

UNIVERZITA KARLOVA

Filozofická fakulta

Katedra Blízkého východu

Bakalářská práce

Aleksandar Columby

IV. ročník prezenční studium

Obor: Blízkovýchodní studia

Integrace arabských výpůjček do češtiny
Integration of Arabic loanwords into Czech

2024

Vedoucí práce: prof. PhDr. Petr Zemánek, CSc.

Prohlášení studenta:

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

Abstrakt

Arabské lexikální výpůjčky jsou přítomny v mnoha slovanských jazycích. I do českého jazyka pronikly, a nadále vstupují, arabská slova. Těmto výpůjčkám v českém jazyce, tedy arabismům, byla věnována pozornost v předešlých výzkumech, ač v malé míře. Tento výzkum se soustřeďuje na současnou vrstvu arabismů, kterým se předešlý výzkum nevěnoval. Konkrétně se pak práce zabývá jejich sémantikou. Autor si klade otázku, zda současné výpůjčky z arabštiny v češtině prochází změnami významu a pakliže prošly změnami, autor se ptá, jakými. Tuto otázku zodpovídá na základě komparace jejich významu vyvozeného z kolokací, do kterých vstupují v korpusech. Pomocí Českého národního korpusu (ČNK) autor nejprve vytvořil seznam současných arabismů a následně jejich kolokace zkoumal v českém korpusu SYN12 a arabském korpusu The Arabic Web Corpus (artenten18). Z výzkumu vyplývá, že většina současných arabismů (12/18) neprošla významnou změnou významu. U všech slov, která změnou prošla (6/18) došlo k *zúžení významu*. Kromě změny významu jeho zúžením, dvě slova (šarī'a a ġihād) nabyly negativního příznaku. I přes jistá omezení, tento výzkum přinesl nové poznatky o současné vrstvě arabismů v českém jazyce a může posloužit jako základ pro další výzkum arabismů v češtině.

Klíčová slova: arabismus, čeština, arabština, kolokace, lexikální výpůjčky

Abstract

Arabic loanwords are present in many Slavic languages. Arabic words have also entered the Czech language, and still do. Arabic loanwords in Czech have been studied previously, however to a small extent. This research focuses on the contemporary layer of Arabic loanwords, which previous research did not deal with. Specifically, the work deals with their semantics. The author asks the question whether contemporary Arabic loanwords in Czech undergo changes in meaning, and if they have undergone changes, the author asks what changes. He bases his answer on a comparison of their meaning deduced from the collocations that they form in corpora. Using the Czech National Corpus (ČNK), the author first created a list of contemporary Arabic loanwords and then examined their collocations in the Czech corpus SYN12 and the Arabic corpus The Arabic Web Corpus (artenten18). The research shows that most contemporary Arabic loanwords in Czech (12/18) have not undergone a significant change in meaning. All the words that have undergone a change (6/18) have narrowed their meaning. In addition to the meaning being changed by narrowing it down, two words (šarī'a and ġihād) have acquired a negative connotation. Despite certain limitations, this research brought new knowledge about the current layer of Arabic loanwords in the Czech language and can serve as a base for further research into Arabic loanwords in the Czech language.

Keywords: Arabic loanword, Czech, Arabic, collocation, lexical borrowings

Obsah

Obsah.....	4
Úvod	6
Formální poznámky	7
1. Stav bádání.....	8
2. Metoda práce	11
2.1. Český národní korpus.....	12
2.2. Korpus SYNV12.....	12
2.3. Arabic Web corpus 2018 (artenten18).....	13
2.4. Práce s korpusem	14
2.4.1. Zadávání dotazů.....	14
2.4.2. Kolokační parametry.....	15
2.5. Zpracování výsledků dotazu, analýza a komparace	15
3. Lexikální výpůjčky.....	17
3.1. Klasifikace lexikálních výpůjček.....	18
3.2. Arabismus jako výpůjčka	19
3.3. Jazykový kontakt Slovanů a Arabů	20
4. Arabismy v českém jazyce	23
4.1. Adaptace arabismů	24
5. Význam podle kolokace.....	27
5.1. Operacionalizace kolokace a kolokační okno.....	28
5.2. Míry asociace a nedostatky.....	29
6. Výsledky analýzy – změny ve významech.....	32
6.1. Zachování významu.....	34
6.2. Zúžení významu.....	35
6.3. Případová studie příznaku – šaría	37
6.4. Novotvar <i>džihádista</i>	40
7. Závěr.....	43
Použitá literatura	45
Slovníky.....	47
Korpusy.....	47
Příloha č. 1	48
Příloha č. 2	49
Příloha č. 3	50

Seznam zkratek

artenten18 – Arabic Web Corpus 2018

BCMS – Bosnian, Croatian, Montenegrin, Serbian

ČNK - Český národní korpus

IC16 – korpus InterCorp v16

i.p.m. – *instances per milion*, jednotka relativní frekvence slova v korpusu

IS- Islámský stát

LCM- Lexical Constellation Model

L3, P3 – rozsah kolokačního okna z levé strany (L) a z pravé (P)

pos – *part of speech*, atribut vyjadřující hodnotu slovního druhu

SYN V12 – synchronní korpus českého jazyka, verze 12

Úvod:

Arabština má nesmírný vliv na ostatní jazyky ve smyslu jazykového kontaktu. Hlavní faktor určující dalekosáhlý vliv arabského jazyka je jeho spjatost s islámem a svatou knihou Koránem, který byl podle tradičního muslimského narativu zjeven v klasické arabštině. Díky tomuto vztahu arabského jazyka a svaté knihy muslimů, nejenže dodnes arabština figuruje v dennodenním životě Arabů a alespoň dvou miliard muslimských věřících, ale od zjevení islámu napříč historií hrála velmi důležitou roli v komunikaci, šíření informací a mezikulturních kontaktech.

Ve většině zemí, do kterých se arabština rozšířila spolu se svými arabsky mluvícími osadníky, se stala jazykem dominantním a zaměnila místní jazyk.¹ Tam, kde se osadníci neusadili, se sice arabština nestala dominantním jazykem, ale sehrála jinou velmi důležitou roli. Versteegh (2015, s. 46-47) tvrdí, že to byl islám, který vytvořil impuls pro postupný přechod arabštiny z podoby lokálního jazyka arabských kmenů do podoby globálního jazyka a následně zmiňuje další aspekt (Ibid, s. 47-50) šíření arabštiny, její tzv. říši učenlivosti (ve Versteeghově originále „*empire of learning*“), která se skrze arabštinu a kolem ní vytvořila. Touto pomyslnou říší Versteegh míní široké pole vědění, duchovního ale i sekulárního, které arabština vytváří a zprostředkovává. Tato říše stále nezanikla a zůstává relevantní převážně v částech Afriky a Asie, a i přesto, že období, kdy evropské myšlení čerpalo z arabských zdrojů, už skončilo, vliv arabštiny na evropské jazyky lze stále rozpoznat. Arabština byla v průběhu historie s různými evropskými jazyky v přímém jazykovém kontaktu v onom obvyklém smyslu. Nicméně také na evropské jazyky působila v rámci nepřímého lingvistického kontaktu, skrze již zmíněný kulturní aspekt.

S příchodem nových kulturních reálií přichází také potřeba tyto nové reálie pojmenovat. Jazykovým zdrojem těchto pojmenování se velmi často stává dominantní jazyk kultury přinášející tyto reálie a právě takto do evropských jazyků vstoupilo plno výpůjček z arabštiny. Arabská slova takto doputovala i do češtiny a dalších slovanských jazyků. Tento proces stále probíhá v současné době, ač v jiné podobě. Právě na tyto současné výpůjčky z arabštiny, se soustřeďuje tato práce.

Motivace napsat tuto práci má původ v kontextu dlouholetého zájmu autora o téma vlivu arabského jazyka na jazyky slovanské. V minulosti se autor již věnoval tématu adaptace arabského

¹ Ovšem v některých oblastech, kde se osadníci usadili, se arabština uchytila pouze na omezenou dobu a následně se obyvatelstvo vrátilo ke svému původnímu jazyku, například Persie (Versteegh 2015, s. 48) či střední Asie.

písmu pro potřeby zápisu jazyka BCMS podobou seminární práce². Nyní se ovšem rozhodl soustředit na jiný aspekt jazykového kontaktu arabštiny a slovanských jazyků a tím je lexikum. Cílem práce je popsat sémantiku současných lexikálních výpůjček z arabštiny v českém jazyce. Autora zajímá, jakých významů tato arabská slova v českém jazyce nabývají. Jaké jsou tedy významové rozdíly mezi těmito arabskými slovy v původním jazyce, arabštině, a jejich vypůjčenými podobami v českém jazyce? Odpověď na tuto otázku formuluje na základě výsledků analýzy komparace sémantiky těchto slov v kontextu původního jazyka a v kontextu jazyka přejímajícího, tedy v češtině. Sleduje Firthovy koncepty o definici významu slova podle jeho kolokací, autor významy vyvozuje z kolokací, které získává z českého a arabského korpusu.

Formální poznámky

Na závěr úvodu chce autor uvést některé formální poznámky k práci. Transliteraci arabštiny autor provádí podle transliteračního systému arabštiny užívaného nakladatelstvím Brill verze 1.0 z prosince roku 2010, ve kterém je každý arabský konsonant přepsán jedním latinským písmenem³ s tím, že autor namísto užití znaku *y* užívá znak *j*. Citace, z textů, které nejsou původně v češtině, přeložil autor sám, pokud není jinak uvedeno. Původní text citace lze vždy nalézt v poznámce pod čarou u dotyčné citace. Citační styl zvolen pro tuto práci je harvardský. V rámci text autor cituje následující formou: (Autor, Rok, s. XX).

² Jazyk BCMS označuje jiho-východo-slovanský jazyk skládající se ze 4 hlavních variet, které tvoří tuto zkratku: *Bosnian, Croatian, Montenegrin, Serbian*.

³ Rietbroek, P. (2010). *Brill's simple Arabic transliteration system, version 1.0*. Dostupné z: https://brill.com/fileasset/downloads_static/static_fonts_simple_arabic_transliteration.pdf

1. Stav bádání

Konkrétní problematika této práce nebyla v dosavadní literatuře přímo reflektována. Dosavadní bádání v oblasti lexikálních výpůjček arabského jazyka v českém jazyce existuje, ovšem není objemné. Výzkumníci se v této problematice zabývali především několika konkrétními oblastmi. Těmi jsou identifikace arabismů v českém lexiku a jejich popis a také mapování historických kontextů vypůjčování slov z arabštiny do češtiny. Problematika arabismů v českém jazyce bývá doposud reflektována v podobě studentských diplomových prací, akademických článků či v podobě jednotlivých kapitol knihy či kompendia, které budou v následujícím textu krátce okomentovány. Nicméně toto téma nikdy nebylo zpracováno v podobě obsáhlejší vědecké publikace.

Výzkum problematik srodných, ač do jisté míry vzdálených, je více rozšířen a má podstatnější výzkumnou historii. Pod těmito odnožemi lingvistického výzkumu má autor na mysli převážně studium arabismů ve světových jazycích Evropy. Zde je důležité poukázat zejména na výzkum arabismů ve španělském jazyce. Kvantita a kvalita tohoto směru bádání vychází převážně z velkého množství výzkumného materiálu (z arabského jazyka) ve španělštině, díky historickému kontextu. Nicméně i vliv a velikost španělského jazyka jakožto jednoho z jazyků s největším počtem mluvčích na světě také jistě hraje roli. Tento faktor se jistě také spoluúčastní na úrovni výzkumu arabismů v anglickém jazyce a dalších světových jazycích. Kromě již zmíněné španělštiny a angličtiny lze zahrnout i francouzštinu a portugalštinu.⁴ I přesto, že se tato oblast výzkumu soustřeďuje na jiné jazyky, některé texty z tohoto proudu bádání posloužily jako inspirace pro napsání této práce. Zejména je třeba zmínit práce soustředící se na arabismy v jiných slovanských jazycích a práce o orientalismech v lexiku slovanských jazyků či češtiny.

Ve slovanských jazycích je arabismů řada. Najdeme je v polštině (Górska 2008), bulharštině (Khayrullina-Valieva 2020, s. 208) a dalších slovanských jazycích. Významný příspěvek do výzkumu arabismů v slovanských jazycích přinesl Blažek (2009) v kapitole *Slavonic Languages*, kde

⁴ S tím, že portugalština je také původně jazykem Pyrenejského poloostrova a podobně jako španělština se těší hojnému arabskému materiálu ve svém lexiku. Obzvláště ve francouzštině jsou zkoumány i arabismy současné, které do francouzského hovorového jazyka vstupují přes arabské diaspory žijící ve Francii. Pro více o vlivu arabštiny na francouzský argot viz Edzard (2012) či Měchurová (2023) pro hlubší pojednání o vlivu arabštiny na francouzštinu od roku 1900 do roku 2022.

představuje přehled arabských výpůjček do slovanských jazyků. Arabismy sledujeme například v jazyku BCMS skrze významný počet lingvistických prací, ve kterých jsou arabismy nejčastěji představeny jako turcismy, občasně jako druh orientalismu a jindy přímo předmětem výzkumu jako arabismy. Škaljić (1966) sestavil nejvýznamnější slovník turcismů a jeho dílo je základním kamenem tohoto bádání. Nicméně byl to Muftić (2002, s. 22), který tvrdil, že ve Škaljićově slovníku je více než polovina slov arabského původu. Proto se rozhodl psát o těchto slovech jako o arabismech, nikoli o turcismech, a popsal především jejich fonetické změny. Do studia arabismů jakožto orientalismů či turcismů vnesli důležité příspěvky také Hanka Glibanović-Vajzović⁵ anebo také Aldin Mustafić, který se převážně zabývá studiem písma arebicy.

Studium orientalismů v českých překladech literárních děl jugoslávských autorů provedli Válek (2011) a Kovar (2015) při Ústavu slavistiky na Masarykově univerzitě. Obě práce jsou převážně translátologické a sledují překlady orientalismů ze srbochorvatského originálu do češtiny a dalších slovanských jazyků.⁶ Je třeba zmínit také příspěvek Karlíkové a Blažka (2008)⁷ ve sborníku *Varia Slavica*. V tomto krátkém příspěvku autoři prezentují seznam arabských slov ve slovanských jazycích, rozdělených podle sémantického okruhu, do kterého spadají. Počet akademických textů zabývajících se přímo problematikou arabismů v českém jazyce je vskutku nízký. Autor si je vědom pouze dvou obsáhlejších prací (Kristýna Hazbunová (2006) a Wail Zahran (2017)), které se zabývají arabismy.

Práce zabývající se arabismy v českém jazyce se téměř zpravidla zabývají arabismy starší vrstvy, tedy těmi, které nejsou současné a které jsou také zdokumentované a popsány v etymologických slovníkách. Dosavadní výzkum nepracuje se slovy ze seznamu autora, některé práce se ovšem alespoň vztahují k sémantice, i když sémantika nikdy není hlavním sledovaným jevem při studiu arabismů. Hazbunová (2006) ve své magisterské práci vytváří seznam arabismů v českém jazyce a poté je analyzuje z hlediska hláskového, etymologického a částečně i sémantického. Hazbunová zmiňuje, že ke změnám ve významu docházelo často, i když byla některá slova převzata beze změny. V rámci 131 studovaných slov Hazbunová identifikuje metonymii jako

⁵ Hanka Glibanović-Vajzović se zabývala zejména sociolingvistickým rozměrem arabismů a různých druhů písma v prostorách Bosny a Hercegoviny. Viz například její práce *“O turcizmima u srpskohrvatskom jeziku sa sociolingvističkog stanovišta”* z roku 1986.

⁶ V případě Válka (2011) i do maďarštiny.

⁷ Sami autoři tento text považují za první text zabývající se tématem arabismů ve slovanských jazycích. Toto tvrzení je jistě diskutabilní, jelikož není jasné, v čem přesně spočívá jeho prvenství.

nejčastější změnu a dále zmiňuje změny následující: metafora, zúžení a rozšíření významu. Zahran (2017) se ve své bakalářské práci zabývá především úlohou al-Andalusu v rozšíření arabismů, frekvencí arabismů v češtině a také hláskovými změnami. K sémantice se Zahran vyjadřuje jen velmi okrajově v rámci odstavce, který důležité poznatky o tématice nepřináší.

Kromě těchto dvou prací se tématem arabismů v českém lexiku zabývá článek *Arabismy v českém lexiku* Heleny Karlíkové, který byl vydán v Novém encyklopedickém slovníku češtiny v roce 2017. Profesorka Karlíková v tomto textu předkládá původy a důvody kontaktu mezi arabskými a evropskými civilizacemi, a způsoby, kterými se jednotlivé arabismy dostávaly do slovanských jazyků a následně také představuje dvě kategorie arabismů vymezených podle sémantických okruhů a poté podává výčet některých těchto sémantických okruhů. Ačkoli se tento článek nezabývá sémantickými změnami při přejímání arabských slov do češtiny, uvádí původní význam arabismu v arabském jazyce a jeho význam v českém lexiku.

Studium arabismů starší vrstvy v českém jazyce zpravidla vychází ze dvou etymologických slovníků českého jazyka. Těmi jsou *Etymologický slovník jazyka českého* Václava Machka (2010) a *Český etymologický slovník* Jiřího Rejzka (2015)⁸. Slovníky zaznamenávají některé arabismy v českém jazyce, nicméně ani nejnovější edice neobsahují úplný seznam arabských výpůjček v českém jazyce. Chybějícím prvkem jsou mimo jiné ony arabismy, které do českého jazyka vstoupily v nedávné době.

Jako obecná literatura o lexikálních výpůjčkách posloužili převážně Haspelmath (2009) a Nekula (2017). Teoretický rámec pro studium významu skrze kolokace se skládá z těchto příspěvků: Firth (1951), Bartsch a Evert (2014), McEnery a Hardie (2019), Xiao (2015), dále Brezina, McEnery a Wattam (2015) a také Pecina (2018).

⁸ Autor uvedl data vydání nejnovějších zredigovaných verzí obou slovníků, se kterými se konzultoval.

2. Metoda práce

Aby mohl autor začít zodpovídat výzkumnou otázku, musel nejprve shromáždit jazykový materiál v podobě arabismů. Jelikož etymologické slovníky češtiny obsahují pouze arabismy staré vrstvy, nemohly posloužit jako zdroj pro sběr arabismů, které do češtiny přešly v nedávné době. Stejně tak tomu je i u dosavadní akademické produkce na toto téma. Kvůli nedostupnosti materiálu v těchto dvou polích se autor rozhodl použít pro jeho určení racionální přístup v kombinaci s hledáním v korpusu. Jelikož je autor rodilý mluvčí češtiny a student arabského jazyka, nejprve identifikoval arabismy, o nichž ví, že se v češtině vyskytují s tím, že vynechal arabismy starší vrstvy. Tato slova poté zadával do aplikace *slovo v kostce*⁹ od Českého národního korpusu (dále už jen ČNK), která umožňuje zobrazit rychlý náhled statistiky slova napříč korpusy českého jazyka v databázi ČNK. Tímto autor potvrdil svou zkušenost s jejich výskytem v českém jazyce, ale také mu tato funkce umožnila sesbírat další arabismy. Funkce *slovo v kostce* obsahuje sekci *podobně používaná slova*, ve které se v podobě dlaždice zobrazují slova kontextově podobná. Tato sekce se tedy stala zdrojem některých arabismů, které autor v předchozí fázi neurčil. Celý seznam zkoumaných arabismů, spolu s jejich relativní frekvencí¹⁰ a orientačním překladem, lze nalézt v příloze č. 1.

Pro zodpovězení výzkumné otázky bylo potřeba zvolit adekvátní jazykový vzorek, ve kterém by bylo možné studovat význam sesbíraných slov. Kromě možnosti studovat dané téma pomocí korpusů, o kterých bude záhy pojednáno podrobněji, existují i další možné vzorky a metody potenciálně vhodné pro tento typ práce. Jednou z nich je práce s webovými vyhledávači, jako je například Google. Tato metoda je sice dostupnou možností poskytující veliký vzorek, nicméně má svá (velká) omezení. Například webový vyhledávač neumožňuje nahlížet na více příkladů výskytu slova paralelně, jako tomu je u korpusových konkordancí. Dále ve vyhledávacích nejsou statistické nástroje, což znamená, že vzorek získaný z nich by byl méně kontrolovatelný. Prohlížeč Google by tedy mohl sloužit nejspíše k vyhledávání a ověřování tipů, nicméně to vše by bylo potřeba poté zpracovat kontrolovanou metodou.

⁹ Aplikace *slovo v kostce* (nebo také WaG) je založena na automatické analýze korpusů a přesnost této aplikace závisí na anotaci slova v daných korpusech. Zdroj pro data je ČNK a data o psaném jazyce pocházejí převážně z korpusu SYN2015.

¹⁰ Relativní frekvence je u slov uvedena v jednotce i.p.m. (z anglického *instances per million*).

Další z možných metod výzkumu je soustředit se na mluvu uživatelů češtiny. Takový lingvistický výzkum by do jisté míry bylo možno provést přímo na datech získaných z přirozené mluvy, nicméně získání mluveného vzorku, který by poskytl reprezentativní jazykový vzorek je velmi obtížné. Všechny arabismy, které tato práce zkoumá, se v češtině užívají pouze od nedávna. Proto znalost těchto slov u mluvčích češtiny nelze očekávat, a tím spíše nelze vůbec očekávat znalost jejich významu. Takovýto výzkum by se významně lišil od výzkumu, kterým se autor zabývá, jelikož by chování arabismů zkoumal v mluveném jazyce, nikoliv v textu. Proto se jako nejlepší možnost pro studium současných výpůjček z arabského jazyka v češtině jeví práce s korpusem.

2.1. Český národní korpus

Český národní korpus (dále již jen ČNK) na svých internetových stránkách (korpus.cz) nabízí širokou škálu korpusů. Kromě korpusů českého jazyka, poskytuje ČNK i korpusy cizojazyčné. Vícejazyčný paralelní korpus *InterCorp v16* (dále už jen IC16) umožňuje zarovnat různojazyčné korpusy vedle sebe a tím slouží jako nástroj pro kontrastivní výzkum. Pro komparaci významů daných arabských slov v arabštině s jejich protějšky-arabismy v češtině by zarovnané korpusy *IC16 - čeština* a *IC16 - arabština* mohly posloužit jako skvělý komparativní nástroj. Nicméně po zadání několika zkušebních dotazů si autor uvědomil, že všechny konkordance měly stejný druh zdroje: „*SUBTITLES*“. Následně po prostudování obsahu korpusu IC16 a jeho individuálních jazykových sekcí bylo zřejmé, že arabština je slabě podporovaný jazyk, jehož obsah v IC16 čítá pouhé tři texty, z nichž pouze jeden je původně psaný arabsky.¹¹ I přesto, že je IC16 nedostatečně podpořen v oblasti arabského jazyka a tím není užitečný v tomto výzkumu, třeba nové verze korpusu *InterCorp* budou rozšířeny o větší počet arabského materiálu a budou uplatnitelnější.

2.2. Korpus SYN12

Jelikož si autor klade otázku o arabismech současných, kritériem pro výběr korpusu byla také jeho současnost. Proto autor zvolil korpus SYN V12, který je nejnovějším korpusem z řady synchronního psaného souhrnného korpusu SYN. Mezi výhody tohoto korpusu a jeho subkorpusů patří relativně objemná jazyková data (v případě SYN V12 se jedná o 5 175 milionů slov¹²),

¹¹ Kromě malého počtu textů je korpus neaplikovatelný v mé práci kvůli nízkému počtu tokenů (126 576) a také nepřesnosti ve srovnání, která vyplývá v první řadě z povahy druhu zdroje textů (filmové titulky) a v druhé řadě ze skutečnosti, že většina arabského materiálu v IC16 je automaticky zarovnána počítačem, nikoli manuálně, čímž se snižuje přesnost zarovnání.

¹² Tento počet slov je nominálně vysoký, nicméně ve světě korpusové lingvistiky je vzorek o 5 milionech spíše malý, obzvláště když mají studovaná slova nízkou frekvenci.

referenčnost a další (příručka ČNK, 2023). SYN V12 obsahuje korpusy SYN2000, SYN2005, SYN2006PUB, SYN2009PUB, SYN2010, SYN2013PUB, SYN2015, SYN2020 a publicistiku z let 2010-2022. Vzhledem k jeho obsahu není reprezentativní, jelikož v něm převažují publicistické texty. Tento fakt nepředstavuje překážku pro tuto práci, jelikož současné lexikální výpůjčky z arabského jazyka lze očekávat právě v publicistických či odborných textech. Nicméně dílčí části SYN V12, tedy subkorpusy SYN200, SYN2005, SYN2010, SYN2015 a SYN202 byly vytvořeny se záměrem být reprezentativní (příručka ČNK, 2023).

2.3. Arabic Web corpus 2018 (artenten18)

Arabský korpus, který autor zvolil pro potřeby této práce, je Arabic Web corpus 2018 (artenten18) webové stránky Sketch Engine. Artenten18 je korpus z řady TenTen korpusů, které se všechny skládají z internetových textů. Korpusy řady TenTen jsou budovány pomocí speciální technologie sběru dat z internetu. Obsahuje 6,4 miliardy slov, je lemmatizovaný a jeho tokeny mají přiřazené morfologické značky¹³, kam spadá také hodnota slovního druhu¹⁴. Je žánrově anotovaný a klasifikovaný dle témat¹⁵. V rozhraní Sketch Engine lze zobrazit kolokáty hledané položky. Kolokační parametry lze nastavit obdobně jako v rozhraní ČNK.

Některé současné arabismy, které se v češtině vyskytují¹⁶, bohužel nemohly být zahrnuty do zkoumaného seznamu arabismů tohoto výzkumu, jelikož je jejich frekvence ve zvoleném korpusu velmi nízká a proto u mnohých z nich nelze určit jejich význam podle kolokací. Je možné, že frekvence těchto slov bude vyšší v nových verzích synchronního korpusu SYN a tím se tato slova stanou použitelná v takovémto typu výzkumu. Seznam současných arabismů, obsahující i slova, která nebylo možno zkoumat v rámci této práce, lze nalézt v příloze č. 2, spolu s jejich relativní frekvencí, která je uvedena vedle slov v jednotce i.p.m. Tento seznam se liší od seznamu v příloze č. 1 tím, že obsahuje slova, která nebylo možno zkoumat. Je třeba poznamenat, že slova *hidžáb*, *burka* a *nikáb* autor začlenil do sémantického okruhu islám, nikoliv do sémantického okruhu oblečení, jelikož jsou výrazně asociovány s islámem.

¹³ Nebo také v korpusovém úzu anglické slovo *tag*.

¹⁴ Hodnotu slovního druhu vyjadřuje atribut nesoucí název *pos*, nebo li "part of speech".

¹⁵ Viz graf klasifikace témat korpusu artenten18 v příloze č. 3.

¹⁶ Jinými slovy: vyskytují se v korpusech českého jazyka, i v korpusu SYN V12.

2.4. Práce s korpusem

V minulé kapitole byly popsány korpusy, se kterými autor pracoval. V této kapitole autor vysvětlí, jakým způsobem s korpusy pracoval. To znamená nejprve vysvětlit, na co musel autor dbát při zadávání dotazů do korpusů a jak nastavil kolokační parametry v obou korpusech. V této kapitole autor paralelně popisuje každé dané téma ve vztahu k českému korpusu i tomu arabskému.

2.4.1. Zadávání dotazů

Dotazy, které autor zadával do korpusu SYN V12, musely zohlednit, že pro některé arabismy existuje více než jeden možný přepis do českého jazyka¹⁷. Aby autor zachytil všechny možné varianty různých slov v jednom dotazu, využil možnosti zápisu dotazu regulárním výrazem. Slova s jedním možným přepisem byla jednoduše zadána do rozhraní v podobě jednoduchého dotazu.

Stejným jednoduchým dotazem tak autor zadával arabská slova do arabského korpusu, když nebylo zapotřebí využít regulárních výrazů, které artentem¹⁸ také dokáže zpracovat. Zadávání některých arabských slov ze seznamu může být poněkud problematické. Tato skutečnost vychází z jistých charakteristik arabského jazyka a písma. Zaprvé slova v arabském jazyce jsou velmi často mnohovýznamová, a proto mohou kolokovat s různými tematickými okruhy. Za druhé arabské písmo ve standardním zápisu, tedy bez vokálových znamének, není morfologicky jednoznačné. Ve vztahu k mnohovýznamovosti se v této práci jistě jedná zejména o arabské ekvivalenty slov *hamás*, *fatah*, *al-Káida* a *hummus* (حمص، القاعدة، فتح، حماس).¹⁸ V případě prvních tří se v českém kontextu jedná o zkratkové názvy politických hnutí a skupin, kdežto v arabském diskursu tato slova také mají zcela jiné a nepolitické významy¹⁹. Co se týče nejednoznačnosti arabského písma, je třeba zmínit slova, která mohou pod vlivem této nejednoznačnosti zkreslit výsledky dotazu. Zde se jedná zejména o arabské ekvivalenty slov *fatah* a *hadždž* (فتح، حج), která v podobě arabského písma mohou být čtena například jako slovesa.

¹⁷ Například arabské slovo *hadždž* se v korpusu vyskytuje ve dvou podobách: *hadždž* a *hadž*.

¹⁸ Je zajímavé se zamyslet, zda by bylo správné sem zařadit i slovo *hummus*, které v arabském originále neodkazuje pouze na onen východo-mediteránský pokrm z cizrny a sezamové pasty, ale odkazuje i na cizrnu jako takovou. Ovšem na rozdíl od prvních tří slov, je zde vztah dvou významů mnohem bližší a kontexty těchto slov jsou si velmi podobné.

¹⁹ Orientační překlady těchto slov lze nalézt v příloze č.1.

2.4.2. Kolokační parametry

Rozhraní funkce kontext ČNK sice disponuje možností upravení hodnot kolokačního okna a také minimální frekvenci zobrazení, nicméně výsledný seznam kolokátů je vždy omezen na pouze 50 položek. Autor se rozhodl nastavit kolokační parametry ČNK na kolokační okno L3/P3 a minimální frekvenci kolokátu 4. Pro toto kolokační okno se autor rozhodl na základě studie Bartsch a Evert (2014). Pro upřesnění výsledků hledání autor využil asociační míry logDice, která v předběžných testech výzkumu stále zobrazovala nejrelevantnější kolokace. V rámci stejných parametrů autor vyhledával i v korpusu artenten18, opět s užitím asociační míry logDice.

2.5. Zpracování výsledků dotazu, analýza a komparace

Ve výsledných seznamech kolokátů autor porovnává kolokace slov. Jelikož se práce soustředí na změny významu arabských slov, poté co byly přejaty do češtiny, v analýze výsledků se autor především soustředí na výsledky českých kolokací. Kvůli morfologické nejednoznačnosti arabského písma autor u arabských kolokátů dbá na rozlišení mezi výsledky, které se pojí k onomu hledanému významu a těmi, které se vyskytují kvůli dalším možným významům slova.

Předtím, než autor analyzuje slova, kolokáty filtruje podle jejich relevance. Slova, která získala skóre míry logDice rovné 0, či vyšší než 0, autor považuje za relevantní kolokace²⁰ a pracuje s nimi v rámci následné analýzy. Slova, pro která byla hodnota logDice nižší než 0 jsou vyřazena z analýzy.

Výsledné kolokace z českého korpusu porovnává s těmi arabskými a určuje, zda došlo ke změně významu slova po transferu do češtiny. Mimo změny významu autor také sleduje změnu příznaku slova. Na základě analýzy výsledků autor vyvozuje závěry o významech současných arabismů v češtině a následně slova agreguje podle sémantických změn do skupin. V rámci závěru autor také poskytuje případovou studii, na které ilustruje rozdíl v příznaku.

Autor měl za záměr doplnit zjištění z korpusů slovníkovými údaji u arabismů v českém jazyce. Nicméně většina arabismů, se kterými autor pracuje, není uvedena v českých etymologických slovnících. Těmito konzultovanými slovníky byli nejnovější verze Machkova

²⁰ Toto minimální skóre autor určil na základě studie Rychlý 2008.

Etymologického slovníku jazyka českého (1971) a Rejzkova Českého etymologického slovníku (2015).

3. Lexikální výpůjčky

Lexikální výpůjčka, nebo také slovo přejaté, je jedním z možných výsledků jazykového kontaktu. Thomason (2001, citovaný v Manfredi a Tosco, 2018, s. 1.) výrazu jazykový kontakt²¹ rozumí následovně:

„nejjednodušeji řečeno, jazykový kontakt je užívání více než jednoho jazyka na stejném místě ve stejnou dobu.“²²

V procesu jazykového kontaktu na sebe dva jazyky působí a navzájem se ovlivňují. Mezi jazyky, které se nacházejí ve stavu kontaktu, dochází k replikám materiálu mezi jednotlivými jazyky. Lexikální výpůjčka je termín pro slovo, jež je přejato z lexika jednoho jazyka do druhého. Nekula (2017a) popisuje, že jazyk, ze kterého je slovo replikováno do jazyka jiného, se označuje za jazyk modelový, či jazyk výchozí nebo zdrojový. Zatímco jazyk, který repliku vytváří je označován za jazyk replikující, cílový či přijímací (Nekula 2017a). Pro potřeby této práce je užitá definice lexikální výpůjčky, která se pokouší vyhnout kontroverzím a být tak co nejpoužitelnější, proto je tato definice poměrně široká:

„Přejímka (či lexikální výpůjčka) je zde definována jako slovo, které v nějakém momentě v historii jazyka vstoupilo do jeho lexika výsledkem půjčování.“ (Haspelmath 2009, s. 36)²³

Z této definice vyplývá pozoruhodný vztah slova přejatého a slova původního. Haspelmath (2009, s. 38) dodává, že nelze vyloučit, že dané slovo není výpůjčkou, tedy že někdy v historii nebylo vypůjčeno z jazyka jiného. Půjčování materiálu jednoho jazyka jazykem druhým je, jak již bylo zmíněno, výsledek jazykového kontaktu. Při jazykovém kontaktu si jazyk přijímající „půjčuje“ materiál či vzor z jazyka zdrojového. Haspelmath (2009, s. 37) podotýká, že termíny *vypůjčka* a *půjčování* jsou, i přes jejich vysokou míru užití, poněkud nepřesné ve svém metaforickém významu, jelikož jazyk modelový slovo neztrácí a jazyk replikující slovo nejspíše nikdy „nevrátí“. Slova a formy, jež přijímající jazyk integruje do sebe, do něj nevstupují nutně na dobu omezenou. Proto

²¹ Pro více o jazykovém kontaktu, viz například *Language Contact* autora Yarona Matrase z roku 2020.

²² Původní text v angličtině: *“in the simplest definition, language contact is the use of more than one language in the same place at the same time”*.

²³ Původní text v angličtině: *“Loanword (or lexical borrowing) is here defined as a word that at some point in the history of the language entered its lexicon as a result of borrowing”*.

Haspelmath navrhuje i jiné termíny jako například anglická slovesa *transfer* či *copy* namísto *borrow*. Stejně tak nacházíme i v češtině ne jeden způsob jak označit tento fenomén. V této práci se proto v tomto kontextu objevují v synonymním významu slovesa *vypůjčit* a *replikovat*. Také slovo *replika* je v tomto kontextu synonymní slovu *vypůjčka*.

Již bylo poukázáno na to, že lexikální výpůjčky jsou výsledkem jazykového kontaktu. Nicméně autor pokládá za věcné a významné se soustředit na možné důvody replikace slov. Obecně platí, že k lexikálním výpůjčkám dochází tam, kde vzniká pojmenovací potřeba (Nekula 2017a). Toto tvrzení je velmi logické, nicméně zakrývá nám fakt, že každý jazyk má prostředky pro tvoření nových slov dle potřeb. Proč tedy sledujeme, že pojmenovací potřebu jazyka řeší prostřednictvím replikace prvků cizího jazyka a nikoliv inovací v rámci jazyka? Haspelmath (2009, s. 47) poukázal na to, že zde jde o jakousi efektivnost v používání jazyka: slovo, jež je více rozšířené mezi mluvčími, bude vždy upřednostněno před jeho ekvivalentem v nějakém jiném jazyce (i kdyby toto známější slovo patřilo do nemateřského jazyka mluvčího). Také podotýká, že jazykové komunity, které jsou etnicky definovány svým jazykem, mohou mít tendenci z důvodů zachování etnické identity nová slova nečerpat z nových jazyků, nýbrž vytvářet kalky neologismů z původních slov (Ibid, s. 47).

3.1. Klasifikace lexikálních výpůjček

Obecně lze výpůjčky klasifikovat do dvou skupin podle toho, zda je přejaté slovo do cílového jazyka přejato spolu s jistým novým kulturním konceptem, pro jehož označení bude použito slovo ze zdrojového jazyka, či zda je přejaté slovo duplikací významu, který již v cílovém jazyce označuje původní slovo. I přesto, že je tato dichotomie základem klasifikace lexikálních výpůjček, všechna přejatá slova v daném jazyce nelze jednoznačně roztřídit do těchto dvou skupin.

Do první skupiny patří takzvané kulturní výpůjčky, které vznikají na základě potřeby mluvčího přejímajícího jazyka pojmenovat novinku ze světa mluvčích jazyka zdrojového. Tyto nové kulturní objekty s sebou často přinášejí i své označení, neboli název, z jazyka zdrojového. Mimo jiné mohou spadat do sémantických polí přírody, gastronomie, kultury, společnosti či technologie (Nekula 2017a).

Druhou skupinou výpůjček jsou takzvané „jádrové výpůjčky“, které při vstupu do cílového jazyka vytváří analog rodnému slovu cílového jazyka, se stejným významem, či rodné slovo zcela

zaměňují. Haspelmath (2009, s. 48-50) tvrdí, že dominantním vysvětlením příčiny existence tohoto druhu lingvistických výpůjček je prestiž, kterou se sebou slovo z modelového jazyka potenciálně nese. Ta je stěžejní pro pochopení jazykovou komunitou vnímanou „správnost“ či nadřazenost jisté jazykové variety, či jevu, nad druhou. V tomto případě se jedná o přejímání slov z cizího jazyka na úkor užívání slov „rodných“, jelikož je jazyková komunita cizího jazyka vnímána jako nadřazená.

3.2. Arabismus jako výpůjčka

Arabismus je v obecném významu jazykový prvek přejatý z arabštiny. Nicméně pro potřeby práce je arabismus koncipován v poněkud užším významu, a to konkrétně ve významu lexikální výpůjčky neboli slova přejatého z arabského jazyka. Nejedná se zde tedy o jiné, potenciální, jazykové prvky jako například fráze, které by byly přejaty z arabského jazyka do jazyka českého. Předpokládejme, že takovýchto nelexikálních arabismů je v češtině velmi málo, pokud vůbec nějaké existují.

V této práci je arabismus definován jako lexikální výpůjčka z arabského jazyka do jazyka jiného. Arabismy byly v evropském vědeckém diskursu tradičně označovány za podmnožinu výpůjček z jazyku východních, tedy orientalismů. Do této nadmnožiny spadají výpůjčky z jazyků Asie a Afriky. Každý jeden diskurs o orientalismech (jakožto lingvistických výpůjčkách z Východu) definuje orientalismy, a pracuje s nimi, podle svých potřeb. Například za orientalismy v jiho-východoslovanských jazycích jsou považovány výpůjčky z turečtiny, perštiny a arabštiny. S tím, že v tomto kontextu je výraz turcismus často užíván jako synonymum slovu orientalismus, nejspíše kvůli tomu, že turečtina byla hlavním jazykovým zprostředkovatelem mezi jiho-východoslovanskými jazyky a perštinou či arabštinou. Nicméně termín orientalismus je v lingvistickém kontextu nadmíru nepřesný, či spíše nicneříkající a vychází ze zastaralého evropského orientalistického přemýšlení, které má tendenci takzvaný „Východ“ vnímat mnohem jednotvárněji, než popravdě je.²⁴

Slova, jež jsou považována za arabismy napříč světovými jazyky nejsou vždy nutně ryze arabského původu. Jejich první záznamy nemusí být v jazyce arabském, nýbrž mohou být poprvé zaznamenána v jiném jazyce a být tedy nearabského původu. Nicméně posléze se v arabském jazyce míra jejich užití zvýšila či byla arabským jazykem přenesena do jazyků dalších, které tyto

²⁴ Diskuzi o orientalismu započal Said v knize *Orientalismus* (1978), pro více o orientalismu (a balkanismu) a jeho podobách specificky v rámci jiho-východo-evropského kontextu, viz Todorova (1997) či Bakic-Hayden (1995).

výpůjčky přijaly jako přicházející z arabštiny a označili je za arabismy. Mezi tato slova patří mj. některá slova řeckého (*karát*) či perského původu (*tác*). Z tohoto důvodu nelze tvrdit, že všechny arabismy jsou slova arabského původu, i přesto, že velká část z nich arabského původu je. Arabština je jedním z 5 jazyků, které měly obrovský význam jako nositelé kultury (Hoffer 2002, s.4). Arabská *říše učenlivosti*²⁵, stejně tak jako první muslimské chalífáty, nevznikla ve vakuu, ale v oblasti, ve které již několik tisíciletí kultury vznikaly a zanikaly. Zejména pak v období Abbásovského chalífátu docházelo k překladům významných děl a na základě vědění zpřístupněného takzvaným překladatelským hnutím v Abbásovském chalífátu²⁶ se abbásovští myslitelé, a následně i ti evropští, významně podíleli ve vědeckém pokroku. Mohadi (2019, s. 1299-1300) podotýká, že po období překladatelského hnutí měli myslitelé abbásovského chalífátu velmi významnou příležitost asimilovat perské a řecké vědění, adaptovat jej a na těchto poznáních tvořit nové vědění. Lze předpokládat, že v procesu seznamování se s těmito cizími texty a také v procesu jejich překladu došlo k akumulaci cizích slov, perského či řeckého původu, která se následně skrze arabštinu rozšířila dále. Například Belkin (1987 zmíněn v Asatrian 2006, s. 98) považuje období abbásovského chalífátu za období největšího přílivu persianismů do arabštiny. Nicméně Asatrian (2008, s. 99) Belkinovu periodizaci považuje za nepodpořenou daty.

3.3.Jazykový kontakt Slovanů a Arabů

První zaznamenaný fyzický kontakt mezi Slovy a Araby byl styk slovanských žoldněřů ve Východořímské říši v bitvách s Araby. Za první písemnou zmínku o těchto kontaktech se považuje text *Chronographia* byzantského kronikáře Theofana Homologeta (Karlíková 2017).

Dalším prostorem styku Arabů a Slovanů byl Pyrenejský poloostrov od příchodu Arabů na poloostrov a vzniku al-Andalusu. Slované plnili různé společenské funkce napříč stoletími existence Al-Andalusu. Jejich úlohy ve společnosti se původně odvíjely od jejich otrocké identity, nicméně později byli mnozí statusů otroka zbaveni a plnili významnější společenské role, například ty politické. Podle Jankowiaka (2017, s. 170) lze na základě archeologických nálezů tvrdit, že slovanští

²⁵ V originále se jedná o termín *empire of learning* z článku *An empire of learning: Arabic as a global language* (Versteegh 2015).

²⁶ Pro více o tomto tématu viz Dimitri Gutas. *Greek Thought, Arabic Culture : The Graeco-Arabic Translation Movement in Baghdad and Early 'Abbasid Society (2nd-4th/5th-10th C.)*. Routledge; 1998. Accessed April 12, 2024.
<https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=460204&lang=cs&site=ehost-live>

otroci byli převážně dováženi z území mezi řekou Labe a ruskými řekami. Zajímavostí je, že na obchodní cestě z těchto krajin směrem na Pyrenejský poloostrov hrálo důležitou roli město Praha²⁷, které bylo místem pro prodej otroků. I přes to, že je styk Arabů a Slovanů v al-Andalusu zdokumentovaný, to samé o jejich jazykovém kontaktu říci nelze. Vycházejí z aktuálního stavu výzkumu autorovi nezbývá, než se domnívat, že jazykový kontakt probíhal ve velmi omezené míře, která nebyla jakkoli významná. Jazykový odkaz al-Andalusu ve slovanských jazycích není výsledkem jazykového kontaktu s arabštinou, nýbrž s jazyky, které španělské arabismy zprostředkovaly. Hluboký otisk, který zanechala arabština na španělštině, má své kořeny v lingvistické situaci al-Andalusu, která je velmi komplexní a nabývá různých podob v různých obdobích. V této multi-etnické a nábožensky rozmanité společnosti byly různé variety hovorové arabštiny a variety románštiny v blízkém jazykovém kontaktu, jehož výsledkem bylo významné vzájemné ovlivnění těchto jazyků a také vznik různých kontaktních variet (Vicente 2020, s. 228-240). Jelikož byla arabština považována za prestižní jazyk, hlavní směr výpůjček byl směrem z arabštiny do románských variet.

Následně považuji za velmi důležité zmínit kontakt slovanských jazyků s arabštinou v osmanské společnosti. Zde docházelo za prvé ke kontaktu zprostředkovanému osmanštinou, a za druhé k přímému kontaktu mezi slovanskými jazyky a lingvistickým materiálem arabštiny. K přímému kontaktu mezi slovanskými jazyky a arabštinou docházelo nejčastěji ve slovanských muslimských komunitách, jejichž členové se vzdělávali v islámských vzdělávacích institucích, neboli *medresách* a *mektebech*. Zde zcela přirozeně přicházeli do kontaktu s arabským jazykem, který jakožto jazyk Koránu zaujímá výjimečnou pozici v islámském vzdělání. Mnozí slovanští muslimští intelektuálové své vysokoškolské vzdělání získali v Istanbulu (Alispahić, 2014, s. 14) a během svých studií se naučili ovládat osmanštinu, perštinu i arabštinu. V těchto jazycích také psali a tvořili, nicméně právě arabština a arabské písmo sehrálo důležitou roli v kulturní produkci v jejich rodném jazyku. První pokusy o standardizaci persko-arabského písma pro zápis jazyku BCMS na území Bosny byly provedeny v 19. století (Hadzijačić 1955), nicméně písmo bylo volně užíváno pro zápis místních variet již od začátku 16. století (Mustafić 2013, s. 69).

²⁷ Pro více o tomto tématu viz Charvát a Prosecký (1996).

V moderní době, se všemi technologickými výdobytky v oblasti komunikace, geografie není rozhodujícím faktorem v uskutečňování přímého jazykového kontaktu. V době, kdy informace mezi jedinci proudí čím dál rychleji, není divu, že se spolu s obecným trendem globalizace ve 21. století globalizují i jazyky novými způsoby. Klein-Franke (1980, citovaný v Versteegh, 2015, s. 46) tvrdí, že arabská *říše učenlivosti* ztratila svou klíčovou pozici v produkci západo-evropského vědění po pádu Konstantinopole v roce 1453. S postupným šířením koloniálních imperií a zavedením koloniálních vzdělávacích systémů v Africe a Asii si nelze nepovšimnout také rozmachu oněch evropských *říší učenlivosti*. Ty nejenže suplovaly pozici oné říše arabské v mnoha kontextech a tím její vliv oslabily, ale hlavně se samy posílily a zafixovaly své pozice (Versteegh 2015, s. 51). Důkazem nám může být anglická *říše učenlivosti*, která je bez pochyb dominantním kanálem šíření informací a vědeckých poznatků současné doby. Jazykový kontakt slovanských jazyků a arabštiny je v kontextu současné doby výrazně jiný ve způsobech, jakými se k informacím mluvčí jazyků dostávají. Moderní slovní zásoba přichází v zásadě napřímo, například díky turismu či faktu, že lze čerpat přímo z původních zdrojů.

4. Arabismy v českém jazyce

Arabština není velkým zdrojem cizího lexika češtiny. Arabismy tvoří, v porovnání například s německými výpůjčkami či francouzskými, malou část českého lexika. Některé arabismy v českém jazyce označují kulturní reálie či náboženské koncepty spjaté s arabskou kulturou a arabským světem. Na rozdíl od těchto slov (*mešita, imám*), která lze poměrně jednoduše identifikovat jako arabismy, jsou v češtině také arabismy, které jsou tak frekventované a zažitě v češtině (*cukr, župan*), že by se mnozí rodilí mluvčí jistě divili původu těchto slov. Zde je potřeba zdůraznit, jak již bylo zmíněno v kapitole *Arabismus jako výpůjčka*, že ne všechna slova, která považujeme za arabismy či arabské výpůjčky jsou vždy nutně ryze arabského původu. To, že je označujeme za arabismy vyplývá alespoň ze dvou skutečností. Buďto se významné rozšíření těchto slov do dalších jazyků odehrálo až po převzetí do arabštiny²⁸ anebo byla tato slova za arabismy označena mluvčími daného jazyka, jelikož do jejich jazyka byla přejata z arabštiny. Také je třeba podotknout, že si autor není vědom jiných výpůjček z arabštiny než těch lexikálních (například přejímek gramatického materiálu)²⁹. Tuto skutečnost lze také vysvětlit tím, že čeština nebyla s arabštinou v přímém kontaktu po delší dobu.

Jelikož mluvčí češtiny nebyli spolu s mluvčími arabštiny v situaci významného jazykového kontaktu, čeština přijala relativně nízký počet lexikálních položek, a většinu z nich zprostředkovaně. Zprostředkovateli těchto výpůjček byly druhé jazyky, skrze které daný arabismus vstoupil do češtiny. Do slovanských jazyků arabismy vstoupily jednou ze dvou cest: západní či východní. Schéma těchto dvou cest bylo načrtnuto následně:

*„Západní – (španělština) → francouzština či italština → (němčina či nizozemština) →
(zejména) západní & východní slovanské jazyky;
Východní – (perština) → turečtina (výjimečně tatarština) → (střední řečtina) → (zejména)
jižní & východní slovanské jazyky.“*

(Karlíková a Blažek 2008, s. 121)

²⁸ Slovo *cukr* je jedním z těchto slov. V tomto případě bylo dříve než v arabštině zaznamenáno v sanskrtu a v perštině (Hazbunová 2006), nicméně protože se do evropských jazyků rozšířilo skrze arabštinu, je považováno za arabismus.

²⁹ Z arabštiny nicméně byla přejata některá slova, která jsou v arabském jazyce ve stavu genitivní konstrukce tvořené více než jedním slovem. Pro příklad viz poznámka pod čarou č. 35.

Toto binární schéma je praktické ve své jednoduchosti, nicméně některá slova v českém jazyce do tohoto schématu nezapadají a schéma je vcelku obecné.³⁰ Hazbunová (2006, s. 167-169) při popisu cest arabismů do českého jazyka sleduje podobné tendence jako Karlíková a Blažek (2008), zejména u cesty západní s tím, že se zmiňuje o trasách méně častých, jako například o trase přes západní Afriku či nedostatečně probádané trase, jež vedla přes Černé moře a přes ruštinu. Hazbunová (2006, s. 168-169) dále dodává, že arabismy ve slovanských jazycích nejsou velmi probádané téma a pro zmapování dalších cest je potřeba téma více zkoumat.

4.1. Adaptace arabismů

Adaptace cizích slov do daného jazyka má vícero aspektů. Adaptace může mít podobu výslovnostní, pravopisnou, syntaktickou i morfologickou (Nekula 2017b). Plně adaptované výpůjčky, nebo také *pravé výpůjčky* jsou ty, které jsou adaptovány na všech zmíněných úrovních a jsou vnímány jako součást slovní zásoby (Nekula 2017b). Na druhé straně škály přizpůsobenosti lexikální výpůjčky stojí *cizí slovo*, což je výpůjčka, která není zcela adaptována. Nekula (2017b) tvrdí, že klasifikace druhů výpůjček by se nicméně neměla určovat formálně, ale spíše z hlediska sociolingvistiky, jelikož pociťovaná cizost slova je spíše to, co odlišuje cizí slovo od plně adaptované výpůjčky. Na základě těchto informací, je zřejmé, že arabismy v českém jazyce lze rozdělit do těchto dvou skupin. O to se zatím nikdo nepokusil, podle informací dostupných autorovi.³¹ Stojí nicméně za zmínku, že současné arabismy, kterými se autor v práci zabývá, mohou být s velkou pravděpodobností pociťovány jako více cizí, nežli arabismy vrstvy starší, které jsou po výrazně delší dobu používány mluvčími češtiny nežli slova nově příchozí, a tím jsou spíše pociťovány jako slova domácí, původní.

Arabské výpůjčky v českém jazyce jsou všechny adaptovány syntakticky a morfologicky. Stejně tak je i většina arabismů v českém jazyce plně adaptovaná pro fonologický systém češtiny. Zde je potřeba zdůraznit, že arabismy v českém jazyce nebyly přizpůsobeny pouze češtině, nýbrž se postupně foneticky přizpůsobovaly každému nearabskému jazyku, ve kterém se vyskytly, a český jazyk je přijímal již ve změněné podobě, proto je obtížné pozorovat změny pouze v češtině (Hazbunová 2006, Karlíková 2017). Karlíková (2017) podotýká, že mnohé arabismy v češtině lze

³⁰ Tato obecnost je nejspíše také způsobena obecností kategorie „slovanské jazyky“.

³¹ Tato práce nicméně poskytne podklady pro takovýto směr badání.

považovat i za internacionalismy či evropeismy, jelikož jsou tato slova rozšířená v mnoha jazycích Evropy i světa.

Co se týče změn u samohlásek, dochází ke změnám jejich kvality i kvantity.³² Většina změn kvantity nabývá podoby krácení, a k dloužení dochází v mnohem menší míře (Hazbunová 2006, s. 163). Je více způsobů, jak se změnila kvalita hlásek. I když mezi nejčastější změny patří změna $a > e$, změna $u > o$ (alkohol [al-kuḥūl], Korán [qurʔān]) a změna $i > e$ (hena [ḥinnāʔ]), Hazbunová (2006, s. 164) pozoruje také změny méně časté: $a > e > i$ (cibetka [zabad]) či $i > e > a$ (damašek [dimašq]).

U souhlásek došlo k změnám výraznějším a četnějším, nejen protože se arabský systém konsonantů od toho českého liší ve větší míře nežli systém vokálů. Mezi pravidelné změny konsonantů lze zahrnout ztrátu emfáze (*tác* [tās(a)]), ztrátu geminace u starších arabismů (*župan* [ḡubba]), kompletní ztrátu znělé faryngály $ʕ$ (*loutna* [al-ʕūd]), změnu znělé faryngály $ħ > k$ (*kamelot* [ḥammāl]) či $ħ > h$ (*saharský* [ṣaḥrāʔ]), změnu znělé uvuláry $q > g$ (*gáza* [qazz]) či $q > k$ (*káva* [qahwa]), změnu neznělé uvulární frikativy $x > g$ (*magazín* [maxāzin]) anebo $x > k$ (*moka* [al-muxā]) a změnu frikativní uvuláry $ǧ > g$ (*gazela* [ǧazāl]) či $ǧ > g > k$ (*karafa* [ǧarrāfa]) (Hazbunová 2006, s. 164-166).

Z hlediska morfologického prošla arabská slova v češtině četnými změnami. Hazbunová (2006, s. 169-172) za časté (a standardní) změny považuje změnu rodu, změnu čísla a změnu třídy. Za zmínku stojí zajímavé změny, sice méně časté, ale mající vztah ke specifikům arabské morfologie. Například v případě slova *admirál* došlo k převzetí genitivní konstrukce³³, která se přeměnila v podstatné jméno (Hazbunová 2006, s. 171). Další zajímavou změnou je arabský určitý člen, který se s určitými hláskami v arabštině asimiluje. Kvůli těmto dvěma možnostem výslovnosti určitého členu dochází také k rozmanitosti změn při přejímání arabských slov s určitým členem (Hazbunová 2006, s. 171). Hazbunová (2006, *ibid*) pozoruje, že pokud určitý člen není v arabském originále asimilován, může se člen u některých výpůjček jasně projevit (*albatros*, *alkovna*) či po něm může zbýt pouze laterála *l* (*loutna* [al-ʕūd]). V oněch případech, kdy k asimilaci došlo, je těžší určitý člen

³² Je důležité zmínit, že změny zmíněné v této kapitole se zakládají na předpokladu idealizované formy spisovné arabštiny. Zdrojem výpůjčky je nicméně vyslovovaná forma, která může být potenciálně ovlivněna dialektem.

³³ Původní arabská genitivní konstrukce, ze které vzniklo slovo *admirál*, je *amīr al-baḥr*, neboli princ moří.

v češtině rozpoznat, protože také dochází ke ztrátě geminace, která je doprovodným jevem asimilace určitého členu.

5. Význam podle kolokace

Kolokace je lingvistický koncept denotující juxtapozici dvou slov, která není náhodná. Takováto definice je poněkud nepřesná a nedostatečná. Je tomu tak, protože podat jednoznačnou definici slova kolokace v jedné větě by vyloučilo její další možné významy. Tento lingvistický koncept v minulosti nabýval různé podoby a dodnes tento termín není jednoznačný. Se stále se prohlubujícím výzkumem v oblasti kolokací se některé historické definice a kritéria pro jejich určení proměnily a vyvinuly. V následujícím úseku autor podává zběžnou historii termínu kolokace v rámci lingvistického výzkumu, s důrazem na práci britského lingvisty Johna R. Firtha, z jehož koncepce kolokace tato práce vychází.

Podle Robinse (1967, s. 21) ideu kolokací a jejich vlivu na změnu významu daného slova chápali již antičtí stoičtí anomalisté, kteří odmítali tvrzení, že každé jedno slovo má jeden význam. Xiao (2015, s. 106) tvrdí, že ačkoliv se systematickému studiu kolokací v rámci moderní lingvistiky věnoval John R. Firth jako první, studium společného výskytu slov prováděl Alexander Cruden již v polovině 18. století. Konceptualizace kolokace jsou dnes velmi rozmanité a než přeneseme pozornost k Firthovskému pojetí kolokace, které je pro tuto práci stěžejní, autor představí Xiaovu (2015) stručnou klasifikaci kolokací. Xiao během svého úvodního exkurzu do historie kolokace určuje její tři hlavní druhy. Tím prvním je kolokace logické soudržnosti³⁴, která se zakládá na konceptuální asociaci mezi slovy, která se vyskytují v podobných prostředích. Od druhého typu kolokace, kterou Xiao označuje za *neighborhood collocation* neboli kolokaci sousedství, se první druh liší prakticky tím, že nevychází z konkrétního a bezprostředního kontextu daného slova v daném textu jako je tomu u kolokací sousedství. Tento typ kolokací je onen, se kterým pracují vědci v rámci korpusové lingvistiky. Třetí typ kolokace, který Xiao uvádí, může nést různé názvy, nicméně jedná se o kolokace ve smyslu víceslovných výrazů, také označovány anglickými termíny *word cluster*, *lexical bundle*, *multiword unit* či *n-gram* (Xiao 2015, s. 107). Pro více o rozmanitosti definic kolokace, viz sekce 2.1.3, v kapitole 2.1 *Notion of Collocation* doktorské práce Pavla Peciny.³⁵

³⁴ V anglickém originále *coherence collocation*. (Xiao 2015, s. 107)

³⁵ Pecina ve své doktorské práci (2008, s. 15) představuje rozmanitost definic na základě 5 rozličných aspektů: gramatická omezenost, lexikální výběr, sémantická soudržnost, jazyková institucionalizace a také frekvence a opětovný výskyt (jako jeden aspekt).

Konceptualizace kolokace jakožto zdroje pro význam jednoho daného slova se tradičně váže k Johnu R. Firthovi³⁶, který je některými (Bartsch & Evert, 2014, Xiao, 2015 či Robins 1997) považován za prvního, kdo termín kolokace použil v roce 1951 či 1957. Ovšem jiní (npř. Pecina, 2018) se domnívají, že koncept kolokace do lingvistiky vnesl již dříve jiný britský lingvista: Harold. E Palmer v roce 1938.

I přesto, že mezi lingvisty nepanuje shoda o tom, zda jím byl Firth, kdo vymyslel termín kolokace, je zřejmé, že Firthova koncepce kolokací jakožto zdroje pro pochopení významu slova byla stěžejní pro tento směr bádání a je jím dodnes. Slavný Firthův citát z roku 1957 představující kolokaci zní:

„Navrhuji přednést technický termín, význam podle kolokace, a uplatnit test kolokability.“

(Firth 1951, s. 194)³⁷

Autor v této práci pracuje s kolokací ve výše popsaném smyslu firthovském. Tento typ kolokace lze na základě již zmíněného třídění podle Xiaoa označit za kolokaci sousedství. S takovouto definicí kolokace se standardně pracuje v rámci vyvozování významu z kolokací při práci s korpusem.

5.1. Operacionalizace kolokace a kolokační okno

Firthova operacionalizace konceptu kolokace byla omezena okolnostmi doby před vznikem počítačů, které by byly schopné s velkou rychlostí provést velký počet operací. Firthův manuální výzkum se nicméně díky neustálému technologickému vývoji přeměnil v nové podoby, které byly již od příchodu prvních počítačů mnohem podobnější současným metodám výzkumu a nástrojům. Pecina (2008, s. 14) zmiňuje, že klíčovou roli ve vývoji operacionalizace kolokace sehrál lingvista John Sinclair, jemuž lingvistika vděčí za popsání struktury kolokací. Sinclairovo (1991, s. 170) pojetí struktury kolokací se zakládá na třech termínech: *node*, *span* a *collocates*.³⁸ Node je slovo, jehož kolokace jsou studovány. Podle ČNK lze v českém jazyce termín *node* označit výrazem *základové slovo/fráze*. Termín *span* (anglické slovo pro rozpětí, rozsah, interval) označuje kontextové okno

³⁶ John Rupert Firth (1890-1960) byl profesorem obecné lingvistiky na University College v Londýně.

³⁷ Originál v angličtině: „I propose to bring forward as a technical term, meaning by „collocation“, and to apply the test of collocability.“

³⁸ Je nutno zmínit, že toto Sinclairovo pojetí struktury kolokace je obecně přijímané a stalo se standardem. V citovaném díle z roku 1991 opakuje již dříve představenou strukturu.

kolem základového slova, v rámci kterého najdeme kolokáty (collocates), neboli slova v blízkém kontextu slova základového, která mohou vypovídat o kolokabilitě³⁹ (collocability) základového slova. Toto kontextové rozpětí (span) je definováno dvěma stranami (levá a pravá, označeny L a P) ve vztahu k základovému slovu a také hodnotou n , která označuje rozpětí do levé a pravé strany.⁴⁰ V současných korpusech, které umožňují vyhledávat kolokáty si ten, kdo pokládá korpusu dotaz, nastavuje rozpětí podle svých výzkumných potřeb. Tradičně se kontextové rozpětí nastavuje mezi hodnotami 3 a 5 včetně. Nicméně rozpětí, pokud to softwarový aparát daného korpusu umožní, lze nastavit podle potřeb výzkumu.

V předcházejících odstavcích již bylo vysvětleno, že identifikace kolokací se zakládá na frekvenci jejich společného výskytu v daném textu. Podle Xiaoa (2015, s. 108) se Firthova práce s kolokacemi zakládala pouze na tomto způsobu vyhledávání, tedy na jednoduchém výpočtu frekvence společného výskytu v textu. Takováto metoda nicméně nezohledňovala ve výpočtech faktor pravděpodobnosti. Halliday (1966, citován v Xiao, 2015, s. 108) tvrdí, že:

“o výskytu položky [neboli slova] v kolokačním prostředí lze hovořit pouze z hlediska pravděpodobnosti,”⁴¹

Tato myšlenka byla důležitá pro další vývoj výzkumu kolokací v korpusech. Kvůli předpokladu, že kolokace nelze studovat bez zohlednění vlivu pravděpodobnosti na výsledky dotazů, se začaly užívat statistické rovnice, které měly za úkol tento vliv korigovat a poskytnout přesnější výsledek. Tyto statistické pomůcky se staly standardní výbavou lingvisty při práci s kolokáty v korpusu. Nesou název míry asociace, jelikož měří míru asociace mezi daným základovým slovem a jeho kolokáty.

5.2. Míry asociace a nedostatky

Míry asociace⁴² jsou neodmyslitelnou součástí zkoumání kolokací. Tyto matematické rovnice, jak již bylo nastíněno, původně vznikly, aby pomohly výzkumníkovi přesněji zaměřit na kolokace zohledněním faktoru pravděpodobnosti. Funkci těchto rovnic si lze představit jako filtraci

³⁹ Kolokabilitu vyjadřuje schopnost slova být kombinováno s jinými slovy.

⁴⁰ Rozpětí se tedy označuje například následovně: „L3/P3“. S tím, že 3 je hodnota n .

⁴¹ Původní text v angličtině: “[the] occurrence of an item in a collocational environment can only be discussed in terms of probability,”

⁴² V anglickém jazyce známé jako *association measures*.

nevýznamných či nerelevantních souvškytů. Každá míra má svou specifickou rovnici, která přináší specifické výsledky. Většina měř se nicméně obecně zakládá na podobné myšlence. Zakládají se na matematickém vztahu předpokládané frekvence⁴³ spolu-vyskytujícího se slova a pozorované frekvence toho samého slova. Tímto způsobem tedy míry umožňují ve výzkumu zohlednit pravděpodobnost. V každé rovnici je s těmito dvěma hodnotami frekvencí naloženo jinak a jsou uvedeny do různých vztahů. Proto se také jejich výsledky liší. Mezi nejčastěji užívané míry asociace lze zahrnout následující: mutual information (MI), log likelihood (LL či G2), t-score (t) a dice coefficient (Dice). Míra *mutual information* efektivně filtruje spolu-vškyty, nicméně je obecně známo, že upřednostňuje data s nízkou frekvencí (Xiao 2015, s. 110, Bartsch a Evert 2014, s. 51-52). Jak již bylo nastíněno, každá míra asociace přináší své výhody a nevýhody a jiným způsobem filtruje spolu-vškyty. Nelze tedy říci, která míra je nejefektivnější či nejpresnější. Efektivita a přesnost míry může také záviset na daném jazyku, či korpusu, který je zkoumán (Garcia, García-Salido a Alonso-Ramos 2019).

Míry asociace jsou sice skvělý nástroj pro zpřesnění výsledků dotazů na kolokace, nicméně úroveň nástrojů pro práci s kolokacemi je výrazně nižší než úroveň teorie o kolokacích (Brezina, McEnery a Wattam 2015, s. 141). Kromě faktoru pravděpodobnosti, na jehož zohlednění při práci s kolokacemi poukázal Halliday, je také potřeba brát v potaz fakt, že ne všechny kolokáty, které korpus na základě dotazu zobrazí, jsou zobrazeny, kvůli jejich vztahu se základovým slovem, nýbrž mohou být zobrazeny kvůli jejich vztahu k ostatním kolokátům základového slova (Cantos a Sánchez 2001 citovaný v Almela 2011, s. 86). Moises Almela (2011) tuto teorii ilustruje na studiu kolokací anglického slova *incidence* a na základě jeho ilustrativního výzkumu dochází k tomu, že je potřeba rozlišovat mezi různými kategoriemi kolokací. Předkládá názvosloví *collocate* a *subcollocate*. *Collocate* odkazuje na běžný kolokát, zatímco *subcollocate* je termín, který má za úkol odlišit ty kolokáty, které se vyskytují ve výsledku dotazu, kvůli výskytu jiných kolokátů. Almela (ibid, s. 87-88) netvrdí, že je potřeba tyto subkolokáty vyfiltrovat, považuje je za chyby v dotazu, nýbrž tvrdí, že je třeba je zohlednit tak, že jim bude připsána jiná hodnota asociace se základovým slovem, jelikož subkolokáty stále vykazují o významu základového slova, ale zprostředkovaně skrze jeho (pravé) kolokáty. Proto Almela (ibid) přichází s novým modelem zobrazení kolokačních vztahů *Lexical*

⁴³ Předpokládaná frekvence je zcela jednoduše frekvence, kterou by slovo mělo, kdyby neexistovala žádná asociace či kolokační vztah mezi slovy.

Constellation Model (LCM)⁴⁴, který zobrazuje kolokace jako konstelaci, v níž lze zobrazit, které kolokáty jsou přímé a které jsou subkolokáty.⁴⁵ Současné nástroje pro studium kolokací trpí dalšími nedostatky, které nejsou tak relevantní pro tuto práci, jelikož se týkají studia kolokačních sítí⁴⁶.

⁴⁴ Model lexikální konstelace.

⁴⁵ Almela svou teorii potvrzuje výpočty podmíněné pravděpodobnosti mezi základovým slovem *incidence* a jeho kolokáty v korpusu Bank of English.

⁴⁶ Studium kolokačních sítí je spjato se studiem významu skrze kolokace, nicméně spíše než slovy se tento směr studia zabývá vztahem kolokátů a makrostruktury textu. Ptá se po tom, do jakých lexikálních vzorců je text organizován (Philips 1989 citován v Brezina, McEnery a Wattam 2015). Pro více o nedostacích nástrojů pro studium kolokačních sítí viz Brezina, McEnery a Wattam 2015.

6. Výsledky analýzy – změny ve významech

Před představením výsledků analýzy a závěrů z ní vyvozených, je nutno zmínit, že některá data jsou omezena a podmíněna výběrem textů, které se v korpusech nachází, či relativně malou velikostí zvoleného českého korpusu SYN12. Například u slov označujících ženské muslimské oděvy (*hidžáb, nikáb, burka*) byly výsledky podmíněny významným počtem textů korpusu SYN12, které reflektují francouzskou a belgickou politiku zákazu nošení náboženských oděvů v některých společenských kontextech. Časové vymezení textů korpusu SYN12 sice dopomohlo autorovi soustředit výzkum na arabismy současné, nicméně toto vymezení mělo i zkreslující vliv na data například právě ve vztahu k výše popsaným slovům. Je tedy třeba zdůraznit a poznamenat, že kolokační data získaná z korpusu SYN12, a na nich založené závěry o chování současných arabismů v českém jazyce, jsou podmíněná výběrem korpusu a dostupnými daty. Lze předpokládat, že kolokační data stejných slov v jiných českých korpusech by se do jisté míry lišila od dat získaných autorem v rámci této práce. I přes zmíněná omezení, ze získaných dat lze vyvést závěry o chování současných arabismů v českém jazyce z hlediska jejich významu, jelikož frekvence studovaných slov je v korpusu SYN12 dostatečná pro získání relevantních kolokátů, které dosahují kladných hodnot logDice.

Z analýzy výsledků autor vyvozuje, že 6 z 19 zkoumaných současných arabismů vykazují známky změny významu, zatímco u 12 zbylých slov⁴⁷ je původní arabský význam ponechán. Změna významu, kterou autor identifikuje u všech 6 zkoumaných arabismů, je zúžení významu. Na následující straně v tabulce č. 1 lze nahlédnout na analyzované arabismy a sémantické změny, kterými prošly. V následujících oddílech jsou slova agregována v rámci 2 skupin podle toho, zda prošla změnou či byl jejich původní význam ponechán. Zúžení významu je signifikováno nepřítomností některých kolokací u arabismů v českém jazyce, které se v kolokačním seznamu arabských slov vyskytovaly.

⁴⁷ Slovo džihádista nebylo zkoumáno pro změny významu, takže je z tohoto počtu vyřazeno, i když figuruje v seznamu zkoumaných slov v rámci českého korpusu. O něm je pojednáno v samostatném oddíle.

tabulka č. 1							
seznam zkoumaných arabismů podle sémantického pole a frekvence, sémantické změny							
slovo	orientační překlad	sémantická změna	frekvence (v i.p.m.)	slovo	orientační překlad	sémantická změna	frekvence (v i.p.m.)
sémantický okruh: islám				sémantický okruh: politika			
džihád	úsilí, svatá válka, boj	zúžení významu, negativní příznak	1,61	hamás	palestinské islamistické politické hnutí či nadšení, zápal, zaujetí	zachování významu	4,54
džihádista	člověk usilující o obranu a šíření islámu	x	1,19	al-káida (káida)	militantní teroristická islamistická organizace	zachování významu	2,54
ramadán	devátý měsíc v islámském kalendáři, období půstu	zachování významu	1,02	hizballáh, hizbulláh,	libanonská šíitská politická organizace	zachování významu	2,43
burka	pokryvka celého těla nošená ženami	zachování významu	0,67	fatah	palestinská politická strana či otevření, započeti, dobytí	zachování významu	0,79
šaría	system islámského náboženského práva	zúžení významu, negativní příznak	0,66	chalífa	zástupce, nástupce, titul pro nástupce proroka Mohameda	zúžení významu	0,48
mudžáhid	ten, kdo o něco usiluje či bojuje v džihádu	zúžení významu	0,45	intifáda	protestní a ozbrojené povstání palestinců	zachování významu	0,43
hidžáb	clona, závěs, muslimský šátek nošený ženami	zachování významu	0,35	chalífát	vláda chalífy, hlavy muslimské obce	zúžení významu	0,34
nikáb	závoj zakrývající obličej mimo oči, nošený ženami	zachování významu	0,16	sémantický okruh: jídlo			
hadždž/hadž	muslimská pout'	zachování významu	0,13	kebab	masový pokrm	zúžení významu	0,74
mudžahedín/mudžahedín	ten, kdo o něco usiluje či bojuje v džihádu	zachování významu	0,12	hummus	pokrm z cizrny a sezamové pasty	zachování významu	0,26

Autor také považuje za důležité podotknout, že i přesto, že na základě dat lze rozhodnout, která slova si svůj význam ponechala a která nabyla významů nových, existují rozdíly v kolokacích u slov, která si původní význam ponechala. Tyto rozdíly nejsou natolik významné, aby se dalo hovořit o změně významu, a nejspíše mají původ v jejich jedinečných společenských a sociolingvistických kontextech. Za zmínku zde stojí například kolokát *al-muqāwama*⁴⁸ (المقاومة), označující odpor, rezistenci a boj proti. Tento kolokát nabývá vysokých hodnot *logDice*⁴⁹ v kolokačních seznamech slov *ğihād* (جهاد), *ħizb-alláh* (حزب الله), *muğāhid* (مجاهد) a *ħamás* (حماس). Nicméně ani jeden z jejich ekvivalentů v českém korpusu nekolokuje se slovy označujícími odpor, rezistenci či boj proti.

V rámci analýzy autor uvádí také v samostatném oddílu případovou studii příznaku slova *šaría*, které v češtině nabývá příznaku, který v arabštině přítomný není. Samostatný oddíl je také

⁴⁸ Vyskytují se také se spojkou *wa* v podobě *wa al-muqāwama* (والمقاومة). Můžeme předpokládat, že výskyt tohoto kolokátu pramení z jiného úhlu pohledu, ze kterého lze na tento boj a úsilí nahlížet jako na úsilí defenzivní a ospravedlněné.

⁴⁹ Kolokát *al-muqāwama* se vyskytuje v seznamu kolokátů slova *ğihād* dvakrát jednou bez spojky *wa* s *logDice* hodnotou 6,32 a podruhé se spojkou *wa* s hodnotou 7,42. U dalších slov byl kolokační vztah se slovem *al-muqāwama* na základě míry *logDice* následující: *ħizb-alláh* (6,95), *muğāhid* (6,31), *ħamás* (9,31).

vymezen pro české slovo džihádista, které je jediným zkoumaným arabismem, jež nemá arabský ekvivalent a je českým novotvarem.

6.1. Zachování významu

Arabismy, které v jazykovém prostředí češtiny zachovaly svůj původní význam, jsou následující: *ramadán*, *hadždž*, *hidžáb*, *nikáb*, *burka*, *Hamás*, *Hizballáh*, *mudžáhedín/mudžahedín*, *intifáda*, *hummus*, *Fatah* a *(al-)Káida*. Obecně u těchto slov autor nezaznamenal významné rozdíly v kolokacích, které slova vytvářela v arabském a českém kontextu.

Arabské ekvivalenty slov *hummus* a *Fatah* nicméně představují speciální případy slov, jejichž zápis arabským písmem není jednoznačný, jelikož konsonantní skelet, ze kterého se skládají, lze vokalizovat různými způsoby a tím získat rozlišná slova s rozlišnými významy. Tento fakt se odrazil v jejich kolokátech, jelikož zadaná slova vytvářela kolokace jednak se slovy pojícími se s významy, které autor vyhledával, ale také se slovy, která se pojila na jiné významy slov, která mají jinou vokalizaci. V případě arabského slova *ḥummus* (حُمُّص) se kromě kolokátů vázaných na toto slovo zobrazily ve výrazně větším počtu kolokáty vázající se na název syrského města Ḥims (حفص)⁵⁰.

Slovo *Fataḥ* (فَتْح) je název palestinské politické organizace, které je invertovanou zkratkou arabského názvu *Ḥarakat al-Taḥrīr al-Waṭanī al-Filasṭīnī* (حركة التحرير الوطني الفلسطيني), nicméně stejný konsonantní skelet s nulovým vokálem po druhém konsonantu *fath* (فَتْح) představující maṣdar⁵¹ slovesa *fataḥa* (فَتَّح)⁵². Kolokace se slovem *fath* byly výrazně více zastoupeny než kolokace se slovem *Fataḥ*. I přesto se u obou slov vyskytovaly i kolokáty vázající se na zkoumaná slova, které se významně nelišily od kolokátů v českém korpusu a proto jsou tato slova zařazena do skupiny slov, u kterých se význam zachoval.

V tomto oddílu stojí za zmínku i rozdíl mezi kolokáty slov *hidžáb* a *burka* oproti slovu *nikáb*. Zatímco slovo *hidžáb* kolokuje se slovy *pokrývku* (logDice: 4,07) a *šátek* (logDice: 3,83) a slovo *burka*

⁵⁰ V českém jazyce známé jako Homs, dříve Emesa.

⁵¹ Maṣdar je podstatné jméno slovesné, které může plnit funkci slovesa i jména. Užívá se namísto infinitivu, který arabština nemá.

⁵² Sloveso *fataḥa* (فَتَّح) má následující významy: otevřít, zahájit, nastínit, dobyt. Jeho maṣdar *fath* (فَتْح) má významy: otevření, započítí, zahájení, začátek, dobytí, almužna, dar, příspěvek. (Arabsko-český slovník, 2009)

silně kolokuje se slovy odkazujícími na zahalování⁵³, slovo *nikáb* tyto kolokace nemá. Ovšem mezi kolokáty slov *nikáb* a *burka* najdeme sdílený kolokát *obličej*, který v seznamu kolokátů slova *hidžáb* nenajdeme. Nicméně, tato slova byla z analýzy vyřazena z důvodu soustředění výskytu jen do několika textů v českém korpusu.

6.2. Zúžení významu

K zúžení významu došlo u následujících šesti zkoumaných slov: *chalífa*, *chalífát*, *džihád*, *kebab*, *mudžáhíd* a *šaría*. U slova *chalífa* sledujeme v českém korpusu převážně kolokace vztahující se ke vlastním jménům, mužským arabským titulům *emír*, *šajch* a *šejk*, ke městu Dubaj skrze různé formy slova *dubajský*, nepřímo také k stavbě Burdž Chalífa skrze kolokace *mrakodrap* a *mrakodrapu* a *gigantický*. Nicméně za kolokace vztahující se k původnímu významu slova *chalífa*, jakožto titul pro nástupce proroka Mohameda, by bylo možné považovat pouze *vládce* a *islámské*. Po konzultaci těchto slov v širším kontextu v rámci jejich vyčlenění pozitivním filtrem, je zřejmé, že slovo *islámské* kolokuje se slovem *chalífa* v původním významu, zatímco *vládce* kolokuje s vlastními jmény ve všech případech (6 výskytů) kromě jednoho, který kolokuje se slovem *chalífa* v původním významu.

Původní arabský význam je tedy zachován ve velmi omezené míře, zatímco kolokacím dominují jiné významy. Je otázkou, zda se zde jedná o zkreslení dat výběrem textů či tato data věrohodně reflektují užití daného slova. Jistě by bylo věcné porovnat kolokace slova *chalífa* s kolokacemi starší podoby stejného slova *kalif*. To bohužel není možné v korpusu SYN12, jelikož má slovo *kalif* moc nízkou frekvenci (0,07 i.p.m.) a nelze z něj vyvodit relevantní kolokáty.

Slovo *chalífát* v korpusu SYN12 kolokuje téměř výhradně se slovy vztahujícími se k *Islámskému státu* (IS), jako například *Bakr*⁵⁴, *Bagdádí*, *islámský*, *bojovníky*, *IS*, *Sýrii* či *Iráku*. Na základě nahlédnutí do kontextu skrze pozitivní filtraci kolokací *chalífát-říše* a *chalífát-muslimské* autor potvrdil, že tyto dva kolokáty (*říše*, *muslimské*) jako jediné kolokují se slovem *chalífát* v jeho historickém původním významu. Na základě těchto dat autor usoudil, že je význam téměř zcela zúžený na současný *chalífát IS*.

⁵³ Kolokáty slova *burka* vztahující se k zahalování jsou následující: *zahalené* (logDice: 5,48), *zahalování* (logDice: 5,4), *zahalená* (logDice: 4,96), *zahalenou* (logDice: 4,82), *zahalovat* (logDice: 4,76), *zahalena* (4,54). Tyto kolokáty zabírají 6 z prvních 10 pozic v kolokačním seznamu tohoto slova.

⁵⁴ U slov *Bakr* a *Bagdádí* se jedná o části jména iráckého vůdce hnutí IS, který se jmenoval *Abū Bakr al-Bagdādī*.

Slovo *džihád* v českém jazyce a jeho arabský protějšek *ǧihād* (جهاد) většinou vytvářela významově podobné kolokace. V češtině byl nicméně význam slova ochuzen o významy vyplývající z arabských kolokátů *al-nafs*, *bi-l-māl*.⁵⁵ Tyto položky postavené v genitivní konstrukci se slovem *ǧihād* upřesňují druh džihádu a nevztahují se k válce, boji a radikálním politicko-náboženským uskopením, která představují v obou korpusech hlavní zdroj kolokátů. I přes podobnost významu výsledků se česká data liší od těch jim podobných v arabštině. V datech z českého korpusu dominují přídavná jména *militantní*, *extremistická*, *radikální*, *teroristická*, *palestinská* a *svatá*. Až na poslední dva kolokáty v arabských datech nenajdeme obdobné významy. Džihád se tedy v českém kontextu projevuje nejen v zúženém významu slova na takzvaný *džihád mečem*, ale nabývá také negativní příznak.

Slovo *kebab* v korpusu SYN12 kolokuje zvláštním způsobem, který si autor vysvětluje podmíněností zvolených textů. Nejfrekventovanější slovo, které mělo také nejvyšší hodnotu logDice je *přejetý*. Po pozitivní filtraci této položky bylo z konkordancí jasné, že se jedná o citaci kontroverzního výroku, uvedené v mnoha článkách. V SYN12 nalezneme 18 výskytů této citace.

Slovo *kebab* nejvíce kolokuje se slovy, které asociují rychlou stravu (*gyros*, *hamburgery*, *pizzerie*, *polotovarů*, *rychlého*, *rychlé*), dále kolokuje i s některými slovy, které se blíží arabskému významu (*grilované*, *turecké*, *mleté*, *tureckého*, *masem*, *masa*, *maso*). Podle dat se tedy jedná o masový pokrm, který má asociaci s tureckem (a islámem skrze kolokát *mešit*). Slovo *kebab* je nicméně zařazeno v oddílu *zúžení významu*, jelikož v porovnání s arabskými kolokáty mu chybí varieta druhů pokrmu, kterou v arabských datech jasně vyčteme. Arabská data obsahují celkem 8 různých druhů tohoto pokrmu⁵⁶, podle následujících kolokátů: *šiš* (شيش), *dūnir*⁵⁷ (دونر), *sīḥ* (سيخ), *tāwūq* (طاووق), *aḍana* (أضنة), *iskandar* (اسكندر), *al-mīrū* (الميرو) a *al-ḥilla* (الحلة).

Slovo *mudžáhid* je zařazeno v tomto oddíle, protože v českém korpusu SYN12 se nevyskytují kolokace, které by se vázaly na význam slova jako vlastního jména. Zatímco v korpusu

⁵⁵ U výrazu *ǧihād al-nafs* se jedná o vnitřní boj člověka se sebou samým a pokušeními. *ǧihād bi-l-māl* odkazuje na úsilí šířit islám finančními prostředky.

⁵⁶ Druhy uvedeny se týkají přípravy pokrmu, autor vynechal kolokáty odkazující na druhy pokrmu podle druhu masa.

⁵⁷ Zde se jedná o slovo *dōner*, jehož výslovnost v arabském jazyce bude jistě mezi mluvčími variovat. Autor slovo v latinském přepisu vokalizoval podle pravidel moderní spisovné arabštiny.

artenten¹⁸ tyto kolokace najdeme. Kromě toho rozdílu, na základě kterého autor slovo dotyčně agregoval, se vyskytují kolokace rovnající se těm arabským. U slova *mudžáhid* je opět položka s nejvyšší hodnotou logDice (5,44) neočekávaná. Jedná se o slovo *Hekmatjára*, část jména afghánského velitele. Tento fakt je způsoben relativně nízkou frekvencí slova *mudžáhid* (0,45) v korpusu a výběru textů a z nich vyplývající nahodilosti.

O slově *šaría* je pojednáno v následujícím samostatném oddíle.

6.3. Případová studie příznaku – šaría

U slova *šaría* jasně pozorujeme, že nabylo příznaku, který v arabském korpusu není obsažen. Taktéž se význam slova výrazně zúžil, nejspíše nejvýrazněji ze všech zkoumaných slov. Na druhou stranu je třeba zmínit, že některé kolokáty se v českém a arabském kontextu přímo shodují. Slovo *islámský* a jeho varianty odpovídají arabskému *al-islāmīja* (الإسلامية) a slovo z českého korpusu *Ansar/Ansár* koresponduje s arabským *anṣār* (أنصار). Zúžení významu lze ilustrovat na faktu, že zatímco v arabském korpusu slovo *šarīʿa* také kolokuje například se slovy *aḥkām*⁵⁸, *maqāṣid*⁵⁹, *kullīya*⁶⁰, *taṭbīq*⁶¹ a dalšími (viz tabulka č. 2 na další straně), které odkazují na význam slova *šarīʿa* jako práva ne nutně islámského, české kolokáty v korpusu SYN12 (tabulka č. 3) asociují pouze na uplatnění šaría ve smyslu islámského práva, a to radikálními a extremistickými skupinami. Tato asociace je také jeden z důvodů pro příznak, které slovo nabylo v českém jazyce. Tento negativní příznak je ovšem formovaný i mnoha dalšími kolokáty, jako například *nastolení*, *bičování*, *uplatňování*, *striktní* či *přísném*. Příznak je také tvořený kolokacemi s hnutím Talibán a přídavnými jmény od něj odvozenými, které posilují vazbu na extremismus.

⁵⁸ Jedná se o plurální formu slova *ḥukm* znamenající rozsudek, rozhodnutí, posouzení či zhodnocení.

⁵⁹ Zde se jedná o plurál slova *maqṣad* znamenající cíl, záměr, úmysl.

⁶⁰ Slovo *kullīja* v tomto kontextu označuje fakultu, zde tedy fakultu islámského práva.

⁶¹ Slovo *taṭbīq* znamená uplatnění, aplikaci, praxi.

tabulka č. 2							
kolokační seznam slova šarī'a (artentn18)							
	kolokáty	frekvence kolokace	logDice skóre		kolokáty	frekvence kolokace	logDice skóre
1	الإسلامية	96,157	10,68	26	الفقه	3,092	7,18
2	أحكام	15,712	9,38	27	أصول	3,128	7,16
3	الإسلامية	16,801	9,31	28	السمحاء	2,226	7,11
4	أنصار	12,165	9,18	29	بأحكام	2,282	7,08
5	مقاصد	9,05	9,07	30	وتطبيق	2,34	7,01
6	كلية	14,887	8,8	31	علماء	3,601	7
7	تطبيق	16,094	8,8	32	تحكيم	2,229	7
8	السمائية	6,721	8,56	33	بخالف	2,367	6,99
9	والقانون	7,072	8,56	34	العقيدة	2,621	6,98
10	الإسلام	16,205	8,13	35	تطبيق	2,224	6,97
11	الغراء	4,184	8	36	عقده	2,354	6,95
12	بكلية	4,561	7,85	37	علل	2,008	6,93
13	شريعة	3,794	7,77	38	انصار	2,061	6,91
14	مبادئ	4,254	7,72	39	اليهودية	2,445	6,91
15	الخاب	3,119	7,57	40	مخالفة	2,499	6,84
16	جاءت	5,807	7,54	41	الدين	8,64	6,81
17	بجامعة	4,222	7,47	42	وفق	4,425	6,77
18	بتطبيق	3,183	7,45	43	نصوص	2,08	6,75
19	الشريعة	4,379	7,43	44	موسى	3,376	6,74
20	لأحكام	2,924	7,36	45	الإسلام	2,736	6,7
21	السمحة	2,689	7,36	46	أساذ	2,288	6,7
22	علوم	3,467	7,33	47	كليات	1,889	6,69
23	والدراسات	2,831	7,25	48	لتطبيق	1,8	6,65
24	الموافقة	2,417	7,22	49	القوانين	2,578	6,63
25	والقوانين	2,687	7,22	50	وأحكام	1,56	6,55

Slovo *šarī'a* nabývá negativního příznaku v češtině. Kromě toho, jak autor již zmínil v předchozím oddíle, negativního příznaku nabylo také slovo *džihád*. Autor obě slova klasifikuje do sémantického okruhu islámu, nicméně některá slova ze stejného okruhu mají příznak neutrální s tím, že *šarī'a* má také významný přesah mimo islám, jelikož se týká legislativy obecně.

tabulka č. 3							
kolokační seznam slova šaría (SYNV12)							
	kolokáty	frekvence kolokace	logDice skóre		kolokáty	frekvence kolokace	logDice skóre
1	Ansar	45	8,25	26	zavírá	4	4,02
2	islámského	56	6,96	27	výkladu	19	3,99
3	Ansár	10	6,27	28	zavádějí	6	3,95
4	přísného	18	5,84	29	zavedení	81	3,92
5	islámském	11	5,8	30	striktní	8	3,9
6	uplatňováno	8	5,77	31	přísnými	4	3,84
7	islámské	31	5,65	32	islámu	15	3,77
8	islámským	11	5,5	33	zavádění	14	3,52
9	talibanská	5	5,3	34	nastolit	4	3,35
10	šaría	8	5,25	35	interpretace	9	3,3
11	šariu	4	4,93	36	muslimové	6	3,24
12	nastolení	11	4,89	37	muslimy	5	3,21
13	islámskému	5	4,88	38	uplatňovat	8	3,19
14	afghánská	5	4,73	39	zásad	11	3,15
15	ortodoxního	4	4,63	40	zavést	21	3,13
16	bičování	4	4,6	41	zaveden	5	2,98
17	uplatňování	13	4,55	42	uplatňuje	7	2,95
18	striktním	4	4,45	43	zákonům	4	2,90
19	Tálibánem	4	4,39	44	výklad	13	2,84
20	islám	13	4,35	45	interpretaci	7	2,82
21	islámský	6	4,35	46	přísným	4	2,8
22	přísném	4	4,25	47	dodržování	20	2,76
23	božím	4	4,12	48	zavádět	4	2,73
24	zavedeno	6	4,05	49	Zavedení	4	2,59
25	přísné	22	4,02	50	středověké	7	2,51

Srovnáme příznak slov *džihád* a *šaría* s příznakem slov *hadždž/hadž* v tabulce č. 4 a *ramadán* v tabulce č. 5, která jsou v korpusu SYNV12 bez příznaku.

tabulka č. 4							
kolokační seznam slova hadždž/hadž (SYNV12)							
	kolokáty	frekvence kolokace	logDice skóre		kolokáty	frekvence kolokace	logDice skóre
1	Wad	4	6,94	18	kteří	4	-5,31
2	Mekky	13	6,33	19	při	5	-5,35
3	satana	4	5,93	20	aby	4	-5,77
4	každoroční	14	3,56	21	ze	4	-5,96
5	muslimské	6	3,53	22	pro	7	-6,43
6	muslimů	4	2,64	23	za	6	-6,61
7	výroční	6	2,05	24	s	7	-7,39
8	poutí	7	1,34	25	že	7	-7,48
9	letošní	9	-1,42	26	-	4	-7,5
10	tradiční	6	-1,42	27	z	6	-7,58
11	velkou	8	-1,64	28	.	48	-7,87
12	celého	4	-2,15	29	na	11	-8,03
13	velké	7	-2,66	30	,	42	-8,04
14	během	6	-3,17	31	se	12	-8,11
15	do	76	-3,79	32	je	5	-8,27
16	Jak	4	-4,36	33	a	11	-8,48
17	roce	4	-5,27	34	v	6	-9,09

Tabulka č. 5 svědčí o neutralitě příznaku slova *ramadán* skrze významné kolokáty (například *muslimský, islámský, posvátný, půst*) shodující se s měsícem ramadánem. Zdá se, že se slova *viceprezident, viceprezidentem, Viceprezident* a *vicepremiér* vymykají významu slova ramadán. Proto je také autor podrobil hlubší analýze. Po konzultaci výsledků pozitivní filtrace kontextu těchto čtyř slov se autor ujistil, že se vyskytují jakožto titul zesnulého iráckého politika Tāhy Jasína Ramadāna, jehož jméno v sobě obsahuje slovo *ramadán*.

tabulka č. 5							
kolokační seznam slova ramadán(SYNV12)							
	kolokáty	frekvence kolokace	logDice skóre		kolokáty	frekvence kolokace	logDice skóre
1	muslimského	200	9,02	26	ZAČAL	4	3,92
2	muslimský	73	7,82	27	Viceprezident	5	3,91
3	islámského	62	6,9	28	muslimů	12	3,89
4	islámský	43	6,89	29	mešity	7	3,73
5	viceprezident	145	6,86	30	islámským	4	3,68
6	posvátného	35	6,78	31	arabský	5	3,62
7	muslimském	21	6,24	32	muslimských	8	3,59
8	posvátný	25	6,23	33	vicepremiér	10	3,49
9	Muslimové	21	5,84	34	nadcházejícího	5	3,18
10	půst	39	5,77	35	muslimské	6	3,13
11	islámském	12	5,49	36	modlí	4	3,02
12	u	8	5,2	37	soumraku	4	2,97
13	postního	8	5,1	38	Islámský	4	2,65
14	viceprezidentem	14	5,1	39	probíhající	10	2,53
15	postní	13	5,09	40	denního	7	2,41
16	muslim	12	4,98	41	končící	4	2,11
17	ramadán	8	4,97	42	věřící	6	1,92
18	muslimové	16	4,5	43	skončení	21	1,81
19	Půst	5	4,35	44	islámu	4	1,77
20	Muslimský	4	4,22	45	konec	42	1,74
21	Posvátný	4	4,22	46	útočit	4	1,73
22	půstu	10	4,19	47	2022	8	1,66
23	muslimům	6	4,13	48	vrcholí	5	1,58
24	ramadānu	5	4,12	49	Vánoc	6	1,27
25	muslimy	10	4,04	50	kalendáře	4	1,24

6.4. Novotvar *džihádista*

Slovo *džihádista* je v této studii vyjímečným případem slova, které je českým novotvarem, utvořeným z arabského slova *džihád* a českého afixu – *ista*, kterým jsou tvořeny názvy stoupenců myšlenkových proudů (Vondráček, 2018). Z nezkoumaného seznamu arabismů⁶² do této skupiny slov patří také přechýlený ekvivalent *džihádista* a slovo *salafista*. Jelikož slovo *džihádista*

⁶² Viz příloha č. 2.

v arabštině neexistuje, autor nemůže provést komparaci jejich kolokací a významů, nicméně může poskytnout kolokace (v následující tabulce č. 6) slova z korpusu SYN12.

tabulka č. 6							
kolokační seznam slova džihádista (SYNV12)							
	kolokáty	frekvence kolokace	logDice skóre		kolokáty	frekvence kolokace	logDice skóre
1	Emwazi	7	4,93	26	tajných	4	1,4
2	Islámskéh o	18	4,93	27	uprchlíků	5	1,16
3	IS	28	4,58	28	bojuje	7	0,86
4	chalífátu	6	4,51	29	radikální	4	0,82
5	teroristick ého	12	4,42	30	považová n	4	0,76
6	ISIL	5	4,29	31	Izrael	4	0,58
7	Mosul	5	4,19	32	přidat	5	0,36
8	ů	4	3,78	33	letecké	4	0,33
9	Hebdo	4	3,58	34	kontakty	4	0,32
10	Kurdové	5	3,58	35	válku	4	0,24
11	syřských	5	3,57	36	občanské	5	0,14
12	Mosulu	4	3,51	37	útoky	4	0,11
13	Asad	5	3,44	38	Bruselu	4	-0,18
14	Sýrii	15	3,33	39	armáda	4	-0,19
15	rebelové	4	3,11	40	redaktor	4	-0,23
16	zabito	4	2,91	41	vojenské	5	-0,38
17	nálety	4	2,9	42	boje	6	-0,44
18	Islámský	5	2,9	43	bojovat	6	-0,51
19	Sýrie	8	2,61	44	dostaly	4	-0,51
20	světovými	4	2,3	45	zastavit	4	-0,53
21	bojové	4	1,73	46	státu	14	-0,64
22	padne	7	1,66	47	státy	6	-0,68
23	Iráku	14	1,61	48	boji	5	-0,7
24	návod	4	1,49	49	skupinu	4	-0,77
25	Irák	4	1,47	50	obětí	4	-0,82

V porovnání s kolokacemi, do kterých vstupují slova *mudžáhid* a *mudžahedín*, která sdílí významy se slovem *džihádista*, lze rozeznat rozdílné asociace s určitými regiony. Slovo *džihádista* (tabulka č. 6) má silnější asociaci s oblastí Sýrie a Iráku, a islámským státem, zatímco slova *mudžáhid* a *mudžahedín* kolokují spíše s Afghánistánem (viz tabulky č. 7 a 8). Je otázkou, zda tyto rozdílné asociace s regiony pramení z jejich reálného užití v jiných kontextech či jsou tyto data způsobena omezenostmi výzkumu. Pro zodpovězení této otázky by jistě bylo věcné provést podobný výzkum v jiných korpusech či na jiném materiálu a zjištění o těchto asociacích konzultovat s dalšími daty.

tabulka č. 7							
kolokační seznam slova mudžáhid (SYNV12)							
	kolokáty	frekvence kolokace	logDice skóre		kolokáty	frekvence kolokace	logDice skóre
1	Hekmatjára	4	5,44		1992	4	-0,48
2	sjednocených	4	5,32	27	rada	5	-0,6
3	islamistického	5	5,23	28	americké	5	-1,08
4	revolucionářů	4	4,68	29	tzv	5	-1,8
5	islámská	4	4,48	30	vlády	6	-2,1
6	islámských	7	4	31	zemi	5	-2,45
7	Mluvčí	36	3,56	32	Jeho	7	-2,52
8	Ládin	5	3,55	33	letech	12	-2,69
9	Kábulu	5	3,24	34	mezi	20	-2,89
10	přihlásila	9	2,84	35	svého	8	-2,97
11	neznámá	7	2,67	36	proti	13	-3,01
12	opoziční	17	2,21	37	díky	4	-3,61
13	bojích	6	2,16	38	ze	13	-4,26
14	mluvčího	13	1,95	39	řekl	6	-4,34
15	armádou	4	1,91	40	nich	4	-4,39
16	Afghánistánu	10	1,85	41	"	56	-4,55
17	bojovníků	4	1,78	42	ke	4	-4,61
18	bin	4	1,77	43	kde	6	-4,85
19	bojoval	6	1,42	44	Je	5	-4,95
20	svaté	4	0,82	45	že	38	-5,04
21	známá	6	0,39	46	kdy	4	-5,06
22	mluvčí	74	0,21	47	po	14	-5,08
23	bojů	4	0,16	48	při	6	-5,08
24	listu	5	0,06	49	byl	8	-5,1
25	občanské	4	-0,14	50	jsou	10	-5,17

tabulka č. 8							
kolokační seznam slova mudžahedín/mudžáhedín (SYNV12)							
	kolokáty	frekvence kolokace	logDice skóre		kolokáty	frekvence kolokace	logDice skóre
1	Harkat	4	6,95	16	o	9	-6,7
2	neznámá	6	2,52	17	z	11	-6,71
3	složku	4	2,37	18	s	10	-6,88
4	Válka	6	2,31	19	ale	6	-6,92
5	Afghánistánu	6	1,14	20	"	10	-7,03
6	bojovali	4	0,41	21	se	24	-7,11
7	známá	4	-0,18	22	si	5	-7,28
8	letech	7	-3,47	23	že	8	-7,29
9	proti	5	-4,39	24	i	6	-7,39
10	mezi	5	-4,89	25	,	60	-7,53
11)	16	-5,17	26	v	16	-7,68
12	"	11	-5,97	27	je	6	-8,01
13	po	6	-6,3	28	na	11	-8,03
14	jako	5	-6,35	29	.	40	-8,13
15	-	8	-6,5	30	a	13	-8,23

7. Závěr

Současné arabismy v češtině jsou takřka neprobádaným tématem. Autor v rámci této práce tuto oblast výzkumu otevřel, skrze výzkum jejich významů, a přispěl novými poznatky o arabismech v českém jazyce. Menší podíl autorova příspěvku k tématu je vytvoření seznamu současných arabismů a uvedení jejich frekvence v korpusu SYN12. Příspěvek nicméně převážně spočívá v nových poznatcích o sémantice současných arabských lexikálních výpůjček v českém jazyce. Zdejší výzkum se zakládá na datech získaných pouze z jednoho českého korpusu, který je relativně malý (5 175 mil. slov) a je podmíněn výběrem textů, což se projevilo neobvyklými kolokacemi, které se autor pokusil pomocí pozitivní filtrace kolokací vysvětlit.

Výsledky výzkumu ukazují, že zatímco některé arabismy (6) prošly změnou významu, konkrétně jeho zúžením, jiné arabismy (12) si svůj význam ponechaly. Zúžení významu dvou arabismů (*džihád* a *šaría*) bylo doprovázeno nabytím negativního příznaku v českém jazyce. Obě tato slova jsou zahrnuta v sémantickém okruhu *islám*, ovšem některá jiná slova (*hadždž/hadž* a *ramadán*) ze stejného okruhu mají příznak neutrální. Český novotvar *džihádista* nebyl zkoumán z hlediska změny významu, nicméně je zahrnut v seznamu zkoumaných současných arabismů. Analýza jeho kolokací a komparace dat slova *džihádista* se slovy *mudžáhid* a *mudžáhedín/mudžahedín* naznačuje možné rozdíly v asociaci s regionem, kterým je potřeba lépe porozumět. Z některých zvláštních kolokací u určitých slov, způsobených omezenostmi korpusu, vyplývají další otázky. Například jaké kolokace by vytvářela tato slova: *hadždž/hadž*, *mudžáhid*, *kebab* či *chalífa* v korpusu, kde mají vyšší frekvenci či v korpusu s jinými zdrojovými texty.

Autor se domnívá, že je velký prostor pro doplnění poznatků, které táto práce přinesla. Provedením podobné studie na jiném jazykovém materiálu, případně dalších českých korpusech, by bylo možné doplnit a rozšířit zdejší poznatky o významech těchto arabských slov. Kromě výzkumu sémantiky slov, by jistě bylo smysluplné budoucí výzkum soustředit na další aspekty chování současných arabismů v českém jazyce. Slova ze seznamů vytvořených v rámci této práce lze podrobit například výzkumu etymologickému pro bližší pochopení transferu těchto slov z arabštiny do češtiny a jejich tras. Stejně tak by bylo zajímavé hlouběji studovat morfologii slov, obzvláště pak českých novotvarů, které mají základ v arabismech.

V neposlední řadě je jistě potřeba seznamy rozšířit o další současné arabismy, které autor nezaznamenal, přičemž si je autor vědom toho, že seznam slov není kompletní a podrobný. Vzorek, který autor v rámci práce studoval, je částečný a nereprezentuje chování všech současných arabismů v českém jazyce. Ostatně kvůli nízké frekvenci určitých slov v korpusu SYN12 musela některá slova být vyřazena z výzkumu. Není vyloučeno, že v jiném korpusu by některá slova ze seznamu v příloze č. 2 nabyla vyšších frekvencí a mohla se tak stát předmětem studia kolokací a jejich významu.

Práce je v současném stavu dostupnosti dat sondou, která ověřila použitelnost využitých metod pro tento typ zkoumání, a je zřejmé, že i při relativně omezeném materiálu na české straně je možno dojít k relevantním závěrům.

Použitá literatura

- Algeriani, A. M. A. and Mohadi, M. (2019) ‘The House of Wisdom (Bayt al-Hikmah), an Educational Institution during the Time of the Abbasid Dynasty. A Historical Perspective’, *Pertanika Journal of Social Science and Humanities* 27. Dostupné z: <https://api.semanticscholar.org/CorpusID:208196043>.
- Almela, M. (2011) ‘Improving Corpus-Driven Methods of Semantic Analysis: A Case Study of the Collocational Profile of *Incidence*’, *English Studies*, 92(1), s. 84–99. doi: 10.1080/0013838X.2010.537050.
- *ArTenTen – Arabic corpus from the web. Sketch Engine*. Dostupné z: <https://www.sketchengine.eu/artenten-arabic-corpus/#toggle-id-3> (citováno: 20214-04-22).
- Asatrian, M. (2006) ‘Iranian Elements in Arabic: The State of Research’, *Iran & the Caucasus*. Brill, 10(1), s. 87–106. Dostupné z: <http://www.jstor.org/stable/4030944> (citováno: 2024-04-22).
- Bartsch, S., Evert, S. and Erlangen-Nürnberg, F.-A.-U. (2014) ‘Towards a Firthian Notion of Collocation’, Dostupné z: <https://api.semanticscholar.org/CorpusID:4816866>.
- Blažek, V. (2009) ‘Slavonic languages’, in *Encyclopedia of Arabic Language and Linguistics*, Vol. IV (Q-Z), Kees Versteegh (ed.). Leiden-Boston: Brill, s. 259–267.
- Blažek, V. and Karlíková, H. (2008) ‘Arabic Borrowings in Slavic’, I. Janyšková and H. Karlíková (eds.) *Varia Slavica. Sborník příspěvků k 80. narozeninám Radoslava Večerky*. Praha: Nakladatelství Lidové noviny, s. 121–132.
- Brezina, V., McEnery, T. and Wattam, S. (2015) ‘Collocations in context: a new perspective on collocation networks’, *International Journal of Corpus Linguistics*, 20, s. 139–173. Dostupné z: <https://api.semanticscholar.org/CorpusID:54178135>. (citováno: 2024-04-17).
- Charvát, P. a Prosecký, J. (1996) *Ibrahim ibn Ya‘qub at-turtushi: Christianity, Islam and Judaism meet in East-Central Europe, c. 800-1300 A.D.* Proceedings of the International Colloquy, 25-29 April 1994. Praha: Academy of Sciences of the Czech Republic, Oriental Institute.
- Cvrček, V. a Richterová, O. (2023). cnk:intercorp:verze16. *Průručka ČNK*. Dostupné z: <http://wiki.korpus.cz/doku.php?id=cnk:intercorp:verze16&rev=1697033307> (citováno 2024-04-18).
- Dworkin, S. N. (2012) *A history of the spanish lexicon: A linguistic perspective*, Oxford Academic. Oxford, United Kingdom: Oxford University Press. Dostupné z: <https://doi.org/10.1093/acprof:oso/9780199541140.001.0001> (citováno: 2024-04-17).
- Firth, J. R. (1951). Modes of meanings, *Papers in Linguistics 1934–1951*. Oxford University Press, s. 190-215.
- Garcia, M., García Salido, M. and Alonso-Ramos, M. (2019) ‘A comparison of statistical association measures for identifying dependency-based collocations in various languages’, Savary, A. et al. (eds) *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*. Florence, Italy: Association for Computational Linguistics, s.

49–59. doi: 10.18653/v1/W19-5107.

- Glibanović-Vajzović, H. (1986) ‘Turkish Borrowings in Serbo-Croat Analyzed from Sociolinguistic Point of View’, *Književni jezik*, (2), s. 141–147. Dostupné z: <https://www.ceeol.com/search/article-detail?id=767878> (citováno: 2024-04-17).
- Górska, E. (2018) ‘Arabic-Polish Language Contacts – an Overview’, *Folia Orientalia*. Commission of Oriental Studies of the Polish Academy of Sciences – Kraków Branch, Bibliotheca1, s. 49–59. doi: 10.24425/for.2019.126124.
- Gutas, D. (1998) *Greek thought, Arabic culture: The Graeco-arabic translation movement in Baghdad and early 'Abbasid Society (2nd-4th/5th-10th centuries)*. London: Routledge.
- Hadzijahic, M. (1955) Književnost na Arabici, *Enciklopedija Jugoslavije*, I., A-Bosk. Zagreb: Izdanje i naklada leksikografskog zavoda FNRJ. s. 144.
- Haspelmath, M., Haspelmath, M. and Tadmor, U. (2009) ‘Lexical borrowing: Concepts and issues’, *Loanwords in the World's Languages: A Comparative Handbook*. Berlin, Germany: De Gruyter Mouton, s. 35–54.
- Hazbunová, K. (2006) *Arabské výpůjčky v češtině*. Diplomová práce. Univerzita Karlova, Filozofická fakulta. Dostupné z: <http://hdl.handle.net/20.500.11956/7388> (citováno: 2024-04-17).
- Jankowiak, M. (2017) ‘What does the slave trade in the Saqaliba tell us about early Islamic slavery?’, *International Journal of Middle East Studies*, 49(1), s. 169–172. doi:10.1017/s0020743816001240.
- Karlíková, H. (2017) *Arabismy v českém lexiku*, czechEncy | Nový encyklopedický slovník češtiny. P. Karlík, M. Nekula, and J. Pleskalová (eds). Dostupné z: <https://www.czechency.org/slovník/ARABISMY%20V%20%C4%8CESK%C3%89M%20LEXIKU> (citováno: 2024-04-17).
- Khayrullina-Valieva, A. G. (2020) ‘Turkic lexical elements in the Bulgarian language’, *Litteraria Copernicana*, 33(1), s. 205–211. doi: 10.12775/LC.2020.015.
- Machálek, T. (2019). *Slovo v kostce – agregátor slovních profilů*. Praha: FF UK. Dostupné z: <http://korpus.cz/slovo-v-kostce/>
- Měchurová, M. (2023) *Arabismy ve francouzštině od roku 1900 do roku 2022*. Bakalářská práce. Univerzita Karlova, Filozofická fakulta. Dostupné z: <http://hdl.handle.net/20.500.11956/183159> (citováno: 2024-04-17).
- Muftić, T. (1969) O arebici i njenom pravopisu, *Prilozi za orijentalnu filologiju sv. XIV–XV*. Sarajevo. s. 101-102.
- Mustafić, A. (2013) *Epoha fonetske misli kod arapa i arebica*. D. Babić.
- Nekula, M. (2017a) *Výpůjčka*, czechEncy | Nový encyklopedický slovník češtiny. J. Pleskalová, M. Nekula, and P. Karlík (eds.). Dostupné z: <https://www.czechency.org/slovník/V%C3%9D%C5%AEJ%C4%8CKA> (citováno: 2024-04-17).
- Nekula, M. (2017b) *Adaptace výpůjček*, czechEncy | Nový encyklopedický slovník češtiny. P. Karlík, M. Nekula, and J. Pleskalová (eds). Dostupné z:

<https://www.czechency.org/slovník/ADAPTACE%20V%C3%9DP%C5%AEJ%C4%8CEK> (citováno: 2024-04-17).

- Pecina, P. (2009) *Lexical Association Measures: Collocation Extraction*. Praha, Czechia: ÚFAL (Studies in Computational and Theoretical Linguistics).
- Rietbroek, P. (2010). Brill's simple Arabic transliteration system, version 1.0. Dostupné z: https://brill.com/fileasset/downloads_static/static_fonts_simple_arabic_transliteration.pdf
- Robins, R. H. (1967) *A Short History of Linguistics*. London, England: Longmans
- Rychlý, P. (2008). A Lexicographer-Friendly Association Score, *RASLAN 2008*. 2. vyd. Brno: Masarykova Univerzita. s. 6-9.
- Škaljić, A. (1966) *Turcizmi u srpskohrvatskom jeziku*. Sarajevo, SFRJ: Izdavačko preduzeće Svjetlost.
- Vicente, Á. (2019) 'Andalusi Arabic', S. Manfredi and C. Lucas (eds.) *Arabic and contact-induced change*. Berlin, Germany: Language Science Press, s. 225–244.
- Vondráček, M. (2018) -ista, *Heslář - Slovník afixů*. J. Šmindl (ed.). Dostupné z: <https://www.slovníkafixu.cz/heslar/-ista> (citováno 2024-04-19).
- Xiao, R. (2015) 'Collocation', D. Biber and R. Reppen (eds.) *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press (Cambridge Handbooks in Language and Linguistics), s. 106–124.
- Zahran, W. (2017) *Význam a frekvence přejatých slov z arabštiny v současné češtině*. Bakalářská práce. Univerzita Karlova, Filozofická fakulta. Dostupné z: <http://hdl.handle.net/20.500.11956/87349> (citováno: 2024-04-17).

Slovníky

- Machek, V. (2012) *Etymologický slovník jazyka českého*. 5. vyd. Praha : NLN, Nakladatelství Lidové noviny.
- Moustafa, A., Obadalová, N., Ondráš, F., Zemánek, P. (2009). *Arabsko-český slovník, testovací verze*. Online. Set Out. Dostupné z: <http://arczslovník.aspone.cz/> (citováno 2024-04-25)
- Rejzek, J. (2015) *Český etymologický slovník*. 3. vyd. Praha: Leda.

Korpusy

- Křen, M. – Cvrček, V. – Čapka, T. – Hnátková, M. – Jelínek, T. – Kocek, J. – Kovářiková, D. – Krivan, J. – Milička, J. – Petkevič, V. – Skoumalová, H. – Šindlerová, J. – Škrabal, M. (2023): *Korpus SYN, verze 12 ze 29. 12. 2023*. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupné z: <https://www.korpus.cz>
- The Arabic Web Corpus (artenten18).(2018). Sketch Engine.

Příloha č. 1

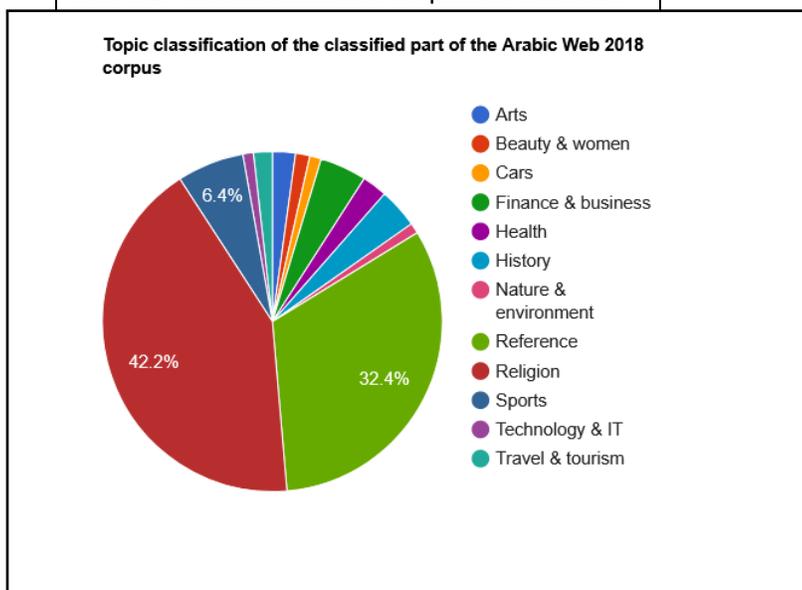
seznam zkoumaných arabismů podle sémantického pole a frekvence					
slovo	orientační překlad	frekvence (v i.p.m.)	slovo	orientační překlad	frekvence (v i.p.m.)
sémantický okruh: islám			sémantický okruh: politika		
džihád	úsilí, svatá válka, boj	1,61	hamás	palestinské islamistické politické hnutí	4,54
džihádista	člověk usilující o obranu a šíření islámu	1,19	al-káida (káida)	militantní teroristická islamistická organizace	2,54
ramadán	devátý měsíc v islámském kalendáři, období půstu	1,02	hizballáh, hizbulláh, hezbollah	libanonská šiitská politická organizace	2,43
burka	pokrývka celého těla nošena ženami	0,67	fatah	palestinská politická strana	0,79
šaría	systém islámského náboženského práva	0,66	chalífa	zástupce, nástupce, titul pro nástupce proroka Mohameda	0,48
mudžáhid	ten, kdo o něco usiluje či bojuje v džihádu	0,45	intifáda	protestní a ozbrojené povstání palestinců	0,43
hidžáb	clona závěs, muslimský šátek nošený ženami	0,35	chalífát	vláda chalífy, hlavy muslimské obce	0,34
nikáb	závoj zakrývající obličej mimo oči, nošený ženami	0,16	sémantický okruh: jídlo		
hadždž/hadž	muslimská pouť	0,13	kebab	masový pokrm	0,74
mudžahedín/mudžahedín	ten, kdo o něco usiluje či bojuje v džihádu (v ar. plurálová forma slova mudžáhid)	0,12	hummus	pokrm z cizrny a sezamové pasty	0,26

Příloha č. 2

seznam současných arabismů podle sémantického pole a frekvence					
slovo	frekvence (v i.p.m.)	orientační překlad	slovo	orientační překlad	frekvence (v i.p.m.)
sémantický okruh: islám			sémantický okruh: politika		
džihád	1,61	úsilí, svatá válka, boj	hamás	palestinské islamistické politické hnutí či nadšení, zápal, zaujetí	4,54
džihádista	1,19	člověk usilující o obranu a šíření islámu	al-káida (káida)	militantní teroristická islamistická organizace	2,54
ramadán	1,02	devátý měsíc v islámském kalendáři, období půstu	hizballáh, hizbulláh, hezbollah	libanonská šíitská politická organizace	2,43
burka	0,67	pokrývka celého těla nošena ženami	fatah	palestinská politická strana	0,79
šaría	0,66	systém islámského náboženského práva	chalífa	zástupce, nástupce, titul pro nástupce proroka Mohameda	0,48
mudžáhíd	0,45	ten, kdo o něco usiluje či bojuje v džihádu	intifáda	protestní a ozbrojené povstání palestinců	0,43
hidžáb	0,35	clona, závěs, muslimský šátek nošený ženami	chalífát	vláda chalífy	0,34
nikáb	0,16	závoj zakrývající obličej mimo oči, nošený ženami	nakba	katastrofa, užíváno pro popis události roku 1948 v Palestině	0,01
hadždž/hadž	0,13	muslimská pouť	sémantický okruh: jídlo		
sala fista	0,13	stoupenec salafiji	kebab	masový pokrm	0,74
mudžahedín/ mudžahedín	0,12	ten, kdo o něco usiluje či bojuje v džihádu	hummus	pokrm z cizrny a sezamové pasty	0,26
fatwa/fetva	0,11	nábožensko-právní stanovisko	falafel	smažený pokrm z luštěnin	0,13
sunna	0,1	zvyk, jeden z pramenů islámského práva	musaka	pokrm s mj. lílkem a bramborami	0,12
umma	0,08	muslimská obec	baklava	dezert z těsta a ořechů	0,1
hadís	0,04	původně ústně uchovaná výpověď o činech a skutcích proroka	tabouleh	svěží salát z bulguru a bylin	0,04
ulamá	0,03	islámští učenci, duchovenstvo	mezze	předkrm	0,02
šaháda	0,02	islámské vyznání víry	tažín	severoafriický pokrm připravovaný ve stejnomenné nádobě	0,01
tauhíd	0,02	nauka o jedinství Boží	baba ganuš	pokrm z pečeného lítku a sezamové pasty	0
zakát	0,02	almužna, jeden z pěti pilířů islámu	sémantický okruh: oblečení		
masdžid	0,01	mešita	džellába	dlouhý oděv	0
šía	0,01	šíitský islám, jeden ze dvou hlavních proudů islámu	abája	dlohuý oděv	0,08
wahhábovci	0,01	stoupenci wahhabismu, ortodoxního nábožensko-politického směru	burnus	dlouhý oděv s kapucí	0,06
džihádista	0,01	ž. forma džihádisty			
sala fíja	0	reformistické hnutí v islámu			
saum	0	půst			
širk	0	modloslužba, přidružování k Bohu,			

Příloha č. 3

Graf č.1 - klasifikace témat korpusu artenten18



- Zdroj: ArTenTen – Arabic corpus from the web. Sketch Engine. Dostupné z: <https://www.sketchengine.eu/artenten-arabic-corpus/#toggle-id-3> (citováno: 20214-04-22).