

Čo môžu „povedat“ vektorové reprezentácie slov (word embeddings) o jazyku a politike?

Jana Wachtarczyková

Abstract

What Can Word Embeddings "Say" About Language and Politics? *The article explores the new possibilities that corpus linguistics offers to other humanities research, specifically what word embeddings allow in clarifying the understanding and usage of linguistic resources in language practice. Results of vector representations of words are presented using the web interface Semä, a publicly available tool developed at the L. Štúr Institute of Linguistics, Slovak Academy of Sciences. The material basis - language corpora utilized by the Semä application - is briefly outlined. We are dedicated to comparing dictionary, encyclopaedic and vector information. Results obtained through vector models for three key self-identifying expressions in the discourse on politics: politika (politics), politický (political), and politik (politician) - are described. These results are understood as a linguistic reflection of a world that differs slightly in each linguacultural environment. Cross-linguistic comparison of vector models shows what is universal and specific for political discourse in the chosen language, the hierarchy of contextually relevant terms, and their structure in the semantic field. The advantages of word embeddings, such as the exactness of results, clarity of outputs, and traceability of specific examples, are highlighted. We note that the in-depth applicability of this method requires interdisciplinarity and competence in both language and the subject discourse, specifically in the area of politics.*

Keywords: word embeddings, contextual semantics, politics, politician.

Kľúčové slová: vektorové reprezentácie slov, kontextová sémantika, politika, politický, politik.

Úvod

Prepojenie jazyka a politiky je bytostne podmienené a existenčných podôb tohto prepojenia je nepreberné množstvo. Celý cyklus konferencií Jazyk a politika dokladá, že „svet politiky je svetom jazyka, že jazyk je najdôleži-

tejším politickým nástrojom a sprievodným javom politiky, pretože každý politický akt musí byť sprostredkovaný jazykom“ (Javorčíková, Kundra 2023: 77). Ide o všeobecne akceptované predpoklady, pri ktorých niet pochybností. Ale neplatí to v určitom zmysle a v istej miere aj naopak? Zjednodušene povedané, že nielen celá politika je „o jazyku“, ale že aj jazyk samotný a komunikácia je vo veľkej miere aj „o politike“? Že politika v najvšeobecnejšom zmysle ako správa vecí verejných, ktorá sa dotýka každého individua v ľudskej spoločnosti, sa nejakým spôsobom premieňa do jazyka, do chápania jednotlivých prvkov jazyka – slov, fráz, nadvetných konštrukcií. Že každý prvok jazyka tým, ako je zasadený do reťazca medziľudskej komunikácie v konkrétnom politickom lingvokultúrnom priestore, je súčasťou naratívov, konceptuálnych schém a mechanizmov skupinovej kognície prebiehajúcich v rámci tohto priestoru. Môžeme sa takto pýtať práve preto, že nielen jazyk politiky, ale celkovo súčasná verbálna komunikácia sa vo verejnom priestore uskutočňuje s dôležitou účasťou funkcií s mocenskou a regulačnou intenciou (napr. takto pôsobí persúázia v mienkotvorných médiách, vytvára sa názorová a postojová polarizácia, prebieha nastavovanie regulatívov v obsahu aj forme komunikovaných obsahov, porov. Cingerová, Dulebová, Štefančík 2021: 39-40).

Jednotlivé prvky jazyka je možné skúmať a definovať ako izolované, teda nezávislé od situačného kontextu, alebo naopak, ako prvky vyslovene determinované kontextom, ako výsledky vyhodnocovania konkrétneho jazykovosituačného kontextu. Jazykové korpuse a aplikácie, ktoré rozličným spôsobom spracúvajú a vyhodnocujú veľké jazykové dáta poskytujú nové pohľady na fungovanie jazykových jednotiek, a to nielen na otázky ich výskytu v jazyku (frekvenciu), ale napríklad aj na otázky distribúcie slov a otázky ich vzájomnej kontextovej súvislosti. Rozsiahle a reprezentatívne korpuse s dobrou anotáciou umožňujú svojmu používateľovi nazerať cez prizmu jazykových dát na rôzne oblasti ľudskej spoločnosti a sledovať, čo tieto dáta vypovedajú o medziľudskej komunikácii, o chápaní jednotlivých slov, ale aj o špecifických spôsoboch verbalizácie myslenia, cítenia a prežívania skutočnosti v každom jazyku.

Čo sú vektorové reprezentácie slov (word embeddings)

V tomto príspevku chceme priblížiť jeden z mnohých pohľadov na jazyk prostredníctvom vektorových modelov slovenčiny (word embeddings¹) a predstaviť, aký typ informácie prinášajú a čo môžu „povedať“ na tému jazyka a politiky v porovnaní s tým, čo „hovoria“ lexikografické alebo encyklopedické uchopenia témy. Vektorové reprezentácie slov predstavujú taký model spracovania slovnej zásoby, v ktorom sú slová alebo frázy prirodzeného jazyka reprezentované vektormi reálnych čísel. Word embeddings patrí k často používaným a progresívnym metódam využitia veľkých jazykových dát. Myšlienka takejto reprezentácie sa od počiatku spája s tzv. distribučnou sémantikou, ktorá sa venuje kvantifikácii a kategorizácii sémantických podobností medzi jazykovými položkami na základe ich distribučných vlastností vo veľkých vzorkách jazykového materiálu. Vychádza pritom z distribučnej hypotézy, ktorá predpokladá, že jazykové položky s podobným výskytom (rozmiestnením v texte, teda distribúciou), majú aj podobný význam. Od prvých úvah v teoretickej rovine (Harris 1954) sa len pomerne nedávno (Mikolov 2013) dospelo ku konkrétnej a praktickej implementácii v prirodzenom jazyku, ktorá dáva veľmi dobré a použiteľné výsledky. Táto vektorová reprezentácia korpusových dát nepotrebuje lingvisticky označovaný korpus, stačia jej dostatočne konzistentne tokenizované textové údaje jazyka, preto sa táto reprezentácia často využíva v oblasti data miningu, vyhovuje aj bežným používateľom, ktorí nemajú k dispozícii nástroje na klasické počítačové spracovanie prirodzeného jazyka alebo potrebné časové možnosti. V našom jazykovednom prostredí využívame na tento účel rozhranie s názvom *Semä*.² To poskytuje nielen prístup k modelom slovenského jazyka, ale aj mnohých ďalších jazykov. Momentálne sú dostupné modely pre angličtinu, arabčinu, bulharčinu, češtinu, čínštinu, estónčinu, fínčinu, francúzštinu, holandčinu, chorvátčinu, latinčinu, lotyštinu, maďarčinu, nemčinu, poľštinu, portugalčinu, ruštinu, španielčinu, taliančinu, slovinčinu

¹ Slovenský termín nie je celkom ustálený, ale v našich prácach používame spojenie „vektorové reprezentácie slov“ ako domáce synonymné spojenie k originálnemu „word embeddings“.

² Prečo *Semä*? Motivácia tohto pomenovania vystihuje zameranie aplikácie, ktorým je zisťovanie vzťahov sémantickej podobnosti medzi jazykovými jednotkami určitého jazyka. V pomenovaní je odlišený názov rozhrania od termínu *séma* a zároveň je ozvláštnený grafémou *ä*, ktorá je v URL adresách neobvyklá, avšak technicky realizovateľná. Rozmanité konotácie archaizmu *semä* v rozličných použitiach (*naše Semä*, *čo hovorí Semä*) sú želaným bočným efektom pomenovania.

a švédčinu.³ V testovacej fáze sú modely pre gruzínčinu a nórčinu. Materiálové zdroje pre vektorové modely vo väčšine jazykov sú založené na rodine korpusov Aranea (Benko 2014). Chorvátsky model je založený na korpuse hrWac 2.0 (Ljubešić, Klubička 2014), slovinský na slWac 2.1 (Erjavec, Ljubešić, Logar 2015) a čínsky na Hanku 1.0 (Gajdoš, Garabík, Benická 2016). Funkcie a možnosti rozhrania sú podrobne opísané v práci Garabík, 2020, jeho využiteľnosť pri onomastickom výskume aj v práci Wachtarczyková, Garabík 2022, jeho možnosti pri medzijazykovom porovnávaní v práci Garabík, Wachtarczyková 2022. Názorne vysvetlenie a opis toho ako vyzerá a funguje Semä (jeho funkcie a moduly) možno nájsť najmä v texte Wachtarczyková, Garabík 2022. Na tomto mieste preto len pripomenieme základné informácie o jazykových korpusoch a ďalej sa budeme venovať interpretácii vektorových modelov výrazov kľúčových výrazov jedného slovotvorného hniezda: *politika – politický – politik*.

Slovenské modely sú natrénované na kombinácii korpusov Araneum Slovacum V Maximum a prim-8.0-juls-all. Ten prvý je webový korpus slovenčiny, spracovaný v rámci rodiny korpusov Aranea (Benko 2014), ktorý zachytáva obraz textov na internete v slovenskom jazyku (nie nevyhnutne z územia Slovenska). Použitá verzia vznikla v januári 2020, teda reflektuje aktuálny jazykový obraz v danom čase a v predchádzajúcich obdobiach. Veľkosť korpusu je 4 miliardy tokenov.

Druhý korpus (prim-8.0-juls-all) je jedným z najväčších korpusov z korpusov Slovenského národného korpusu (Slovenský národný korpus 2018), obsahuje texty beletrie, žurnalistiky a odborné texty, z časového obdobia rokov 1955 – 2018, pričom pochopiteľne množstvo textov výrazne narastá v súčasnosti. storočí či tisícročia. Veľkosť korpusu je 1.6 miliardy tokenov.

Prezentované vektorové modely používajú skip-gram model (okrem FastText modelov) s dimenziou 200 rozmerov a kontextovým oknom 7 slov. Limit pre počet výskytov slova v korpuse je dynamický a pohybuje sa v závislosti od veľkosti korpusu od 20 po 400.

Lexikografická, encyklopedická a vektorová informácia

Kľúčové pojmy politického diskurzu bývajú vágne a v spoločenskej kognícii sa chápu intuitívne. Ak by sme sa pýtali napr. Čo je politika?, Aký je to politický? alebo Kto je to politik? a chceli poznať reprezentatívny (výstižný,

³ Voľbu niektorého zo spomínaných jazykov možno uskutočniť v poli „Jazyk“ na stránke Semä, kde sa v stĺpci zobrazia dostupné jazyky (označené kódom podľa normy ISO 639-1, napr. *cs, pl, ru...*). Pre slovenčinu sú dve voľby – sk alebo sk ll (ktorá vyhľadá aj okrajové a málo frekventované výskyty).

komplexný, všeobecne platný) význam týchto slov, zrejme by sme siahli po slovníku alebo encyklopédii. Aby sme mohli porovnať výpovednú hodnotu rôznych druhov výkladu slov, názorne pripomeňme, akú podobu má informácia o význame výrazu *politika* v slovníku, encyklopédii a vo vektorovom modeli. Slovník slovenského jazyka⁴ ponúka definíciu, ktorá politiku definuje ako reláciu vlády, štátu, spoločenských činností vo vzťahu k obyvateľom (ako „činnosť týkajúca sa... zameraná na ochranu záujmov, uskutočňovanie úloh“).

Obr. 1: Heslová stať politika podľa SSJ

Slovník slovenského jazyka z r. 1959 – 1963⁷.

politika, -y ž.

1. činnosť vlády štátu, politickej strany al. určitej spoločenskej organizácie v oblasti vnútorných a medzinárodných vzťahov zameraná na ochranu triednych záujmov a na uskutočňovanie triednych úloh príslušného štátu, strany al. organizácie: *mierová p.* SSSR; *vnútorná, medzinárodná, zahraničná p.*; *p. z pozície sily* mocenská; *vládna p.*; *stranická, proletárska, robotnícka p.*; *národnostná, školská p.* týkajúca sa národnosti, školstva; *právna, daňová, roľnícka p.*; *kádrová p.* výber odborne a politicky vyspelých pracovníkov; *cenová, mzdová, finančná, hospodárska p., kultúrna p.*; *robiť p-u, miešať sa do p-y*;

2. činnosť zameraná na dosiahnutie niečoho, charakter, spôsob, forma tejto činnosti: *špinavá p.* nečestná; *On, syn, robí si svoju politiku a oni svoju.* (Tat.) *Počínal chápať panskú politiku.* (Taj.) *Býva to najlepšia politika proti bláznom, dať sa im vybláznit' do vôle.* (Záb.)

• *pštrošia p.* neprezieravá, krátkozraké počínanie, majúce tendenciu úmyselne niečo nevidieť;

3. *hovor:* otázka, udalosti spoločenského života, verejného života v štáte: *Maister odišiel do krčmy filozofovať a rozoberať vysokú politiku.* (Grát) *Ich veselý hovor vyhyýbal sa politike a denným biedam.* (Vaj.) *Doma vše čítaval otcovi z novin politiku.* (Taj.)

4. *hovor:* *expr.* chytráctvo, vypočítavosť, prefikanosť: *Viac z politiky než z počtu povinností konala služby.* (Vans) *Čože ty vieš, akú politiku majú takí páni!* (Ligej) *Mišo — vy ste politik!* (Kuk.) *chytrák; v tom je (bude) nejaká p.*

Zdroj: Slovník slovenského jazyka 1963

Encyklopédie prinášajú rozsiahly výklad, ktorý sa vzájomne odlišuje hĺbkou historického exkurzu a množstvom referencií na rôzne prístupy, školy a smery. Definícia politiky napr. v Sociologickej encyklopédii je rozsiahla, vyčerpávajúca, reflektujúca tri odlišné myšlienkové prúdy vymedzenia politiky – ako uplatňovanie moci, ako fungovanie štátu a ako regulatívny systém celospoločenských procesov. Informácia v tejto komplexnosti je však použiteľná skôr na akademické účely a nevyplýva o aktuálnom a typickom význame, v akom sa politika v danom lingvokultúrnom prostredí a v jeho jazykovokomunikačnej praxi najčastejšie chápe a používa.

⁴ Aj keď je definícia z uvedeného slovníka (1963) dobovo poznačená (prívlastkom *triedny* – *triedne záujmy, triedne úlohy*) celkovo je výstižná, rozvinutá, dobre štruktúrovaná a použiteľná aj v súčasnosti.

Obr. 2: Typ akademickej encyklopedickej informácie



politika – (z řec. politiké = správa) – pojem vymezovaný trojím způsobem v návaznosti na odlišné myšlenkové proudy. 1. V intencích „kratocentrické“ tradice, která přiznává primát moci ve spol. a často také v morálně-lidských otázkách a která je spojena se jmény *Thrasymacha* a *Th. Hobbesa*, z moderních sociologů s takovými autory, jako je *W. Ch. Mills*, *I. L. Horowitz*, *R. G. Dahrendorf* nebo i *R. Aron*, je p. „bojem o moc“. Odkaz na moc však problém vymezení p. neřeší, neboť v s-gickém myšlení je kategorie moci jednou z nejsložitějších a nejspornějších, kromě toho tento přístup vyžaduje odlišení polit. moci od ostatních forem moci a navíc moc není nezbytně sama o sobě cílem, má také své instrumentální a cílové aspekty, plní určité funkce, jež je třeba vymežit.

Zdroj: Sociologická encyklopedie

Populárna definícia vo Wikipédii sa podobne snaží predstaviť najčastejšie prístupy k téme a predstavuje politiku ako umenie možného, umenie riadiť štát alebo ako proces a metódu plánovitej činnosti správy celospoločenského poriadku alebo ako úsilie o získanie moci či podielu na moci pri rozhodovaní. Reflektuje históriu aj modernitu politiky a uvádza, že politickými subjektmi môžu byť v demokratickej spoločnosti nielen politici a kolektívne zoskupenia (strany), ale aj občania a kolektívne organizácie.

Obr. 3: Typ populárnej encyklopedickej informácie

Definície [upraviť | upraviť zdroj]

Definícií politiky je mnoho. Mnohí politici a teoretici prízvukujú známu myšlienku nemeckého politika a vojvodu z 19. storočia **Bismarcka** – „Politika nie je exaktná veda... je to umenie možného“. Mnohí považujú politiku za plánovité, organizované a cieľavedomé sociálne konanie, ktoré je zamerané na vybudovanie, udržiavanie alebo zmenu spoločenského poriadku. Iní chápu politiku ako umenie spravovať veci verejné, resp. riadiť štát a realizovať jeho ciele. Podľa ďalších je politika úsilím o získanie moci, jej udržanie, alebo úsilím o získanie podielu na moci.

Zdroj: Wikipédia (<https://sk.wikipedia.org/wiki/Politika>)

Vektorový model – kontextuálny obraz sveta v jazyku

Akú informáciu prinášajú vektorové reprezentácie slov? Dalo by sa povedať, že schematicky načrtávajú kontúry jazykového obrazu sveta a rôznych aspektov chápania a používania konkrétneho výrazu. Vo všeobecnosti sa

vektorové modely jazyka snažia reprezentovať slova ako vektory v mnoho-rozmernom priestore tak, aby vlastnosti vektorov (ako je ich poloha či orientácia) zodpovedali vlastnostiam slov v jazyku.

Konkrétne v prípade word embeddings sa ukazuje, že poloha vektorov odráža sémantické vlastnosti slov alebo ich kontextovú distributívnu ekvivalenciu. Pri vyhľadávaní využívajúc rozhranie Semä (viac Garabík 2020) dostávame kontextuálny obraz hľadaného slova v danom jazyku. Toto automatizované modelovanie vychádza z už spomínanej distribučnej hypotézy, ktorá predpokladá, že výskyt slov v určitom kontexte odráža ich kontextovo podmienenú sémantickú podobnosť (semantics similarity), resp. významovú príbuznosť (semantics relevance). Preto vzťahy medzi získanými vektormi môžeme považovať za dobrý model obsahovej súvislosti a podobnosti. V širšom zmysle (pri širšej interpretácii) je vektorový model zároveň výsekom jazykového obrazu sveta. Manifestuje sa v ňom jazykovo-kognitívna prax používateľov konkrétneho jazyka („prax spriemerovaného autora“ textov v korpuse, ktoré sú materiálovým zdrojom vyhľadávania). Rozhranie Semä pri vyhľadávaní vygeneruje výstup pozostávajúci z dvoch častí:

1. Tabuľka sémanticky najbližších výrazov s počtom výskytov v korpuse (frekvenciou) a odkazmi na externé zdroje.

2. Vizualizácia sémantického poľa podľa zvoleného typu vizualizácie.

Semä – výsledková tabuľka

Prehľadná tabuľka je vygenerovaná pri vyhľadávaní slov a obsahuje 4 stĺpce; riadky tabuľky zodpovedajú slovám z významového okolia vyhľadávaného slova alebo výrazu a sú usporiadané podľa sémantickej vzdialenosti (najbližšie slová sú na začiatku). V stĺpcoch sa uvádzajú zľava: 1. číselné hodnoty sémantickej príbuznosti, 2. slovo, 3. celkový počet výskytov slova v danom korpuse, 4. odkazy na externé zdroje. Stĺpec obsahujúci odkazy na externé zdroje je špecifický pre každý jazykový model v závislosti od dostupnosti zdrojov pre daný jazyk. V prípade slovenských modelov sú tu obsiahnuté nasledujúce odkazy, každý z nich priamo vyhľadá konkrétne slovo z daného riadku tabuľky v danom externom zdroji:

S (slovníkový portál Jazykovedného ústavu Ľ. Štúra SAV),

G (Google, vyhľadávanie obmedzené na doménu .sk),

W (Wiktionary),

a (Araneum Slovacom V Minus),

A (Araneum Slovaca V Maximum),

P (korpus prim-8.0-public-sane oddelenia Slovenského národného korpusu JÚĽS SAV).

Hierarchia výrazov zoradených pod sebou v tabuľke je daná na základe kosínusovej podobnosti (rozhoduje kosínus uhla, ktorý zvierajú vektory, pričom ak je kosínus uhla 0, ide o sémantickú blízkosť, ak je 1, ide o sémantickú vzdialenosť. Vektory, ktoré zodpovedajú sémanticky blízkym slovám zvierajú malý uhol a v tabuľke sú najbližšie pod vyhľadávaným slovom (*politika*).

Obr. 3a: Výsledková tabuľka slovenského vektorového modelu

Obr. 3a: Výsledková tabuľka českého vektorového modelu

☒	slovo	počet	☒	slovo	počet
0.000	politika	862637 S G W A A P	0.000	politika	862348 S G A A W
0.561	reforma	263204 S G W A A P	0.340	politka	918 S G A A W
0.565	stratégia	316019 S G W A A P	0.354	politikum	5771 S G A A W
0.580	diplomacia	57080 S G W A A P	0.580	reforma	235223 S G A A W
0.605	politická	608959 S G W A A P	0.583	zahraničnopolitický	3609 S G A A W
0.610	imigračná_politika	2081 S G W A A P	0.591	hospodársko-politický	632 S G A A W
0.610	zahraničnopolitická	5070 S G W A A P	0.597	diplomacie	28956 S G A A W
0.626	azylová_politika	2113 S G W A A P	0.599	zahraničně	3207 S G A A W
0.628	ekonomika	535639 S G W A A P	0.602	zahraničně-politický	1968 S G A A W
0.630	kohézná_politika	2734 S G W A A P	0.603	strategie	483815 S G A A W
0.632	agenda	87775 S G W A A P	0.607	monetární	9178 S G A A W
0.632	legislativa	169259 S G W A A P	0.616	politický	1101701 S G A A W
0.634	geopolitika	2798 S G W A A P	0.617	ekonomika	435794 S G A A W
0.644	Politika	1894 S G W A A P	0.622	proteccionismus	3056 S G A A W
0.671	presadzovanie	48222 S G W A A P	0.625	politik	530178 S G A A W
0.673	makroekonomická	8671 S G W A A P	0.626	politicko-ekonomický	2164 S G A A W
0.673	inštitucionálna	14506 S G W A A P	0.633	EU	829409 S G A A W
0.674	Spoločný_polnohospodárska	1648 S G W A A P	0.635	zákonodárství	7377 S G A A W
0.674	demokracia	263683 S G W A A P	0.636	vládnutí	23540 S G A A W
0.677	doktrína	22519 S G W A A P	0.639	unie	520885 S G A A W
0.678	štrukturálna_reforma	6449 S G W A A P	0.647	ekonomický	728675 S G A A W
0.698	menová_politika	23869 S G W A A P	0.652	geopolitika	3005 S G A A W
0.699	únia	257077 S G W A A P	0.657	proexportní	1996 S G A A W
			0.658	appeasement	1348 S G A A W
			0.659	vláda	1506140 S G A A W

Zdroj: SEMÄ = <https://www.juls.savba.sk/semä.html>

Je evidentné, že tento výstup nekonkuruje ani nenahrádza slovníkovú alebo encyklopedickú definíciu a vyžaduje ďalšiu interpretáciu a prácu so zdrojmi. Veľmi objektívne (kvantifikovane), konkrétne a spoľahlivo však odpovedá na otázku čo chápeme pod pojmom politika v slovenskom diskurze, s čím sa najvýraznejšie a najtesnejšie *politika* spája, čo evokuje. V slovenskom jazykovom priestore je to výraz *reforma* a následne aj *stratégia* a *diplomacia*. Je pozoruhodné, aj keď nie prekvapivé, že podobne je to aj v českom jazykovom priestore, kde najbližším výrazom s dostatočnou frekvenciou je tiež *reforma*. Na rozdiel od slovenského modelu, v českom modeli sa objavuje aj výraz *politikum*. Pri rozkliknutí zdrojov môžeme pre-

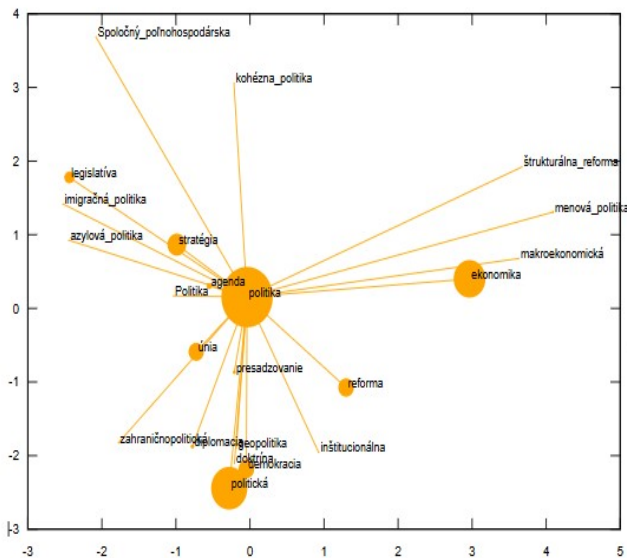
veriť, v akom kontexte sa výraz objavuje a odvodiť z toho ďalšiu interpretáciu: „Školné je nezbytné jako samotná reforma. To, jakým způsobem bylo zapracováno do návrhu reformy, je ale bohužel tristně špatné a jen dosvědčuje, že reforma je politikum, nikoliv odborná snaha za větší kvalitu univerzitního vzdělání.“ Aj tento príklad ukazuje, akým spôsobom je v českom kontexte prepojená politika a reforma. Nemôžeme povedať, že výraz *politika* je ekvivalentný alebo významovo blízky výrazu *reforma*, ale že *reforma* je *politikum*, teda téma politickej agendy (v českom diskurze). V slovenskom modeli sa politika spája s tromi najdôležitejšími pojmami – *reforma* (napr. „Reforma spoločnej poľnohospodárskej politiky je orientovaná na posilnenie konkurencieschopnosti poľnohospodárstva EÚ...“), *stratégia* („...táto aktivita si kladie za cieľ vytvoriť aj priestor na šírenie a zavádzanie ideí komplexnej sociálnej starostlivosti o klientov tak na úrovni ďalších podobných poskytovateľov sociálnych služieb ako aj pre presadenie tohto projektového konceptu na úrovni politik a stratégií.“) a *diplomacia* („Diplomacia je nadstavba spoločnosti, je o tom, že presadzujete politiku a záujmy štátu.“). Kvantitatívne je však *politika* najčastejšie prepojená s výrazom *ekonomika*, a to v slovenskom aj českom vektorovom modeli. Pohľad na praktickú distribúciu výrazu *politika* v citovaných kontextoch potvrdzuje, že jej spojitosť s uvedenými pojmami je natoľko častá a prirodzená, že pôsobí ako všeobecný a univerzálny jav aj v iných jazykoch a vektorových modeloch. Vektorové reprezentácie výrazov len štatisticky vyhodnocujú a približujú výrazy, ktoré sú kontextovo blízke hľadanému výrazu a tvoria jeho kontextové dominanty (vyskytujú sa často) a tiež výrazy, ktoré tvoria tzv. kontextové konštanty.

Za kontextové konštanty môžeme považovať napríklad tie výrazy, ktoré síce nepredstavujú kontextové ekvivalenty alebo významovo podobné slová, ale konštantne (takmer vždy a predvídateľne) sa vyskytujú v daných spojeniach, pričom hľadaný výraz špecifikujú, dotvárajú, precizujú. V prípade výrazu *politika* ide o vyjadrenie typu či oblasti politiky prostredníctvom prívlastkov ako *imigračná politika*, *zahraničná p.*, *azylová p.*, *kohezná*, *makroekonomická p.*, *menová p.* atď. V českom vektorovom modeli sa objavujú podobné prívlastky, aj keď ich vzájomná hierarchia je trocha iná ako v slovenskom modeli. Konštantami sú tu prívlastky *zahraničněpolitický*, *hospodářsko-politický*, *monetární*, *politicko-ekonomický*, *proexportní* a kontextovo najfrekvencovanejší je výraz *ekonomický* (obr. 4b), čo korešponduje s bytostnou previazanosťou *politiky* a *ekonomiky* (obr. 4a).

Semä - vizualizácia sémantického poľa

Výsledkovú tabuľku vo výstupe rozhrania Semä dopĺňa vizualizácia sémantického poľa vektorového modelu,⁵ ktoré rýchlo (na prvý pohľad) poskytuje informácie o sémantických vzťahoch a zoskupeniach slov. Rozhranie môže zobraziť priamo dvojrozmerný obraz, v perspektívnej projekcii trojrozmerný a s využitím farebnej informácie aj ortogonálny priemet štvorrozmerného grafu do trojrozmerného obrazu, ktorý je potom zobrazený na monitore v dvojrozmernej perspektíve. Semä môže vygenerovať aj vizualizáciu v podobe cloudu (statického mračna súvisiacich slov alebo pohyblivého cloudu). V nasledujúcich obrázkoch približujeme vizualizáciu 2D modelu výrazu *politika*.

Obr. 4a: Vizualizácia sémantického poľa výrazu *politika* v slovenskom diskurze



⁵ Keďže model je prirodzene mnohorozmerný, na zobrazenie na obvyklých dvojrozmerných obrazovkách nevyhnutne potrebujeme redukovať počty rozmerov. Na redukciu sa využíva metóda ISOMAP (Tenenbaum, De Silva and Langford 2000).

Vpravo (obr. 4a) leží zóna ekonomiky a sémantické línie, ktoré reprezentujú ekonomický kontext politiky s výrazmi *ekonomika, štrukturálna reforma, menová politika a makroekonomická (politika)*. Zóna v modeli na spodku predstavuje kontexty realizácie a presadzovania politiky v nadnárodnom spoločenstve ako to indikujú výrazy *únia, presadzovanie, reforma, inštitucionálna, geopolitika, reforma*. Zóna v slovenskom vektorovom modeli v strede vľavo predstavuje aspekty politiky ako agendy a legislatívy, teda reprezentuje subkategoriálnu os politiky ako súčasti regulatívneho systému celospoločenských procesov s výrazmi ako *legislatíva, agenda, stratégia* a konkretizáciami ako *imigračná politika* a *azylová politika*. Dve osobitné blízko ležiace osi vo vrchnej časti modelu – *spoločný poľnohospodárska a kohézna politika* predstavujú zrejme osobitný diskurz týkajúci sa agrárnej a environmentálnej politiky EÚ (kohézne fondy). Výhodou porovnávania vektorových modelov je nielen odhaľovanie sémantických konštánt a dominánt, ktoré sa spájajú s používaním hľadaného výrazu v jazykovej praxi (*reforma, ekonomika, diplomacia, stratégia, EÚ, vláda...*) a opakujú sa aj v modeloch iných jazykov, ale aj odhaľovanie špecifik, ktoré niektoré modely majú, ale iné nie. Napríklad v tomto prípade ide o rozdielový výraz *appeasement*, ktorý označuje zmierovaciú politiku a politiku ústupkov voči iným subjektom aj za cenu nevýhodnej situácie s cieľom nevyprovokovať vojenský či politický stret. Tento výraz v slovenskom modeli úplne chýba, ale v iných modeloch (v českom, poľskom, anglickom atď.) je prítomný. Prítomnosť alebo neprítomnosť výrazu v modeli odráža aj vecnú situáciu v realite, teda prítomnosť alebo neprítomnosť situácie *appeasementu* v politike, resp. absenciu tejto tematizácie v slovenskom politickom diskurze.

Súhrnne možno povedať, že informácia o význame výrazov, ktorú prinášajú vektorové modely slov, je komplementárna vo vzťahu k slovníkovej aj encyklopedickej informácii. Zásluhou toho, že slovníky aj encyklopédie sú súčasťou jazykových korpusov, na základe ktorých sa generujú vektorové modely, vektorové reprezentácie tieto informácie inherentne obsahujú a dobre korešponujú s výkladmi v slovníkoch a encyklopédiách konkrétneho lingvokultúrneho priestoru. Na rozdiel od slovníkových a encyklopedických výkladov, kde je hierarchia významov určovaná na základe prijatého autorského konceptu (odráža jazykovú prax cez prizmu tohto konceptu a prijatého systému výkladu), tak vektorové modely automatizovane vyhodnocujú veľké objemy dát, kvantifikujú jednotlivé výskyty a zároveň vytvárajú hierarchiu sémantickej blízkosti slov v danom kontexte bez apriórneho autorského konceptu. Vyhodnocujú tým len samotnú jazykovú prax a umožňujú jednoducho, ale zároveň objektívne odpovedať na otázky typu „čo sa chápe ako politika?“ v konkrétnom diskurze. Slovenský vektorový

model ukazuje, že *politika* sa v slovenskom jazykovom priestore chápe najmä v súvislosti s *reformou*, *politickou stratégiou* a *diplomaciou*.

Pohľad na český vektorový model toto chápanie potvrdzuje a kopíruje – v českom modeli sa *politika* tiež dominantne spája s *reformou*, *diplomaciou* a *stratégiou*. Vektorové modely umožňujú rozsiahlu interpretáciu distribúcie súvzťažných pojmov na naznačených kontextových líniách a kategoriálnych osiach modelu, ktoré sú zovšeobecnením trendov jazykovej praxe. Práca s externými zdrojmi zas umožňuje precizovať a konkretizovať zistenia, vyhľadávať konkrétne použitia výrazov a vytvoriť si komplexnú predstavu o spôsobe používania analyzovaného pojmu napr. prostredníctvom autentických ukážok. Veľkou výhodou využitia vektorových modelov je možnosť medzijazykovej komparácie výsledkov vyhľadávania. Na príklade výrazov *politika* (slov.) – *politika* (čes.) sme predstavili slovenský a český vektorový model a mohli pozorovať, že štruktúra sémantického poľa aj kontextové dominanty sú veľmi podobné (*reforma*, *diplomacia*, *stratégia*), že ako kontextová konštanta vystupuje *ekonomika*, ktorá vytvára osobitnú sémantickú zónu a subkategoriálne línie a na nich súvisiace relevantné výrazy (*hospodársko-politický*, *ekonomický*, *politicko-ekonomický*, *monetárni*, *pro-exportní*). Pri porovnaní výsledkov vektorových reprezentácií slov môže ísť o medzijazykovú komparáciu, ale aj o vnútroparadigmatickú komparáciu viacerých výrazov s rovnakým fundujúcim výrazom (gr. *politiké* od gr. *polis* – občianske spoločenstvo, porov Kardoš 2024), ale s odlišným významom (*politik*, *politický*, *politikum* atď.). Opäť je možné zisťovať a sledovať mieru podobnosti a odlišnosti v štruktúre vektorových modelov a následne sa podujat' na komplexnejšiu interpretáciu zistení. V nasledujúcich pasážach sa v záujme dokreslenia situácie budeme pýtať na význam ďalších autoidentifikačných bodov diskurzu o politike – výrazov *politický* a *politik*.

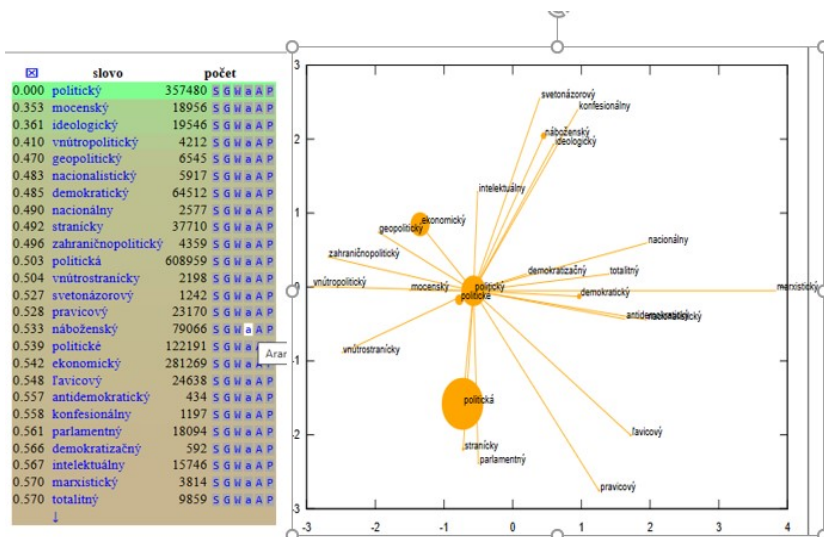
Aký je to „politický“?

V Slovníku súčasného jazyka (1964) je výraz *politický* definovaný ako relácia (prostredníctvom vzťahu k politike), teda ako „týkajúci sa politiky ako činnosti vlády, štátu ai. Slovník⁶ ponúka aj niekoľko výkladov: *verejnoprávny*, *mocenskoprávny*, *štátoprávny* s kvalifikátorom zastaralý. Je ešte niektorý z týchto výrazov v obehú alebo sú kontextovými alternatívami iné výrazy, napr. *reformný*, *diplomatický* alebo *strategický* ako pri slove *politika*?

⁶ Aktuálne synonymá nie sú lexikograficky spracované, pretože Synonymický slovník slovenského jazyka neobsahuje heslo *politický*.

Slovenský vektorový model naznačuje, že adjektívum *politický* má omnoho viac a jemnejšie rozlíšených kontextových synonym než výraz *politika* a sú odlišné. *Politický* znamená predovšetkým *mocenský* („Až doteraz platilo, a ešte väčšinou v mocenskej politike aj platí ono známe Rozdeľuj a panuj! Zaujímavé je, že v prípade Európy sa bude toto heslo meniť zrejme na Spájaj a panuj!“), ale aj *ideologický* („Úlohou SIS je získavať, sústreďovať a vyhodnocovať informácie o extrémistických prejavoch a aktivitách radikálnych skupín tzv. športových fanúšikov, ktoré vychádzajú z rasovej, národnostnej, ideologickej, politickej, sociálnej alebo náboženskej nenávisti alebo nadradenosti.“) a *vnútropolitický* („Prognóza ponúka možné varianty vývoja krajiny v Európe, vnútropolitického vývoja, demografie, ekonomiky, regionálneho rozvoja.“). Uvedené výrazy môžeme považovať za kontextové dominanty vektorového modelu slova *politický* v slovenskom diskurze. Ako konštanty sa javia výrazy *demokratický* a *ekonomický*, pretože previazanosť ekonomiky a demokracie s politickými javmi je v európskom priestore štandardom, ktorý sa premieta aj do vektorového modelu vysokými číslami výskytu v kontexte s výrazom *politický*. Pri pohľade na doklady použitia výrazov v korpusoch sa ukazuje, že uvedené dominanty sa často objavujú v enumeratívnych konštrukciách, takže výraz *politický* nenahrádzajú, ale kontextuálne ho špecifikujú a dopĺňajú o aspekt moci, ideológie a vnútorných záležitostí. Vektorový model zároveň ukazuje, že konceptualizácia hesiel v Slovníku slovenského jazyka bola výstižná a nadčasová (najmä v prípade výkladu *politický* ako *mocenskoprávny*) a zároveň veľmi systematická (dôsledné používanie segmentu *-právny* v definícii), no jazyková prax je predsalen živlejšia a neľadá na koncepty.

Obr. 5: Výsledková tabuľka a vizualizačné pole vektorového modelu slova *politický*



Zdroj: SEMÄ = <https://www.juls.savba.sk/semä.html>

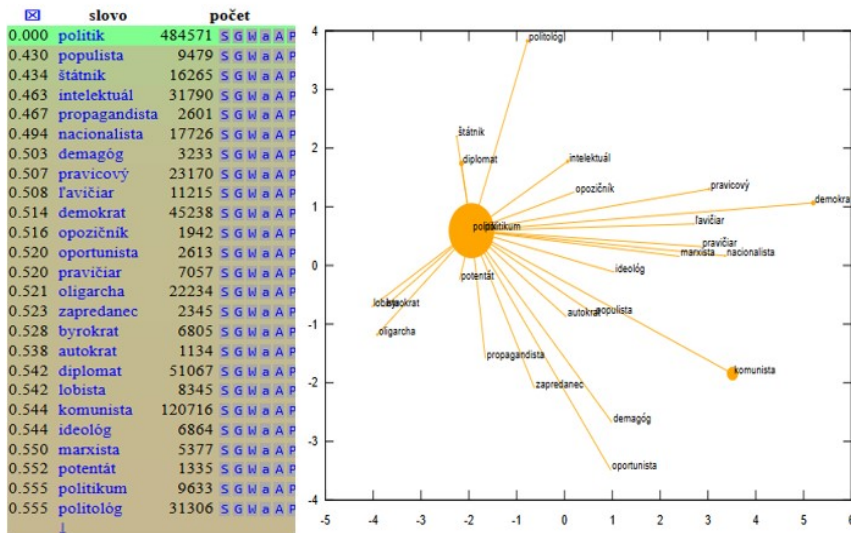
Kto je to „politik“?

Výsledky podľa hierarchie sémantickej príbuznosti ukazujú, že najtesnejšie sa k výrazu *politik* v slovenskom diskurze viaže *populista* (napr. „Je to smutné, ale v politike platí „Havlov paradox“, že demokracia oveľa viac obmedzuje demokratov dodržiavajúcich pravidlá než populistov a autokratov, ktorí sa rešpektovaním pravidiel neunúvajú.“). Nemôžeme ho však považovať za kontextovú dominantu vzhľadom na nízku početnosť výskytov a vzhľadom na negatívnu konotáciu, keďže predpokladáme, že sa vyskytuje v špecifických kontextoch. Kontextové dominanty by sme hľadali skôr v neutrálnych pomenovaniach a vizualizačné pole naznačuje, že by sa mohli nachádzať na kategoriálnej osi s výrazmi *štátnik*, *diplomat* (v hornej časti poľa). Doklady použitia však ukazujú, že *politik* a *štátnik* sa veľmi často vyskytujú v koordinatívnej syntagme, takže sa chápu ako pojmy, ktoré nie sú totožné. *Štátnik* je síce formálne neutrálne pomenovanie, ale zrejme má inherentne pozitívnu expresivitu a viaže sa na reprezentatívne postavy našej histórie („Cieľom takejto mediálnej udalosti by malo byť vniesť do našich ubitých duší aspoň kvapku živej vody – reálnej hrdosti na skutočného slo-

venského štátnika a krásneho človeka“⁷). Podobne je to s výrazom *diplomat* – vyskytuje sa tiež v spojení *politik a diplomat* a má pozitívne konotácie („Vzdelanie, skúsenosti, predchádzajúca kariéra aj sieť personálnych kontaktov dáva každému diplomatovi jeho osobitú individuálnu hodnotu.“).

Vysoký výskyt má aj výraz *komunista* a v tomto prípade môže ísť o aktuálny typ politika, ale skôr o aktéra minulého režimu. Príznačné je, že komunista sa nachádza v rovnakej zóne ako *autokrat* a *populista*, ktorá tak reprezentuje os s negatívnymi konotáciami. Pre slovenský vektorový model je charakteristické, že obsahuje rozsiahlu zónu (spodná časť vizualizačného poľa) s hodnotiacimi výrazmi s inherentnou negatívnou expresivitou ako *zapredanec*, *propagandista*, *oportunista*, *demagóg*, *potentát*, *byrokrat*, *lobbista*, *oligarcha*. Ďalšia zóna (vpravo) obsahuje vnúropolitické špecifikácie k typu politika, klasickú typológiu prostredníctvom výrazov *pravicový*, *ľavičiar*, *marxista*, *nacionalista*, *intelektuál*, *opozičník*.

Obr. 6: Výsledková tabuľka a vizualizačné pole vektorového modelu výrazu *politik*



Zdroj: SEMÄ = <https://www.juls.savba.sk/semä.html>

⁷ M. R. Štefánika.

Obr. 7a, b: Výsledková tabuľka sémantickej blízkosti vektorového modelu na materiáli anglického jazyka a výrazu *politician* (*politik*). Na prvom obrázku (7a) sú slovenské ekvivalenty vyhľadávania, na obr. 7b sú výsledky vyhľadávania v origináli.

☒	slovo	počítač		☒	word	count	
0,000	politik	371309	G a A W	0.000	politician	371309	G a A W
0,304	politik	1383	G a A W	0.304	politican	1383	G a A W
0,414	byrokrat	41715	G a A W	0.414	bureaucrat	41715	G a A W
0,471	demagóg	5699	G a A W	0.471	demagogue	5699	G a A W
0,501	podnikateľ	102727	G a A W	0.501	businessman	102727	G a A W
0,509	pravičiar	4405	G a A W	0.509	right-winger	4405	G a A W
0,511	technokrat	7732	G a A W	0.511	technocrat	7732	G a A W
0,512	liberálny	111224	G a A W	0.512	liberal	111224	G a A W
0,517	zákonodarca	93061	G a A W	0.517	legislator	93061	G a A W
0,523	ideológ	10137	G a A W	0.523	ideologue	10137	G a A W
0,525	ľavicový	16984	G a A W	0.525	leftist	16984	G a A W
0,531	mienkotvorca	701	G a A W	0.531	opinion-maker	701	G a A W
0,531	intelektuál	71135	G a A W	0.531	intellectual	71135	G a A W
0,533	podnikateľov	5854	G a A W	0.533	businesspeople	5854	G a A W
0,534	propagandista	10068	G a A W	0.534	propagandist	10068	G a A W
0,535	patolízal	3884	G a A W	0.535	sycophant	3884	G a A W
0,538	konzervatívny	106314	G a A W	0.538	conservative	106314	G a A W
0,542	štátnik	38698	G a A W	0.542	statesman	38698	G a A W
0,546	učenec	31725	G a A W	0.546	pundit	31725	G a A W
0,550	socialistický	32412	G a A W	0.550	socialist	32412	G a A W
0,560	populista	5043	G a A W	0.560	populist	5043	G a A W
0,563	oligarcha	7658	G a A W	0.563	oligarch	7658	G a A W
0,564	aparátnik	1718	G a A W	0.564	apparatchik	1718	G a A W
0,565	funkcionár	1891	G a A W	0.565	officeholder	1891	G a A W
0,565	poslanca	22443	G a A W	0.565	parliamentarian	22443	G a A W

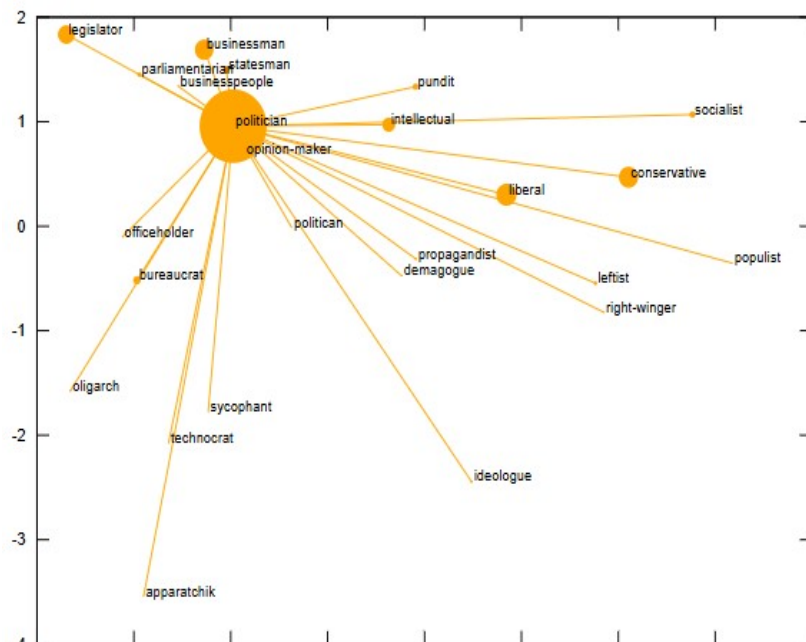
Zdroj: SEMÄ = <https://www.juls.savba.sk/semä.html>

Obr. 8: Vizualizačné pole vektorového modelu *politician*.

Pohľad na výsledkovú tabuľku a hierarchiu významovo relevantných výrazov k slovu *politik* a na štruktúru vizualizačného poľa tohto výrazu na základe korpusového materiálu v angličtine dokladá, že chápanie statusu politika je anglickom lingvokultúrnym prostredím iné a kontextovo blízke sú odlišné ako výrazy v našom diskurze. Najbližšími výrazmi sú *bureocrat*, *demagogue*, *businessman*. Distribúcia kontextovo relevantných výrazov vo vizualizačnom poli však ukazuje, že aj tu sa nachádza zóna koncentrujúca výrazy s negatívnou konotáciou (4 subkategorické osi v spodnej časti poľa ako *burreocrat*, *oligarch*, *technocrat*, *officeholder*, *sycophant* (*pätolízal*), *apparatchik*). Frekventovanejšie a neutrálne kontextové ekvivalenty nachádzame v ľavej hornej časti poľa. Sú to výrazy označujúce politika s rôznymi funkčnými aspektmi, napr. zákonodarným aspektom

tom (*legislator*), s ekonomicko-podnikateľským aspektom (**businessman**), reprezentatívnym a. (*statesman*), s aspektom výkonu politickej moci (*parliamentarian*). Vpravo v strede obr. 9 možno vidieť odraz politického spektra – koexistenciu a typickú kontextovú relevanciu dvojice *liberal – conservative*, ale aj *populist, socialist, leftist a right-winger* a nad nimi výrazy označujúce aktérov z radov intelektuálov (*pundit, intellectual*). V spodnej časti tejto zóny je subkategorická línia politika ako mienkotvorcu (*opinion maker*), *ideológa, propagandistu či demagóga*. Vektorové modely teda odrážajú situačné špecifiká každého lingvokultúrneho priestoru a to, ako sa jednotlivé výrazy chápu a používajú v jazykovej praxi. Prostredníctvom zdrojov poskytujú aj materiál na hlbšiu analýzu tejto praxe, ale kladú nároky na znalosť skúmaného jazyka, ako aj na znalosť mimojazykových okolností, najmä politickej scény a sociokultúrnych podmienok prostredia, v ktorom daný diskurz prebieha. Na elementárnej úrovni však preukazujú, že v každom lingvokultúrnom prostredí sú kontexty s vysokou mierou expresivity (výrazy inherentne expresívne alebo s kontextovo podmienenými negatívnymi konotáciami), osobitne sa formujú neutrálne kontexty a v nich kontextové dominanty výrazu a špecifikačné zóny, ktoré ďalšími výrazmi precizujú, rozširujú a dopĺňajú celkový obraz fungovania daného výrazu v diskurze.

Obr. 9: Vizualizačné pole vektorového modelu výrazu *politician*



Zdroj: SEMÄ = <https://www.juls.savba.sk/semä.html>

Záver

Aký typ informácie teda prinášajú vektorové reprezentácie slov? Sú to primárne informácie o jazyku – o používaní a chápaní jeho jednotiek, ale zároveň o mimojazykovej realite objektov, ktoré pomenávajú. Predstavujú akýsi komprimát všeobecného aktuálneho diskurzu o téme, ako aj lexikografickej a encyklopedickej informácie o objekte, pretože tie sú prostredníctvom definícií vnorené do materiálu, z ktorého sa model generuje. Podstatnou vlastnosťou modelu je jeho názornosť. Výsledky vyhľadávania sa dajú rozmanito vizualizovať. Umožňuje to lepšiu percepciu zistených konfigurácií slov a ich ľahšiu projekciu do vzťahov podobnosti, príbuznosti či iného typu distributívnej ekvivalencie

Prednosťou vektorových modelov je ich exaktnosť, keďže výsledky vyhľadávania sú presne kvantifikované, presne lokalizované, dohľadateľné a znázornené ako celok so špecifickou štruktúrou. Kvantifikácia jednotlivých položiek vo vektorových modeloch a ich exaktnosť približujú jazyko-

vedu ako tradičného zastupcu tzv. mäkkých vied (spoločenských a humanitných) k tzv. tvrdým vedám a pomáhajú objektivizovať výsledky takto orientovaného výskumu a zvyšovať ich sociovedeckú relevanciu.

Výzvou pri využívaní vektorových reprezentácií slov je ich interdisciplinarita a multispektrálnosť. Výsledky vyhľadávania prepájajú technológie korpusovej lingvistiky s diskurzívne a sociolingvisticky orientovanými výskumami, ale aj s výskumami sociálnej kognície a ďalšími humanitnými smermi. Vektorové modely slov odrážajú kontextuálny obraz sveta v jazyku – tak, ako ho ľudia v danej kultúre vidia, chápu a verbalizujú. Prostredníctvom vektorových reprezentácií je možné sledovať apelatíva, ale aj propriá. Pri analyzovaní vlastných mien v konkrétnom lingvokultúrnom prostredí sa rysuje tzv. onymická scéna (porov. Wachtarczyková, Garabik 2022), jej aktéri a ich vzájomná konfigurácia na tejto scéne. Práve to by mohlo byť zaujímavé pre politológov a sociolingvistov, ktorí by pri znalosti politickej scény ľahko a adekvátne dokázali interpretovať vektorové zistenia.

Rovnako podnetné sa ukazuje sledovanie kontextovej distribúcie a sémantických väzieb abstraktných a vágnych výrazov. Na príklade trojice sebaidentifikačných výrazov (*politika*, *politický*, *politik*) sme sa snažili ukázať, že vektorové modely dávajú aktuálne, reálne a jednoduché (nie však povrchné) odpovede na otázky o podstate spomínaných kľúčových výrazov. Zásluhou možnosti medzijazykovej komparácie vektorových modelov toho istého slova v rôznych jazykoch sa dá pozorovať to, čo pre sledovaný fenomén politiky, jej aktérov a kvalít zvaných „politické“ je spoločné a čo odlišné v jednotlivých lingvokultúrnych priestoroch. Vymedzenie a opis kontextových dominánt, konštant, kategoriálnych a subkategoriálnych osí vo vektorových modeloch dáva zas trochu iný pohľad na otázky vzťahu jazyka a politiky v jeho praktickom rozmere.

Vektorové reprezentácie slov môžu byť výzvou pre ich ďalšie využitie napríklad aj pri politologickej reflexii ťažko uchopiteľných vágnych, nejednoznačných alebo inak problematických pojmov ako sú rôzne -izmy (napr. *liberalizmus* a jeho deriváty ako *liberál*, *liberálny* alebo jeho opozitá ako *konzervativizmus*), pri reflexii lingvokultúrnych stereotypov (napr. *Slovák*, *slovenský*) alebo pri skúmaní toho, ako sa chápu územnopolitické celky (napr. *Stredná Európa*, *Felvidék*), zoskupenia (*Visegrád*) alebo udalostí (napr. vojnové konflikty). A napokon, prostredníctvom vektorových reprezentácií slov je možné sledovať jednotlivé výrazy v ich rôznych súvislostiach a interakciách, odhaľovať ich kontextovú expresivitu a techniky verbálnej manipulácie, persúazie a účelovej sugescie v komunikátoch.

Literatúra

- BENKO, V. (2014). Aranea: Yet Another Family of (Comparable) Web Corpora. In: SOJKA, P., HORÁK, A., KOPEČEK, I., PALA, K. (eds.). *Text, Speech and Dialogue. TSD 2014. Lecture Notes in Computer Science*. Cham: Springer: pp. 257–264. https://doi.org/10.1007/978-3-319-10816-2_31
- CINGEROVÁ, N., DULEBOVÁ, I., ŠTEFANČÍK, R. (2021). *Politická lingvistika*. Bratislava: EKONÓM.
- ERJAVEC, T., LJUBEŠIĆ, N., LOGAR, N. (2015). The slWaC corpus of the Slovene Web. *Informatica: an international journal of computing and informatics*, 39(1): 35–42.
- FRANCIS, G. K., BRINKMANN, P. (2009). Human four-dimensional spatial intuition in virtual reality. *Psychonomic Bulletin & Review*, 16(5): 818–823. <https://doi.org/10.3758/PBR.16.5.818>
- GAJDOŠ, Ľ., GARABÍK, R., BENICKÁ, J. (2016). The New Chinese Webcorpus Hanku – Origin, Parameters, Usage. *Studia Orientalia Slovaca*, 15(1): 21–33.
- HARRIS, Z. S. (1954). Distributional structure. *Word*, 10(2 – 3): 146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- GARABÍK, R. (2008). Storing morphology information in a Wiki. In: *Lexicographic Tools and Techniques*. Moscow: IITP RAS, pp. 55–59.
- GARABÍK, R. (2020). Word Embedding Based on Large-Scale Web Corpora as a Powerful Lexicographic Tool. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 46(2): 603–618. <https://doi.org/10.31724/rihji.46.2.8>
- GARABÍK, R., WACHTARCZYKOVÁ, J. (2022). Gender asymmetry of Visegrád group languages as reflected by word embeddings. *Jazykovedný časopis*, 73(3): 354–379. <https://doi.org/10.2478/jazcas-2023-0013>
- Politika (heslo): <<https://sk.wikipedia.org/wiki/Politika>> [23. 1. 2023].
- ISO 639-1:2002 *Codes for the representation of names of languages – Part 1: Alpha-2 code*. International Organization for Standardization.
- JANERT, Ph. K. (2010). *Gnuplot in action: understanding data with graphs*. Greenwich: Manning Publications Co.
- JAVORČIKOVÁ, J., KUNDRÁ, J. (2023). Kritické myslenie a jazyk politiky. In: ŠTEFANČÍK, R. (ed). *Jazyk a politika. Na pomedzi lingvistiky a politológie 8*. Bratislava: EKONÓM, s. 77–85.
- LJUBEŠIĆ, N. – KLUBIČKA, F. (2014). {bs, hr, sr}Wac – Web corpora of Bosnian, Croatian and Serbian. In: *Proceedings of the 9th Web as Corpus*

- Workshop (WaC-9)*, Association for Computational Linguistics, pp. 29–35. <https://doi.org/10.3115/v1/W14-0405>
- KORDOŠ, J. (2024). *Stopy starej gréčtiny v slovenčine*, <<https://www.truni.sk/news/stopy-starej-grectiny-v-slovenčine>> [25. 10. 2022].
- MIKOLOV, T., CHEN, K., CORRADO, G., DEAN, J. (2013). Efficient Estimation of Word Representations in Vector Space. In: *ArXiv*. <https://arxiv.org/abs/1301.3781>
- MIKOLOV, T., GRAVE, E., BOJANOWSKI, E., PUHRSCH, P., ARMAND, Ch. J. (2018). Advances in Pre-Training Distributed Word Representations. In: *arXiv*. <https://doi.org/10.48550/arXiv.1712.09405>
- ŘEHŮŘEK, R., SOJKA, P. (2010). Software framework for topic modelling with large corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Malta.
- PECIAR, Š. (ed.). (1963). *Slovník slovenského jazyka*. III. P – R. (1963). Bratislava: SAV.
- STRMISKA, Z. *Sociologická encyklopedie. Heslo politika*, <<https://encyklopedie.soc.cas.cz/w/Politika>> [23. 10. 2024].
- TENENBAUM, J. B., De SILVA, V., LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500): 2319–2323. <https://doi.org/10.1126/science.290.5500.2319>
- WACHTARCZYKOVÁ, J., GARABÍK, R. (2022). Vlastné mená vo vektorových priestoroch ako výsledky korpusovej inštrumentalizácie. In: *Vlastné mená v interdisciplinárnom kontexte*. Nitra: Univerzita Konštantína Filozofa, s. 78–102.

Poznámka:

Príspevok bol vypracovaný v rámci riešenia projektu VEGA č. 2/0107/23 *Vektorové reprezentácie vágnych výrazov v slovenčine*.

Kontakt:

Mgr. Jana Wachtarczyková, PhD.

Jazykovedný ústav Ľudovíta
Štúra SAV, v. v. i.
Slovenská akadémia vied

L. Štúr Institute of Linguistics
Slovak Academy of Sciences
Bratislava

Email Address: jana.wachtarczykova@juls.savba.sk