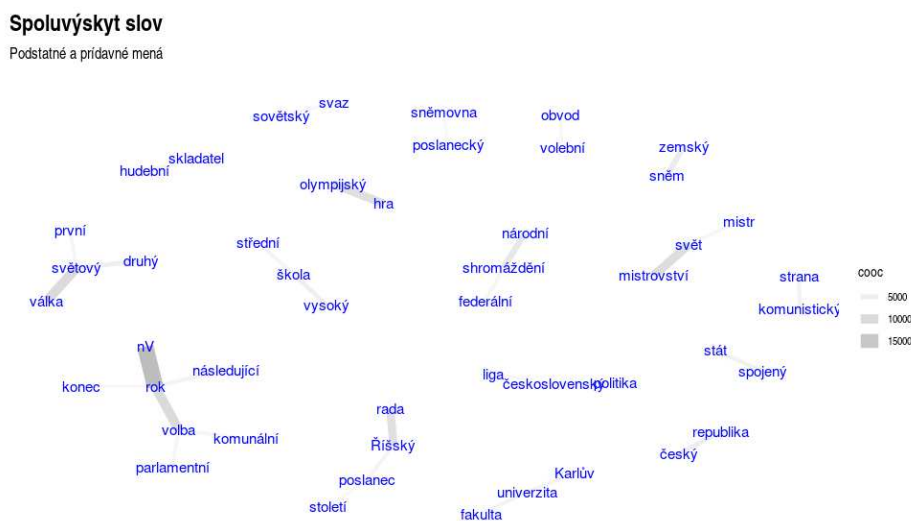
Tieto výsledky vyvracajú náš predpoklad z hypotézy H2, keďže ani české ani slovenské výsledky neukazujú signifikantné rozdiely v priemernom počte slov, pri ktorých by sme mohli konštatovať nižšiu kvalitu spracovania článkov v ženskom datasete, ako ju definoval J. E.

Blumenstock (2008). **Výsledky výskumu teda v prípade oboch jazykov nepotvrdili pravdivosť nami stanovenej hypotézy H2.**

7.2 Frekvencie výskytu slov

Predtým, ako sa špecificky zameriame na výskyt jednotlivých slov zo zvolených kategórii, si pre lepší prehľad ukážeme vizualizáciu, ktorá zachytáva 30 podstatných a 30 prídavných mien⁵¹, ktoré sa najčastejšie spolu vyskytujú (*Co-occurrence*), a ktorá nám poskytne náhľad do najpoužívanejších slovných spojení v českom aj slovenskom datase a celkovo poslúži ako ukážka najpodstatnejších tematických celkov v rámci ženského a mužského datasetu.

7.2.1 Česká verzia Wikipédie



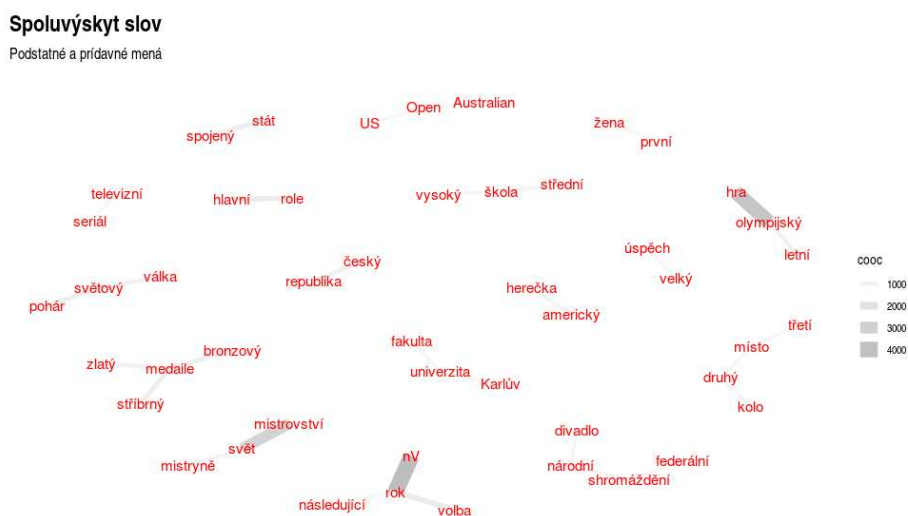
Graf č. 8 Spoluvýskyt slov v českom mužskom datase

V českom mužskom datase vidíme, že väčšina najčastejších spojení sa týka **politiky**, **vzdelávania** (prirodzený odkaz na dosiahnuté vzdelanie), **športu**, **geografie**, s výnimkou spojenia „hudební“ a „skladatel“ (**umenie**), a , ktorý sa vytvoril okolo podstatného mena „rok“, ktorého výskyt je v datase prirodzene vysoký, vzhľadom na povahu článkov (biografie). Najčastejšie (15 000) sa spolu vyskytujú zaužívané slovné spojenia a označenia ako „světová válka“, „mistrovství světa“ a „olympijské hry“⁵². Ostatné slovné spojenia majú relatívne nižšiu frekvenciu výskytu, ktorá sa pohybuje od 10 000 po 5 000 (vid’ Graf č. 7 vyššie). Spojenie „nV“

⁵¹ Zvolili sme si počet 30, keďže predstavuje kompromis, kedy je vizualizácia ešte dobre čitateľná, a zároveň má výpovednú hodnotu. Vizualizované sú tie najčastejšie kombinácie podstatných a prídavných mien, ktoré v texte nasledujú priamo za sebou (skipgram=0).

⁵² V rámci prepočtov spoluvýskytu sa berú do úvahy základné lematizované tvary slov, preto sú slová uvedené v neurčitku.

a „rok“ v tomto prípade predstavuje chybu pri automatickom spracovávaní textu; n predstavuje značku nového riadku⁵³.



Graf č. 9 Spoluvýskyt slov v českém ženském datase

Ženský dataset ukazuje oproti mužskému väčší dôraz na **šport**, ďalej na **umenie**, predovšetkým film, televíznu tvorbu a divadlo. Viditeľne sú v datase zdôrazňované aj športové a iné úspechy, resp. úspech všeobecne („velký úspech“, „hlavní role“). Ďalej nájdeme tematické okruhy týkajúce sa **vzdelávania** (prirodzený odkaz na dosiahnuté vzdelanie) a v malej miere aj **politiku a geografiu**. Takisto vidíme klaster, ktorý sa vytvoril okolo podstatného mena „rok“, ktorého výskyt je biografických článkoch prirodzene vysoký. Z genderového pohľadu je zaujímavý výskyt spojenia „první“ a „žena“. Toto slovné spojenie je jedným z kritérií Finkbeinerovej testu⁵⁴, ktoré podľa autorky testu (Curtis, 2013) naznačuje zaujatosť voči ženám (viď kapitola 5 Stručný exkurz do genderovej lingvistiky). V mužskom datase nenájdeme žiadnu priamu referenciu na pohlavie. V prípade ženského datase sa frekvencie spoluvýskytu pohybujú v rozmedzí 1 000 až 4 000, a teda značne nižšie ako v mužskom datase. Najčastejšie (4 000) sa spolu vyskytujú zaužívané slovné spojenia a označenia ako „mistrovství světa“ a „olympijské hry“, podobne ako tomu je v mužskom datase. Ostatné slovné spojenia majú relatívne nižšiu frekvenciu výskytu, ktorá sa pohybuje od 3000 po 1 000 (viď Graf č. 8 vyššie).

Na vizualizácii si môžeme všimnúť, že aj ženské prídavné mená sú uvádzané v mužskom rode, nie je to chyba datase, český aj slovenský jazykový model, ktorý bol

⁵³ Táto chyba vznikla pri anotovaní textu s pomocou knižnice UDPipe, ktorá si dataset v rámci anotácie rozdeľuje na vety a slová a ďalej spracováva sama automaticky (čierna skrinka). Takéto typy chýb sme neodstraňovali, aby sme nenarušili štruktúru dokumentu, a tým pádom neznehodnotili všetky dáta.

⁵⁴ Podľa Finkbeinerovej testu by text nemal obsahovať slovné spojenie „prvá žena, ktorá“.

použitý, automaticky berú za lematizovaný (slovníkový) tvar slova nominatív singuláru mužského rodu. Podobne ako v mužskom datsete vidíme rovnakú chybu pri anotácii v prípade spojenia „nV“ a „rok“. Táto anotačná chyba sa nachádza tak v českom, ako aj slovenskom datsete, a spoločne s automatickým prevodom na mužský rod ju vidíme ako nedostatok predtrénovaných modelov.

7.2.2 Frekvencie výskytu vybraných kategórií slov

V tejto časti sa špeciálne zameriame na výskyt vybraných kategórií slov, a to konkrétne na slová, ktoré sa týkajú genderu (aj referencie na opačné pohlavie), sociálneho statusu a rodiny a nakoniec aj kariéry. Sledovali sme, koľko krát sa dané slovo nachádza v datsete (počty aj poradie) a frekvenciu uvádzanú v percentách. Poradie a frekvencia sa viažu na konkrétny typ slova (slovný druh, resp. medzi podstatnými menami alebo prídavnými alebo slovesami) nie na celý dataset. Takto sme získali výsledky, ktoré nám presnejšie ukazujú dôležitosť a používanosť slova v datsete, keďže sa špeciálne zameriavame len na plnovýznamové slová (podstatné, prídavné mená a slovesá). Úplný zoznam sledovaných slov uvádzame v Prílohe 2.

Kategória gender

Prvou sledovanou kategóriou je prirodzene gender, resp. slová vyjadrujúce pohlavie opisovanej osoby. Súčasný výskum naznačuje, že v rámci žurnalistických a iných typov online a offline textov je stále aktuálna tendencia (neprimerane) prízvukovať pri ženách ich pohlavie, resp. špecificky zdôrazňovať, že sa jedná o ženu (podrobnejší rozbor aktuálnej situácie nielen na Wikipédii uvádzame v kapitole 4 Ako sa hovorí o ženách).

V českom mužskom datsete sa najčastejšie vyskytovalo podstatné meno **muž** (5 923, 0,087 %) a v rámci celkového poradia sa ukázalo byť ako **175. najčastejšie podstatné meno**. Zvyšné sledované slová sa už v datsete nevyskytujú tak často a dosahujú početnosť od 2 000 nižšie.

GENDER (PODSTATNÉ MENÁ)			
	Poradie	Počet	%⁵⁵
muž	175	5 923	0,087
pán	532	2 158	0,032
chlapec	797	1 417	0,021
chlap	4 871	92	0,001
gentleman	7 369	45	0,001
Spolu %			0,141

Tabuľka č. 9 Frekvencie slov (českí muži - gender - podst. mená)

Český ženský dataset ukazuje značne odlišné výsledky. Podstatné meno **žena** sa v datasete vyskytuje 6 311 (0,509 %), čo z neho robí **11. najpoužívanejšie podstatné meno**. V porovnaní so slovom muž z mužského datasetu, ktorému patrí až 175. priečka, sa v datasete vyskytuje približne 6x častejšie ako jeho ekvivalent v mužskom datasete. Celkovo pri porovnaní s mužským datasetom sa na pohlavie dotyčnej osoby odkazuje v článkoch výrazne častejšie aj pri ostatných slovách vyjadrujúcich pohlavie dotyčnej osoby. Spoločne sa na pohlavie vzťahuje u žien 0,7 %⁵⁶ z celkového počtu podstatných mien v datasete, u mužov je toto číslo **6x menšie** (len 0,14 %). Sumár výsledkov uvádzame v Tabuľke č. 9.

GENDER (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
žena	11	6 311	0,509
ďívka	172	1 133	0,091
dáma	282	699	0,056
paní	417	484	0,039
slečna	1 765	80	0,006
lady	1 853	75	0,006
Spolu %			0,708

Tabuľka č. 10 Frekvencie slov (české ženy - gender - podst. mená)

Rovnako aj pri pohľade na prídavné mená sa slovo **ženský** ukazuje častejšie ako jeho ekvivalent v mužskom datasete; je to **27. najpoužívanejšie prídavné meno, v porovnaní so slovom mužský, ktorému patrí až 288 priečka**. Vyskytuje sa teda 9x častejšie ako prídavné meno mužský z druhého českého datasetu. Ak spočítame percentá jednotlivých prídavných mien, tak zistíme, že slová prídavné mená odkazujúce na gender sa vyskytujú **8x častejšie**

⁵⁵ Ako sme už uviedli, jedná sa o percentuálne vyjadrenie frekvencie výskytu konkrétneho slova v rámci vybraného slovného druhu (frequency percentage). Podobne ako aj poradie sa vzťahuje len na konkrétny slovný druh a nie celý dataset, napr. slovo muž je 175. najčastejšie podstatné meno v mužskom datasete a tomu zodpovedá 0,087 % z celkového počtu podstatných mien v datasete.

⁵⁶ Na prvý pohľad sa čísla môžu zdať malé, pracujeme však s naozaj veľkým počtom slov v jednotlivých datasetoch, preto sú tieto čísla prirodzene malé. Zároveň sa čísla, pre vyššiu výpovednú hodnotu, počítajú len pre konkrétny slovný druh a nie počet slov v celom datasete. V tom prípade by boli čísla prakticky zanedbateľné.

v **ženských článkoch** (0,564 %) ako ich ekvivalenty v mužskom datasete (0,064 %). Z prídavných mien sa najčastejšie v mužskom datasete (288. miesto) vyskytovalo slovo **mužský** (288, 0,054 %). Celkové výsledky sú zhrnuté v Tabuľke č. 10.

GENDER (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
mužský	288	1 767	0,054
chlapecký	1 199	352	0,011
Spolu %			0,064

Tabuľka č. 11 Frekvencie slov (českí muži - gender - prid. mená)

GENDER (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
ženský	27	2 902	0,482
dívčí	191	490	0,081
Spolu %			0,564

Tabuľka č. 12 Frekvencie slov (české ženy - gender - prid. mená)

Kategória gender - referencia na opačné pohlavie

V rámci tejto kategórie sa bližšie pozrieme na spôsob, akým sa prezentujú muži a ženy v závislosti na opačnom pohlaví. V minulosti bola žena často priamo spájaná so svojim mužským príbuzným, prípadne manželom, takže pri pohľade na výsledky je potrebné zohľadniť aj historické fakty. Súčasné výskumy (Wagner, et al., 2015) však naďalej naznačujú, predovšetkým na Wikipédii, že aj dnes v prípade žien pretrváva častejšie odkazovanie (slovné alebo v podobe prelinkovania článkov) nielen na svoje, ale aj na opačné pohlavie (na muža, manžela, otca, brata, syna a podobne). V prípade českého mužského datasetu sú výsledky nasledovné:

REFERENCIE NA OPAČNÉ POHLAVIE (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
dcera	79	12 094	0,178
manželka	89	11 160	0,164
matka	107	9 218	0,135
žena	120	8 398	0,123
sestra	251	4 435	0,065
milenka	1 935	418	0,006
partnerka	1 990	401	0,006
snoubenka	3 992	129	0,002
exmanželka	11 745	20	0,000
Spolu %			0,697

Tabuľka č. 13 Frekvencie slov (českí muži – opačné pohlavie - podst. mená)

V prípade ženského datasetu sú si výsledky až prekvapivo podobné, hlavne čo sa týka obsadenia prvých priečok. V ženskom datasete sú však vybrané slová používané celkovo častejšie, prvé dve slová **syn** a **manžel** dokonca patria medzi **10 najčastejších slov**, naznačujúc tradičné spájanie ženy predovšetkým s rolou matky a manželky. V mužskom datasete túto tendenciu nevidíme, slová sa umiestňujú až v rámci prvej stovky. Najčastejšie sa v prípade žien odkazuje na ich **synov** (3x častejšie ako na dcéry v mužskom datasete) a **manželov** (3x častejšie ako na manželku v mužskom datasete), prípadne **otcov** (3x častejšie ako na matky v mužskom datasete), čo je v porovnaní s mužským datasetom veľký nepomer. Aj menej frekventované slová ako brat, muž, partner, milenec majú v ženskom datasete vyššie zastúpenie ako v mužskom, kde sú tie percentá skoro zanedbateľné. V prípade oboch datasetov sa najčastejšie odkazuje na opačné pohlavie v kontexte tradičných rodinných rolí, u žien sa celkovo spomína opačné pohlavie častejšie ako u mužov, **v priemere až 3x častejšie**. Naznačuje to väčšie spájanie žien a ich úspechov s ich mužskými príbuznými. Celkový prehľad výsledkov uvádzame v Tabuľke č. 13 nižšie.

REFERENCIE NA OPAČNÉ POHLAVIE (POSTATNÉ MENÁ)			
	Poradie	Počet	%
syn	6	7 003	0,565
manžel	8	6 591	0,531
otec	18	5 566	0,449
bratr	52	3 663	0,295
muž	124	1 527	0,123
partner	335	596	0,048
milenc	701	274	0,022
snoubenec	2135	61	0,005
exmanžel	25000	1	0,000
Spolu %			2,038

Tabuľka č. 14 Frekvencie slov (české ženy – opačné pohlavie - podst. mená)

Katégória sociálny status a rodina

Ďalšou zo sledovaných kategórií je sociálny status a rodina, čo je doména, ktorá je tradične (či stereotypne) považovaná za ženskú. Český mužský dataset ukazuje nasledovné výsledky:

RODINA A SOC. STATUS (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
syn	14	27 081	0,398
otec	35	21 620	0,317
rodina	42	20 536	0,301
dítě	65	13 849	0,203
dcera	79	12 094	0,178
vztah	157	6 562	0,096
rodič	229	4 750	0,070
manželství	300	3 825	0,056
manžel	452	2 540	0,037
poměr	759	1 517	0,022
rozvod	1 536	589	0,009
vdovec	4 717	97	0,001
novomanžel	5 567	73	0,001
otčím	9 985	26	0,000
snoubenec	11 654	20	0,000
exmanžel	43 738	2	0,000
Spolu %			1,691

Tabuľka č. 15 Frekvencie slov (českí muži - rodina - podst. mená)

V rámci mužského datasetu patria prvé priečky slov odkazujúcim na **rodinu** (syn, otec, rodina, dítě, dcera), čo je z časti prirodzené, keďže sa v rámci biografických článkov takéto

informácie bežne vyskytujú. Zaujímavé sú však aj zmienky o **manželstve** (3 825, 0,056 %, 300. priečka) a označenie muža ako **manžela** (2 540, 0,037 %, 452. priečka). Zvyšok sledovaných slov – odkazujúcich práve na sociálny stav - sa v datasete vyskytujú pomerne zriedkavo.

Naopak pri pohľade na výsledky ženského datasetu, sa u žien výrazne častejšie akcentuje ich rola **matky**, predovšetkým v rámci podstatných mien, kde sa na prvých miestach, podobne ako u mužov, vyskytujú práve slová súvisiace s rodinou a deťmi, avšak podstatne častejšie ako pri mužoch (viď. Tabuľka č. 17). Podstatné meno „dcera“ sa v ženskom datasete vyskytuje 8 664-krát, čo je 4x častejšie ako v mužskom, „dítě“ zas až 16x častejšie, podstatné meno syn raz tak často, matka 1,5x častejšie ako otec v mužskom datasete, rovnako ako aj slovo rodina.

Ďalej sa kladie oveľa väčší dôraz na **status manželky**, ktoré je 30. najčastejšie podstatné meno (v porovnaní s manželom, ktorý je až 452. v poradí, a teda je manželka v ženskom datasete 9,8x častejšie, prípadne so slovom manželstvo, ktoré je 300. v poradí u mužov, u žien sa však vyskytuje 4,5x častejšie).

Takisto je zaujímavý aj pohľad na slovo **rozvod**, ktoré sa u žien vyskytuje 5,5x menej často ako u mužov, u žien je teda téma rozvodu prominentnejšia, podobne ako aj spoločenský status **vdova**, ktorá sa oproti vdovcovi spomína 24x častejšie. Samozrejme, môžeme argumentovať historickým vývojom postavenia ženy v spoločnosti, wikipedické články sú však písané v dnešnej dobe a ideálne by nemuseli kopírovať predstavy o ženách z dôb minulých.

Zvyšné slová, označujúce rodinný status, sú už menej prominentné v oboch datasetoch – snúbenica/snúbenec, exmanželka/exmanžel, otčim/macocha a podobne, čo sú role, ktoré sa prirodzene vyskytujú menej často, prípadne nie sú až tak zaujímavé pre potreby textu. Celkovo sa podstatné mená z kategórie gender vyskytujú **v ženskom datasete 2,25x častejšie, ako v tom mužskom**. Sumár výsledkov uvádzame nižšie, v tabuľke č. 15.

RODINA A SOC. STATUS (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
dcera	5	8 664	0,698
dítě	6	7 003	0,565
syn	10	6 485	0,523
matka	12	5 776	0,466
rodina	13	5 657	0,456
manželka	30	4 490	0,362
manželství	61	3 096	0,250
vztah	73	2 415	0,195
rodič	83	2 107	0,170
rozvod	323	606	0,049
poměr	381	525	0,042
vdova	653	296	0,024
snoubenka	2 814	39	0,003
macecha	4 718	18	0,001
exmanželka	7 595	8	0,001
mladomanželka	7 595	8	0,001
Spolu %			3,804

Tabuľka č. 16 Frekvencie slov (české ženy - rodina - podst. mená)

V prípade prídavných mien v mužskom datasete sa na rozdiel od podstatných mien skôr akcentuje sociálny status spojený s uzavretím či neuzavretím manželstva – prídavné meno **ženatý** dosiahlo 145. priečku (3 452, 0,0105 %), nasledované prídavným menom **rodinný** (169. priečka, 3 011, 0,091 %) a **slobodný** (176. priečka, 2 914, 0,088 %), resp. najčastejších 5 prídavných mien je mix medzi statusom rodinným a manželským (viď Tabuľka č. 16, nižšie).

RODINA A SOC. STATUS (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
ženatý	142	3 452	0,105
rodinný	169	3 011	0,091
svobodný	176	2 914	0,088
manželský	589	848	0,026
bezďetný	707	685	0,021
rozvedený	1 139	374	0,011
ovdovčľý	2 281	150	0,005
zasnoubený	2 746	115	0,003
otcovský	4 305	58	0,002
rodičovský	5 454	41	0,001
mimomanželský	6 516	31	0,001
nemanželský	21 191	5	0,000
Spolu %			0,354

Tabuľka č. 17 Frekvencie slov (českí muži - rodina - prid. mená)

Prídavné mená v ženskom datasete ukazujú zaujímavú tendenciu, v porovnaní s mužským, kde pri ženách nájdeme relatívne vysokú početnosť pri prvých 9 sledovaných slovách, ktoré vyjadrujú **manželský a rodinný stav** (viď Tabuľka č. 17 nižšie).

V mužskom datasete dosahujú porovnateľne vysokú početnosť výskytu len slová **ženatý, rodinný a slobodný**, a teda sa na rozdiel u žien, tak neakcentuje fakt, že sú bezďetní, otcovský, prípadne zasnúbení, rozvedený alebo vdovci. U žien sa zdôrazňujú všetky relatívne často sa v spoločnosti vyskytujúce rodinné stavy, u mužov len tie najtypickejšie (ženatý – slobodný). Čo sa týka prídavných mien, tak u žien je výskyt týchto slov **1,7x častejší**.

RODINA A SOC. STATUS (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
provdaný	91	990	0,165
rodinný	112	779	0,130
svobodný	180	514	0,085
bezďetný	252	362	0,060
mateľský	287	318	0,053
rozvedený	343	265	0,044
zasnoubený	398	225	0,037
ovdovĕlý	527	163	0,027
manželský	1 840	35	0,006
rodičovský	1 840	35	0,006
mimomanželský	2 419	24	0,004
nemanželský	33 412	1	0,000
Spolu %			0,618

Tabuľka č. 18 Frekvencie slov (české ženy - rodina - prid. mená)

V ďalšom kroku sa pozrieme na výskyt sloviac v rámci tejto kategórie. Výsledky z českého mužského datasetu sú nasledovné:

RODINA A SOC. STATUS (SLOVESÁ)			
	Poradie	Počet	%
oženit	34	8 812	0,441
rozvést	374	892	0,045
rozejít	450	709	0,036
zasnoubit	975	217	0,011
ovdovĕt	1 476	104	0,005
Spolu %			0,538

Tabuľka č. 19 Frekvencie slov (českí muži - rodina - slovesá)

Výsledky z českého ženského datasetu sú zobrazené v tabuľke nižšie:

RODINA A SOC. STATUS (SLOVESÁ)			
	Poradie	Počet	%
provdát	19	3 155	0,817
rozvést	113	648	0,168
rozejít	219	316	0,082
zasnoubit	252	268	0,069
ovdovĕt	338	185	0,048
Spolu %			1,184

Tabuľka č. 20 Frekvencie slov (české ženy - rodina - slovesá)

Podobné výsledky ako predchádzajúce slovné druhy ukazuje aj porovnanie najčastejších sloviac z kategórie. U mužského datasetu sa výrazne často v textoch vyskytuje len

sloveso **oženit' sa**, ktoré dosiahlo 34. miesto, sloveso **provdat se** dosiahlo 19. miesto, a teda sa v ženskom datasete v rámci sloviac vyskytuje **1,8x častejšie**. Zvyšné sledované slovesá z mužského datasetu sa umiestňujú od 374. po 1475. miesto, resp. v % dosahujú výskyt od 0,045 – 0,005 %. U žien sa pohybujeme od 59. do 338. miesta a teda v rozmedzí výrazne vyššom (0,313 – 0,048 %). Celkovo je u ženského datasetu **výskyt týchto sloviac 2x tak častý**.

Kategória kariéra

V rámci tejto kategórie sa budeme venovať spôsobu, akým sa v textoch uvádzajú úspechy a prezentuje kariéra žien a mužov a či je možné z nasledujúcich dát vyvodit' určité závery. V českom mužskom datasete sú výsledky nasledovné:

KARIÉRA (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
práce	25	23 388	0,343
titul	45	19 810	0,291
kariéra	48	18 660	0,274
studium	63	15 247	0,224
cena	67	13 569	0,199
úspěch	116	8 727	0,128
vítězství	182	5 825	0,086
vítěz	235	4 653	0,068
vedoucí	239	4 625	0,068
vzdělání	287	3 932	0,058
ocenění	346	3 434	0,050
profese	596	1 929	0,028
talent	704	1 652	0,024
šéf	720	1 615	0,024
odborník	756	1 521	0,022
uznání	783	1 454	0,021
nadání	1 363	690	0,010
profesionál	2 117	362	0,005
respekt	2 160	352	0,005
Spolu %			1,929

Tabuľka č. 21 Frekvencie slov (českí muži - kariéra - podst. mená)

Ako najčastejšie sa v mužskom datasete ukazujú slová **práca** (23 388, 25. miesto, 0,343 %), **titul** (19 810, 45. miesto, 0,291 %) a **kariéra** (18 660, 48. miesto, 0,274 %). Zvyšné sledované podstatné mená sa pohybujú od 63. miesta po 2 160., resp. od počtu výskytov 15 247 po 352. Veľmi podobné výsledky ukazuje aj český ženský dataset, na rozdiel od predchádzajúcich kategórii, v kategórii kariéra sú ich výsledky značne podobné, v prípade

podstatných mien sme dokonca zaznamenali častejší výskyt týchto slov, predovšetkým v rámci prvej trojky slov, ktoré sú **titul** (1,5x častejšie), **kariéra** (1,3x častejšie) a **cena** (1,76x častejšie). Naopak výrazne horšie výsledky ako v mužskom datasete dosahujú slová, ktoré v sebe nesú **odkaz na profesionalitu a odbornosť**, a to **šéfka** (12x menej často), **odborníčka** (5,5x menej často), **profesionálka** (2,5x menej často) a **vedúca** (1,65x menej často). V rámci tejto kategórie sa ukázali pri celkovom porovnaní opačné tendencie, ako sme si mysleli, v českom ženskom datasete sa podstatné mená s odkazom na kariéru **vyskytujú 1,2x častejšie ako u mužov**. Úplný zoznam slov pre ženský dataset uvádzame nižšie, v Tabuľke č. 21.

KARIÉRA (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
titul	16	5 586	0,450
kariéra	28	4 592	0,370
cena	34	4 360	0,351
práce	56	3 386	0,273
studium	70	2 524	0,203
úspěch	75	2 379	0,192
vítězka	121	1 557	0,126
ocenění	183	1 069	0,086
vítězství	186	1 051	0,085
vzdělání	268	751	0,061
vedoucí	391	504	0,041
talent	452	451	0,036
uznání	828	224	0,018
profese	1 040	171	0,014
nadání	1 454	109	0,009
odbornice	2 324	54	0,004
respekt	2 529	46	0,004
šéfka	4 486	19	0,002
profesionálka	4 521	19	0,002
Spolu %			2,326

Tabuľka č. 22 Frekvencie slov (české ženy - kariéra - podst. mená)

Výsledky pre prídavné mená a slovesá pre oba datasety dopadli porovnateľne, bez zaujímavejších rozdielov medzi mužským a ženským datasetom (viď Tabuľka č. 22, Tabuľka č. 23, Tabuľka č. 24 a Tabuľka č. 25). Teda v prípade, že sa hovorí o mužovi a o žene v súvislosti s kariérou a pracovnou oblasťou, nevidíme také výrazné rozdiely, ako pri rodine či genderi. Naopak, porovnanie ukazujú, že v ženskom datasete sa v prípade všetkých 3 sledovaných slovných druhov **vyskytujú mierne častejšie ako v mužskom**. Rozdiel

u podstatných mien je 0,397 % v prospech žien, u prídavných mien 0,110 % a sloviess 0,419 %.
 Výsledky pre jednotlivé časti sú sumarizované nižšie v tabuľkách.

KARIÉRA (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
známý	12	21 531	0,653
úspěšný	55	7 887	0,239
odborný	90	5 331	0,162
profesionální	98	4 704	0,143
slavný	116	4 204	0,127
vítězný	272	1 890	0,057
pracovní	293	1 756	0,053
uznávaný	359	1 511	0,046
nadaný	774	613	0,019
vzdělaný	805	754	0,017
oceňovaný	990	449	0,014
telentovaný	1 042	418	0,013
kariérní	1 691	232	0,007
respektovaný	2 139	163	0,005
Spolu %			1,556

Tabuľka č. 23 Frekvencie slov (českí muži - kariéra - prid. mená)

KARIÉRA (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
známý	13	4 137	0,688
úspěšný	54	1 662	0,276
profesionální	70	1 305	0,217
odborný	117	748	0,124
slavný	134	666	0,111
vítězný	232	396	0,066
pracovní	281	322	0,054
uznávaný	428	206	0,034
vzdělaný	559	151	0,025
kariérní	626	133	0,022
telentovaný	807	99	0,016
nadaný	877	88	0,015
oceňovaný	1 097	69	0,011
respektovaný	2 266	26	0,004
Spolu %			1,664

Tabuľka č. 24 Frekvencie slov (české ženy - kariéra - prid. mená)

KARIÉRA (SLOVESÁ)			
	Poradie	Počet	%
pracovať	7	26 400	1,322
vyhráť	23	11 590	0,580
zvíťaziť	79	4 546	0,228
uspieť	118	3 024	0,151
vzdelávať	635	442	0,022
uznávať	702	386	0,019
respektovať	844	270	0,014
Spolu %			2,336

Tabuľka č. 25 Frekvencie slov (českí muži - kariéra - slovesá)

KARIÉRA (SLOVESÁ)			
	Poradie	Počet	%
pracovať	9	4 953	1.283
vyhráť	14	3 582	0.928
zvíťaziť	53	1315	0.341
uspieť	120	602	0.156
vzdelávať	524	98	0.025
respektovať	854	45	0.012
uznávať	904	41	0.011
Spolu %			2,756

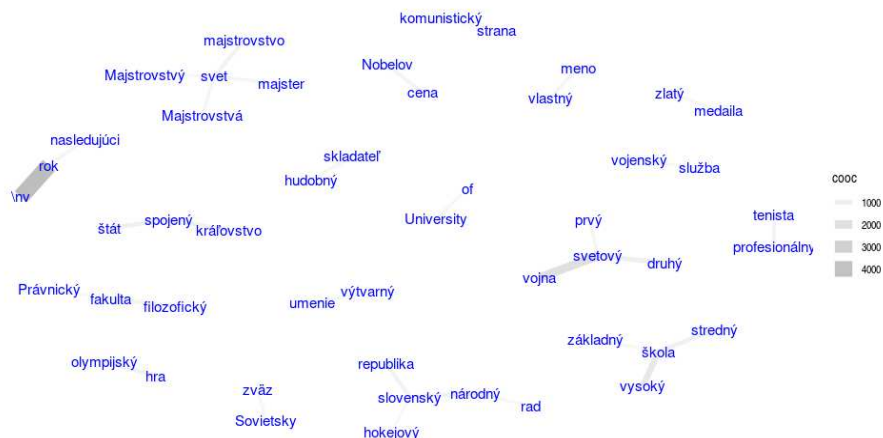
Tabuľka č. 26 Frekvencie slov (české ženy - kariéra - slovesá)

7.2.3 Slovenská verzia Wikipédie

V slovenskom mužskom datasete sa najčastejšie spolu vyskytujú nasledovné podstatné a prídavné mená:

Spoluvýskyt slov

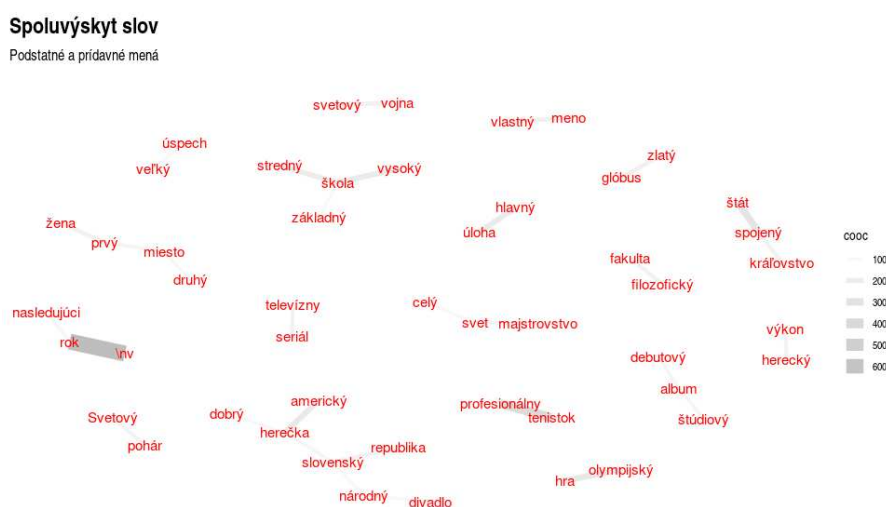
Podstatné a prídavné mená



Graf č. 10 Spoluvýskyt slov v slovenskom mužskom datasete

Pri pohľade na slovenský mužský dataset vidíme, že výsledky sú veľmi rôznorodé. Najvýraznejšie je zastúpená kategória **šport**, ktorej sa týkajú 4 klastre a čiastočne 5, ďalej tu nájdeme **vzdelávanie** (3 klastre), v menšom merítku je tu zastúpené **umenie**, **geografia**, **politika**, **vojenstvo** a nakoniec aj klaster okolo slova „rok“, podobne, ako je to vo všetkých predchádzajúcich výsledkoch. Na rozdiel od českých výsledkov, v slovenskom mužskom dataseťe nájdeme prevažne kategóriu šport, politika, ktorá prevažuje v českom dataseťe, je skôr okrajová. Vidíme takisto dôraz na dosiahnuté úspechy; v športe ale aj všeobecnejšie vo vede. Najčastejšie sa spolu opäť vyskytujú zaužívané slovné spojenia a názvy (4 000 výskytov) ako „svetová vojna“, o niečo menší výskyt zase „Nobelova cena“, „majstrovstvá sveta“ a podobne. Rozpätie, ktoré sa v rámci mužského datasetu dosahuje je od 4 000 po 1 000 spoločných výskytov.

Anotácia slovenského datsetu od pohľadu neprebehla úplne bezchybne, vidíme napríklad opakujúce sa slová (majstrovstvá – majstrovstvo – Majstrovstvý), chýbajúce písmenká (Nobelova), veľké, malé písmená a opäť raz opakujúci sa problém s predložkou V.



Graf č. 11 Spoluvýskyt slov v slovenskom ženskom dataseťe

Ženy sú v tomto prípade podobné tak českým ženským výsledkom ako aj slovenským mužským. V rámci datasetu sa najčastejšie vyskytujú spojenia, ktoré môžeme zaradiť do kategórie **šport**, podobne ako v českom dataseťe. Po športe nasleduje kategória **umenie**, spoločne so **vzdelávaním** (dosiahnutým vzdelaním), **geografiou** a klasterom okolo slova „rok“. S českým ženským dataseťom vidíme spriaznenosť predovšetkým s prevahou kategórií umenie a šport, podobnosť so slovenským mužským dataseťom vidíme zas v zdôrazňovaní dosiahnutých úspechov; v športe, umení. Podobné tendencie vidíme aj medzi ženským slovenským a českým dataseťom, kde sa okrem športových úspechov akcentuje aj úspech ako taký (napr. „dobrá herečka“, „veľký úspech“, „hlavná rola“). Podobne ako v českom ženskom

datasete, aj v slovenskom nájdeme spojenie „prvá žena“, a teda explicitný odkaz na gender, ktorého ekvivalent absentuje v oboch mužských datasetoch. Početnosť výskytu sa v prípade slovenského ženského datasetu pohybuje len v rozmedzí 600 – 100, pričom podrobné výsledky sú vizualizované vyššie, na Grafe č. 10.

Nakoniec ešte pripomenieme, že aj v ženskom datasete nájdeme chyby, ktoré nastali pri anotácií, a to napríklad pri prevode na základný tvar (tenistok namiesto tenistka), veľké a malé písmená a tradične problém s predložkou V.

7.2.4 Frekvencie výskytu vybraných kategórií slov

Kategória gender

V kategórii gender sú výsledky v slovenskom mužskom datasete nasledovné:

GENDER (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
muž	195	1 110	0,083
chlapec	625	358	0,027
pán	768	281	0,021
chlap	2 952	46	0,003
gentleman	8 968	9	0,001
chlapček	29 166	2	0,000
Spolu %			0,134

Tabuľka č. 27 Frekvencie slov (slovenskí muži - gender - podst. mená)

V porovnaní so slovenským ženským datasetom vidíme prvé rozdiely. U žien sa častejšie odkazuje na ich pohlavie, čo predovšetkým ilustruje porovnanie **žena vs. muž**. Podstatné meno žena sa v datasete objavuje ako **7. najčastejšie podstatné meno**, naopak označenie **muž** sa v mužskom datasete ukazuje len ako **197. najčastejšie**, a teda až **6x častejšie**. Podobne sú na tom aj **zvyšné označenia príslušnosti k genderu** uvádzané v Tabuľke č. 26 a 27. U žien sú výrazne častejšie – **až 5x častejšie**, ako ich zodpovedajúce ekvivalenty u mužov.

GENDER (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
žena	7	1 224	0,516
dievča	155	220	0,093
dáma	383	93	0,039
pani	575	61	0,026
slečna	1 617	18	0,008
lady	X	X	X
Spolu %			0,681

Tabuľka č. 28 Frekvencie slov (slovenské ženy - gender - podst. mená)

Pri pohľade na prídavné mená sú výsledky rovnaké. U žien sa prídavné meno **ženský** vyskytuje častejšie, ako **26. najčastejšie prídavné meno**, u mužov (mužský) je to naopak až 326. miesto, a teda sa vyskytuje až **10x častejšie**. Podobne ako v českom datasete sa aj v slovenskom nemiestne prízvukuje ženský gender, **približne 10x častejšie**. Sumár výsledkov pre mužský a ženský dataset uvádzame v Tabuľke č. 28 a 29 nižšie.

GENDER (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
mužský	326	285	0,043
chlapčenský	1 107	80	0,012
Spolu %			0,055

Tabuľka č. 29 Frekvencie slov (slovenskí muži - gender - príd. mená)

GENDER (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
ženský	26	497	0,428
dievčenský	125	12	0,103
Spolu %			0,531

Tabuľka č. 30 Frekvencie slov (slovenské ženy - gender - príd. mená)

Kategória gender (referencia na opačné pohlavie)

Výsledky z tejto kategórie sú sumarizované v Tabuľke č. 30 a Tabuľke č. 31. V rámci mužského datasetu sa najčastejšie odkazuje tradične na **manželku** (2 394, 69. miesto, 0,178 %), matku a dcéru, ostatní ženský členovia rodiny či blízki sa spomínajú sporadickejšie. V ženskom datasete však vidíme trochu iné tendencie, u žien sa najčastejšie z mužských blízkych či príbuzných spomína **otec**, ďalej **syn**, **brat** a **manžel**. Tento stav môže naznačovať, že sa žena často dáva (či dávala v minulosti) do **kontextu s významnejším mužským predstaviteľom**, ako dcéra významného otca, matka slávneho syna, sestra známejšieho brata

či jednoducho manželka (viď Tabuľka č. 30 a 31). Celkovo sa na opačný gender odkazuje v ženskom datase 2x častejšie ako v tom mužskom.

REFERENCIE NA OPAČNÉ POHLAVIE (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
manželka	69	2 394	0,178
matka	77	2 225	0,166
dcéra	88	2 044	0,152
žena	167	1 211	0,090
sestra	243	889	0,067
partnerka	1 320	142	0,011
milenka	1 614	112	0,008
snúbenica	4 752	23	0,002
exmanželka	9 727	8	0,001
Spolu %2,			0,674

Tabuľka č. 31 Frekvencie slov (slovenskí muži - opačný gender - podst. mená)

REFERENCIE NA OPAČNÉ POHLAVIE (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
otec	10	1 100	0,463
syn	25	907	0,382
brat	48	660	0,278
manžel	54	597	0,251
muž	123	289	0,122
partner	226	157	0,066
mileneč	969	35	0,015
snúbenec	1 193	27	0,011
exmanžel	21 812	1	0,000
Spolu %			1,589

Tabuľka č. 32 Frekvencie slov (slovenské ženy - opačný gender - podst. mená)

Katégoria rodina a sociálny status

Ak sa pozrieme bližšie na slová z katégorie rodina a sociálny status, tak v mužskom datase na popredných priečkach vidíme predovšetkým slová týkajúce sa najbližšej rodiny, predovšetkým **otec** (4 848, 23. miesto, 0,361 %) a **rodina** (4 389, 28. miesto, 0,327 %). V mužskom datase je **viditeľná hierarchia**, kde slová ohľadom rodiny sú na vrchole nasledované manželstvom a partnerstvom. Ženský datase je na tom podobne, s rovnakou postupnosťou slov od rodiny na vrchole až po manželstvo a partnerstvo.

RODINA A SOC. STATUS (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
otec	23	4 848	0,361
rodina	28	4 389	0,327
syn	41	3 568	0,266
dieťa	51	3 089	0,230
dcéra	88	2 044	0,152
vzťah	110	1 712	0,127
rodič	185	1 137	0,085
manželstvo	286	796	0,059
pomer	498	447	0,033
manžel	915	223	0,017
rozvod	1 166	163	0,012
vdovec	6 326	15	0,001
snúbenec	10 802	7	0,001
otčim	11 497	6	0,000
mladomanžel	48 114	1	0,000
Spolu %			1,671

Tabuľka č. 33 Frekvencie slov (slovenskí muži - rodina - podst. mená)

Vidíme však aj rozdiely. V prípade žien sa viac **akcentuje potomstvo** (viď. Tabuľka č. 33), a teda nepriamo aj jej rola ako matky. Vidíme ďalej rozdiel aj v pohľade na manžela a manželku. **Manžel** sa v mužskom datase ukazuje na **915. mieste**, u žien sa naopak **manželka** uvádza na **45. mieste, čo je až 18x častejšie**. U žien je rovnako aj častejší výskyt slov, ktoré zodpovedajú **najbežnejším označeniam sociálneho statusu**, akým sú teda manželka (18x častejšie), vdova (16x častejšie), ale aj snúbenica (11x častejšie), prípadne také udalosti ako rozvod (4,8x častejšie). Celkovo sa podstatné mená z kategórie rodina vyskytujú **v ženskom datase 2x častejšie ako v mužskom**.

RODINA A SOC. STATUS (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
dieťa	6	1 356	0,571
dcéra	8	1 224	0,516
rodina	12	1 060	0,447
matka	13	1 059	0,446
syn	25	907	0,382
vdova	38	812	0,016
manželka	45	753	0,317
manželstvo	53	610	0,257
vzťah	59	547	0,230
rodič	72	477	0,201
rozvod	270	137	0,058
pomer	524	66	0,028
snúbenica	1 193	27	0,011
macocha	2 923	8	0,003
exmanželka	13 999	1	0,000
mladomanželka	16 022	1	0,000
Spolu %			3,484

Tabuľka č. 34 Frekvencie výskytu slov (slovenské ženy - rodina - podst. mená)

Pri prídavných menách sú výsledky podobné. V rovnakom poradí sa umiestnila prvá trojica výsledkov z mužského aj ženského datasetu – ženatý / vydatá, rodinný, slobodný, ktoré sa vyskytujú najčastejšie, za nimi nasledujú postupne zvyšné prídavné mená, ktoré však predstavujú menej tradičné a všeobecne menej používané prídavné mená. Rozdiely medzi datasetmi však stále pretrvávajú. U žien sa rodinné prídavné mená vyskytujú **1,7x častejšie ako u mužov**. Výsledky za oba datasety sú zhrnuté nižšie.

RODINA A SOC. STATUS (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
ženatý	103	792	0,118
rodinný	163	530	0,079
slobodný	204	438	0,065
manželský	592	161	0,024
bezdetný	1 010	89	0,013
rozvedený	1 421	61	0,009
rodičovský	3 066	24	0,004
otcovský	4 055	17	0,003
mimomanželský	5 386	11	0,002
nemanželský	5 896	10	0,001
ovdovený	11 011	4	0,001
zasnúbený	15 346	3	0,000
Spolu %			0,320

Tabuľka č. 35 Frekvencie výskytu slov (slovenskí muži - rodina - príd. mená)

RODINA A SOC. STATUS (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
vydatý	85	163	0,140
rodinný	102	142	0,122
slobodný	244	61	0,052
manželský	252	59	0,051
materský	256	58	0,050
bezdetný	313	48	0,041
zasnúbený	401	39	0,034
rozvedený	580	28	0,024
ovdovený	1 245	12	0,010
rodičovský	1 285	12	0,010
mimomanželský	1 345	11	0,009
nemanželský	2 231	6	0,005
Spolu %			0,550

Tabuľka č. 36 Frekvencie výskytu slov (slovenské ženy - rodina - príd. mená)

Slovesá ukazujú rovnakú tendenciu, pričom pri niektorých slovách je nepomer podstatne výraznejší. Mužským výsledkom dominuje sloveso **oženit' sa** (1 412, 49. miesto, 0,304 %), rovnako ako ženským sloveso **vydat' sa** (1 517, 6. miesto, 1,746 %), s tým rozdielom, že v ženskom datasete sa vyskytuje až **5,7x častejšie**. Početnosť sloviac postupne pozvoľna klesá v oboch datasetoch, celkovo pri ich porovnaní však vidíme, že u žien je **výskyt týchto slov 5x častejší**. Výsledky pre obidva datasety sumarizujeme v tabuľkách nižšie.

RODINA A SOC. STATUS (SLOVESÁ)			
	Poradie	Počet	%
oženit'	49	1 412	0,304
rozviest'	286	258	0,055
rozíšť	502	138	0,030
zasnúbit'	1 287	37	0,008
ovdovieť	3 251	8	0,002
Spolu %			0,399

Tabuľka č. 37 Frekvencie výskytu slov (slovenskí muži - rodina - slovesá)

RODINA A SOC. STATUS (SLOVESÁ)			
	Poradie	Počet	%
vydat'	6	1 517	1,746
rozviest'	80	182	0,209
rozíšť	212	63	0,073
zasnúbit'	281	48	0,055
ovdovieť	682	15	0,017
Spolu %			2,100

Tabuľka č. 38 Frekvencie výskytu slov (slovenské ženy - rodina - slovesá)

Kategória kariéra

Poslednou sledovanou kategóriou slov je kariéra. Výsledky slovenského datasetu do značnej miery kopírujú výsledky toho českého, a to tak pre ženy ako aj pre mužov. Mužský a ženský dataset vykazujú podobný výskyt slov súvisiacich s kariérou, úspechom v pracovnej oblasti, len s miernym rozdielom. Výsledky za slovenský mužský dataset sú sumarizované nižšie, v Tabuľke č. 38.

KARIÉRA (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
práca	6	6 158	0,459
štúdium	7	5 926	0,441
kariéra	39	3 641	0,271
cena	42	3 520	0,262
titul	75	2 236	0,166
úspech	87	2 054	0,153
víťazstvo	150	1 336	0,099
vedúci	183	1 150	0,086
ocenenie	213	1 012	0,075
vzdelanie	230	955	0,071
víťaz	360	622	0,046
talent	457	491	0,037
odborník	499	446	0,033
šéf	806	263	0,020
uznanie	822	260	0,019
nadanie	1 301	146	0,011
profesia	1 969	85	0,006
profesionál	2 181	73	0,005
rešpekt	2 242	70	0,005
Spolu %			2,267

Tabuľka č. 39 Frekvencie výskytu slov (slovenskí muži - kariéra - podst. mená)

Ženský dataset vo veľkej miere kopíruje výsledky mužského datasetu, bez zaujímavejších odchýlok. Podobne ako v českom ženskom datasete aj tu sa pri pohľade na slová – šéfká, vedúca, profesionálka a odborníčka, teda slov, ktoré súvisia s odbornosťou a pracovným postavením, črtá výraznejší nepomer s mužským datasetom. V porovnaní s mužským datasetom sa tieto slová vyskytujú u žien v značne menšej miere; **šéfká** (3x menej často ako šéf), **vedúca** (6x menej ako vedúci), **profesionálka** (1,6x menej ako profesionál), **odborníčka** (33x menej ako odborník).

Už pri letmom pohľade na tabuľku vidíme, že nie sú veľké rozdiely v prezentácii ženského a mužského úspechu v pracovnej oblasti, výsledky sú veľmi vyrovnané, dokonca navzdory očakávaniam, **mierne prevyšuje ženský dataset** (2,42 % zastúpenie u žien vs. 2,26 % zastúpenie slov u mužov).

KARIÉRA (PODSTATNÉ MENÁ)			
	Poradie	Počet	%
kariéra	11	1 081	0,455
cena	16	1 005	0,423
štúdium	18	951	0,401
práca	43	768	0,324
uznanie	45	776	0,019
úspech	62	518	0,218
títul	79	416	0,175
ocenenie	94	363	0,153
vzdelanie	175	198	0,083
víťazstvo	214	169	0,071
víťazka	254	144	0,001
talent	319	117	0,049
vedúca	992	34	0,014
nadanie	1 392	22	0,009
profesia	1 397	22	0,009
šéfka	1 917	15	0,006
rešpekt	2 548	10	0,004
profesionálka	3 424	7	0,003
odborníčka	8 548	2	0,001
Spolu %			2,420

Tabuľka č. 40 Frekvencie výskytu slov (slovenské ženy - kariéra - podst. mená)

Pri porovnaní prídavných mien (viď Tabuľka č. 40 a 41) a slovies (viď Tabuľka č. 42 a 43) v mužskom a ženskom datasete vidíme, že v rámci nich nie sú významnejšie rozdiely, napriek predpokladom z hypotézy H4. Percentuálny výskyt slov v datasetoch je porovnateľný; pri prídavných menách **mierne vedie mužský dataset** (1,078 % vs. 0,826 %). Aj usporiadanie slov v rámci tabuľky početnosti je takmer rovnaké, bez väčších odlišností.

KARIÉRA (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
úspešný	34	1 913	0,286
profesionálny	42	1 717	0,257
odborný	50	1 440	0,215
slávny	119	683	0,102
pracovný	218	413	0,062
víťazný	309	298	0,045
uznávaný	314	298	0,045
vzdelaný	725	131	0,020
nadaný	737	129	0,019
telentovaný	739	129	0,019
oceňovaný	1 831	45	0,007
kariérny	4 936	13	0,002
Spolu %			1,078

Tabuľka č. 41 Frekvencie výskytu slov (slovenskí muži - kariéra - prid. mená)

KARIÉRA (PRÍDAVNÉ MENÁ)			
	Poradie	Počet	%
profesionálny	21	565	0,486
úspešný	33	431	0,371
odborný	71	187	0,161
slávny	129	115	0,099
pracovný	196	80	0,069
uznávaný	349	43	0,037
vzdelaný	486	32	0,028
víťazný	491	32	0,028
telentovaný	604	26	0,022
oceňovaný	1 283	12	0,010
kariérny	5 100	2	0,002
Spolu %			0,826

Tabuľka č. 42 Frekvencie výskytu slov (slovenské ženy - kariéra - prid. mená)

Aj pri pohľade na slovesá vidíme rovnaké usporiadanie slov v rámci tabuľky, percentuálny rozdiel medzi jednotlivými datastami **medzi slovesami nie je signifikantný** (1,531 % v mužskom vs. 1,524 % v ženskom datase).

KARIÉRA (SLOVESÁ)			
	Poradie	Počet	%
pracovať	7	6 122	1,317
zvít'aziť	134	548	0,118
uspieť	391	181	0,039
vzdelávať	686	91	0,020
uznávať	743	82	0,018
rešpektovať	978	56	0,012
vyhrávať	1 264	38	0,008
Spolu %			1,531

Tabuľka č. 43 Frekvencie výskytu slov (slovenskí muži - kariéra - slovesá)

KARIÉRA (SLOVESÁ)			
	Poradie	Počet	%
pracovať	8	1 141	1,313
zvít'aziť	136	16	0,122
uspieť	445	28	0,032
vzdelávať	587	19	0,022
uznávať	824	12	0,014
vyhrávať	1 005	9	0,010
rešpektovať	1 032	9	0,010
Spolu %			1,524

Tabuľka č. 44 Frekvencie výskytu slov (slovenské ženy - kariéra - slovesá)

7.2.5 Vyhodnotenie hypotéz H3 a H4 pre českú aj slovenskú Wikipédiu

V tejto časti zhrnieme výsledky za predchádzajúcu časť vzhľadom na stanovené hypotézy H3 a H4.

H3: v slovenskom a českom ženskom datase bude častejší výskyt slov odkazujúcich na ich gender, sociálny status a rodinu a slov odkazujúcich na opačné pohlavie, ako v mužskom datase.

Dáta ukazujú, že naozaj je častejší výskyt slov odkazujúcich na gender, špecificky slova **žena** v slovenskom aj českom ženskom datase, rovnako ako aj prídavné meno **ženský**. Celkovo sa na gender odkazovalo v českom ženskom datase **6x častejšie pri podstatných menách a 8x častejšie pri prídavných menách**. V slovenskom datase vyzerajú výsledky podobne; v ženskom datase sa podstatné mená vyskytujú **5x častejšie, prídavné mená zase 10x častejšie** ako v mužskom.

V rámci analýzy sa v oboch ženských datasetoch ukazovala aj tendencia **častejšie odkazovať v texte na opačné pohlavie**, predovšetkým na mužských priamych príbuzných (otec, syn, brat a vo veľkej miere aj manžel). V prípade mužských datasetov sa tieto slová nevyskytovali tak často ako u žien, a to ani v jednom z jazykov. Slová z tejto kategórie sa v českom ženskom datasete ukazujú **3x častejšie** ako v mužskom. Rovnako ako aj v slovenskom, u žien sa spomína opačné pohlavie **2x častejšie** ako u mužov.

Čo sa týka rodiny a sociálneho statusu tak u žien sa výrazne častejšie vyskytujú slová, ktoré vypovedajú o jej **spoločenskom postavení** (manželka, vdova, a podobne), čo platí rovnako pre oba jazyky. Zároveň sa však častejšie vyskytujú všetky sledované slová z kategórie, dáta ukazujú, že sa u žien naozaj akcentuje rodina a sociálny status viac ako u mužov. V slovenskom ženskom datasete sa spomínajú **podstatné mená z tejto kategórie 2x častejšie, prídavné mená 1,7x častejšie a slovesá 5x častejšie**. České výsledky ukazujú podobný nepomer; **podstatné mená sa v ženskom datasete objavujú 1,7x častejšie ako v mužskom, prídavné mená 2x častejšie a slovesá raz tak často**. Výsledky sme interpretovali a zvýraznili tie najzaujímavejšie v predchádzajúcich podkapitolách.

Na základe týchto zistení, ktoré podrobnejšie rozvádzame v predchádzajúcich podkapitolách môžeme povedať, že **hypotéza H3 bola potvrdená**.

Percentá a frekvencie sa v RStudiu počítajú v rámci konkrétneho slovného druhu a nie na celý dataset, tak výsledky ukazujú relevantnejšie porovnania v rámci slovných druhov, avšak pre celkové zhodnotenie nie sú najvhodnejšie. Preto sa pre celkové overenie si výsledkov ešte pozrieme na prepočet výsledkov jednotlivých kategórií v rámci celého datasetu (viď Tabuľka č. 44 a 45).

Podľa výsledkov pre českú Wikipédiu vidíme, že ženský dataset má celkovo vyššie percentuálne zastúpenie jednotlivých kategórií (kategórii kariéra sa budeme venovať v ďalšej časti) v rámci celého datasetu, a teda sa potvrdili predchádzajúce výsledky o prízvukovaní genderu, rodiny/sociálneho statusu a časté odkazovanie na opačné pohlavie.

PERCENTUÁLNE ZASTÚPENIE KATEGÓRIÍ V CELOM DATASETE ⁵⁷				
ČESKO	GENDER	GENDER II	RODINA	KARIÉRA
ŽENY	0,3 %	0,6%	1,2%	1,1%
MUŽI	0,05 %	0,2 %	0,6%	1%
ROZDIEL	0,25 %	0,4 %	0,6 %	0,1 %

Tabuľka č. 45 Percentuálne zastúpenie kategórií – Česko

Slovenská Wikipédia dopadla podobne, resp. aj tu vidíme rovnaký nepomer, ktorý potvrdzuje výsledky z českej Wikipédie, vid' tabuľka nižšie.

PERCENTUÁLNE ZASTÚPENIE KATEGÓRIÍ V CELOM DATASETE ⁵⁸				
SLOVENSKO	GENDER	GENDER II	RODINA	KARIÉRA
ŽENY	0,2 %	0,4 %	1 %	1 %
MUŽI	0,04 %	0,2 %	0,5 %	0,9 %
ROZDIEL	0,16 %	0,2 %	0,5 %	0,1 %

Tabuľka č. 46 Percentuálne zastúpenie kategórií - Slovensko

Na základe týchto zistení, ktoré podrobnejšie rozvádzame v predchádzajúcich podkapitolách môžeme povedať, že **hypotéza H3 bola potvrdená**.

H4: v ženskom datasete bude menej častý výskyt slov odkazujúcich na kariéru. V prípade mužského datasetu budú tieto zistenia opačné, a to v prípade oboch jazykov.

V rámci kategórie kariéra sme v oboch prípadoch získali **výsledky s najmenšími rozdielmi v početnosti výskytu jednotlivých slov**. Výsledky z českej Wikipédie sú nasledovné: pri podstatných menách majú častejší výskyt ženy o 0,397 %, u prídavných mien je to o 0,11 % a u sloviess prevažuje ženský dataset o 0,419 %. Na slovenskej Wikipédii s výnimkou prídavných mien, kde o 0,25 % dominujú muži, prevažuje ženský dataset; u podstatných mien o 0,153 % a u sloviess o 1,395 %. Muži a ženy na českej Wikipédii sú z hľadiska výskytu slov odkazujúcich na kariéru a **pracovné úspechy skoro vyrovnaní, na oboch Wikipédiách pritom nečakane ženy dosahujú jemnú prevahu**.

Na základe výsledkov tak môžeme povedať, že **hypotéza H4 bola vyvrátená**.

⁵⁷ Percentá sú počítané z celého počtu všetkých slov v českom ženskom a českom mužskom datasete. Predchádzajúce výsledky frekvenčnej analýzy boli počítané len v rámci vybraného slovného druhu (podstatné mená, prídavné mená, slovesá).

⁵⁸ Percentá sú počítané z celého počtu všetkých slov v slovenskom ženskom a slovenskom mužskom datasete. Predchádzajúce výsledky frekvenčnej analýzy boli počítané len v rámci vybraného slovného druhu (podstatné mená, prídavné mená, slovesá).

Tieto výsledky nám potvrdil aj kontrolný prepočet v Tabuľkách č. 44 a 45, ktoré v oboch jazykoch ukazujú jasnú prevahu ženského datasetu (0,1 %). Na základe výsledkov tak môžeme povedať, že **hypotéza H4 bola vyvrátená**.

7.3 Porovnanie českej a slovenskej Wikipédie

V rámci predchádzajúcej analýzy výsledkov sme z veľkej časti už výsledky českej a slovenskej Wikipédie porovnali. Teraz predchádzajúcu analýzu a závery z hypotéz H1 až H4 zhrnieme vo výsledku hypotézy H5.

H5: pri porovnávaní výsledkov analýzy dosiahne česká Wikipédia celkovo lepšie výsledky, a teda preukáže menšiu mieru stereotypizácie žien ako slovenská.

Najskôr teda zhodnotíme počty článkov a celkovo kvalitu článkov o ženách na českej a slovenskej Wikipédii. Počet článkov o ženách na českej Wikipédii približne odpovedá počtu článkov o ženách na slovenskej Wikipédii vzhľadom na fakt, že slovenská Wikipédia je celkovo menšia, rozdiel činí spolu len **1 %**. Biografie o ženách na slovenskej Wikipédii tvoria **16 % z celkového počtu** slovenských biografii, čo je v prepočte 3 803 článkov z 24 510. Česká Wikipédia obsahuje celkovo 110 866 biografických článkov a 18 349 biografii o ženách, resp. **17 % z celkového počtu českých biografii**. V tomto porovnaní **vychádza lepšie česká verzia Wikipédie**.

Takisto sa zásadne nelíšia ani v priemernom počte slov na článok. Na slovenskej Wikipédii nie je rozdiel medzi mužským a ženským článkom **ani 1 slovo** a rozdiel v hodnote mediánu činí **36 slov, resp. 4,19 %**. Na českej Wiki je rozdiel približne **10 slov, resp. 4 %** a v **mediáne 49 slov, resp. 4,49 %**. Teda v tomto porovnaní **vychádza lepšie slovenská Wikipédia**.

Čo sa týka pohľadu do vybraných kategórii slov, už v rámci predchádzajúcej analýzy sme si mohli všimnúť, že rozdiely medzi jazykmi nie sú nijak výrazné. Percentá a frekvencie sa však v RStudiu počítajú v rámci konkrétneho slovného druhu a nie na celý dataset. Takéto výsledky ukazujú relevantnejšie porovnania v rámci slovných druhov, kedy odfiltrujeme neplnovýznamové slová a vieme sa presnejšie zamerať. Pre celkové zhodnotenie však nie sú najvhodnejšie, keďže jednotlivé percentá nie je možné spájať pre celkové výsledky kategórie⁵⁹.

⁵⁹ Percentá sú počítané z počtu jednotlivých slovných druhov v datasete a nepracuje sa pri výpočte s celkovým počtom slov, teda vieme len ich percentuálny podiel pre napr. slovesá, ale vzhľadom na odlišný základ tieto výsledky nemôžeme spájať.

Preto sa pre celkové overenie si výsledkov ešte pozrieme na prepočet percentuálnych výsledkov jednotlivých kategórii na celý dataset (viď Tabuľka č. 44 a 45).

PERCENTUÁLNE ZASTÚPENIE KATEGÓRIÍ V CELOM DATASETE				
ČESKO	GENDER	GENDER II	RODINA	KARIÉRA
ŽENY	0,3 %	0,6%	1,2%	1,1%
MUŽI	0,05 %	0,2 %	0,6%	1%
ROZDIEL	0,25 %	0,4 %	0,6 %	0,1 %

Tabuľka č. 47 Percentuálne zastúpenie kategórií – Česko

PERCENTUÁLNE ZASTÚPENIE KATEGÓRIÍ V CELOM DATASETE				
SLOVENSKO	GENDER	GENDER II	RODINA	KARIÉRA
ŽENY	0,2 %	0,4 %	1 %	1 %
MUŽI	0,04 %	0,2 %	0,5 %	0,9 %
ROZDIEL	0,16 %	0,2 %	0,5 %	0,1 %

Tabuľka č. 48 Percentuálne zastúpenie kategórií - Slovensko

Sumár výsledkov v tabuľkách vyšiel v súlade so závermi z hypotéz H3 a H4, a teda ukázal v oboch jazykových verziách vyšší výskyt slov odkazujúcich na gender, opačné pohlavie a rodinu, a takisto aj pri kategórii Kariéra.

Výsledky teda ukázali, že na českej Wikipédii sa slová z **kategórie Gender vyskytujú 6x častejšie** u žien ako u mužov, do kontextu s **opačným pohlavím** sú takisto častejšie dávané ženy ako muži, a to **3x častejšie**, **rodina a sociálny status** sa u žien **zdôrazňuje 2x tak často**, jedine v pracovnej oblasti sú muži a ženy **vykresľovaní približne rovnako**.

Na slovenskej Wikipédii sa u žien **gender spomína 5x častejšie**, do kontextu s **opačným pohlavím** sa uvádzajú takisto **2x častejšie**, **rodina a sociálny status** sa vyskytuje **2x častejšie** ako pri mužoch, jedine v rámci kariéry je **rozdiel zanedbateľný**, resp. o niečo málo vedú ženy. Pri porovnaní týchto výsledkov s českými teda **mierne vedie slovenská Wikipédia**.

Výsledky našej analýzy teda ukazujú, že česká a slovenská Wikipédia prezentuje ženy veľmi podobným až rovnakým spôsobom čo sa týka stereotypizácie v rámci biografii. Rozdiely medzi nimi sú, nie sú však veľké, výsledky sú nakoniec veľmi tesné, no ukazujú v prospech slovenskej Wikipédie. Na základe týchto zistení môžeme konštatovať, že **hypotéza H5 sa nepreukázala ako pravdivá**.

7.4 Sumarizácia výsledkov hypotéz

Celkovo sa nám podarilo overiť všetkých 5 sformulovaných hypotéz s nasledujúcim výsledkom:

Hypotéza H1 sa nám potvrdila, a teda dokázala predpokladanú nerovnováhu v počte biografických článkov o ženách a o mužoch pri oboch jazykových verziách (len 16 %, resp. 17 % z celkového počtu biografických článkov je o ženách).

Hypotéza H2 sa nám nepotvrdila, a teda sme nezistili signifikantný rozdiel v kvalite článkov o ženách a o mužoch, resp. medzi priemerným počtom slov na článok v mužskom a v ženskom datasete pri obidvoch jazykoch bol rozdiel len minimálny. To isté platí aj pre medián.

Hypotéza H3 sa potvrdila a dotýkala sa konkrétne výskytu slov z určitej kategórie. Výsledky výskumu potvrdili predpoklad z hypotézy H3, keďže sme zistili signifikantný rozdiel v tom, ako často sa vyskytujú slová z kategórie gender, referencia na opačné pohlavie a rodina a sociálny status. V oboch jazykoch sú tieto slová výrazne prominentnejšie u žien ako u mužov.

Hypotéza H4 sa nám nepotvrdila. V rámci slov z kategórie kariéra sme naopak zistili, že úspechy žien a mužov sú prezentované porovnateľne, pričom ženy majú v oboch jazykových verziách o niečo málo častejší výskyt takýchto slov.

Posledná **hypotéza H5** sa takisto navzdory predpokladom nepotvrdila. Dáta pri porovnávaní nepreukázali nižšiu úroveň stereotypizácie žien na českej Wikipédii, naopak slovenská Wikipédia vychádzala z porovnania o niečo lepšie.

ZÁVER

Predkladaná práca si kládla za cieľ preskúmať spôsob, akým sú na českej a slovenskej Wikipédii prezentované ženy a muži, a zároveň poukázať na možnú nerovnosť v tejto prezentácii. Náš výskum samozrejme nie je prvý, ktorý sa touto tematikou zaoberal. Pri formulácii hypotéz a celkovo pri vytváraní si vlastného pohľadu na problematiku sme vychádzali predovšetkým zo zahraničných výskumov a prieskumov. V súčasnosti sa najaktívnejšie venuje výskumu Wikipédie z hľadiska genderu nemecká akademička Claudia Wagner, z ktorej výskumov sme primárne čerpali. Ďalšou známou postavou je aj Eduardo Graells-Garrida, ktorého výskum *First Women, Second Sex: Gender Bias in Wikipedia* (2015) sme z časti replikovali. Poslednou známejšou dvojicou v súčasnom výskume genderu na Wikipédii je Jahna Otterbacher a Olga Zagovora, ktorých výskumy sme si takisto predstavili v rámci práce. Uvedení akademici, a môžeme povedať, že aj výskumy všeobecne, sa však vo veľkom zameriavajú viac menej len na anglickú verziu Wikipédie, prípadne na jazykové verzie krajín, z ktorých pochádzajú (Nemecko, Grécko a pod.). Názor na česko-slovenské wikipedické prostredie sme si preto v rámci prípravy výskumu formovali predovšetkým z dostupných štúdií (offline aj online) mediálneho prostredia, i keď takých výskumov bolo pomenej, prípadne staršieho dáta. Výskumy zamerané špecificky na české a slovenské wikipedické prostredie, prípadne v rámci V4, aktuálne absentujú. Preto sme sa uchýlili k mediálnym výskumom, ktoré aspoň z časti dodali práci česko-slovenský kontext.

Absenciu wikipedických a obzvlášť výskumov s genderovým fókusom v našom regióne sme však videli aj príležitosť začať týmto výskumom diskusiu o genderovej problematike na Wikipédii aj v rámci nášho geografického priestoru. Tento výskum je akousi prvou sondou do prostredia českej a slovenskej Wikipédie, na ktorú sa pozeráme z hľadiska genderu a lexiky. To bol takisto jeden z dôvodov, prečo sme do práce zaradili aj výsledky nad rámec sledovaných hypotéz (spoluvýskyt slov a podobne), keďže sa jedná o jeden z prvých takýchto výskumov, chceli sme poskytnúť aj všeobecnejší pohľad na situáciu v rámci Wikipédie a dať čitateľovi základný prehľad pred tým, ako mu predstavíme konkrétne výsledky.

Od výskumov, z ktorých sme vychádzali sa naša práca líši predovšetkým špecifickými jazykmi textu a vybraným nástrojom spracovania. V rámci výskumu sme síce vychádzali z metód použitých vo výskumoch vyššie uvedených autorov a autoriek, avšak vzhľadom na to, že ani čeština ani slovenčina nepatria medzi celosvetovo rozšírené jazyky, museli sme vybrať nástroj, ktorý dokáže s týmito jazykmi pracovať. Preto sme si zvolili softvér RStudio

a jazykový model UDPipe, ktorý to dokáže. Takisto sme na rozdiel od niektorých uvedených výskumov (konkrétne Otterbacher, et al., 2019 a čiastočne aj Graells-Garrido, et al., 2015) pracovali so všetkými vtedy dostupnými biografickými textami v ich plnom znení (bez sprostredkovania, resp. nie len anotácia, plný text všetkých článkov).

Sumár výsledkov výskumu

Výsledky nášho snaženia zapadajú do kontextu výsledkov predchádzajúcich autorov a autoriek. Sú v súlade s výsledkami nami replikovaného výskumu E. Graells-Garridu a kolegov *First Women, Second Sex: Gender Bias in Wikipedia* (2015), keďže sme identifikovali rovnaký (ne)pomer v počte ženských biografí; výskum Graells-Garridu a kolegov uvádza 15,5 % (2015), my sme identifikovali 16, resp. 17 % a tiež aj menšie rozdiely v priemernom počte slov. Rovnako sa naše výsledky zhodujú aj s výskumom C. Wagnerovej a kolegov (2015), ktorí v rámci výskumu *It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia* zistili, že síce je článkov o ženách na Wikipédii menej, sú však v porovnateľnej kvalite ako tie o mužoch (čo do dĺžky aj celkovo kvality spracovania). Zároveň sa pri pohľade na dĺžku, resp. kvalitu článkov o ženách ukazuje zaujímavý detail, na ktorý z časti poukazuje Reagle a Rhue (2011), a ktorým je existencia skleneného stropu. Ženy sa všeobecne do Wikipédie dostávajú ťažšie, čo platí tak o textoch (Reagle a Rhue, 2011; Wagner, et al., 2016) ako aj o editorskej činnosti (Shyong K. Lam, et al., 2011). No ak sa už o nich na Wikipédii píše, tak články majú porovnateľnú úroveň, ako tie o mužoch (Reagle a Rhue, 2011). O čom svedčia aj výsledky nášho bádania, predovšetkým zanedbateľné rozdiely medzi mediánmi a priemernými počtami slov, prípadne takisto aj veľmi dobré výsledky porovnania v rámci kategórie slov Kariéra.

Ďalej sa nám postupne potvrdili aj výsledky z výskumov E. Graells-Garridu a kolegov (2015) a C. Wagnerovej a kolegov (2015; 2016) týkajúce sa použitej lexiky. Aj v našich textoch naprieč oboma jazykmi sa častejšie odkazovalo na ženský gender. Takisto vidíme paralelu aj s viacerými výskumami, ktoré poukazujú na dominanciu súkromnej (rodinnej) sféry v textoch a mediálnych výstupoch o ženách (Otterbacher, 2019). Naopak, čo sa týka kariéry a pracovnej oblasti, tak tam boli rozdiely medzi lexikou použitou v článkoch o mužoch a o ženách len minimálne, dokonca v mierny prospech žien, a to v oboch jazykových verziách. Na zamyslenie je aj spojenie častého prízvukovania genderu u žien v spojení s vyrovnanou prezentáciou pracovných úspechov. Je možné, že sa jedná o ilustráciu tzv. pozitívnej diskriminácie, v rámci ktorej sa naopak s ohľadom na oklieštené historické možnosti sebarealizácie vyzdvihuje fakt, že sa jedná o profesne úspešnú a uznávanú ženu.

Poslednou z tém, ktorej sme sa dotkli boli predpokladané kultúrne rozdiely. Česko a Slovensko sú krajiny, ktoré majú dlhú spoločnú históriu, ikeď sa už vyše štvrt'oročia vyvíjame samostatne. Určite existujú viaceré kultúrne rozdiely, no v rámci našej práce sa nepotvrdili. Výsledky z analýzy českej a slovenskej Wikipédie sú si natoľko podobné, že nie je na mieste hovoriť o väčších kultúrnych rozdieloch. Miera stereotypizácie žien, ktorú sme v rámci výskumu potvrdili, je si v oboch jazykoch podobná, pričom mierne lepšie z porovnania vychádza slovenská Wikipédia. Jedno z možných vysvetlení sa podľa nás môže skrývať v rozdielnych veľkostiach českej a slovenskej Wikipédie, ktoré tak môžu mať inak naladenú editorskú komunitu.

Celkovo sa nám v rámci práce potvrdili dve hypotézy, zvyšné tri sa nám nepotvrdili. V rámci hypotézy H1 sa nám potvrdil predpoklad celkovo menšieho zastúpenia biografí o ženách v porovnaní s biografiami o mužoch na českej aj slovenskej Wikipédii.

Hypotéza H2 nám nepotvrdila zásadný rozdiel v kvalite, resp. v dĺžke biografí medzi mužským a ženským datasetom, a to ani v jednom jazyku. Porovnanie priemerného počtu slov na článok a mediánu medzi ženským a mužským datasetom v oboch jazykoch vykazuje len malé, nevýznamné rozdiely.

V rámci hypotézy H3 sme si potvrdili predpoklad, že v rámci oboch ženských datasetov bude viditeľné zvýrazňovanie kategórií genderu, rodiny/sociálneho stavu a opačného pohlavia. V slovenskom a českom mužskom datasete sa uvedené kategórie slov celkovo vyskytovali menej často ako u žien.

Hypotéza H4 sa nám nepotvrdila, keďže výsledky ukázali naopak mierne častejší výskyt slov z kategórie Kariéra tak u českých ako aj u slovenských žien. Prezentácia pracovných úspechov žien a mužov bola nakoniec najvyrovnanejšia zo sledovaných oblastí, a takisto jediná, v ktorej boli ženy o niečo zvýhodňovanejšie pred mužmi, a to v oboch jazykoch.

A nakoniec hypotéza H5, ktorá sa síce nepotvrdila, ale výsledky medzi Wikipédiami boli naozaj značne vyrovnané. Nakoniec výsledky ukázali v prospech slovenskej Wikipédie, ktorá ukazuje o niečo menšiu mieru stereotypizácie žien ako česká.

Výsledky výskumu teda celkovo ukazujú nerovnosť v tom, ako sú prezentovaní muži a ženy na jednom z celosvetovo najnavštevovanejších portálov. Sú to detaily, ktoré si bežne možno ani nevšimneme, no ako obsah, ktorý vytvárajú samotní používatelia, do určitej miery odrážajú nastavenie spoločnosti, prípadne neideálne pomery vo wiki editorskej komunite, ako

naznačujú aj doterajšie výskumy. Dáta ukazujú, že na ženy sa v oboch jazykoch stále pozerá tradičnou optikou genderu a rodinno-sociálnych stereotypov, prízvukuje sa ich vzťah k deťom, k partnerom, rodine, prípadne sa v texte často odvolávajú na ich mužských príbuzných. Nie je to však situácia, ktorú by nebolo možné adresovať. Ako vidíme, v rámci prezentácie pracovného života a úspechu sú si muži a ženy na Wikipédii rovní, s malou prevahou žien. Podobne ako disproporcia, čo do počtu článkov, ktorá je aktuálne celkom značná, nie je konečná a so vznikom (Slovensko) či podporou (Česko) projektov ako je *WikiGap 2019: Oslava MDŽ*, sa situácia časom môže posunúť k väčšej inklúzii a inému, menej stereotypmi poznačenému pohľadu na ženy. Naše výsledky poskytujú prvý bližší pohľad na situáciu v rámci Wikipédie v našom kultúrnom a geografickom priestore a je možné na nich ďalej stavať.

Hlavný prínos práce a odporúčania pre prax

Hlavný prínos našej práce vidíme v otvorení diskusie o genderovej problematike na Wikipédii, ktorá je obľúbeným a často používaným zdrojom informácií v Česku aj na Slovensku. Miera stereotypizácie, na ktorú sme v rámci výskumu narazili, nie je ľahko postrehnuteľná, články sami o sebe nie sú napísané zle či chybné, ide skôr o ďalšiu úroveň kvality článku, aby bol napísaný čo najlepšie, aj vzhľadom na popularitu, a tým aj veľkú návštevnosť Wikipédie. Je to zdroj, z ktorého čerpajú informácie milióny ľudí, medzi nimi aj mladá generácia, a práve preto je na mieste stopnúť zaužívané vzorce a baviť sa o kvalite článkov z trochu nezvyčajného, no potrebného pohľadu genderu.

Z výsledkov našej analýzy môžeme vyvodiť viaceré praktické odporúčania, predovšetkým čo sa týka používania genderovo neutrálnejšieho jazyka (vysoký výskyt označenia žena a celkovo genderovo podfarbených slov) a väčšiu vyváženosť pri opisovaní súkromnej a pracovnej sféry, keďže sa stále u žien akcentuje práve rodina a sociálny status. Zároveň je v celku problematický aj častý výskyt slov odkazujúcich na opačné pohlavie, ktoré stavia ženy do prílišnej závislosti od jej mužských príbuzných či partnerov, a z časti degraduje úspechy, ktoré samy dosiahli. Prípadne častý výskyt slovného spojenia „prvá žena“, ktorý sa preukázal pri oboch ženských datasetoch v rámci vizualizácie spoluvýskytu slov. Z časti, samozrejme, sa môže jednať o tzv. pozitívnu diskrimináciu, to je však ustupujúci trend, minimálne v písanom slove (viď. Finkbeiner test), ktorý z dlhodobého hľadiska môže byť škodlivý. Jedným z praktických odporúčaní je teda predovšetkým vytvorenie pomocného návodu pre tvorbu textov a celkovo prácu s genderovou problematikou (všeobecne alebo špecificky v prípade žien), ktorý chýba na slovenskej aj na českej Wikipédii.

Limity výskumu

Náš výskum má svoje limity. Viaceré z nich vyplývajú práve z češtiny a slovenčiny ako jazyka spracovávaného textu. Prvé limity sa ukazujú už pri formulácii výskumného kontextu v teoretickej časti, keďže výskumy venujúce sa špecificky kombinácii Wikipédie, genderu a česko-slovenskému jazykovému priestoru aktuálne viac menej absentujú. Preto sme na vytvorenie kontextu pracovali s anglickými výskumami či európskym mediálnym kontextom, a tak vytvárali paralelu so situáciou v Česku a na Slovensku. Z tohto pohľadu môžu vzniknúť medzery v kontexte výskumu.

Ďalší problematický aspekt bolo samotné sťahovanie textov a odfiltrovanie biografických článkov, obzvlášť na slovenskej Wikipédii, ktorá pre naše potreby nemala dobrú štruktúru kategórii. Ako sme už uviedli v rámci metodológie, pri sťahovaní a následnom triedení dát mohlo dôjsť ku strate článkov a chybnému zatriedeniu, čo mohlo skresliť výsledky. Vzhľadom na veľkosť datasetu však ručné overenie si správnosti zadelenia nebolo možné, resp. veľmi časovo náročné. Veríme, že použitými metódami, ktoré sme detailne popísali v rámci metodológie výskumu, sme obmedzili toto riziko na minimum.

Pri použití predtrénovaného jazykového modelu UDPipe v rámci anotácie datasetu došlo k viacerým nedokonalostiam, resp. chybám. Text nebol anotovaný dokonalo, došlo k zlej lematizácii, zlému zaradeniu do kategórie a podobne. Niektoré chyby sa neskôr ukázali aj priamo na vizualizácii. Ako problematický sa tiež môže ukázať aj jednak čisto subjektívny výber slov do kategórii, a takisto aj fakt, že sme sa primárne zamerali len čisto na počet a výskyt slov a nebrali sme do úvahy kontext, v ktorom bolo slovo použité, takže pre ucelený obraz o situácii je potrebné genderovú problematiku preskúmať aj inými (kvalitatívnymi) metódami na to, aby sme potvrdili naše výsledky, keďže sa jedná o prvý výskum tohto charakteru v našom geografickom prostredí.

Napriek uvedeným limitom veríme, že predkladaná práca priniesla výsledky, ktoré hodnoverne odrážajú aktuálny stav genderovej problematiky na českej a slovenskej Wikipédii a sú schopné poslúžiť ako styčný bod pri formulácii a zavádzaní praktických opatrení na zmiernenie tohto stavu.

Zoznam použitej literatúry

AGUIAR, M. a E. HURST, 2007. Measuring Trends in Leisure: The Allocation of Time over Five Decades. In: *The Quarterly Journal of Economics* [online]. 2007, roč. 122, č. 3, s. 969–1006 [cit. 6. jún 2019]. Dostupné na: <http://www.jstor.org/stable/25098866>

AHMED, S., 2012. *On being included: Racism and Diversity in Institutional Life*. Raleigh, NC: Duke University Press. 2012, s. 255. ISBN 978-0-8223-5236-5.

Ako používať rodovo vyvážený jazyk: Možnosti, otázky, príklady, 2014. [online]. Bratislava: Centrum vzdelávania MPSVR SR. [cit. 6. jún 2019]. Dostupné na: http://www.ruzovyamodrysvet.sk/chillout5_items/1/5/9/9/1599_8a6542.pdf

BALÚCHOVÁ, B. M., 2010. Rodová (ne)citlivosť na Slovensku (a úloha médií v tomto probléme). In: *Mezinárodní Masarykova konference pro doktorandy a mladé vědecké pracovníky*. Hradec Králové: MAGNANIMITAS [online]. 2010 [cit. 6. jún 2019]. Dostupné na: https://www.academia.edu/8052656/Rodov%C3%A1_ne_citlivost%C5%A5_na_Slovensku_a_%C3%BAloha_m%C3%A9di%C3%AD_v_tomto_prob%C3%A9me

Basic Analytical Use Cases I · NLP with R and UDPipe, 2019 [online] [cit. 20. jún 2019]. Dostupné na: <https://bnosac.github.io/udpipe/index.html>

BAZELY, D., 2018. Perspective: Why Nobel winner Donna Strickland didn't have a Wikipedia page. In: *Washington Post* [online]. 2018 [cit. 29. máj 2019]. ISSN 0190-8286. Dostupné na: <https://www.washingtonpost.com/outlook/2018/10/08/why-nobel-winner-donna-strickland-didnt-have-wikipedia-page/>

BEAR, J. B. a B. COLLIER, 2016. Where are the women in Wikipedia? Understanding the different psychological experiences of men and women in Wikipedia. In: *Sex Roles* [online]. 2016, roč. 74, č. 5-6, s. 254-265 [cit. 7. december 2017]. Dostupné na: <https://link.springer.com/article/10.1007/s11199-015-0573-y>

BEUKEBOOM, C. J., et al., 2010. The negation bias: When negations signal stereotypic expectancies. In: *Journal of Personality and Social Psychology* [online]. 2010, roč. 99, č. 6, s. 978–992 [cit. 30. apríl 2019]. ISSN 0022-3514. Dostupné na: <https://research.vu.nl/en/publications/the-negation-bias-when-negations-signal-stereotypic-expectancies>

BEUKEBOOM, C. J., 2016. *Mechanisms of linguistic bias: How words reflect and maintain stereotypic expectancies* [online]. 2016 [cit. 16. apríl 2019]. Dostupné na: https://www.researchgate.net/publication/258290822_Mechanisms_of_linguistic_bias_How_words_reflect_and_maintain_stereotypic_expectancies

BEUKEBOOM, C. J. a Ch. BURGERS, 2017. *Linguistic Bias* [online]. 2017 [cit. 7. december 2017]. Dostupné na: <http://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-439>

BLIKSTAD-BALAS, Marte, 2016. “You get what you need”: A study of students' attitudes towards using Wikipedia when doing school assignments. In: *Scandinavian Journal of Educational Research* [online]. 2016, roč. 60, č. 6, s. 594–608 [cit. 30. apríl 2019]. ISSN

0031-3831. Dostupné na:

<https://www.tandfonline.com/doi/abs/10.1080/00313831.2015.1066428>

BLUMENSTOCK, Joshua E., 2008. Size matters: word count as a measure of quality on Wikipedia. In: *Proceeding of the 17th international conference: Proceeding of the 17th international conference on World Wide Web - WWW '08* [online]. Beijing, China: ACM Press. 2008, s. 1095 [cit. 8. máj 2019]. ISBN 978-1-60558-085-2. Dostupné na: https://www.researchgate.net/publication/221023915_Size_matters_Word_count_as_a_measure_of_quality_on_Wikipedia

BOSCH, T., 2012. How Kate Middleton's Wedding Gown Demonstrates Wikipedia's Woman Problem. In: *Slate Magazine* [online]. 2012 [cit. 29. máj 2019]. Dostupné na: <https://slate.com/technology/2012/07/kate-middleton-s-wedding-gown-and-wikipedia-s-gender-gap.html>

CAMERON, D., 2007. *The Myth of Mars and Venus*. B.m.: Oxford University Press. 2007. ISBN 978-0-19-921447-1.

CARNAGHI, A., et al., 2008. Nomina Sunt Omina: On the Inductive Potential of Nouns and Adjectives in Person Perception. In: *Journal of personality and social psychology* [online]. 2008, roč. 94, s. 839–59 [cit. 30. apríl 2019]. Dostupné na: https://www.researchgate.net/publication/5407184_Nomina_Sunt_Omina_On_the_Inductive_Potential_of_Nouns_and_Adjectives_in_Person_Perception

CECI, S. J. a W. M. Williams, 2011. Understanding current causes of women's underrepresentation in science. In: *Proceedings of the National Academy of Sciences* [online]. 2011, roč. 108, č. 8, s. 3157-3162 [cit. 7. december 2017]. Dostupné na: <http://www.pnas.org/content/108/8/3157>

COATES, J., 2016. *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language*. B.m.: Routledge. 2016. ISBN 978-1-317-29253-1.

COHEN, N., 2011. Define gender gap? Look up Wikipedia's contributor list. In: *New York Times* [online]. 2011 [cit. 7. december 2017]. Dostupné na: http://www.nytimes.com/2011/01/31/business/media/31link.html?_r=0

COLLIER B. a J. B. BEAR, 2012. *Conflict, Confidence, or criticism: An Empirical Examination of the Gender Gap in Wikipedia* [online]. 2012 [cit. 7. december 2017]. Dostupné na: https://www.researchgate.net/publication/220879022_Conflict_confidence_or_criticism_An_empirical_examination_of_the_gender_gap_in_wikipedia

CONNAWAY, L. S., et al., 2011. „If it is too inconvenient, I'm not going after it”: Convenience as a critical factor in information-seeking behaviours. In: *Library & Information Science Research* [online]. 2011, roč. 33, č. 3, s.179–190. [cit. 7. december 2017]. Dostupné na: <https://doi.org/10.1016/j.lisr.2010.12.002>

CURTIS, B., 2013. The Finkbeiner Test: Seven rules to avoid gratuitous gender profiles of female scientists. In: *Columbia Journalism Review* [online]. 2013 [cit. 7. december 2017]. Dostupné na: http://archives.cjr.org/the_observatory/finkbeiner_test_gender_gap_fem.php

- Česká Wikipedie, 2019. In: *Wikipédie* [online]. 2019 [cit. 9. jún 2019]. Dostupné na: https://cs.wikipedia.org/w/index.php?title=%C4%8Cesk%C3%A1_Wikipedie&oldid=17241437
- MAHDALOVÁ, K., 2014. Analýza: Jaký je veřejný obraz žen a jaké doopravdy jsou? In: *Datová žurnalistika* [online]. 2014 [cit. 5. jún 2019]. Dostupné na: <https://www.datovazurnalistika.cz/analyza-jaky-je-verejny-obraz-zen-a-jake-dopravdy-jsou/>
- MAHDALOVÁ, K., 2015. Obraz žen ve veřejném diskurzu. In: *Datová žurnalistika* [online]. 2015 [cit. 5. jún 2019]. Dostupné na: <https://www.datovazurnalistika.cz/obraz-zen-ve-verejnem-diskurzu/>
- DEVINE, P. G. a S. M. BAKER, 1991. Measurement of Racial Stereotype Subtyping. In: *Personality and Social Psychology Bulletin* [online]. 1991, roč. 17, č. 1, s. 44–50 [cit. 30. apríl 2019]. ISSN 0146-1672. Dostupné na: <https://journals.sagepub.com/doi/abs/10.1177/0146167291171007>
- DIJCK, J. van, 2013. *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press. 2013, s. 482. ISBN 978-0-19-997079-7.
- D'HEER, J., et al., 2019. The bits and bytes of gender bias in online news: a quantitative content analysis of the representation of women in Vice.com. In: *Feminist Media Studies* [online]. 2019 [cit. 2. jún 2019]. ISSN 1468-0777. Dostupné na: <https://doi.org/10.1080/14680777.2019.1574858>
- DOBŠOVIČOVÁ, K., 2011. *Dôveryhodnosť encyklopedických zdrojov vo webovom prostredí* [Bakalárska práca] [online]. Bratislava: Univerzita Komenského v Bratislave FiF KKIV. 2011, s. 85 [cit. 2. jún 2019]. Dostupné na: <https://opac.crzp.sk/?fn=detailBiblioForm&sid=100EE58FA53739A6085A5C1E99B3>
- EUROPEAN JOURNALISM OBSERVATORY (EJO), 2018. *Where Are the Women Journalists in Europe's Media?* European Journalism Observatory - EJO [online]. 2018 [cit. 5. jún 2019]. Dostupné na: <https://en.ejo.ch/research/where-are-all-the-women-journalists-in-europes-media>
- EUROSTAT, 2018. *Štatistika digitálnej ekonomiky a spoločnosti – domácnosti a jednotlivci - Statistics Explained* [online]. 2018 [cit. 30. apríl 2019]. Dostupné na: https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals/sk#Pr.C3.ADstup_na_internet
- EKŞİ, G., 2009. Gender Stereotypes and Gender Roles Revealed in the Students Written Work. In: *Journal of Language and Linguistic Studies* [online]. 2009, roč. 5, č. 1, s. 37–54. [cit. 7. december 2017]. Dostupné na: <http://dergipark.ulakbim.gov.tr/jlls/article/view/5000084185>
- FERIKOVÁ, E., 2018. Ženy vs. muži: Kto tvorí domáce a zahraničné spravodajstvo? Newton Media [online]. 2018 [cit. 5. jún 2019]. Dostupné na: <https://www.medan.sk/zeny-vs-muzi-kto-tvori-domace-a-zahranicne-spravodajstvo/>

FILOVÁ, M. 2015. *Wikipédia a jej využívanie v akademickom prostredí* [Bakalárska práca] [online]. Bratislava: Univerzita Komenského v Bratislave FiF FiF.KKIV. 2015, s. 51 [cit. 9. jún 2019]. Dostupné na: <http://opac.crzp.sk/?fn=detailBiblioForm&sid=4A220BA9C567A8491B013AD1351D&seo=CRZP-detail-kniha>

FORD, H. a J. WAJCMAN, 2017. 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. In: *Social Studies of Science* [online]. 2017, roč. 47. [cit. 9. jún 2019]. Dostupné na: https://www.researchgate.net/publication/311769445_'Anyone_can_edit'_not_everyone_does_Wikipedia's_infrastructure_and_the_gender_gap

GARCIA, D., I. WEBER a V. RAMA KIRAN GARIMELLA, 2014. Gender Asymmetries and Fiction: The Bechdel Test of Social Media. In: *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014* [online]. 2014 [cit. 7. december 2017]. Dostupné na: https://www.researchgate.net/publication/261289271_Gender_Asymmetries_in_Reality_and_Fiction_The_Bechdel_Test_of_Social_Media

GARRISON, J. C., 2015. Getting a "quick fix": First-year college students' use of Wikipedia. In: *First Monday* [online]. 2015, roč. 20, č. 10 [cit. 28. marec 2019]. ISSN 13960466. Dostupné na: <https://journals.uic.edu/ojs/index.php/fm/article/view/5401>

GAZDÍKOVÁ, L., 2014. *Wikipédia ako nástroj vzdelávania* [Bakalárska práca] [online]. .Dubnica nad Váhom: Dubnický technologický inštitút v Dubnici nad Váhom DTI. 2014, s. 61 [cit. 9. jún 2019]. Dostupné na: <https://opac.crzp.sk/?fn=detailBiblioForm&sid=635D8AA4F414A2800B15F4AD10F7&seo=CRZP-detail-kniha>

FINKBEINER, A., 2013. *The Last Word on Nothing | What I'm Not Going to Do* [online]. 2013 [cit. 8. december 2017]. Dostupné na: <http://www.lastwordonnothing.com/2013/01/17/5266/>

GERKEN, J., 2010. How Courts Use Wikipedia. In: *The Journal of Appellate Practice and Process* [online]. 2010, roč. 11, č. 1, s. 191 [cit. 4. apríl 2019]. Dostupné na: <https://lawrepository.ualr.edu/appellatepracticeprocess/vol11/iss1/8>

GITHUB, 2017. Open Source Survey [online]. 2017 [cit. 4. apríl 2019]. Dostupné na: <https://opensourcesurvey.org/2017/>

GRAELLS-GARRIDO, E., et al., 2015. *First Women, Second Sex: Gender Bias in Wikipedia* [online]. 2015 [cit. 8. december 2017]. Dostupné na: https://www.researchgate.net/publication/272195361_First_Women_Second_Sex_Gender_Bias_in_Wikipedia

GLOBAL MEDIA MONITORING PROJECT (GMMP), 2015. *Who Makes the News?* [online]. London: World Association for Christian Communication. 2015 [cit. 6. jún 2019]. Dostupné na: <http://whomakesthenews.org/gmmp/gmmp-reports/gmmp-2015-reports>

- GLOTT, R. a GHOSH, R., 2010. *Analysis of Wikipedia survey data: Age and gender differences* [online]. 2010 [cit. 8. december 2017]. Dostupné na: <http://www.wikipediasurvey.org>
- GRAY, J., 1992. *Men are from Mars and Women are from Venus*. New York: Harper Collins. 1992, s. 286. ISBN 9780060574215.
- HARGITTAI, E. a S. SHAFER, 2006. Differences in Actual and Perceived Online Skills: The Role of Gender*. In: *Social Science Quarterly* [online]. 2006, roč. 87, č. 2, s. 432–448 [cit. 29. máj 2019]. ISSN 1540-6237. Dostupné na: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6237.2006.00389.x>
- HELLINGER, M., 1984. Effecting Social Change through Group Action, Feminine Occupational Titles in Transition. In: *Language and Power*. London: Sage. 1984, s. 136-153.
- HINNOSAAR, M., 2019. *Gender Inequality in New Media: Evidence from Wikipedia* [online]. 2019 [cit. 29. máj 2019]. Dostupné na: <https://ssrn.com/abstract=2617021>
- KOMANICKÁ, D., 2016. Patronymia a patrilinearita vo vzťahu k nerovnému postaveniu žien v spoločnosti. In: *Prechýľovanie: áno – nie?* [online]. 2016 [cit. 9. jún 2019]. Dostupné na: https://www.academia.edu/38174070/Patronymia_a_patrilinearita_vo_vz%C5%A5ahu_k_nerovn%C3%A9mu_postaveniu_%C5%BEien_v_spolo%C4%8Dnosti
- KORSMEYER, C., 2017. Feminist Aesthetics. V: Edward N. ZALTA, ed. *The Stanford Encyclopaedia of Philosophy* [online]. Metaphysics Research Lab, Stanford University. 2017 [cit. 7. december 2017]. Dostupné na: <https://plato.stanford.edu/archives/spr2017/entries/feminism-aesthetics/>
- KUCHYNÁŘOVÁ, M., 2010. Prečo ľudia na Slovensku robia FLOSS? In: *Informatika a spoločnosť* [online]. Bratislava: Univerzita Komenského v Bratislave. 2010 [cit. 9. jún 2019]. Dostupné na: http://new.dcs.fmph.uniba.sk/files/ias/ProjektyIaS2010/4_Preco_ludia_v_SK_robja_FOSS/4_sprava.pdf
- LAM, S. K., et al., 2011. WP: clubhouse? an exploration of Wikipedia's gender imbalance. In: *The 7th International Symposium: Proceedings of the 7th International Symposium on Wikis and Open Collaboration - WikiSym '11* [online]. Mountain View, California: ACM Press. 2011, s. 1 [cit. 2. apríl 2019]. ISBN 978-1-4503-0909-7. Dostupné na: https://www.researchgate.net/publication/221367798_WPClubhouse_An_exploration_of_Wikipedia's_gender_imbalance
- LANIADO, D., et al., 2012. Emotions and dialogue in a peer-production community: the case of Wikipedia. In: *WikiSym '12* [online]. 2012 [cit. 30. apríl 2019]. Dostupné na: http://chato.cl/papers/laniado_kaltenbrunner_castillo_fuster_2012_emotions_wikipedia.pdf
- LEACH, J., et al., 2006. *Free/Libre and Open Source Software: Policy Support, Gender: Integrated Report of Findings* [online]. 2006 [cit. 30. apríl 2019]. Dostupné na: https://www.jamesleach.net/downloads/FLOSSPOLSD16-Gender_Integrated_Report_of_Findings.pdf

LEMMERICH, F., et al., 2019. Why the World Reads Wikipedia: Beyond English Speakers. In: *The Twelfth ACM International Conference: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining - WSDM '19* [online]. Melbourne VIC, Australia: ACM Press. 2019, s. 618–626 [cit. 28. marec 2019]. ISBN 978-1-4503-5940-5. Dostupné na: <https://dl.acm.org/citation.cfm?id=3291021>

LIH, A., 2004. Wikipedia as Participatory journalism: reliable sources? metrics for evaluating collaborative media as a news resource. In: *Proceedings of the 5th International Symposium on Online Journalism* [online]. 2004, s. 16–17. [cit. 30. apríl 2019]. Dostupné na: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.9104>

LIM, S., 2009. How and why do college students use Wikipedia? In: *Journal of the American Society for Information Science and Technology* [online]. 2009, roč. 60, č. 11, s. 2189–2202 [cit. 7. december 2017]. Dostupné na: <https://doi.org/10.1002/asi.21142>

LIM, S. a N. KWON, 2010. Gender differences in information behaviour concerning Wikipedia, an unorthodox information source? In: *Library & information science research* [online]. 2010, roč. 32, č. 3, s. 212–220 [cit. 7. december 2017]. Dostupné na: <http://www.sciencedirect.com/science/article/pii/S0740818810000356>

LIN, Y.-W. a M. den BESTEN, 2018. Gendered work culture in free/libre open source software development. In: *Gender, Work & Organization* [online]. 2018 [cit. 2. jún 2019]. ISSN 1468-0432. Dostupné na: <https://onlinelibrary.wiley.com/doi/abs/10.1111/gwao.12255>

LIN, Y., 2005. Gender Dimensions of Floss Development. In: *Mute* [online]. 2005 [cit. 30. apríl 2019]. Dostupné na: <http://www.metamute.org/editorial/articles/gender-dimensions-floss-development>

LUKIN, A., 2016. “He” vs “she” in Australian media coverage: what the language of news tells us about gender imbalance. In: *The Conversation* [online]. 2016 [cit. 3. jún 2019]. Dostupné na: <http://theconversation.com/he-vs-she-in-australian-media-coverage-what-the-language-of-news-tells-us-about-gender-imbalance-67139>

MAASS, A., et al., 1989. Language use in intergroup contexts: The Linguistic Intergroup Bias. In: *Journal of personality and social psychology* [online]. 1989, roč. 57, s. 981–93 [cit. 30. apríl 2019]. Dostupné na: https://www.researchgate.net/publication/20540879_Language_use_in_intergroup_contexts_the_Linguistic_Intergroup_Bias

MAASS, A., et al., 1995. Linguistic Intergroup Bias: Differential Expectancies or In-Group Protection? In: *Journal of personality and social psychology* [online]. 1995, roč. 68, s. 116–26 [cit. 30. apríl 2019]. Dostupné na: https://www.researchgate.net/publication/15338537_Linguistic_Intergroup_Bias_Differential_Expectancies_or_In-Group_Protection

MACAULAY, M. a R. VISSER, 2016. Editing Diversity in Reading Diversity Discourses on Wikipedia. In: *Ada: A Journal of Gender, New Media, and Technology* [online]. 2016, roč. 9 [cit. 2. jún 2019]. Dostupné na: <https://adanewmedia.org/2016/05/issue09-macaulay-and-visser/>

MATEOS DE CABO, R., et al., 2014. Perpetuating Gender Inequality via the Internet? An Analysis of Women's Presence in Spanish Online Newspapers. In: *Sex Roles* [online]. 2014, roč. 70, č. 1, s. 57–71 [cit. 2. jún 2019]. ISSN 1573-2762. Dostupné na: <https://doi.org/10.1007/s11199-013-0331-y>

MOLNÁR SATINSKÁ, L., 2015. Problematika rodovo citlivého vyjadrovania v maďarčine (v Maďarsku a na Slovensku) [online]. 2015 [cit. 9. jún 2019]. Dostupné na: https://www.academia.edu/31272315/Problematika_rodovo_citliv%C3%A9ho_vyjadrovania_v_ma%C4%8Far%C4%8Dine_v_Ma%C4%8Farsku_a_na_Slovensku

NAFUS, D., 2012. 'Patches don't have gender': What is not open in open source software. In: *New Media & Society* [online]. 2012, roč. 14, č. 4, s. 669–683 [cit. 30. apríl 2019]. Dostupné na: <http://nms.sagepub.com/content/14/4/669>

NIELSEN, F. A., 2007. Scientific citations in Wikipedia. In: *First Monday* [online]. 2007, roč. 12, č. 8 [cit. 24. apríl 2019]. ISSN 13960466. Dostupné na: <https://firstmonday.org/ojs/index.php/fm/article/view/1997>

OTTERBACHER, J., 2019. Linguistic Bias in Crowdsourced Biographies: A Cross-Lingual Examination. In: *Multilingual Text Analysis: Challenges, Models, And Approaches* [online]. World Scientific. 2019 [cit. 24. apríl 2019]. ISBN 978-981-327-489-1. Dostupné na: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/download/10539/10513>

OTTERBACHER, J., et al., 2015. Linguistic Bias in Collaboratively Produced Biographies: Crowdsourcing Social Stereotypes? In: *Proceedings of the Ninth International AAAI Conference on Web and Social Media* [online]. 2015 [cit. 24. apríl 2019]. Dostupné na: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10539/10513>

PANDE, A., 2004. Undoing Gender Stereotypes in Hindi. In: *Linguistic Online* [online]. 2004, roč. 21, č. 4, s. 47–64. [cit. 7. december 2017]. Dostupné na: http://www.linguistik-online.com/21_04/pande.pdf

PEW RESEARCH CENTER, 2018. *Internet Fact Sheet* [online]. 2018 [cit. 30. apríl 2019]. Dostupné na: <http://www.pewinternet.org/fact-sheet/internet-broadband/>

PROTONOTARIOS, I., et al., 2016. Similar Gaps, Different Origins? Women Readers and Editors at Greek Wikipedia. In: *Tenth International AAAI Conference on Web and Social Media* [online]. 2016 [cit. 30. apríl 2019]. Dostupné na: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13192>

RAVAL, Noopur, 2014. The Encyclopaedia Must Fail! Notes on Queering Wikipedia. In: *Ada: A Journal of Gender, New Media, and Technology* [online]. 2014, roč. 5 [cit. 2. jún 2019]. Dostupné na: <https://adanewmedia.org/2014/07/issue5-raval/>

REAGLE, J., 2013. "Free as in sexist?" Free culture and the gender gap. In: *First Monday* [online]. 2013, roč. 18, č. 1 [cit. 30. marec 2019]. ISSN 13960466. Dostupné na: <https://firstmonday.org/ojs/index.php/fm/article/view/4291>

REAGLE, J., 2017. Naive Meritocracy and the Meanings of Myth. In: *Ada: A Journal of Gender, New Media, and Technology* [online]. 2017, roč. 11 [cit. 2. jún 2019]. Dostupné na: <https://adanewmedia.org/2017/05/issue11-reagle/>

- REAGLE, J. a L. RHUE, 2011. Gender bias in Wikipedia and Britannica. In: *International Journal of Communication* [online]. 2011, roč. 5, s. 1138–1158. [cit. 7. december 2017]. Dostupné na: <http://ijoc.org/index.php/ijoc/article/viewFile/777/631>
- RIDGEWAY, C. a S. J. CORRELL, 2004. *Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations* [online]. 2004 [cit. 7. december 2017]. Dostupné na: <https://journals.sagepub.com/doi/10.1177/0891243204265269>
- ROSS, K., et al., 2018. Women, Men and News. In: *Journalism Studies* [online]. 2018, roč. 19, č. 6, s. 824–845 [cit. 1. apríl 2019]. ISSN 1461-670X. Dostupné na: <https://doi.org/10.1080/1461670X.2016.1222884>
- ROSS, S., 2011. *Wikipedia's gender gap – Wikimedia Blog* [online]. 2011 [cit. 30. apríl 2019]. Dostupné na: <https://blog.wikimedia.org/2011/02/01/wikipedias-gender-gap/>
- ROSENZWEIG, R., 2006. Can History Be Open Source? Wikipedia and the Future of the Past. In: *Journal of American History* [online]. 2006, roč. 93, č. 1, s. 117–146 [cit. 9. jún 2019]. ISSN 0021-8723. Dostupné na: <https://www.sfu.ca/cmns/courses/2012/801/1-Readings/Rosenzweig-%20Can%20history%20be%20open%20source%20.pdf>
- ROWLANDS I., et al., 2008. *The Google generation: The information behaviour of the researcher of the future* [online]. 2008 [cit. 7. december 2017]. Dostupné na: <https://doi.org/10.1108/00012530810887953>
- RUGHINIS, C., 2016. Impromptu Crowd Science and the Mystery of the Bechdel-Wallace Test Movement. In: *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems 2016* [online]. 2016 [cit. 30. apríl 2019]. Dostupné na: https://www.academia.edu/27892427/Impromptu_Crowd_Science_and_the_Mystery_of_the_Bechdel-Wallace_Test_Movement
- SAKASTROVÁ, L., 2011. Měření a hodnocení kvality obsahů moderních kolektivně tvořených encyklopedických děl. In: *Ikaros: elektronický časopis o informační společnosti* [online]. 2011, roč. 15, č. 3 [cit. 30. apríl 2019]. Dostupné na: <http://eprints.rclis.org/15425/>
- SEKAQUAPTEWA, D., et al., 2003. Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. In: *Journal of Experimental Social Psychology* [online]. 2003, roč. 39, č. 1, s. 75–82 [cit. 30. apríl 2019]. Dostupné na: <https://linkinghub.elsevier.com/retrieve/pii/S0022103102005127>
- SELWYN, N. a S. GORARD, 2016. Students' use of Wikipedia as an academic resource — Patterns of use and perceptions of usefulness. In: *The Internet and Higher Education* [online]. 2016, roč. 28, s. 28–34 [cit. 30. apríl 2019]. ISSN 1096-7516. Dostupné na: <https://www.sciencedirect.com/science/article/abs/pii/S1096751615000585>
- SEMIN, G. R. a K. FIEDLER, 1988. The cognitive functions of linguistic categories in describing persons: Social cognition and language. In: *Journal of Personality and Social Psychology* [online]. 1988? roč. 54, č. 4, s. 558-568 [cit. 30. apríl 2019]. Dostupné na: <http://dx.doi.org/10.1037/0022-3514.54.4.558>
- SHANE-SIMPSON, Ch. a K. GILLESPIE-LYNCH, 2017. Examining potential mechanisms underlying the Wikipedia gender gap through a collaborative editing task. In: *Computers in*

Human Behavior [online]. 2017, roč. 66, s. 312–328 [cit. 28. marec 2019]. ISSN 0747-5632. Dostupné na: <https://hal.archives-ouvertes.fr/hal-01612306>

SHAW, D., 2008. Wikipedia in the newsroom. In: *American Journalism Review* [online]. 2008, s. 40–45 [cit. 7. december 2017]. Dostupné na: <http://www.ajr.org/article.asp?id=4461>

SCHRIJVER, P., 2016. *Gender gap on Wikipedia: visible in all categories?* [Bakalárska práca] [online]. Amsterdam: University of Amsterdam. 2016 [cit. 7. december 2017]. Dostupné na: <https://arno.uva.nl/cgi/arno/show.cgi?fid=638722>

SPOERRI, A., 2007. What Is Popular on Wikipedia and Why? In: *First Monday* [online]. 2007, roč. 12, č. 4 [cit. 30. apríl 2019]. Dostupné na: <https://doi.org/10.5210/fm.v12i4.1765>

Stack Overflow 2018 survey: age, gender, sexuality, 2018. [online]. 2018 [cit. 30. máj 2019]. Dostupné na: <https://kaggle.com/heesoo37/stack-overflow-2018-survey-age-gender-sexuality>

ECKERT, S. a L. STEINER, 2013. (Re)triggering Backlash: Responses to News About Wikipedia's Gender Gap. In: *Journal of Communication Inquiry* [online]. 2013, roč. 37, č. 4, s. 284–303 [cit. 24. apríl 2019]. ISSN 0196-8599. Dostupné na: <https://journals.sagepub.com/doi/abs/10.1177/0196859913505618?journalCode=jcia>

TANNEN, D., 1990. *You Just Don't Understand: Women and Men in Conversation*. New York: Ballantine. 1990, s. 330. ISBN 978-0-345-37205-5.

TERRELL, J., et al., 2017. Gender differences and bias in open source: pull request acceptance of women versus men. In: *PeerJ Computer Science* [online]. 2017, roč. 3, s. 111 [cit. 2. jún 2019]. ISSN 2376-5992. Dostupné na: <https://peerj.com/articles/cs-111>

TVARDZÍK, J., 2015. Šéf slovenskej Wikipédie: Vandali nám berú veľa energie a času. In: *Trend* [online]. 2015 [cit. 9. jún 2019]. Dostupné na: <https://www.etrend.sk/ekonomika/sef-slovenskej-wikipedie-vandali-nam-beru-vela-energie-a-casu.html>

VALDROVÁ, J., 2001. Stereotypy a kliše v mediálnej projekci genderu. In: *Sociologický časopis* [online]. 2001, roč. 21, č. 2, s. 23 [cit. 5. jún 2019]. Dostupné na: http://sreview.soc.cas.cz/uploads/a0f454f75498cc92c039c409e5787f65ecff6a6e_149_01-2VALDR.pdf

VALDROVÁ, J., 2018. *Reprezentace ženství z perspektivy lingvistiky genderových a sexuálních identit*. Praha: SLON. 2018, s. 464. ISBN 9788074192616.

VASILESCU, B., et al., 2014. Gender, Representation and Online Participation: A Quantitative Study. In: *Interacting with Computers* [online]. 2014, roč. 26, č. 5, s. 488–511 [cit. 30. máj 2019]. ISSN 0953-5438, 1873-7951. Dostupné na: <https://academic.oup.com/iwc/article-lookup/doi/10.1093/iwc/iwt047>

WALETZKO, A., 2017. Why the Bechdel Test Fails Feminism. In: *HuffPost* [online]. 2017 [cit. 19. jún 2019]. Dostupné na: https://www.huffpost.com/entry/why-the-bechdel-test-fails-feminism_b_7139510

WOOLF, V., 1929. *A Room of One's Own*. The University of Adelaide Library. University of Adelaide Press. 1929, s. 172. ISBN 978-80-272-3567-4.

WAGNER, C., et al., 2015. It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopaedia. In: *ICWSM* [online]. 2015, s. 454-463 [cit. 7. december 2017]. Dostupné na: <http://arxiv.org/abs/1501.06307>

WAGNER, C., et al., 2016. Women through the glass ceiling: Asymmetries in Wikipedia. In: *EPJ Data Science* [online]. 2016 [cit. 7. december 2017]. Dostupné na: <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0066-4>

WIGBOLDUS, D. H., et al., 2000. How do we communicate stereotypes? Linguistic bases and inferential consequences. In: *Journal of Personality and Social Psychology* [online]. 2000, roč. 78, č. 1, s. 5–18 [cit. 24. apríl 2019]. Dostupné na: https://www.researchgate.net/publication/12660299_How_do_we_communicate_stereotypes_Linguistic_bases_and_inferential_consequences

WikiGap 2019: Oslava MDŽ, 2019. [online]. 2019 [cit. 9. jún 2019]. Dostupné na: <http://genderaveda.cz/wikigap-2019-oslava-mdz/>

WIKIMEDIA, 2018. Community Engagement Insights/2018 Report - Meta [online]. 2018 [cit. 30. apríl 2019]. Dostupné na: https://meta.wikimedia.org/wiki/Community_Engagement_Insights/2018_Report

WIKIMEDIA, 2011. *Wikipedia Editors Survey* [online]. 2011 [cit. 30. apríl 2019]. Dostupné na: https://meta.wikimedia.org/wiki/Editor_Survey_2011

WIKIPEDIA, 2019. *Wikipedia: Writing about women* [online]. 2019 [cit. 8. máj 2019]. Dostupné na: https://en.wikipedia.org/w/index.php?title=Wikipedia:Writing_about_women&oldid=894906533

WOOLLEY, A. W., et al., 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. In: *Science* [online]. 2010, roč. 330, č. 6004, s. 686–688 [cit. 4. apríl 2019]. ISSN 0036-8075, 1095-9203. Dostupné na: <https://science.sciencemag.org/content/330/6004/686>

YUN, H. J., et al., 2007. Where Is She? In: *Journalism Studies* [online]. 2007, roč. 8, č. 6, s. 930–947 [cit. 1. apríl 2019]. ISSN 1461-670X. Dostupné na: <https://www.tandfonline.com/doi/abs/10.1080/14616700701556823>

ZAGOVORA, O., et al., 2017. “(Weitergeleitet von Journalistin)”: The Gendered Presentation of Professions on Wikipedia. In: *Proceedings of the 2017 ACM on Web Science Conference - WebSci '17* [online]. 2017, s. 83–92 [cit. 30. marec 2019]. Dostupné na: <http://arxiv.org/abs/1706.03848>

ZICKUHR, K., 2011. Wikipedia, past and present. In: *Pew Internet & American Life Project* [online]. 2011 [cit. 7. december 2017]. Dostupné na: <http://pewinternet.org/Reports/2011/Wikipedia.asp>

ZIMMERMAN, D. H. a C. WEST, 1987. Doing Gender. In: *Gender & Society* [online]. 1987, roč. 1, č. 2, s. 125–151 [cit. 30. apríl 2019]. ISSN 0891-2432. Dostupné na: <https://journals.sagepub.com/doi/10.1177/0891243287001002002>

ŽITNÁ, L., 2013. *Problematika nekredibilných vedeckých a odborných webových sídiel* [Bakalárska práca] [online]. Bratislava: Univerzita Komenského v Bratislave FiF FiF.KKIV. 2013, s. 80 [cit. 9. jún 2019]. Dostupné na:
<http://opac.crzp.sk/?fn=detailBiblioForm&sid=5D21D4E84343AF6ACD944D636BBE&seo=CRZP-detail-kniha>

Zoznam použitých tabuliek

Tabuľka č. 1 Percentuálne rozdelenie používateľov Wikipédie podľa genderu (Protonotarios, et al., 2016, s. 81).....	16
Tabuľka č. 2 Základné porovnanie štatistík pre českú, slovenskú a anglickú Wikipédiu	37
Tabuľka č. 3 Počet biografických článkov na českej Wikipédii.....	54
Tabuľka č. 4 Počet slov v českom mužskom datasete	55
Tabuľka č. 5 Počet slov v českom ženskom datasete	55
Tabuľka č. 6 Počet biografických článkov na slovenskej Wikipédii.....	57
Tabuľka č. 7 Počet slov v slovenskom mužskom datasete	58
Tabuľka č. 8 Počet slov v slovenskom ženskom datasete	58
Tabuľka č. 9 Frekvencie slov (českí muži - gender - podst. mená).....	64
Tabuľka č. 10 Frekvencie slov (české ženy - gender - podst. mená).....	64
Tabuľka č. 11 Frekvencie slov (českí muži - gender - príd. mená).....	65
Tabuľka č. 12 Frekvencie slov (české ženy - gender - príd. mená).....	65
Tabuľka č. 13 Frekvencie slov (českí muži – opačné pohlavie - podst. mená).....	66
Tabuľka č. 14 Frekvencie slov (české ženy – opačné pohlavie - podst. mená).....	67
Tabuľka č. 15 Frekvencie slov (českí muži - rodina - podst. mená).....	67
Tabuľka č. 16 Frekvencie slov (české ženy - rodina - podst. mená)	69
Tabuľka č. 17 Frekvencie slov (českí muži - rodina - príd. mená).....	70
Tabuľka č. 18 Frekvencie slov (české ženy - rodina - príd. mená)	71
Tabuľka č. 19 Frekvencie slov (českí muži - rodina - slovesá)	71
Tabuľka č. 20 Frekvencie slov (české ženy - rodina - slovesá).....	71
Tabuľka č. 21 Frekvencie slov (českí muži - kariéra - podst. mená).....	72
Tabuľka č. 22 Frekvencie slov (české ženy - kariéra - podst. mená)	73
Tabuľka č. 23 Frekvencie slov (českí muži - kariéra - príd. mená).....	74
Tabuľka č. 24 Frekvencie slov (české ženy - kariéra - príd. mená).....	74
Tabuľka č. 25 Frekvencie slov (českí muži - kariéra - slovesá)	75
Tabuľka č. 26 Frekvencie slov (české ženy - kariéra - slovesá).....	75
Tabuľka č. 27 Frekvencie slov (slovenskí muži - gender - podst. mená).....	77
Tabuľka č. 28 Frekvencie slov (slovenské ženy - gender - podst. mená).....	78
Tabuľka č. 29 Frekvencie slov (slovenskí muži - gender - príd. mená)	78
Tabuľka č. 30 Frekvencie slov (slovenské ženy - gender - príd. mená).....	78
Tabuľka č. 31 Frekvencie slov (slovenskí muži - opačný gender - podst. mená)	79

Tabuľka č. 32 Frekvencie slov (slovenské ženy - opačný gender - podst. mená)	79
Tabuľka č. 33 Frekvencie slov (slovenskí muži - rodina - podst. mená).....	80
Tabuľka č. 34 Frekvencie výskytu slov (slovenské ženy - rodina - podst. mená).....	81
Tabuľka č. 35 Frekvencie výskytu slov (slovenskí muži - rodina - príd. mená)	82
Tabuľka č. 36 Frekvencie výskytu slov (slovenské ženy - rodina - príd. mená).....	82
Tabuľka č. 37 Frekvencie výskytu slov (slovenskí muži - rodina - slovesá)	83
Tabuľka č. 38 Frekvencie výskytu slov (slovenské ženy - rodina - slovesá)	83
Tabuľka č. 39 Frekvencie výskytu slov (slovenskí muži - kariéra - podst. mená)	84
Tabuľka č. 40 Frekvencie výskytu slov (slovenské ženy - kariéra - podst. mená).....	85
Tabuľka č. 41 Frekvencie výskytu slov (slovenskí muži - kariéra - príd. mená)	86
Tabuľka č. 42 Frekvencie výskytu slov (slovenské ženy - kariéra - príd. mená).....	86
Tabuľka č. 43 Frekvencie výskytu slov (slovenskí muži - kariéra - slovesá).....	87
Tabuľka č. 44 Frekvencie výskytu slov (slovenské ženy - kariéra - slovesá)	87
Tabuľka č. 45 Percentuálne zastúpenie kategórií – Česko	89
Tabuľka č. 46 Percentuálne zastúpenie kategórií - Slovensko	89
Tabuľka č. 47 Percentuálne zastúpenie kategórií – Česko	91
Tabuľka č. 48 Percentuálne zastúpenie kategórií - Slovensko	91

Zoznam použitých grafov

Graf č. 1 Výrazy, ktorých výskyt je najčastejší v blízkosti slova „žena“ (Mahdalová, 2015) .	35
Graf č. 2 Percentuálne zastúpenie mužských a ženských biografíí na českej Wikipédii	54
Graf č. 3 Tematické kategórie 30 najpočetnejších článkov - český mužský dataset	56
Graf č. 4 Tematické kategórie 30 najpočetnejších článkov - český ženský dataset	56
Graf č. 5 Percentuálne zastúpenie mužských a ženských biografíí na slovenskej Wikipédii ..	57
Graf č. 6 Tematické kategórie 30 najpočetnejších článkov - slovenský mužský dataset	59
Graf č. 7 Tematické kategórie 30 najpočetnejších článkov - slovenský ženský dataset	59
Graf č. 8 Spoluvýskyt slov v českom mužskom datase	61
Graf č. 9 Spoluvýskyt slov v českom ženskom datase	62
Graf č. 10 Spoluvýskyt slov v slovenskom mužskom datase	75
Graf č. 11 Spoluvýskyt slov v slovenskom ženskom datase	76

Zoznam použitých obrázkov

Obrázok č. 1 Náhľad výslednej podoby spracovaného datasetu pomocou jazykového modelu knižnice UDPipe	52
Obrázok č. 2 Ukážka filtrovania v rámci frekvenčnej analýzy.....	52

Príloha 1

Muži sk		
30 najpočetnejších článkov		
Názov článku	Počet slov	Kategória
Bob Dylan	14 395	umenie
Michael Jackson	13 717	umenie
Michelangelo Buonarroti	11 478	umenie
Charlie Chaplin	11 270	umenie
Frank Zappa	10 460	umenie
George W. Bush	8 617	politika
Juan Perón	8 049	politika
Alan Bartlett Shepard	7 508	veda
Zdeňek Svěrák	7 487	umenie
Ján Pavol II.	7 312	náboženstvo
Ludvík Svoboda	7 118	politika
Bedřich Smetana	6 949	umenie
Napoleon Bonaparte	6 890	politika
Charles-Maurice de Talleyrand-Périgord	6 884	politika
Gaius Julius Caesar	6 310	politika
János Esterházy	6 007	politika
Alex Ferguson	5 433	šport
Jicchak Šamir	5 360	politika
Josif Vissarionovič Stalin	5 359	politika
CM Punk	5 086	umenie
Néstor Kirchner	4 952	politika
Wayne Gretzky	4 886	šport
Adolf Hitler	4 844	politika
Henrich VIII.	4 751	politika
Mahmud II.	4 731	politika
Freddie Mercury	4 673	umenie
Mick Jagger	4 515	umenie
Jozef Tiso	4 486	politika
Usain Bolt	4 426	šport
Marian Rejewski	4 289	veda

Ženy sk		
30 najpočetnejších článkov		
Názov článku	Počet slov	Kategória
Marilyn Monroe	10 059	umenie
Maria Callasová	9 403	umenie
Eva Perónová	8 369	umenie
Mária Terézia	5 904	politika
Taylor Swiftová	5 111	umenie
Olivia Newton-Johnová	4 822	umenie
Avril Lavignová	4 777	umenie
Britney Spearsová	4 663	umenie
Cristina Fernándezová Kirchnerová	4 550	politika
Angelique Kerberová	4 546	šport
Beyoncé	4 255	umenie
Belinda Benčíčová	4 033	šport
Tarja Turunen	3 893	umenie
Demi Lovatová	3 882	umenie
Mariah Careyová	3 723	umenie
Selena Gomezová	3 413	umenie
Misty Copelandová	3 383	umenie
Paulína Fialková	3 337	šport
Lindsey Stirlingová	3 145	umenie
Katy Perry	3 087	umenie
Lady Gaga	3 050	umenie
Adele	2 979	umenie
Madame de Pompadour	2 832	politika
Anastasiya Kuzmina	2 791	šport
Dorothy L. Sayersová	2 618	umenie
Mária Lujza Bourbonská	2 486	politika
Katarína Aragónska	2 457	politika
Elizabeth Short	2 449	umenie
Thalía	2 418	umenie
Madonna	2 329	umenie

Muži cs		
30 najpočetnejších článkov		
Názov článku	Počet slov	Kategória
Napoleon Bonaparte	16 634	politika
Ariel Šaron	13 841	politika
Miloš Zeman	13 831	politika
Ned Kelly	13 264	kriminalistika
Bob Dylan	10 888	umenie
Andrej Babiš	10 747	politika
Bedřich Smetana	10 540	umenie
Belisar	10 460	politika
Kyle Busch	10 353	šport
Roger Federer	9 696	šport
Sigmund Freud	9 661	veda
Józef Piłsudski	9 660	politika
Jozef Tiso	9 410	politika
Donald Trump	9 398	politika
Menachem Begin	9 303	politika
Jung-le Hwang	9 245	umenie
Novak Djoković	9 220	šport
Vladimír Franz	8 848	umenie
John Cale	8 754	umenie
Fabrizio De André	8 693	umenie
Václav Klaus	8 642	politika
Tiberius	8 552	politika
Jan Pavel II.	8 541	náboženstvo
Kimi Räikkönen	8 484	šport
Karel IV.	8 450	politika
George W. Bush	8 442	politika
Konstantin I. Veliký	8 384	politika
Alexandr Veliký	8 202	politika
Jan Žižka	8 185	náboženstvo
Nero	8 052	politika

Ženy cs		
30 najpočetnejších článkov		
Názov článku	Počet slov	Kategória
Serena Williamsová	13 702	šport
Petra Kvitová	12 048	šport
Taylor Swift	9 579	umenie
Chana Senešová	9 550	umenie
Ana Ivanovićová	8 844	šport
Maria Šarapovová	8 112	šport
Lucie Šafářová	7 955	šport
Cher	7 349	umenie
Simona Halepová	7 282	šport
Martina Hingisová	6 814	šport
Kim Clijstersová	6 643	šport
Karolína Plíšková	6 106	šport
Kylie Minogue	5 823	umenie
Angelique Kerberová	5 565	šport
Hypatia z Alexandrie	5 514	veda
Katy Perry	5 473	umenie
Angelina Jolie	5 420	umenie
Alžběta Bavorská	5 418	politika
Britney Spears	5 214	umenie
Evženie z Montijo	5 048	politika
Marilyn Monroe	4 988	umenie
Beyoncé	4 831	umenie
Martina Sáblíková	4 792	šport
Lady Gaga	4 762	umenie
Marie Terezie	4 731	politika
Božena Němcová	4 652	umenie
Eugenie Bouchardová	4 644	šport
Edith Whartonová	4 608	umenie
Marie Curie Skłodovská	4 580	veda
Venus Williamsová	4 481	šport

Príloha 2

ČESKÝ SLOVNÍČEK - KATEGÓRIE

GENDER

žena/muž, dívka/chlapec, chlap, slečna, lady/gentleman, paní/pán, dáma

ženský/mužský, dívčí/chlapecký

SOCIÁLNY STATUS A RODINA

dcera/syn, dítě, matka/otec, rodina, manželka/manžel, manželství, vztah, rodič, rozvod, poměr, vdova/vdovec, snoubenka/snoubenec, macecha/otčím, exmanželka/exmanžel, mladomanželka/mladomanžel

provdaný/ženatý, rodinný, svobodný, bezdětný, mateřský/otcovský, rozvedený, zasnoubený, ovdovělý, manželský, rodičovský, mimomanželský, nemanželský

provdát/oženit, rozvést, rozejít, zasnoubit, ovdovět

KARIÉRA

titul, kariéra, cena, práce, studium, úspěch, vítězka/vítěz, ocenění, vítězství, vzdělání, talent, uznání, profese, nadání, odbornice/odborník, respekt, profesionálka/profesionál, šéfka/šéf, vedoucí

úspěšný, profesionální, odborný, vítězný, pracovní, uznávaný, vzdělaný, kariérní, telentovaný, nadaný, oceňovaný, respektovaný, slavný, známý

pracovat, vyhrát, zvítězit, uspět, vzdělávat, respektovat, uznávat

REFERNCIA NA OPAČNÉ POHLAVIE

syn/dcera, manžel/manželka, otec/matka, bratr/sestra, muž/žena, partner/partnerka, milenec/milenka, snoubenec/snoubenka, exmanžel/exmanželka

SLOVENSKÝ SLOVNÍČEK - KATEGÓRIE

GENDER

žena/muž, dievča/chlapec, pani/pán, slečna, lady/gentleman, dáma
ženský/mužský, dievčenský/chlapčenský

SOCIÁLNY STATUS A RODINA

dieťa, dcéra/syn, rodina, matka/otec, vdova/vdovec, manželka/manžel, manželstvo, vzťah,
rodič, rozvod, pomer, snúbenica/snúbenec, macocha/otčim, exmanželka/exmanžel,
mladomanželka/mladomanžel
vydatý/ženatý, rodinný, slobodný, manželský, materský/otcovský, bezdetný, zasnúbený,
rozvedený, ovdovený, rodičovský, mimomanželský, nemanželský
vydať/oženiť, rozviesť, rozísť, zasnúbiť, ovdovieť

KARIÉRA

kariéra, cena, štúdium, práca, uznanie, úspech, titul, ocenenie, vzdelanie, víťazstvo,
víťazka/víťaz, talent, nadanie, profesia, rešpekt, profesionálka/profesionál,
odborníčka/odborník, šéfka/šéf, vedúca/vedúci
profesionálny, úspešný, odborný, pracovný, uznávaný, vzdelaný, víťazný, talentovaný,
oceňovaný, kariérny, známy, slávny
pracovať, zvíťaziť, uspieť, vzdelávať, uznávať, vyhrávať, rešpektovať

REFERNCIA NA OPAČNÉ POHLAVIE

otec/matka, syn/dcéra, brat/sestra, manžel/manželka, muž/žena, partner/partnerka,
milenec/milenka, snúbenec/snúbenica, exmanžel/exmanželka