

Philipp Wasserscheidt

Südslawische Sprach- und Kulturwissenschaft, Institut für Slawistik und Hungarologie,

Humboldt-Universität zu Berlin, Njemačka

philipp.wasserscheidt@hu-berlin.de

The concept Muslim: discourses in Bosnian, Croatian, German and Serbian webcorpora

The aim of this contribution is to examine the meaning of the word *musliman* or German *Muslim* and to answer the question whether the meaning of this word in a dictionary can be mapped by means of classical distributional analysis, and whether further meanings can be captured by this method. In addition, a simple method for a distributional semantic analysis is presented. The basis of the investigation is the assumption common in corpus-based semantics and lexicography that the meaning of words can be interpreted from the context and thus polysemy and synonymy can be recognized. The article examines four web corpora for Bosnian, Croatian, Serbian and German. First, the context is analysed, assuming that the syntactic structures in the context are themselves inherently meaningful. Specifically, adjective attributes are used to analyse properties of the keyword, coordination partners as indicators for relevant categories, verbs with the keyword as subject and accusative object as indicators for typical actions the keyword performs or is target of. In a second step, the lexemes that occur in these four contexts are clustered and the semantic dimensions of the keyword are determined.

The results suggest that the distributional method can in principle map dictionary definitions well. However, this is mainly true for contexts which involve attributes and coordination but not for action-related syntactic configurations. Overall, the most salient semantic dimensions of the lexeme are those pertaining to members of a religious community, members of a regional ethnic group, and experienced or induced aggression. Here, the four corpora differ considerably: while German usage almost exclusively focuses on religion-related aspects, the other corpora also contain ethnic/regional components, as well as the discourse of aggression. This aggression is mainly verbal in German, but physical in the other corpora. Neutral meaning components for the word *musliman* are salient only in the Bosnian corpus.

1. Introduction

In empirical semantics, there are three different ways to explore the content of a concept or the meaning of a word: experimental studies, the referential method and distributional, corpus-linguistic methods (Croft and Cruse 2004: 242; Stefanowitsch 2010). In the last years corpus linguistic methods have undergone a

considerable development and are today indispensable for the field of semantics. Their advantage is certainly that they are as relieved of subjective assessment or unnatural experimental conditions as they are of the restriction to referable entities. Even if one does not assume that words actually have a meaning, but rather that words merely refer to existing knowledge or knowledge frames (Taylor 2017: 256; Langacker 2008: 42), the distributional method has the potential to make the underlying knowledge empirically accessible (cf. Wasserscheidt and Hrستیć 2020).

The basic idea of distributional semantics since Firth's (1957) and Sinclair's (1995) lexicographical work is that the meaning of a word can be understood from its environment: The context determines, so to speak, the concept. The regularity of the immediate surrounding of a word is used as an indication of its typical use (Müller 2011: 36) which can point to the meaning components of a word, especially when compared with other lexemes. Particularly for abstract nouns, this seems to be a reasonable approach, since abstract nouns cannot be made accessible by referential methods. In fact, also the language users themselves cannot access the meaning of abstract nouns otherwise than through the context.

Nevertheless, the method is to be treated with caution. Statistical evaluations are necessarily based on large amounts of data and therefore do not allow any text-based or user-centred analysis. The obligatory classification of contexts in distributional semantics may also contribute to the concepts appearing less flexible than they might be in reality (Taylor 2017: 256). Moreover, the context itself also consists of words: in order to make it usable, the semantics of the words in the context must ultimately also be analysed.

Whether and to what extent one wants to assume that distribution reflects semantics also depends on what exactly is understood by meaning. In any way, it can be taken for granted that distributional properties and meaning are not simply the same (Dąbrowska 2016: 488). One reason for this is that speakers in the contexts we are studying usually do not use words to simply refer to objects or in order to define them, but to express their attitude towards the references of the words (Taylor 2017: 249). So, an open question is whether and how the distributional method can actually uncover the meaning of a word whatever its definition.

In this paper, I would like to examine this question on the example of the words *musliman* 'Muslim' and German *Muslim* 'Muslim' and at the same time demonstrate an easily applicable method for the distributional analysis of lexical semantics. The method shown is a classical distributional one. The question will be whether this analysis can be used to reproduce a meaning that corresponds to the typical definition of the lexeme in dictionaries. If so, can cultural and lexically conventionalized differences also be accessed using this method? For this purpose, I examine the term in Bosnian, German, Croatian and Serbian corpora. The word *musliman*/*Muslim* is not chosen arbitrarily but because of its importance in the recent public discourse: In the German context the discussion about the status of Islam, about Muslim refugees and Muslim minorities has become a lively discourse since some

years (Kalwa 2013, Josua 2019). The difference to Bosnia–Herzegovina, Croatia and Serbia is obvious here: On the one hand these countries have centuries of experience in the discursive confrontation with Islam (cf. the papers in Telbizova–Sack and Voß 2019). In contrast to Germany, autochthonous Muslims live in Bosnia–Herzegovina and Serbia and to a lesser extent in Croatia, so that a more inclusive discourse could be expected here. In addition, there is the special status of the label *Musliman* as one of the nationalities of former Yugoslavia. This centuries–old discursive integration has recently been challenged anew when the “Balkan route” was used by largely Muslim refugees (Tošić 2022).

In the following section, I would like to introduce the distributional method in more detail, and in particular the use of the tools provided by SketchEngine. I apply this method to four corpora in section 3 and compare the results with each other and with the dictionary meaning of *musliman/Muslim*.

2. Methodology

Since the ground–breaking work of Sinclair (1995) the context of a word in a corpus serves to operationalize its meaning. Essentially, this involves assessing the words to the left or right of a keyword, whereby the syntactic structure is taken into account. This method of context–based meaning analysis (more precisely: co–text–based) is also used in this paper. In contrast to a purely word–based context analysis, as used in vector analyses for example (Mitchell and Lapata 2008), the presence of certain words in the context is not the only factor used here. Rather, the form or type of the context – the linguistic configurations in which the word occurs – is included in the analysis. Thus, different linguistic constructions like attribution, coordination and other syntactic structures are themselves classified according to their function and used as a tool for the semantic analysis (Geeraerts 2016: 242).

The method is similar to the image analysis proposed by Vogel (2010). For the analysis of the image of a country he uses among other things concordances, n–grams, clustering, collocations and co–occurrences as well as different syntactic embeddings of the given concept. For example, he includes definitory expressions like “X is” or “X is not” in his method, the formation of compound nouns with the keyword and other means. Such analyses known as keyword analyses, however, often refer to a broad definition of the concept (Kalwa 2013). Thus Vogel (2010) examines the entire word family of the respective keyword (e.g. German *Türke* ‘Turk’, *Türkei* ‘Turkey’, *türkisch* ‘Turkish’ and *türken* ‘to fake’ for the analysis of *Turkey*). In my example this would correspond to examining the concept of “religious affiliation to Islam” with all its lexical instantiations. But here I would like to examine the lexical concept of the lexeme *musliman/Muslim*. The resulting concept should first of all be comparable with the dictionary meaning of the lemma but should also contain dimensions that go beyond it. Word formations with the roots *musliman* or *Muslim* are not included in the investigation.

This method is contrasted with the result of a qualitative analysis of word senses from monolingual dictionaries of Bosnian, Croatian, German and Serbian.

2.1. Operationalization

The operationalization rests upon a basic analysis of the word *musliman/Muslim*. The word can be identified linguistically as a noun and conceptually as a kind of human being. Therefore, one can assume that the concept *musliman/Muslim* should possess features that humans in general have: They are bearers of properties, form groups or classes, perform actions or are exposed to actions. From the point of view of language structure, the noun *musliman/Muslim* can, for example, have attributes, enter coordination, or function as a subject and object. Obviously, for other types of keywords, such as adjectives or verbs, other features would also be used for conceptual analysis.

In the following I will examine four different features: The properties of the keyword, the categorization of the keyword, actions of the keyword as an actor, and actions to which the keyword is exposed as a target. These are operationalized as follows:

- Properties:
Attribution of the keyword with preceding adjectives. Rationale: Attributive adjectives are the prototypical carriers of properties. (*How* are Muslims)
- Categorization:
Coordination of the keyword with other noun via coordinating conjunctions. Rationale: Coordinative links are conventionally constructed between things of the same perceived category. (*What* are Muslims)
- Agent:
The keyword as subject of sentences with verbal predicates. (*What* Muslims *do*)
- Target:
The keyword as accusative object of transitive verbs. (*What is done to* Muslims)

The individual indicators are measured using the SketchGrammar tool within SketchEngine. I will therefore explain the functionality of this tool in more detail below.

2.2. SketchGrammar

I am using the SketchGrammar (Kilgarriff et al. 2014) provided by SketchEngine¹. This tool can be accessed via the online services of SketchEngine and offers the possibility to carry out an otherwise very complex corpus analysis within a single query. A SketchGrammar is a list of predefined corpus queries in the CQL corpus query language that uses the morphosyntactic annotation of a particular corpus

1 <https://www.sketchengine.eu>

and displays the searched word in different contexts². I utilize four different contexts here. First, I make use of the search for attributive adjectives. It uses the following query for the Bosnian, Croatian and Serbian corpora³:

```
(1) 2:[tag="A.*"] [tag="[ARPMCQ].*" | word="," | word="se" | word="si"]{0,5}
    1:[tag="N.*"]
```

The search is therefore aimed at an adjective (“A.*”) which precedes the keyword (identified by the prefix “1:”) and from which it may be separated by up to five adjectives, adverbs, pronouns, numerals, conjunctions, particles, by commas or the particles *se* or *si*.

The second query searches for the keyword in the context of a coordination and serves here as an indicator for the categorization of the keyword. Five different search queries are used in the SketchGrammar of each corpus for this purpose:

```
(2) 1:[tag="[VARN].*" ] [word="i" | word="ili" | word="te" ] 2:[tag="[VARN].*"
    2:[tag="[VARN].*" ] [word="i" | word="ili" | word="te" ] 1:[tag="[VARN].*" ]
    [word = "niti" ] 1:[] [word = "niti" ] 2:[]
    [word = "ili" ] 1:[] [word = "ili" ] 2:[]
    [word = "kako" ] 1:[] [word = "tako" ] 2:[]
```

The query syntax searches for the keyword either as the first or second member of a coordination with *i* ‘and’, *ili* ‘or’ or *te* ‘and’ or as the first member of the constructions [*niti* X *niti* Y] ‘neither X nor Y’, [*ili* X *ili* Y] ‘either X or Y’ or [*kako* X *tako* Y] ‘X as well as Y’.

The third search query analyses the use of the keyword as a subject. For this purpose, the keyword is searched in the nominative (“Nc..n.*”) in the following three configurations:

```
(3) 1:[tag="Nc..n.*" ] []{0,5} 2:[tag="Vm.r3.*" ]
    1:[tag="Nc..n.*" ] []{0,2} [lemma="biti" ] [tag!="V.*"]{0,3} 2:[tag="Vmp.*" ]
    [lemma="biti" ] [tag!="V.*"]{0,2} 2:[tag="Vmp.*" ] []{0,3} 1:[tag="Nc..n.*" ]
```

The first option is a verb in the third person present, which follows the keyword in nominative in a distance of maximum five tokens. The second variant searches for an auxiliary form of *biti* ‘to be’, which follows the keyword in a distance of maximum 2 tokens and is itself followed by a participle. In the third search query, the verb *biti* and the participle precede the keyword. The second and third queries therefore represent perfect tense.

2 In the Bosnian, Croatian and Serbian corpora (see section 3.2 on the corpora), the morphosyntactic annotation follows the MULTEXT–East specifications (Erjavec 2010), which I will not explain here. See nl.ijs.si/ME/V6/msd/html/msd.html for details.

3 The search queries in the German corpus are principally similar, but naturally differ due to the different tag set (German RFTagger part–of–speech, Schmid and Laws 2008). I will not detail the differences and the search queries here.

The fourth configuration that I include in the analysis is the use of the keyword as an accusative object. Here, too, there are different search queries. The first variant searches for a verb that is not separated from the keyword by another verb or an adposition. The second variant searches for a verb which directly follows the keyword in accusative without a preposition before the keyword or before the first left neighbour. In the third version, the combination of the keyword in accusative and a main verb (“Vm.*”) at the beginning of a sentence is searched for as well.

(4) 2:[tag=“Vm.*”] [tag!=“[VS].*”]{0,5} 1:[tag=“Nc..a.*”]
 [tag!=“S.*”] [tag!=“S.*”] 1:[tag=“Nc..a.*”] 2:[tag=“Vm.*”]
 <s>1:[tag=“Nc..a.*”] 2:[tag=“Vm.*”]

The quality of these search queries can certainly be discussed. It is, for example, obvious that not all tenses were included. However, this is not the subject of consideration here as the queries underlying SketchGrammar are not adjustable for the given corpora.

The operators of SketchEngine point out that the results are in principle independent of corpus size (Lexical Computing 2015). The absolute frequency of the keyword, on the other hand, is very important for a meaningful analysis. It is recommended that the keyword occurs at least several hundred times. In the corpora used here this condition is fulfilled without a doubt. The lemma *musliman* occurs in the hrWaC 69,932, in the hsWaC 68,570, in the srWaC 24,914 times and *Muslim* in the German TenTen corpus 315,657 times.

SketchEngine calculates the logDice for each search query and can output the results sorted accordingly. The formula for the calculation is as follows:

$$(5) \logDice = 14 + \log_2 D = 14 + \log_2 \frac{2f_{xy}}{f_x + f_y}$$

D stands for the dice coefficient, which in our concrete case is calculated by dividing the double of the joint occurrence of *musliman/Muslim* and the respective collocation partner (e.g. the adjective in (1)) by the sum of all occurrences of the word *musliman/Muslim* and all occurrences of the collocation partner. The summation with 14 and the logarithmic calculus serve to make the results easier to read and interpret. The results are usually smaller than 10 (Rychlý 2008: 9).

SketchGrammar and SketchEngine have been developed for lexicographic work, so a semantic analysis should work well here. The advantage is that the tool is already available and there is no need to manually analyse such amounts of data. Therefore, it is also well applicable for laymen. However, the use of a SketchGrammar requires access to the commercial SketchEngine. It is also possible to replicate the analysis with corpora that are run using the free NoSketchEngine, although this is more complicated because SketchGrammar is not available here and all the calculations discussed above would have to be done on a one by one basis.

2.3. Clustering

A list of significant collocates is, of course, not in itself information. A meaningful comparison of the conceptualization of the given word is only possible by comparing the semantic fields the collocates belong to. Therefore, in a third step, the results have to be clustered or categorized. This can be done automatically (cf. Wasserscheidt and Hrستیć 2020) or manually. Especially for a qualitative evaluation of the results as well as for an explorative investigation a manual interpretation seems adequate. Since I want to demonstrate an easily replicable method for a distributional analysis in this study, I perform a manual categorization of the words represented in the output of the SketchGrammar.

3. Data

3.1. Dictionaries

Searching for definitions of the terms *musliman*/*Muslim* in a selection of dictionaries yielded the following results. The definition in WordNet (and in the following BabelNet) was also used for comparison.

Language	Source	Definition
Multilingual	WordNet, BabelNet (babelnet.org)	A believer in or follower of Islam (IS A religious person)
German	Duden (www.duden.de)	Anhänger des Islams ‚Follower of Islam‘
German	DWDS (www.dwds.de)	Anhänger des Islam ‚Follower of Islam‘
German		
Bosnian	Rječnik Bosanskog jezika (Čedić 2007)	sljedbenik Božije objave Kur‘ana
Bosnian	Rječnik Bosanskog jezika (Jahić 2014)	<i>musliman</i> : pripadnik muslimanske vjere, sljedbenik islama <i>Musliman</i> : <i>povj.</i> pripadnik musli- manskog naroda, muslimanske nacije u sklopu socijalističke Jugo- slavije; Bošnjak <i>etn.</i>
Bosnian	Rječnik Bosanskoga jezika (Halilović et al. 2010)	1. onaj koji je muslimanske vjerois- povijesti, sljedbenik islama 2. (Musliman) <i>hist.</i> pripadnik boš- njačkog naroda; Bošnjak

Croatian	Hrvatski jezični portal (www.hjp.znanje.hr)	1. onaj koji je muslimanske vjere, sljedbenik islama; muhamedanac 2. (Musliman) <i>pov. neol.</i> pripadnik muslimanske nacionalnosti u BiH; Bošnjak
Croatian	Veliki rječnik Hrvatskoga standardnog jezika (Jojić et al. 2015)	pripadnik ili sljedbenik islamske vjere [<i>Muslimani su postali konstitutivan narod u Jugoslaviji 1971.</i>] → muhamedanac
Croatian	Hrvatski enciklopedijski rječnik (Anić 2003)	1. onaj koji je muslimanske vjere, sljedbenik islama; muhamedanac 2. (<i>pov. neol.</i> Musliman) pripadnik muslimanske nacionalnosti u BiH, Bošnjak
Serbian	Vokabular (https://vokabular.org)	onaj koji ispoveda islam, tj. muhamedovac
Serbian	Rečnik srpskoga jezika (Vujanić and Nikolić 2011)	pripadnik islamske, muslimanske veroispovesti, muhamedanac

Table 1: Definitions of musliman/Muslim in dictionaries

The German, multilingual and Serbian dictionaries consistently use only the definition “follower of Islam”. Most Bosnian and Croatian dictionaries mention two meanings, namely the religious and the ethnic definitions “member of the Muslim nationality” and “Bosniak” respectively. An interesting point here are Jojić et al. (2015), who only cite the religious definition, but exemplify the ethnic meaning in one of the two example sentences provided (cited in square brackets in Table 1). Beyond that there are differences in the designation of the group as either *muslimanska nacionalnost* ‘Muslim nationality’, *bošnjački narod* ‘Bosniak people’, *muslimanski narod* ‘Muslim people’ or even *muslimanska nacija* ‘Muslim nation’.

3.2. Corpora

For this study, the largest corpora accessible for the respective languages were used. These are web corpora whose data was collected by crawling web pages of a certain domain (.ba, .hr, .rs, and .de). Specifically, I use the Bosnian bsWaC, the Croatian hrWaC, the Serbian srWaC (Ljubešić and Klubička 2014) and the German TenTen corpus (Jakubiček et al. 2013). All corpora were crawled at about the same time, and thus reflect a comparable period. The corpora used differ considerably in their size. While the Bosnian web corpus comprises only about 287 million tokens, the German TenTen corpus has almost 20 billion tokens.

Corpus	„Language“	Size (Tokens)
hrWaC 2.2 (Croatian Web)	Croatian	1,397,757,548
srWaC 1.2 (Serbian Web)	Serbian	554,627,647
bsWaC 1.2 (Bosnian Web)	Bosnian	286,865,790
deTenTen (German TenTen13)	German	19,808,173,163

Table 2: Corpora used in the analysis

Using the method described above, I selected the collocation partners with the largest LogDice values. Only the first ten values have been included in the analysis and are shown below as examples. In the results, I removed some errors in the annotation: In some cases, verbs were interpreted as adjectives (*zarobiti*, *dozvoliti*, *streljati*) or nouns were interpreted as adjectives or as verbs (*hrvati*). The collocation partners are always given as lemmas, but all word forms have of course been included in the analysis.

When interpreting the data, it is important to keep in mind that the corpora are web corpora. These were obtained by crawling the Internet under the respective domain in a mainly uncontrolled manner. Although the idea of web corpora is that the diversity of the Internet itself leads to a high representativeness of the corpora, this should not be interpreted as meaning that the four corpora are representative of the respective languages or ethnic groups. This is most obvious in the case of the Bosnian corpus, which of course also contains texts in Croatian and Serbian or texts written by Bosnian Croats or Serbs, and not only by Bosniaks. At the same time many Bosnian Croats and Serbs certainly also use infrastructures under the domains .hr or .rs, just as there are websites in all languages using other domains (e.g. .com, .org or .edu). Therefore, the Bosnian corpus can neither be considered representative for the country Bosnia–Herzegovina in a strict sense. The same applies to the Croatian and Serbian corpus. Furthermore, the corpora of course exclude all language users who do not use the Internet and all texts that have not been digitised or are not publicly accessible. In this respect, too, we cannot speak of a complete representativeness of web corpora.

3.3. Frequency

The relative frequency of the word *musliman/Muslim* is expressed in tokens (word forms) per million words. For comparison, I have also given the frequency of the German word *Moslem*. Since the frequency is very low, I exclude it from the investigation. I have also investigated the distinction between *musliman* and capitalized *Musliman*. However, since *Musliman* also has a low frequency, I do not include it in the analysis either.

Lemma	bsWaC	hrWaC	srWaC	deTenTen
musliman	229.60	46.00	41.00	
Musliman	9.3	4.05	3.92	
Muslim				15.81
Moslem				3.62

Table 3: Relative frequency of lemmata musliman, Musliman, Muslim, Moslem in tokens per million

The differences are immense: The relative frequency of *musliman* in the Bosnian web corpus is more than five times as high as in the Croatian or Serbian web corpus and even more than fourteen times as high as *Muslim* in the German web corpus.

3.4. Attributes

Attributes here refer exclusively to adjectives preceding the noun. Since no normalization of the spelling forms was carried out in the corpora, there are some double mentions, especially the adjectives *bosanskohercegovački*, *bosansko–hercegovački* and *bh* ‘Bosnian–Herzegovinian’. The adjectives can be roughly divided into three groups: Adjectives that represent Muslims as inhabitants of a region (*bosanski* ‘Bosnian’, *bosansko–hercegovački* ‘Bosian–Herzegovinian’, *srebrenički* ‘from Srebrenica’, *sandžački* ‘from Sandžak’, German *bosnisch* ‘Bosnian’, *lebend*⁴ ‘living’), adjectives that define Muslims as followers of a religious community (*sunitski* ‘Sunni’, *šijitski/šiitski* ‘Shiite’, German *sunnitisch* ‘Sunni’, *shiitisch* ‘Shiite’) and their degree of faith (*umeren* ‘moderate’, *iskren* ‘sincere’, *ahmad* ‘Ahmadiyya’, *ortodoksan* ‘orthodox’, *radikalan* ‘radical’, German *gläubig* ‘believing’, *praktizierend* ‘practicing’, *strenggläubig* ‘orthodox’, *radikal* ‘radical’, *fundamentalistisch* ‘fundamentalist’) as well as adjectives referring to Muslims neutrally as humans (*mlad* ‘young’, *punoljetan* ‘full-aged’, *mnogi* ‘many’, *sav* ‘all’).

bsWaC		hrWaC		srWaC		deTenTen	
bosanski ‘Bosnian’	9.89	bosanski ‘Bosnian’	9.92	bosanski ‘Bosnian’	10.92	gläubig ‘believing’	10.19
mlad ‘young’	8.19	sunitski ‘Sunni’	8.20	srebrenički ‘from Srebrenica’	8.87	bosnisch ‘Bosnian’	8.88

4 This participle is most often used in the construction *in Deutschland lebende Muslime* ‘Muslims living in Germany’

bosanskohercegovачki 'Bosnian–Herzegovinian'	7.71	bh 'Bosnian–Herzegovinian'	7.67	sunitski 'Sunni'	8.16	sunnitisch 'Sunni'	8.76
sunitski 'Sunni'	7.27	šijitski 'Shiite'	7.61	bosanskohercegovачki 'Bosnian–Herzegovinian'	7.95	praktizierend 'practicing'	8.44
mnogi 'many'	7.2	bosanskohercegovачki 'Bosnian–Herzegovinian'	7.31	sandžачki 'from Sandžak'	7.86	strenggläubig 'orthodox'	7.97
iskren 'sincere'	7.1	forumski 'from the forum'	7.20	šiitski 'Shiite'	7.67	radikal 'radical'	7.95
ahmad 'Ahmadiyya'	7.03	radikalan 'radical'	6.96	ubijen 'killed'	7.13	lebend 'living'	7.82
bosanskohercegovачki 'Bosnian–Herzegovinian'	6.86	šiitski 'Shiite'	6.91	bosanskohercegovачki 'Bosnian–Herzegovinian'	6.83	schiitisch 'Shiite'	7.68
sav 'all'	6.85	ortodoksan 'orthodox'	6.88	umeren 'moderate'	6.51	fromm 'devout'	7.54
punoljetan 'full-aged'	6.8	bosanskohercegovачki 'Bosnian–Herzegovinian'	6.63	naoružan 'armed'	6.46	fundamentalistisch 'fundamentalist'	7.40

Table 4: Attributive adjectives of musliman/Muslim sorted by LogDice

The regional affiliation of the Muslims is clearly shown by the salience of the local attribution to Bosnia–Herzegovina. It is interesting that in the German corpus the “Bosnian Muslims” are also a stable phrase. Beyond that, however, there are clear differences in attribution. Thus, the Bosnian corpus is the only one which frequently uses the word *musliman* with neutral attributes such as *mlad* ‘young’, *mnogi* ‘many’, *punoljetan* ‘full-aged’ or *sva* (*svi*) ‘all’. In the other corpora attributes that express the degree of religious practice and at the same time have a negative connotation (*radikalan* ‘radical’, *ortodoksan* ‘orthodox’, German *strenggläubig* ‘or-

thodox', *radikal* 'radical', *fundamentalistisch* 'fundamentalist') are strongly represented. It is also interesting to note that in Serbia, in addition to the local Muslims in Sandžak, an intense discourse about the genocide in Srebrenica (including the adjective *ubijen* 'killed') and about dangerous Muslims (*naoružan* 'armed') also seems to exist. One indication that the source of the corpus can affect the data is the Croatian adjective *forumski* 'from the forum', which refers to Muslims who post in Internet forums, which is certainly not a relevant attribute for Muslims beyond the internet.

3.5. Categorization

The categorization of the word *musliman/Muslim* is assessed based on the use of the word in coordinating constructions, especially with coordinating conjunctions. Again, three broad groups can be distinguished. One group are names of members of religious communities or religions (*kršćanin/hrišćanin* 'Christian', *katolik* 'Catholic', *pravoslavac* 'Orthodox', *žid* 'Jew', *islam* 'Islam', *nemusliman* 'non-Muslim', German *Christ* 'Christian', *Jude* 'Jew', *Islam* 'Islam', *Hindu* 'Hindu', *Buddhist* 'Buddhist', also *Nichtmuslim* 'non-muslim', *Atheist* 'Atheist'). A second, smaller group are ethnic designations (*hrvat* 'Croat', *bošnjak* 'Bosniak', *srbin* 'Serb', *šiptar* 'Albanian', *albanac* 'Albanian', German *Araber* 'Arab', *Türke* 'Turk'). Only a few coordinated nouns refer to *musliman/Muslim* as a male (*muslimanka*, German *Muslimin* 'female Muslim').

bsWaC		hrWaC		srWaC		deTenTen	
islam 'Islam'	11.63	kršćanin 'Christian'	11.04	Hrvat 'Croat'	10.71	Christ 'Christian'	11.64
muslimanka 'female Muslim'	11.05	židov ⁵ 'Jew'	9.95	hrišćanin 'Christian'	10.56	Jude 'Jew'	10.82
kršćanin 'Christian'	10.31	katolik 'Catholic'	9.90	katolik 'Catholic'	10.29	Muslimin 'female Muslim'	9.66
katolik 'Catholic'	9.65	islam 'Islam'	9.67	islam 'Islam'	9.82	Islam 'Islam'	9.58
Hrvat 'Croat'	9.53	Hrvat 'Croat'	9.41	hrvati 'Croats'	9.69	Hindu 'Hindu'	9.45

5 This is an incorrect automatic lemmatization within the corpus based on the plural stem.

nemusli- man 'non- Muslim'	9.45	srbin 'Serb'	9.31	muslimanka 'female Muslim'	8.67	Buddhist 'Buddhist'	8.98
hrvati 'Croats'	8.63	pravosla- vac 'Orthodox'	9.26	pravoslavac 'Orthodox'	8.56	Nichtmuslim 'non- Muslim'	9.95
Bošnjak 'Bosniaks'	8.49	hrvati 'Croats'	8.97	šiptar 'Albanian'	8.43	Araber 'Arab'	8.54
hrišćanin 'Christian'	8.47	Srbin 'Serb'	8.57	Albanac 'Albanian'	8.35	Türke 'Turk'	7.92
pravosla- vac 'Orthodox'	8.38	musliman- ka 'female Muslim'	8.54	nemusliman 'non- Muslim'	8.35	Atheist 'atheist'	7.89

Table 5: Nouns in coordination with musliman/Muslim sorted by LogDice

The classification as a member of a religious community in all corpora plays a more important role in coordination than in attribution. The Bosnian, Croatian and Serbian corpora are mainly limited to the locally represented religious communities, while in the German corpus all world religions occur. The pronounced regional orientation is also reflected in the coordination with ethnic labels: While in the German context Arabs and (also locally relevant) Turks are mentioned here, the term *hrvat* 'Croat' occupies the first place in all three other corpora. Apart from these two predominant uses, the classification as a man in relation to a woman also exists, but only marginal – and again, as expected, much prominent in the Bosnian corpus.

3.6. Agent

As agent of an action, the word *musliman/Muslim* has been analysed through its use as subject of a sentence. As with the previous dimensions, three non-neutral groups of verbs can be defined here. The first group of verbs describes religious actions (*klanjati se* 'adore', *poštovati* 'respect', German *beten* 'pray', *fasten* 'fast', *verehren* 'worship', *glauben* 'believe', *begehen* 'celebrate', *gedenken* 'commemorate'). The second group includes verbs that denote an aggressive action (*ubiti/ubijati* 'kill', *pobiti* 'knock down', *napasti* 'attack', *počiniti* 'perpetrate', *ratovati* 'conduct war', German *erobern* 'conquer'). In the third group there are verbs that signify the acceptance of a faith or the commitment to being a Muslim (*prihvatiti* 'accept', *postati* 'become', *izjašnjavati/izjasniti* 'declare'). There are also several neutral verbs that describe everyday actions or attributions of activity (*znati* 'know', *činiti* 'make', *živeti* 'live').

bsWaC		hrWaC		srWaC		deTenTen	
živjeti 'live'	8.1	vjerovati 'believe'	7.13	ubijati 'kill'	7.61	beten 'pray'	8.14
ubiti 'kill'	7.1	klanjati 'adore'	6.96	pobiti 'knock down'	7.56	fasten 'fast'	7.48
prihvati 'accept'	6.97	ubijati 'kill'	6.58	izjašnjavati 'declare'	7.23	verehren 'worship'	6.29
postati 'become'	6.8	mrziti 'hate'	6.57	počiniti 'perpetrate'	7.02	leben 'live'	6.28
izjašnjavati 'declare'	6.75	ratovati 'conduct war'	6.14	ubiti 'kill'	6.96	glauben 'believe'	6.22
govoriti 'speak'	6.73	izjašnjavati 'declare'	6.14	ratovati 'conduct war'	6.81	begehen 'celebrate'	5.77
trebati 'need'	6.72	ubiti 'kill'	6.04	poštovati 'respect'	6.77	protestieren 'protest'	5.66
klanjati 'adore'	6.63	moliti 'pray'	6.01	boriti 'fight'	6.66	zusammen- leben 'coexist'	5.56
znati 'know'	6.63	živjeti 'live'	5.98	živeti 'live'	6.17	erobern 'conquer'	5.53
preživjeti 'survive'	6.6	poštovati 'respect'	5.94	činiti 'make'	5.79	gedenken 'commemorate'	5.42

Table 6: Verbs with musliman/Muslim as subject sorted by LogDice

Even if the collocations are not as strong as in the case of attribution and categorization, important differences can be found in the corpora. Verbs, which describe religious actions and thus probably serve as a description of the cultural practices of Muslims, can be found almost exclusively in the German corpus. In the Croatian and especially in the Serbian corpus, verbs dominate which seem to represent Muslims as aggressive agents. Here, of course, a further analysis is needed that evaluates the entire context of the verbs in question. In contrast to the German corpus, the Bosnian, Croatian and Serbian corpora also contain verbs that deal with the self-identification (*izjašnjavati/izjasniti* 'declare') as Muslim.

3.7. Target

For the purpose of analysing the concept of Muslim as the target of an action, its use as an object of transitive verbs is examined. It is striking that we find here mainly verbs that describe negative attitudes (*mrziti* ‘hate’), negative verbal actions (*vrijeđati/vređati* ‘insult’, *uvrijediti/uvrediti* ‘offend’, German *beleidigen* ‘offend’, *diskriminieren* ‘discriminate’, *hetzen* ‘agitate’, *provozieren* ‘provoke’, *aufhetzen* ‘incite’, *ausgrenzen* ‘exclude’, *stigmatisieren* ‘stigmatize’), expulsion and assault (*napadati* ‘attack’, *zadesiti* ‘befall’, *pobiti* ‘knock down’, *protjerati/proterati* ‘expel’) or even killing (*ubijati* ‘kill’, *klati* ‘slaughter’, *streljati* ‘shoot’). There are very few verbs that express a positive action (*ujediniti* ‘unite’, *pomoći* ‘help’) and also few neutral verbs.

bsWaC		hrWaC		srWaC		deTenTen	
ubijati ‘kill’	9.37	mrziti ‘hate’	8.00	streljati ‘shoot’	8.15	beleidigen ‘offend’	6.94
mrziti ‘hate’	8.89	ubijati ‘kill’	7.70	ubijati ‘kill’	7.87	bekehren ‘convert’	6.76
vrijeđati ‘insult’	8.27	protjerati ‘expel’	7.40	mrziti ‘hate’	7.61	diskriminieren ‘discriminate’	6.39
pozivati ‘call’	8.11	uvrijediti ‘offend’	7.37	pobiti ‘knock down’	7.56	hetzen ‘agitate’	6.28
napadati ‘attack’	8.1	pobiti ‘beat’	7.26	uvrediti ‘offend’	7.54	radikalisieren ‘radicalize’	6.02
protjerati ‘expel’	7.72	klati ‘slaughter’	7.04	naoružavati ‘arm’	7.26	provocieren ‘provoke’	6.00
ujediniti ‘unite’	7.54	vrijeđati ‘insult’	6.99	proterati ‘expel’	7.02	aufhetzen ‘incite’	5.91
zadesiti ‘befall’	7.36	protjerivati ‘expel’	6.82	mobilisati ‘mobilise’	6.81	ausgrenzen ‘exclude’	5.86
pomoći ‘help’	7.31	naoružavati ‘arm’	6.61	vređati ‘insult’	6.60	heiraten ‘marry’	5.83
upozoravati ‘warn’	7.12	napadati ‘attack’	6.48	ujediniti ‘unite’	6.52	stigmatisieren ‘stigmatise’	5.67

Table 7: Verbs with musliman/Muslim as accusative object sorted by LogDice

Muslims in all corpora seem to be mainly the target of aggressive verbal or physical actions. There is a significant difference between dominance of verbal aggression in the German corpus and the prevalence of physical aggression in the Bosnian, Croatian and Serbian corpora. Only to some extent are there also acts with aggressive intentions from within the group of Muslims (*naoružavati* ‘arm’, *mobilisati* ‘mobilise’, German *radikalisieren* ‘radicalize’, *aufhetzen* ‘incite’).

4. Analysis

It is evident that the word senses found in the dictionaries – i.e. Muslim as a member of a religious group and Muslim as a member of an ethnic group – are also represented in the corpus data. Muslims as ‘followers of Islam’ are by far the most salient conceptual dimension in all four corpora by virtue of religious attributes and coordination with followers of other religions. The conceptualisation of Muslims as conventional agent or target of actions, however, is not reflected in the dictionary definitions.

In order to better compare the individual results, it is useful to aggregate them. Since many clusters that I have identified manually are similar across the different structural contexts, I group them into equally named categories. Thus, I have merged the attributive adjectives indicating members of religion, the adjectives indicating degree of belief, the coordinated nouns for religious communities and the verbs describing religious practices into a common category “religion”. Similarly, I have grouped the verbs with aggression-related meanings into the category of “(verbal or physical) aggression” and so on.

For all these categories, I have summed the respective LogDice values from all attributes, coordinated nouns and verbs respectively. This way the categories represent, so to speak, the most salient semantic dimensions of the concept *musliman*/*Muslim*. In the following table the results are shown, whereby I highlighted all values over 30 bold. All contexts that do not fit into any of the dimensions have not been taken into account.

Meaning dimension		bsWaC	hrWaC	srWaC	deTenTen
Salience		229.60	46.00	41.00	15.81
Attribute	Religion	21.40	36,56	22.34	65.93
	Region	23.46	31.53	42.43	8.88
	Human	29.04	0.00	0.00	7.82
Category	Religion	57.89	49.82	47.58	68.31
	Ethnicity	26.65	36.26	37.18	16.46
	Male	11.05	8.54	8.67	9.66

Agent	Religion	6.63	26.04	6.77	39.32
	Aggression	7.10	25.33	35.96	5.53
	Identity	20.52	6.14	7.23	0.00
Target	Verbal Ag- gress.	17.16	22.36	14.14	30.86
	Phys. Ag- gression	32.55	42.70	30.60	6.28
	Help	14.85	0.00	6.52	0.00

Table 8: Aggregated association strengths (LogDice)

As can be seen, in the Bosnian corpus there are only two dimensions – the categorization as adherents of a religion and the physical aggression experienced – that cross this threshold. Apart from that, we find a rather balanced conceptualization in the Bosnian corpus, especially in the attribution. The situation is quite different in the German corpus, in which the concept very one-sidedly emphasizes the religious interpretation of the word. This coincides with the significantly distinct salience of the lexeme in the two corpora: while it is prominent in the Bosnian web corpus, it is rather marginal in the German corpus.

The conceptualization as an actor reveals diametrical differences between the German, Serbian and Bosnian corpus: While the German concept associates Muslims with religious acts, the word in the Serbian corpus correlates predominantly with aggressive acts – in the Bosnian corpus, on the other hand, with self-identification as a Muslim. The Croatian concept combines the characteristics of the Serbian and the German corpus.

A difference can also be observed in the conceptualisation as target. Although in all corpora aggressive actions against Muslims are highly frequent, in the German corpus these almost exclusively comprise speech acts, in the other corpora mainly physical violence.

However, a semantic analysis should be able to abstract over syntactic embeddings. Reducing the dimensions by removing the syntactic context also facilitates visualisation. In this step, it is certainly an important question how aggregation should be carried out i.e. which categories of the individual syntactic contexts can be merged into a semantic “dimension”. I have chosen religion and aggression as distinct dimensions because they occur saliently in at least two syntactic contexts. Also, since most region-specific attributive adjectives refer to Bosniaks, I combine them with the coordinated nouns denoting ethnic groups to form the dimension “ethnicity”. I group all other properties – especially attributes, coordinations and verbs with a neutral meaning regarding the object of investigation – into the dimension “other”. This group also contains collocations that were not considered in Table 8 as they did not fit into any larger semantic category.

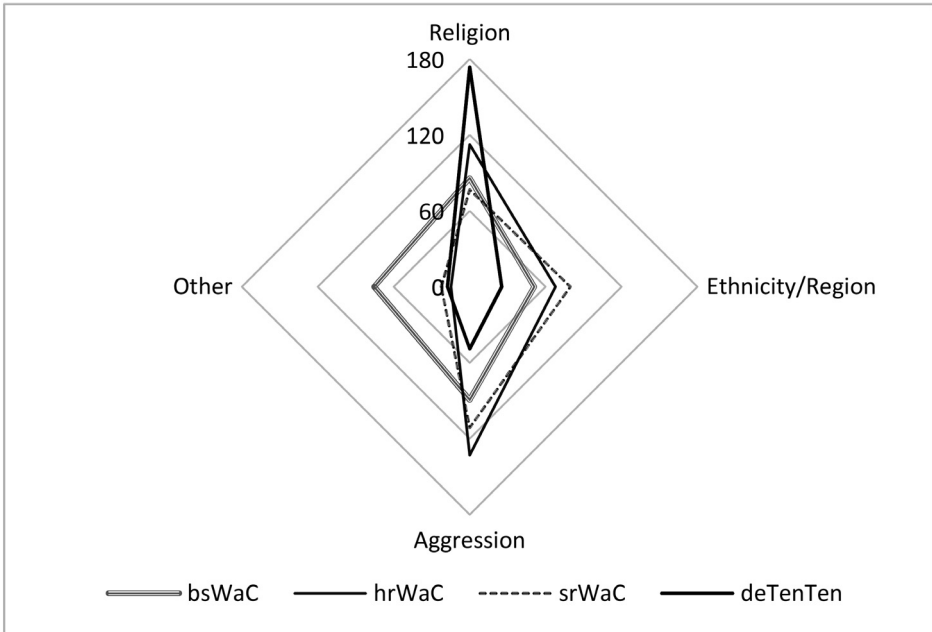


Figure 1: Four dimensions of the meaning of musliman/Muslim

As Figure 1 shows, the dimensions in the Bosnian, Croatian and Serbian corpus are largely similar in respect of the dimensions religion, ethnicity/region and aggression. Bosnian, however, is the only one with a significant share of contexts in the dimension “other” which makes the concept fundamentally different. The German concept differs clearly from the others, since both aggression and ethnicity/region are less pronounced, and the dimension of religion occupies virtually the entire concept.

To summarise, the concept in the German corpus sees Muslims above all as followers of Islam who are exposed to verbal aggression. The concepts in the Serbian and Croatian corpus are very similar and see Muslims simultaneously as followers of Islam and a regionally anchored ethnic group, which is exposed above all to physical aggression, but also acts aggressively itself. In the Bosnian corpus, Muslims are conceptualized as a religious community and regional ethnic group for whom self-identification is important and who often find themselves as victims of aggressive behaviour. In addition, however, the concept also contains a strong neutral component.

5. Discussion

In this paper, a semantic analysis of the word *musliman/Muslim* has been presented using the distributional method. The aim was to compare the results of this analysis with traditional word senses of *musliman/Muslim* in dictionaries. In conclusion, it can be said that the analysis is quite capable of uncovering the meanings codified in dictionaries: It is possible to define *Musliman/Muslim* very clearly as denoting a follower of the religion Islam. This information can be obtained above all from the analysis of adjective attributes and other nouns in coordinating conjunctions. The different histories of denomination in Germany and the former Yugoslavia with the term *Musliman* as an ethnic category can also be identified unambiguously as ethnic connotations hardly play a role in the German corpus and in the dictionaries.

However, the analysis also reveals other aspects of meaning that do not appear in the dictionaries. This applies above all to the contemporary (i.e., at the time of the creation of the web-corpora used and in contrast to the established readings as religious and ethnic group) discourse on Muslims, which deals to a great extent with verbal and physical aggression both experienced and practised. This development, which has been going on for several decades (cf. Tziampiris 2009, papers in Telbizova–Sack & Voß 2019), is also documented in other Western societies (Silva 2017). Furthermore, it can be shown that only in the Bosnian corpus – and thus in the in-group discourse – the word possesses a strong neutral quality.

Even if the results are promising and provide a reasonably clear picture at least within the sample, the method still requires caution. The analysis, which encompasses the entire corpus, necessarily abstracts from the concrete conditions of use, including both the producer and the situation, as well as the broader context. This leads partly to the impression that the concept is relatively stable. But this is, first of all, the result of the abstraction and it should not be concluded that each concrete use cannot also be a very specific one (Taylor 2017: 256). From the analysis above all it cannot be generally concluded what language users actually mean or what recipients understand in an utterance with the word *musliman/Muslim* (Müller 2011: 59).

On all levels, the analysis relies on interpretations whose validity not only depends on the available linguistic data (Roth 2015: 56), as one can see well in the present article. In the case of *musliman/Muslim* this concerns, for example, decisively the pre-selection of syntactic structures, the clustering into certain semantic context types and the aggregation of the latter into semantic dimensions. So while distributional characteristics can provide important indications about the meaning of a word, there are always other methods to be used to validate the results (Dąbrowska 2016: 489). One possible way is to broaden the data basis on the one hand and to replace manual categorisations with quantitative methods on the other hand (cf. Wasserscheidt and Hrستیć 2020).

The advantage of such an analysis, however, is that it creates a background for interpretation and provides a basis for more detailed discourse analyses (Müller 2011: 60).

For example, the restriction of the dictionary definition of *musliman* as a member of a religious community in the Serbian sources can be assessed against this background. This obviously does not correspond to the actual usage in the language. On the contrary, the dimension ethnicity/region is even more pronounced than in the Croatian or Bosnian corpus. The decision regarding the dictionary definition is therefore obviously based on criteria other than language usage.

Another example is the strong religious orientation in the German corpus (cf. also Josua 2019). The knowledge frames of the speaker groups clearly differ in this respect (compare Fillmore 1982 for similar examples). This must be recognised as a baseline when analysing, for example, the understanding of the ethnic situation in Bosnia–Herzegovina or the attitudes towards Muslims in general by speakers of German.

The analysis also leads us to the insight that dictionary definitions are most likely to be mapped via the characteristics that are obtained through (adjective) attribution and categorization in coordination. Conversely, this means that features based on practices or actions do not seem to play a role in dictionary definitions. Attribution and categorization point to what Searle (2010) calls institutional facts: They exist only because of predominantly linguistically mediated social institutions such as religions or ethnic groups, whose ontological status is based on their being accepted. The action-oriented analysis, in contrast, refers to “hard” facts: It deals with expulsions, killings and other aggressions that exist even without institutions. Interestingly enough, the German corpus does not show any hard facts here, but rather institutional, i.e. labels of speech acts, which as such can only exist on the basis of a social acceptance.

Even though the historically conditioned contexts of action or those determined by the current discourse certainly do not represent permanent elements of the concept *musliman/Muslim*, they are nevertheless important. Especially when we regard meanings only as reference to existing knowledge, it is important to grasp this knowledge in our contemporary societies precisely. As *sine qua non*, it is certainly fundamental that Muslims are followers of Islam. For the actual functioning of the word in a historically and discursively charged context, however, the ephemeral attributes (such as aggression) are no less decisive.

References

- Anić, Vladimir (2003). *Hrvatski enciklopedijski rječnik*. Zagreb: Novi Liber
- Čedić, Ibrahim (2007). *Rječnik Bosanskog jezika* (Posebna izdanja / Institut za Jezik u Sarajevu 14). Sarajevo
- Croft, William and Alan Cruse (2004). Categories, concepts and meanings. Croft, William and Alan Cruse, eds. *Cognitive Linguistics*. Cambridge: Cambridge University Press, 74–106
- Dąbrowska, Ewa (2016). Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics* 27(4): 479–491, <https://doi.org/10.1515/cog-2016-0059>
- Erjavec, Tomaž (2010). MULTEXT-East Version 4: Multilingual Morphosyntactic Specifications, Lexicons and Corpora. *Proceedings of the seventh international conference on language resources and evaluation (LREC'06)*, Valetta.
- Fillmore, Charles J. (1982). Frame Semantics. The Linguistic Society of Korea, ed. *Linguistics in the morning calm*. Seoul: Hanshin, 111–137
- Firth, John R. (1957). *Papers in linguistics*. London: Oxford University Press.
- Geeraerts, Dirk (2016). Sense individuation. Riemer, Nick, ed. *The Routledge handbook of Semantics*. London, New York: Routledge Taylor & Francis Group, 233–247
- Halilović, Senahid, Ismail Palić and Amela Šehović (2010). *Rječnik Bosanskoga jezika*. Sarajevo: Filozofski fakultet u Sarajevu
- Jahić, Dževad A. (2014). *Rječnik Bosanskog jezika*. Sarajevo: Bošnjačka Asocijacija 33
- Jakubiček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý and Vít Suchomel (2013). The TenTen corpus family. *7th International Corpus Linguistics Conference CL*, 125–127
- Jojić, Ljiljana, Anuška Nakić, Nada Vajs and Vesna Zevečić (eds.) (2015). *Veliki rječnik Hrvatskoga standardnog jezika*. Zagreb: Školska knjiga
- Josua, Hanna N. (2019). *Die Muslime und der Islam: Wer oder was gehört zu Deutschland?* Leipzig: Evangelische Verlagsanstalt
- Kalwa, Nina (2013). *Das Konzept „Islam“*. Berlin [u.a.]: De Gruyter
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel (2014). The Sketch Engine: Ten years on. *Lexicography* 1(1): 7–36, <https://doi.org/10.1007/s40607-014-0009-9>
- Langacker, Ronald W. (2008). *Cognitive grammar: a basic introduction*. New York: Oxford University Press
- Lexical Computing Ltd (2015). *Statistics used in the SketchEngine*. Online available at <https://www.sketchengine.eu/documentation/statistics-used-in-sketch-engine>. Last accessed 23.08.2019.
- Ljubešić, Nikola and Filip Klubicka (2014). {bs,hr,sr}WaC - Web corpora of Bosnian, Croatian and Serbian. Bildhauer, Felix and Roland Schäfer, eds. *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*. Gothenburg, Sweden: Association for Computational Linguistics, 29–35
- Mitchell, Jeff and Mirella Lapata (2008). Vector-based models of semantic composition. *Proceedings of ACL-08*. Columbus, Ohio: Association for Computational Linguistics, 236–244

- Müller, Marcus (2011). Vom Wort zur Gesellschaft: Kontexte in Korpora. Felder, Ekkehard, Marcus Müller and Friedemann Vogel, eds. *Korpuspragmatik: Thematische Korpora als Basis diskurslinguistischer Analysen* (Linguistik – Impulse & Tendenzen v. 44). Berlin: Walter de Gruyter, 33–82
- Roth, Kersten Sven (2015). Nina Kalwa. 2013. *Das Konzept »Islam«*. Eine diskurslinguistische Untersuchung; Daniela Wehrstein. 2013. *Deutsche und französische Presstexte zum Thema ‚Islam‘. Die Wirkungsmacht impliziter Argumentationsmuster*. *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft* 7(1–2): 55–59, <https://doi.org/10.1515/zrs-2015-0010>
- Rychlý, Pavel (2008). A lexicographer–friendly association score. *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008*: 6–9
- Schmid, Helmut and Florian Laws (2008). Estimation of Conditional Probabilities With Decision Trees and an Application to Fine–Grained POS Tagging. *COLING 2008*, Manchester, United Kingdom: 777–784
- Searle, John (2010). *Making the Social World*: Oxford University Press
- Silva, Derek M. D. (2017). The Othering of Muslims: Discourses of Radicalization in the New York Times, 1969–2014. *Sociological Forum* 32(1): 138–161, <https://doi.org/10.1111/sof.12321>
- Sinclair, John (1995). *Corpus, concordance, collocation*, 3. impr (Describing English language). Oxford: Oxford University Press
- Stefanowitsch, Anatol (2010). Empirical cognitive semantics: Some thoughts. Glynn, Dylan and Kerstin Fischer, eds. *Quantitative Methods in Cognitive Semantics*. Berlin: De Gruyter Mouton, 355–380
- Taylor, John R. (2017). Lexical Semantics. Dancygier, Barbara, ed. *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press, 246–261
- Telbizova–Sack, Jordanka and Christian Voß (eds.) (2019). *Islam auf dem Balkan: Muslimische Traditionen im lokalen, nationalen und transnationalen Kontext* (Südosteuropa–Jahrbuch 44). Berlin: Peter Lang
- Tošić, Jelena. (2022). Populist Variations on Migration: Floating Signifiers of Mobility in the Context of the ‘Balkan Route’ and the COVID–19 Pandemic. *Journal of Balkan and Near Eastern Studies* 24(3): 542–556, <https://doi.org/10.1080/19448953.2021.2015661>
- Tziampiris, Aristotle (2009). Assessing Islamic terrorism in the Western Balkans: The state of the debate. *Journal of Balkan and Near Eastern Studies* 11(2), 209–219, <https://doi.org/10.1080/19448950902921143>
- Vogel, Friedemann (2010). Linguistische Imageanalyse (Lima): Grundlegende Überlegungen und exemplifizierende Studie zum öffentlichen Image von Türken und Türkei in deutschsprachigen Medien. *Deutsche Sprache* (04): 345–377, <https://doi.org/10.37307/j.1868-775X.2010.04.06>
- Vujanić, Milica and Miroslav Nikolić (eds.) (2011). *Rečnik srpskoga jezika*, Izmenjeno i popravljeno izd. Novi Sad: Matica Srpska
- Wasserscheidt, Philipp and Andrea Hrstić (2020). Legal variation? A Frame Analysis of Croatian and Serbian in the Domain of Law. *Mediterranean Language Review* (27): 87–111, <https://doi.org/10.13173/medilangrevi.27.2020.0087>

Koncept musliman: diskursi u mrežnim korpusima bosanskog, hrvatskog, njemačkog i srpskog jezika

Cilj je ovog rada ispitati značenje riječi *musliman* ili njemačkog *Muslim* i odgovoriti na pitanje može li se značenje te riječi u rječnicima mapirati s pomoću klasične distribucijske analize te mogu li se tom metodom odrediti i još neka značenja. K tomu, donosi se i jednostavna metoda distribucijske značenjske analize. Istraživanje kreće od česte pretpostavke u korpusno utemeljenoj semantici i leksikografiji da se značenje riječi razumijeva iz konteksta te se na taj način mogu prepoznati polisemija i sinonimija. Članak proučava četiri mrežna korpusa za bosanski, hrvatski, njemački i srpski jezik. Prvo se analizira kontekst, uz pretpostavku da sintaktičke strukture imaju inherentno značenje u određenom okružju. Tako se pridjevski atributi koriste za analizu obilježja ključne riječi, koordinandi se koriste kao pokazatelji relevantnih kategorija, a glagoli uz koje ključna riječ ima funkciju subjekta ili izravnog objekta koriste se kao pokazatelji tipičnih radnji koje izvršava ili trpi ključna riječ. U drugom se koraku grupiraju leksemi koji se pojavljuju u ta četiri okružja te se određuju značenjske dimenzije ključne riječi.

Rezultati sugeriraju da distribucijska metoda u načelu može prilično dobro mapirati rječničke definicije, no samo kada je riječ o okružju koje uključuje strukture s atributima i koordinaciju, a ne i kada se radi o sintaktičkim konfiguracijama usko vezanima uz radnju, odnosno glagol. Gledano u cjelini, najistaknutije su značenjske dimenzije leksema one koje se tiču članova vjerske zajednice, članova regionalne etničke skupine te iskušene ili potaknute agresije. Po tom se pitanju četiri korpusa značajno razlikuju: uporaba u njemačkome gotovo se isključivo odnosi na vjerske vidove pojma dok preostali korpusi sadrže etničku, odnosno regionalnu komponentu te diskurs o agresiji, koja je u njemačkome uglavnom verbalna, a u drugim korpusima fizička. Neutralne značenjske komponente riječi *musliman* prisutne su samo u korpusu bosanskog jezika.

Keywords: the concept *Muslim*, discourse analysis, semantics, distributional semantics, corpus linguistics, South Slavic languages, German language

Ključne riječi: koncept *musliman*, analiza diskursa, semantika, distribucijska semantika, korpusna lingvistika, južnoslavenski jezici, njemački jezik

