JASMINA JELČIĆ ČOLAKOVAC[1] DOI: 10.15290/CR.2024.45.2.04
University of Rijeka, Faculty of Maritime Studies, Croatia
https://orcid.org/0000-0002-1241-1283
IRENA BOGUNOVIĆ
University of Rijeka, Faculty of Maritime Studies, Croatia
https://orcid.org/0000-0002-2956-7014

Putting languages into perspective: A comprehensive database of English words and their Croatian equivalents

Abstract. Numerous studies have addressed the issue of English words in the context of their adaptation, but there still exists the need for a systematic perspective on English words in terms of their number and frequency of appearance. This article will outline the procedure behind the compilation process of unadapted English words in the Croatian language with a comprehensive description of the final product – an open-access database of single- (SWE) and multi-word (MWE) English expressions extracted from Croatian web corpora (*ENGRI* and *hrWaC*) by means of computational-linguistic tools and manual extraction. The final version of the database contains 2,982 English words in their unadapted form (e.g. *blockbuster*), and 18 words which appear with English orthographic properties in combination with Croatian inflectional affixes (e.g. *downloadati*). Each SWE and MWE entry in the database is accompanied with frequencies of appearance in both corpora as well as its Croatian equivalent where available (29.58% of all entries are listed without an equivalent). The database serves as the first systematic representation of English words in Croatian and provides an indispensable tool for further research into the phenomenon while at the same time opening the door to a new line of research – cognitive processing of English words in Croatian.

Keywords: English words in Croatian, language borrowing, corpus search, database compilation, anglicisms

1. Introduction

Borrowing from English has been documented in many languages. Words and expressions borrowed from English have been investigated in Spanish (e.g. Alvarez-Mellado, 2020), Italian

1 Address for correspondence: Faculty of Maritime Studies, Foreign Languages Department, Studentska 2, 51 000 Rijeka, Croatia. E-mail: jasmina.jelcic@pfri.uniri.hr

(e.g. Pulcini et al., 2012), Norwegian (e.g. Greenall, 2005), Slovenian (e.g. Čepon, 2017), Czech and Slovak (e.g. Entlová & Mala, 2020), Japanese (e.g. Kay, 1995) and South Korean (e.g. Rüdiger, 2018), to mention some of them. Croatian has also become highly receptive to borrowing from English (Mihaljević Djigunović & Geld, 2003). As a result, many English loanwords have become part of Croatian everyday communication (e.g. Nikolić-Hoyt, 2005). The prestigious status of a donor language (e.g. Crystal, 2003) reduces the tendencies of borrowed words to fully adapt to the rules of the recipient language (e.g. McKenzie, 2010; Nikolić-Hoyt, 2005).

Borrowed words are generally described in terms of the degree of their adaptation to the recipient language (e.g. Görlach, 2002; Entlová & Mala, 2020) or their inclusion in the language (e.g. Kay, 1995; Međeral, 2016). A distinction is made between words which have adapted, fully or partially, to the recipient language and those which occur in an original, unadapted form (e.g. *event*, *freelancer*, *bodybuilder*, etc.). Terminology related to unadapted English loanwords is not unified, so terms like 'raw anglicisms' (e.g. Kavgić, 2013), 'English loanwords' (e.g. Görlach, 2002; Kay, 1995; Rüdiger, 2018), 'foreign words' (e.g. Međeral, 2016; Muhvić-Dimanovski & Skelin Horvat, 2006) and 'pseudoanglicisms' (e.g. Filipović, 1990) can be found.

This paper focuses on the latter category, i.e. words borrowed from English which retain the original properties of the donor language, and sometimes take Croatian affixes (e.g. *eventi* (m. nom. pl.), *freelancerima* (m. dat. pl.), *bodybuildera* (m. gen. sg.), etc.). Such words have not become an integral part of Croatian and are perceived as foreign by native speakers, so the term 'foreign words' seems appropriate. For the purpose of precision, the term 'English words' will be used (e.g. Brdar, 2010; Ćoso & Bogunović, 2017).

Borrowed words have long been a subject of discussion among Croatian linguists (Muhvić-Dimanovski & Skelin Horvat, 2006), who generally recommend the use of native words (e.g. Hudeček & Mihaljević, 2005). There are several ways to deal with borrowed words: using multi-word expressions and descriptions, using an existing word and giving it a new meaning, or introducing new words and calques. However, it seems that not all such solutions have been accepted among Croatian speakers (e.g. Drljača, 2006; Patekar, 2019), especially in domains like show business and information technology (e.g. Drljača Margić, 2014). Multi-word expressions and descriptions are often complex to use (e.g. Drljača, 2006). For example, according to the website *Bolje je hrvatski!* (*bolje.hr*), the English word *software* is translated as *programska podrška* (Eng. 'program support'), and *developer* as *razvojni inženjer* (Eng. 'development engineer'). The complexity of these solutions is best illustrated by the translation of the syntagm *software developer* as *razvojni inženjer programske podrške* (Eng. 'program support development engineer'). Giving a new meaning to an already existing word can result in insufficient precision (Drljača, 2006), as in *spravica* (Eng. 'small device') for *gadget* (*bolje.hr*). Finally, the process of introducing a new word or calque is usually slow (e.g. Muhvić-Dimanovski & Skelin Horvat, 2008). For example, the English word *selfie* gained worldwide popularity in 2012, while the Croatian equivalent *sebić* was proposed in 2014 (Halonja & Hudeček, 2014).

In Croatia, a vast body of research has investigated the phenomenon of English words using different theoretical approaches and methods (e.g. Ćoso & Bogunović, 2017; Drljača Margić, 2014; Filipović, 1990; Patekar, 2019). However, most researchers either focus on selectively chosen English words (e.g. Ćoso & Bogunović, 2017; Patekar, 2019) or rely on small-scale, domain-specific corpora (e.g. Brdar, 2010; Hudeček & Mihaljević, 2005). What seems to be neglected is a data-driven approach. One possible reason for that could be the fact that the development of Croatian computational linguistic tools and resources lagged behind those of other languages in the past (e.g. Tadić et al., 2012). However, this is now changing and some new language technologies have been developed in the last decade (Tadić, 2022). Aside from traditional dictionaries (e.g. Filipović, 1990; Görlach, 2002), new resources have emerged. For example, the above-mentioned website *Bolje je hrvatski*!, developed by the *Institute for Croatian language and linguistics*, selectively records the intake of foreign words into Croatian and proposes native equivalents. Borrowed words, including some English words, can also be found in an online dictionary of neologisms (Muhvić-Dimanovski et al., 2016). On the other hand, *Kontekst.io* searches the Croatian web corpus, *hrWaC* (Ljubešić & Klubička, 2016) to find a specific word. The results include information about the word's frequency as well as the frequencies of similar words. Word frequency can also be obtained by searching for a specific word in the available corpora via the platform *Sketch Engine* (Kilgarriff et al., 2004). The results are presented in context, and various options are available to filter them out (e.g. English words occurring in English contexts, names, etc.). However, this method cannot be used to create lists of English words, as the existing corpora are linguistically processed (e.g. tokenized, lemmatized, morphosyntactically tagged, etc.) according to the rules of the Croatian language.

In other languages, researchers have used different methods for the extraction of anglicisms and English words from corpora. Some authors opted for manual search (e.g. Luján García, 2017), and others used the available tools and resources or created new ones (e.g. Alex, 2005; Andersen, 2012). For example, an unsupervised system, based on the idea that there is a relation between Google search results and language membership, was developed for the classification of anglicisms in German (Alex, 2005). Another approach combined lexicon lookup with character N-grams (e.g. Furiassi & Hofland, 2007). Supervised machine learning methods in combination with N-grams has also yielded reliable results (e.g. Alvarez-Mellado, 2020; Serigos, 2017), and it was used to create the Database of English words in Croatian (Bogunović & Kučić, 2022).

The *Database of English words in Croatian* (Bogunović & Kučić, 2022) contains 9,453 English words, some of which (e.g. *summit*, *vintage*, *benefit*) originate from other languages. Although some authors (e.g. Filipović, 1990) state that even words that are not English in origin but were borrowed from English can be considered English loanwords, establishing each word's etymology was not the goal of Bogunović and Kučić's work. The Database represents the result of algorithmic classification and manual evaluation of word lists produced by the algorithm. The results are publicly available on *Figshare.com* as an open source of data. However, it does not provide any information about the availability of Croatian translational equivalents and their frequencies.

Moreover, the database only lists extracted English words with their frequencies, without further elaboration of context-related problems such as polysemy, interlingual cognates, proper names, etc.

The following research aims to fill these gaps by further elaborating the Database of English words in Croatian. The paper presents the results of such an endeavor.

2. Method

The Database of English words and their Croatian equivalents (Bogunović, Jelčić *Čolakovac* & Borucinsky, 2022, hereinafter: the Database) presents an elaboration of Bogunović and Kučić's (2022) database, based on the *ENGRI* corpus (Bogunović et al., 2021; Bogunović & Kučić, 2021), which contains texts from the 12 most popular Croatian news portals between 2014 and 2020. The Database was further updated with data from both *ENGRI* and *hrWaC 2.2* (Ljubešić & Klubička, 2016), built by crawling the .hr top-level domain in 2011 and again in 2014, using the SketchEngine (SkE) platform (Kilgarriff et al., 2004).

2.1. Manual search and evaluation

Manual evaluation of corpus data was used to eliminate the entries from Bogunović and Kučić'l s (2022) database which had appeared in the corpora either in embedded English texts or as part of an English multi-word expression (MWE), the phrase constituents of which the algorithm recognized as single-word (SWEs) entries. The issue of English words appearing in English contexts rather than Croatian sentences was ultimately resolved through SkE search and Xf tagger filtration.

Manual search and evaluation was, however, indispensable in resolving the MWE issue, along with a number of problems which were brought to our attention during corpus search (Jelčić Čolakovac & Borucinsky, 2023). These issues include:

1. the disambiguation of proper names and common nouns (e.g. *PlayStation* as a company vs. *playstation* as a term for a gamer console, etc.);
2. the absence of diacritics from Croatian words appearing in web-crawled sources (e.g. Cro. *vaše* (pro., 2nd pers. pl.) 'yours' vs. Eng. *vase* 'decorative container', etc.);
3. meaning disambiguation (e.g. Cro. *gem* (m. nom. sg.) 'term in a tennis scoring match' vs. Eng. *gem* 'a precious stone', etc.);
4. inflection of Croatian word classes (e.g. Cro. *elaborate* (m. acc. pl.) 'a written elaboration' vs. Eng. *elaborate* (v.) 'to explain' or Eng. *elaborate* (adj.) 'planned in detail', etc.);
5. false cognates (e.g. Cro. *file* (m. nom. sg.) 'chicken breast' vs. Eng. *file* 'document', etc.)
6. adapted English forms (e.g. Cro. *bend* (n.) 'band, a group of musicians' vs. Eng. 'bend' (v.) 'to turn or force from straight or even to curved or angular', etc.).

Once the list of English words was filtered using corpus tools and manual search, our effort was directed towards providing Croatian translational equivalents for each entry in the Database.[2]

2 The term 'entry' will be used interchangeably throughout the paper to refer to both single-word (SWE) and multi-word (MWE) expressions from the Database.

Published sources (dictionaries, books, articles, etc.) on the topic of English loanwords in Croatian served as the stepping-stone to finding adequate equivalents, and, if these proved insufficient, web sources and both corpora were used to aid the search.

Table 1 lists examples from the Database and sources which were used to identify their Croatian equivalents.

Table 1. An exemplification of sources for Croatian equivalents

Entry	Croatian equivalent	Source
ability	*sposobnost* (f. nom. sg.)	Bujas (2019)
afterparty	*zabava nakon posla* (f. nom. sg.; prep.; m. dat. sg.)	Bolje (`https://bolje.hr/`)
bookmark	*straničnik* (m. nom. sg.)	Muhvić-Dimanovski and Skelin Horvat (2008)
	knjižna oznaka (f. adj. sg.; f. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
	dočitnica (f. nom. sg.)	Wiktionary (`https://www.wiktionary.org/`)
composite	*presjek* (m. nom. sg.)	hrWac
	kompozitan (m. adj. sg.)	hrWac
dongle	*hardverski ključ* (m. adj. sg.; m. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
grooming	*uređivanje pasa* (n. nom. sg.; m. gen. pl.)	Bolje (`https://bolje.hr/`)
homepage	*početna stranica* (f. adj. sg.; f. nom. sg.)	Bujas (2019)
	službena stranica (f. adj. sg.; f. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
blind tasting	*kušanje na slijepo* (n. nom. sg.; prep.; n. adv. sg.)	hrWac
acting coach	*učitelj glume* (m. nom. sg.; f. gen. sg.)	Bujas (2019)
	učiteljica glume (f. nom. sg.; f. gen. sg.)	
corkage fee	*naknada za služenje* (f. nom. sg.; prep.; n. nom. sg.)	ENGRI
	čeparina (f. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
remote reality	*udaljena stvarnost (f. nom. sg.)*	ENGRI

2.2. Semantic analysis of sentential context

To evaluate the context in which a particular English word is used in Croatian, the two corpora had to be swept for representative samples of sentential context using the SkE search tools. For those entries appearing in multiple contexts, the predominant context was used in assigning the word to a specific area of human activity, i.e. the semantic field it is usually associated with in the Croatian corpora. The word *design* (*RF* = 7.1402) is one such example, which also appears in contexts related to information and communications technology (ICT), but is predominantly used in contexts related to the fashion industry. Other examples include English words such as *abuse* (*RF* = 0.2683; appearing in law and politics-related context, but predominantly in ICT-related context) and *combat* (*RF* = 0.5191; appearing in sentential contexts related to war, but predominantly in sport and gaming contexts). For those words appearing across multiple contexts with similar frequencies, or whose semantic field could not be determined due to the word's generic reference, the 'OTHER' category was introduced. Such instances include SWEs like *review* (*RF* = 0.7497), *progress* (*RF* = 0.5623), *position* (*RF* = 1.5416) and *reach* (*RF* = 0.4572), and MWEs like *against type* (*RF* = 0.0095), *boom effect* (*RF*= 0.0019) and *extreme ways* (*RF*= 0.0087), which appear in the corpora in various contexts. Based on in-depth semantic analysis of sentential context in which the English words appeared in the Croatian corpora, 12 semantic categories and 12 subcategories have been introduced (Table 2).

Table 2. Representation of semantic categories (n = 12) and subcategories (n =12) for English words in the Database

	Category	Description	Examples
(1)	**ANT**	relating to animals, plants, and non-human entities with human traits	*beast, spider, queen bee*
(2)	**ART**	relating to entertainment and show business, and branches of human creative activities, such as music, dance, literature, etc.	*classic, comeback, casting director*
Subcategory			
	MUSIC	relating to the music industry	*airplay, orchestra, jam session*
	TV	relating to tv, news, and film industry	*binge, spoiler, body horror*
(3)	**PEOPLE**	relating to people, human behavior and activity, and social phenomena in general	*gay, teenager, attention whoring*
Subcategory			
	LANG	relating to language and linguistic phenomena, metaphor, and idiomatic language	*actually, anyway, be on fire*

	Category	Description	Examples
	WAR	relating to combat and human conflict in general	*battle, raid, ground zero*
	LP	relating to government, law, and politics	*e-government, council, cell block*
(4)	**BUSINESS**	relating to business and economy, finance, money, and the world of work in general	*brownfield, offset, debt equity swap*
Subcategory			
	COMMERCE	relating to the act of buying and/or selling, product advertising, and consumerism in general	*delivery, tester, customer loyalty*
(5)	**TECH**	relating to technology and operation of machinery	*clutch, joystick, driver screen*
Subcategory			
	ICT	relating to information and communications technology, Internet, and computer science	*feed, inbox, big data*
	TRANSPORT	relating to means of transport and transport-connected activities	*cargo, landing, economy class*
(6)	**SCIENCE**	relating to science and scientific activity	*molecular, nuclear, case study*
Subcategory			
	EDUCATION	relating to educational activities	*academy, e-learning, action learning*
(7)	**FASHION**	relating to clothing, make-up, style, and the beauty business	*casual, styling, dress code*
(8)	**FOOD**	relating to food and drink, and the act of dining and diet in general	*beef, drive-in, blind tasting*
(9)	**HEALTH**	relating to health, medicine, and the human body	*operation, pill, blood aging*
Subcategory			
	SPORT	relating to sport and games	*draft, playmaker, alpine skiing*
(10)	**TOURISM**	relating to the tourist business and travel for pleasure	*all-inclusive, booking, foot holiday*
Subcategory			
	NATURE	relating to environment and ecology	*emission, winter, hot spring*

	Category	Description	Examples
	LOC	relating to specific places and localities	*penthouse, room, food corner*
(11)	**QUANTITY**	relating to quantity, size, position or duration	*low-level, zero, long term*
(12)	**OTHER**	words with generic references and/or words appearing across multiple contexts	*ancient, progress, free choice*

The proposed categorization is based on the Croatian contexts in which the English words appear, and can by no means be taken to reflect the semantic contexts in which these words are regularly used in English. We would also like to stress that only the most representative semantic categories have been identified. Furthermore, subcategories have been assigned based on available corpus evidence and where repetitive overlap between semantic categories has been observed (e.g. NATURE and LOC have been categorized under TOURISM since a considerable number of words belonging to the two subcategories have repeatedly appeared in contexts relating to tourism and travel, albeit with lower frequencies than in their assigned subcategories).

Finally, after resolving problems through manual search and human evaluation, finding translational equivalents in Croatian, and assigning semantic categories to each entry, the Database has been published as an open-source linguistic resource, with the representation of data in tabular form (row per entry) (Figure 1).

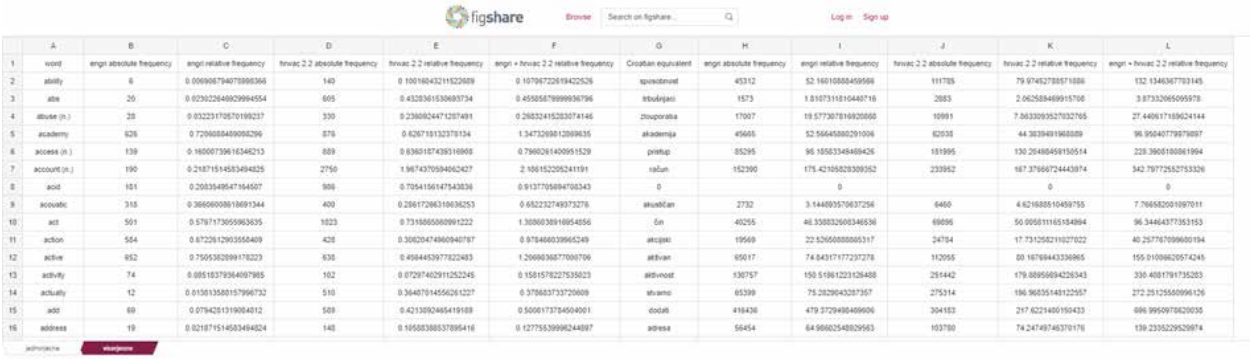

Figure 1. The Database available as open source on *Figshare.com*

3. Results and discussion

The Database contains 2,964 English words and expressions which appear in Croatian texts in their original, unadapted form (e.g. *blockbuster*, *cyberbullying, shopping, zombie, skin*, etc.) and 18 words with English orthographic properties in combination with Croatian inflectional affixes (e.g. *downloadati* (v.t., inf.) ‘to download’, *managerica* (f. nom. sg.) ‘female manager’, etc.).

3.1. Word frequencies

Each database entry is accompanied with a Croatian equivalent if the latter exists in the Croatian language. Absolute frequencies expressing the total number of corpus occurrences for each entry in the database and relative frequencies expressing the proportion of each entry's occurrence in the entire corpus (absolute frequency divided by the total number of words per corpus) are listed for both the English expression and, if applicable, its equivalent. *ENGRI* and *hrWaC 2.2* corpora served as the starting point for the calculation of frequencies which are represented in the Database both per corpus and combined: *ENGRI* absolute frequency (*E*af), *ENGRI* relative frequency (*E*rf), *hrWaC* absolute frequency (*H*af), and *hrWaC* relative frequency (*H*rf). The Database also provides data on combined relative frequencies (*RF*) for both corpora.

Only five entries have been shown to appear in the corpora more than 100,000 times (*web*, *real*, *blog*, *show*, and *post*), while 85 words (2.85%) appear more than 10,000, and less than 100,000 times. SWEs belonging to this frequency band include *link*, *fan*, *e-mail*, *online*, *net*, *mail*, *rock*, *jazz*, etc., with only one MWE appearing in the corpora more than 10,000 times (*big brother*) (cf. Table 2). In total, 709 SWEs and 27 MWEs appear between 1,000 and 10,000 times, which accounts for 24.68% of the Database. If we take the bottom-up perspective on frequencies, 41.78% of all Database entries are recorded 100 times or less in the corpora (184 SWEs and 1062 MWEs respectively), with some MWE entries (e.g. *age verification*, *all girl band*, *anti age effect*, *anti stain effect*, *appearance fee* (*RF* = 0.0012), etc.) and only five SWE entries (*mapmatching*, *mastershot*, *spraypainting* (*RF*= 0.0012), and *personalization* (*RF* = 0.0007)) appearing only once in the Croatian context.[3]

Table 3 illustrates the 10 entries with the highest combined relative frequencies (*RF*) in the Database.

Table 3. Database entries with the highest relative frequencies on the SWE and MWE lists

SWEs					
Entry	**E*af***	**E*rf***	**H*af***	**H*rf***	***RF***
real	86346	99.3957	46730	33.4321	132.8278
web	27648	31.8265	116672	83.4708	115.2973
show	69705	80.2397	36043	25.7863	106.0260
blog	12710	14.6309	112350	80.3787	95.0096
post	18565	21.3708	85431	61.1200	82.4908

3 We would like to note here that all Database entries reflect the spelling of the word(s) as it was used in the Croatian context, which does not necessarily adhere to the standards of the spelling rules for the English language (e.g. *anti age effect* instead of *anti-age effect*, etc.). The same approach was followed in sorting the entries into SWEs and MWEs (e.g. *mapmatching* rather than *map matching*, etc.).

SWEs					
fan	40273	46.3596	39346	28.1494	74.5089
link	6037	6.9494	79778	57.0757	64.0251
online	25053	28.8393	43498	31.1198	59.9592
e-mail	19318	22.2376	49698	35.5555	57.7931
mail	15473	17.8115	42794	30.6162	48.4277
MWEs					
big brother	9657	11.1165	4833	3.4577	14.5742
stand(-)up	2161	2.4876	2055	1.4702	3.9578
fast food	1755	2.0202	2532	1.8115	3.8317
triple(-)double	2920	3.3613	398	0.2847	3.6460
fair play	1606	1.8487	2036	1.4566	3.3053
single	815	0.9382	2329	1.6662	2.6044
made in	663	0.7632	2510	1.7957	2.5589
red carpet	285	0.3281	3061	2.1899	2.5180
open source	141	0.1623	2949	2.1098	2.2721
must have	828	0.9531	1596	1.1418	2.0950

3.2. Single-word and multi-word expressions

The categorization of English words into 1,728 single-word (Cro. *jednorječne*) (SWEs) and 1,254 multi-word (Cro. *višerječne*) expressions (MWEs) represents one of the two major elaborations of Bogunović and Kučić's (2022) database. The restrictions of the original algorithm (Bogunović & Kučić, *under review*), which produced word lists for both databases, prevented it from recognizing English MWEs in the web-crawled sources, hence turning manual evaluation and corpus search into necessary methodological steps in the compilation of our Database.

On the one hand, a detailed manual search of both *hrWaC* and *ENGRI* corpora revealed that many of the words which were initially tagged by the algorithm as SWEs were, in fact, part of an English MWE used in a Croatian context (such examples include English words like *flower* (appearing only as a constituent in the MWEs *flower power* and *flower fashion*) or *cat* (appearing only in MWEs *cat and mouse*, *cat person*, and *cat people*). On the other hand, further examination of corpus examples indicated that some English words were used in Croatian as either a SWE or part of an MWE. These words include, for example, *age* (appearing also in MWEs *age verification*, *anti-age (effect)*, and *coming of age*), *horror* (also in *body horror* and *shock horror*), or *zero*

(also in *ground zero*, *patient zero*, *size zero models*, *zero companies*, *zero hour contract*, and *zero waste*). These entries used as both SWEs and part of English MWEs in Croatian needed to be taken into consideration when absolute and relative frequencies were concerned; it was upon the evaluators to rely on KWIC (*key word in context*) searches in order to distinguish between the SWE and MWE frequencies for words appearing in both wordlists (e.g. the occurrences of the word *coffee* in the MWEs *coffee culture* and *ice coffee* needed to be subtracted from the overall frequencies for the SWE *coffee*). Once these issues had been resolved, the absolute and relative frequencies could be added to the Database for both SWEs and MWEs.

Compounds presented a particular challenge in the process of compilation since some items appeared in the Croatian texts in both hyphenated and non-hyphenated forms. If the English expression appeared in the corpora either as a single word or a hyphenated compound (e.g. *blu(e)-ray*, *all-in-one*, *talk-show*, *co-creation*, *all-inclusive*, *mid-range*, *one-on-one*, *speech-to-text*, *co-production*, *follow-up*, *drive-in*, *pet-friendly*, *ready-made*, etc.), it was categorized as a SWE. The MWE entries used in Croatian as hyphenated MWEs (e.g. *make(-)up artist*, *e(-)book reader*, *regional stand(-)up*, *pop(-)up corner*, etc.) were categorized under MWEs with the hyphen placed in parentheses in order to indicate its optionality. Finally, six entries were listed under both SWEs and MWEs since they appeared in the corpora with and without a hyphen, i.e. as a MWE (*triple(-)double*, *hi(-)tech*, *jet(-)ski*, *head(-)up*, *cut(-)out*, and *stand(-)up*). The final categorization yielded 62 hyphenated compounds on the SWE list, which constitutes 3.59% of the total number of single-word entries in the Database whereas the MWE list included 13 hyphenated entries (1.04% of the total number of multi-word entries). The SWE compound which most frequently appeared in the Croatian corpora is *e-mail*, with a combined relative frequency of 57.79, followed by *start-up* (*RF* = 11.43), *triple(-)double* (*RF* = 3.89), and *blu(e)-ray* (*RF* = 3.07).

3.3. English words with Croatian affixes

Apart from the inclusion of unadapted English words, the Database also lists English words which have taken on Croatian inflectional forms (0.60% of the total number of database entries, 18 entries in total), the majority of which are single-word entries (two inflected MWE entries have been recorded, namely *location manager(ica)* (*managerica*, f. nom. sg.) 'female location manager' and *teen seks comedy* (*seks*, m. nom. sg.) 'teen sex comedy'.

The largest portion of inflected words have taken on the Croatian inflectional suffix *-ica,* which denotes the female gender in Croatian and indicates the female doer of an activity (examples include: Cro. *sprinterica* (f. nom. sg.) 'a female sprinter'; Cro. *managerica* (f. nom. sg.) 'a female manager'; Cro. *youtuberica* (f. nom. sg.) 'a female youtuber'; Cro. *swingerica* (f. nom. sg.) 'a female swinger', etc.). The inflectional suffix was also recorded with *wagsica* (f. nom. sg.), even though the word in English may only refer to women (WAG is literally the acronym of 'wife and girlfriend' and, according to the *Cambridge Diction*ary, stands to denote 'a wife or girlfriend, especially of a well-known sports player'). The *-ica* suffix also appeared in *hoodica* (f. nom. sg. 'a hoodie'), where it does not refer to a female doer, but rather the female gender

of the noun in question, whereby feminine noun properties were added to the English word *hoodie*, probably due to meaning similarities with another Croatian word, *majica* (f. nom. sg., 'any type of T-shirt, blouse, or shirt').

Other Croatian inflectional suffixes which appeared with English words in the corpora include the nominal suffix *-anje* (*swinganje* (n. nom. sg.) 'the act of swinging'), the Croatian verbal suffix *-(a)ti* (*downloadati* (v.t., inf.) 'to download'; *googlati* (v.t., inf.) 'to google') and the adjectival/adverbial suffix *-no* (*maximalno* (n. nom. sg.) 'in the largest or greatest manner'). If an English word was used in the Croatian context in both its unadapted and inflectional form, the two words were listed as separate entries (such was the case with *download* and *downloadati*).

3.4. Croatian equivalents

The second major elaboration of Bogunović and Kučić's (2022) database lies in the addition of Croatian equivalents and their absolute and relative frequencies in both corpora. A total of 29.58% of all the entries in the Database are listed without a Croatian equivalent (296 SWEs and 586 MWEs), while 54 SWE entries and 28 MWE entries are listed with more than one possible equivalent. More than two Croatian equivalents are listed for 11 SWE entries (*bookmark*, *manager*, *managerica*, *maker*, *kickboxer*, *investor*, *hero*, *hater*, *stylist*, *rookie*, and *policy-maker*) and 3 MWE entries (*cloud computing*, *comedy club*, and *cooking class*).

Translational equivalents were found in Croatian for most entries in the Database (e.g. Eng. *ability*/Cro. *sposobnost*, Eng. *air guitar*/Cro. *zračna gitara*, Eng. *zombie*/Cro. *zombi*, Eng. *wild*/Cro. *divlji*, Eng. *winner*/Cro. *pobjednik*, Eng. *city*/Cro. *grad*). In those instances where the English word appeared in the Croatian context bearing more than one meaning, Croatian equivalents were listed separately to account for each of the word's meanings (e.g. Eng. *company*/ Cro. *kompanija* 'an organization that sells goods or services in order to make money', *društvo* 'the fact of being with a person or people, or the person or people you are with'). Croatian equivalents for other meanings of *company* which are found in English are not listed in the Database since the word does not appear to be used in those senses (e.g. Eng. *company* 'a group of actors, singers, or dancers who perform together', 'a large group of soldiers', 'an organized group of young women who are guides', etc.). Similarly, a word was provided with the Croatian equivalent which would belong to the word category in which it was used in the Croatian context. This is to say, English words such as *update*/Cro. *posuvremeniti* (v.t., inf.), *edit*/Cro. *urediti* (v.t., inf.), *ski*/Cro. *skijati* (v.int., inf.), or *record*/Cro. *snimiti* (v.t., inf.) were provided with the translation which reflected its verbal use in Croatian (all of the listed examples are used in Croatian texts as verbs, never as nouns). There are also instances of entries in the Database for which English loanwords in Croatian are listed as translational equivalents due to their high frequency of use among Croatian speakers. In total, 186 database entries (6.24%) are accompanied by a translational equivalent in Croatian that is an English loanword in origin. If we are to analyze each of the two lists separately, SWEs (150 entries, 8.68% of all SWEs) are more frequently accompanied by English loanwords than MWEs (36 entries, 2.87%) in our Database. English loanwords usually appear

in relation to SWEs denoting a doer of an action (Eng. *babysitter*/Cro. *bejbisiter*, *bejbisiterica*, Eng. *blogger*/Cro. *bloger*, *blogerica*, Eng. *breaker*/Cro. *brejker*, *brejkerica*, Eng. *hater*/Cro. *hejter*, *hejterica*, Eng. *leader*/Cro. *lider*, *liderica*, etc.). As expected, English loanwords also frequently appear among English words from the domains of commerce, economy and business (e.g. Eng. *banner*/Cro. *baner*, Eng. *bestseller*/Cro. *bestseler*, Eng. *budget*/Cro. *budžet*, Eng. *consulting*/ Cro. *konzalting*), popular culture (e.g. Eng. *blockbuster*/Cro. *blokbaster*, Eng. *fake*/Cro. *fejk*, Eng. *fancy*/Cro. *fensi*), sports (e.g. Eng. *bridge*/Cro. *bridž*, Eng. *fitness*/Cro. *fitness*, Eng. *jogging*/Cro. *džoging*) and ICT (e.g. Eng. *cluster*/Cro. *klaster*, Eng. *disc*/Cro. *disk*, Eng. *inch*/Cro. *inč*, Eng. *scart*/ Cro. *skart*). The results in the case of MWEs revealed that out of 36 English expressions only 7 were accompanied by an English loanword as a translational equivalent (e.g. Eng. *spin doctor*/ Cro. *spin doktor*, Eng. *shock horror*/Cro. *šok horor*), whereas in the case of the other 29 MWEs only one phrasal constituent was a loanword from English. Such examples include Eng. *gay friend*/Cro. *gej prijatelj*, Eng. *gala opening*/Cro. *gala otvorenje*, Eng. *ultra clear*/Cro. *ultra čist*, or Eng. *travel blog*/Cro. *putopisni blog*.

Multiple Croatian equivalents were oftentimes available for one and the same meaning (e.g. *bookmark* or *corkage fee*), in which cases all Croatian equivalents were listed along with their respective frequencies. Due to the inflectional nature of the Croatian language, English words referring to people were listed with separate Croatian equivalents where one would denote the male, and the other the female doer (e.g. Eng. *advisor*/ Cro. *savjetnik* (m. nom. sg.), *savjetnica* (f. nom. sg.); Eng. *publisher*/ Cro. *izdavač* (m. nom. sg.), *izdavačica* (f. nom. sg.); Eng. *rookie*/ Cro. *početnik* (n. nom. sg.), *početnica* (f. nom. sg.), *novak* (n. nom. sg.), *novakinja* (f. nom. sg.); etc.). The SWE list includes 65 such entries where both the male and female doer were listed under Croatian equivalents; this figure does not include *rapper*/*rapperica*, *manager*/*managerica*, *teenager*/*teenagerica*, *roller*/*rollerica*, *rocker*/*rockerica*, and *youtuber*/*youtuberica*, which are listed as separate database entries since the expressions denoting female doers in Croatian (*rapperica*, *managerica*, etc.) have been adapted to the Croatian language on the morphological level by the addition of the Croatian inflectional suffix, but have retained English orthographic properties. The MWE list includes 37 such entries where both male and female doers are provided as equivalents (e.g. Eng. *decision maker*/Cro. *donositelj odluka* (m. nom. sg.; f. gen. pl.), *donositeljica odluka* (f. nom. sg.; f. gen. pl.); Eng. *dirty cop*/Cro. *korumpirani policajac* (m. adj. sg.; m. nom. sg.), *korumpirana policajka* (f. adj. sg.; f. nom. sg.); Eng. *gay friend*/Cro. *gej prijatelj* (m. adj. sg.; m. nom. sg.), *gej prijateljica* (f. adj. sg.; f. nom. sg.); Eng. *patient zero*/Cro. *nulti pacijent* (m. adj. sg.; m. nom. sg.), *nulta pacijentica* (f. adj. sg.; f. nom. sg.); etc.). In total, Croatian equivalents for male and female doers are listed separately for 102 entries, which comprises 3.19% of the total number of database entries.

3.5. Semantic categorization

An overview of the database entries from the semantic perspective revealed interesting results in terms of the areas of human activity they originate from, i.e. the specific context in which

they usually appear in the Croatian language. The total count of SWEs and MWEs assigned to each of the 12 categories is listed in Table 4.

Table 4. Representation of the total counts (*n*) and percentages (%) of SWEs and MWEs across the 12 semantic categories

Category	SWEs		MWEs		Total
	n	**%**	***n***	**%**	**%**
PEOPLE	273	15.79	343	27.35	20.66
TECH	351	20.31	131	10.45	16.16
OTHER	291	16.84	141	11.24	14.49
BUSINESS	148	8.56	175	13.96	10.83
HEALTH	165	9.49	158	12.59	10.79
ART	190	10.99	108	8.61	9.99
TOURISM	89	5.15	78	6.22	5.60
FASHION	68	0.04	36	2.87	3.49
FOOD	58	3.36	37	2.95	3.19
SCIENCE	42	2.43	26	2.07	2.28
QUANTITY	41	2.37	16	1.28	1.91
ANT	13	0.75	5	0.39	0.60

Differences between single- and multi-word English expressions have also been observed for the 12 subcategories. The most frequent subcategories on the SWE list were ICT (*n* = 284, 16.44%), SPORT (*n* = 126, 7.29%), and MUSIC (*n* = 82, 4.75%), followed by LANG (*n* = 48, 2.78%), TV (*n* = 41, 2.37%) and COMMERCE (*n* = 39, 2.26%). On the other hand, SPORT (*n* = 104, 8.29%), LANG (*n* = 101, 8.05%), and ICT (*n* = 70, 5.58%) were found to be the most frequent subcategories on the MWE list, followed by COMMERCE (*n* = 38, 3.03%), TV (*n* = 35, 2.79%), and LOC (*n* = 32, 2.55%). In total, the most frequent subcategory in the Database was ICT (*N* = 354, 11.87%), followed by SPORT (*N* = 230, 7.71%) and LANG (*N* = 149, 4.99%).

The highest percentage of database entries was found to belong to the PEOPLE category (20.66%), i.e. they were related to human behaviour and social activity, as well as social phenomena in general. Some of the examples of database entries assigned to this category include: words and expressions related to specific people or groups (e.g. *youtuber* (*RF* = 2.9318) and *youtuberica* (*RF* = 0.4364), *millennials* (*RF* = 0.0544), *hooligan* (*RF* = 0.0970), *homeless people* (*RF* = 0.0052), etc.); words related to human (social) activity (e.g. *crowdfunding* (*RF* = 2.1826),

bullying (*RF* = 1.0180), *mobbing* (*RF* = 2.9945), *dating* (*RF* = 0.4373), etc.), and words related to social phenomena, e.g. activism surrounding human sexuality rights (e.g. *straight* (*RF* = 0.5610), *gay* (*RF* = 24.0839), *queer* (*RF* = 2.8626), *drag queen* (*RF* = 0.2211), etc.). These results could be related to the role of the Internet in today's society, where English is the dominant language. Social networking and the Internet in general have been recognized as activities that facilitate spontaneous vocabulary acquisition (e.g. Godwin-Jones, 2019; Zourou, 2012).

The TECH category was the second most frequently recorded category in the Database (16.16%). The recorded results are not surprising if we consider that the inflow of English words in the last few decades closely follows the growth of the ICT industry, namely the Internet. SWEs like *page* (*RF* = 3.5898; Cro. *stranica*, f. nom. sg.), *memory* (*RF* = 1.9727; Cro. *memorija*, f. nom. sg.), and *domain* (*RF* = 0.3248; Cro. *domena,* f. nom. sg.) are used in Croatian contexts only in reference to ICT, and never to refer to their generic denotations (this is why the Croatian word *memorija* (as in 'computer memory') was used as the translational equivalent for *memory*, and not *sjećanje* ('a memory or the act of remembering', n. nom. sg.), which in Croatian can never be used to refer to the ability of a machine to memorize information, but only to the human capacity to remember). The influence of ICT is also evident in terms of borrowed multi-word units, with English MWEs such as *big data* (*RF* = 0.5395), *cloud computing* (*RF* = 0.4135), and *flat rate* (*RF* = 0.5384) frequently appearing in the Croatian corpora. One possible reason for the frequent use of ICT-related English words could be positive attitudes towards English words, especially in this domain (e.g. Drljača Margić, 2014).

3.6. Per-corpus analysis

A per-corpus analysis of database entries revealed significant variations in frequencies collected for some entries. Table 5 shows the 10 SWE and MWE entries with the highest relative frequencies (*rf*) in each corpus.

Table 5. Database entries with the highest per-corpus frequencies on the SWE and MWE lists

ENGRI					
SWE Entry	**E*af***	**E*rf***	**MWE Entry**	**E*af***	**E*rf***
real	86346	99.3957	**big brother**	9657	11.1165
show	69705	80.2397	**triple(-)double**	2920	3.3613
fan	40273	46.3596	**stand(-)up**	2161	2.4876
summit	28575	32.8936	**fast food**	1755	2.0202
web	27648	31.8265	**fair play**	1606	1.8487
online	25053	28.8393	**plus size**	1531	1.7624
rock	20308	23.3772	**open air**	929	1.0694
jazz	20269	23.3323	**must have**	828	0.9531

ENGRI					
e-mail	19318	22.2376	**rock and roll**	820	0.9439
post	18565	21.3708	**street food**	800	0.9209
hrWac					
SWE Entry	**Haf**	**Hrf**	**MWE Entry**	**Haf**	**Hrf**
web	116672	83.4708	**big brother**	4833	3.4577
blog	112350	80.3787	**red carpet**	3061	2.1899
post	85431	61.1200	**open source**	2949	2.1098
link	79778	57.0757	**fast food**	2532	1.8115
net	57462	41.1101	**made in**	2510	1.7957
e-mail	49698	35.5555	**black carpet**	2466	1.7643
real	46730	33.4321	**off topic**	2129	1.5232
online	43498	31.1198	**stand(-)up**	2055	1.4702
mail	42794	30.6162	**fair play**	2036	1.4566
fan	39346	28.1494	**must have**	1596	1.1418

With further comparison of the two corpora it has been established that some entries appeared in one corpus and never in the other. With regard to SWEs, 28 of them appeared in *hrWaC* and never in *ENGRI*, whereas 14 SWEs were found in *ENGRI* that never appeared in KWIC searches in *hrWaC*. Words like *generally* (H*rf*= 0.47), *chain* (H*rf*= 0.37), and *screencast* (H*rf*= 0.09) were never found in the *ENGRI* corpus, despite the word *generally*, for example, appearing 662 times in *hrWaC*. Similarly, *selfie* (E*rf*= 7.08) and *blockchain* (E*rf*= 1.24) appeared more than 1,000 times in *ENGRI*, but were never found in *hrWac*. Similar results were obtained for MWEs in our Database, with 64 of them never appearing in *ENGRI* (e.g. *mind map* (H*rf*= 0.01), *girls' night out* (H*rf*= 0.01), *critical art* (H*rf*= 0.01), etc.), and 319 MWEs found in *ENGRI*, but never in *hrWaC* (e.g. *ticket point* (E*rf*= 0.25), *ice bucket* (E*rf*= 0.15), etc.). These differences may reflect the differences between the two corpora: while *ENGRI* contains texts collected exclusively from news portals, *hrWaC* also includes texts from blogs, forums, etc. Another possible explanation could be the time period in which the texts were collected. Some words, like *selfie*, could have become more popular after the *hrWaC* corpus had been compiled.

3. Concluding remarks

The focus of research on borrowed words in Croatian has primarily been on loanwords which have undergone adaptation to the recipient language. However, the research outlined in this paper highlights the significance of unadapted English words which can also be found in the Croatian language. Their number and frequency of occurrence in the Croatian corpora suggest they have transcended the boundaries of a simple linguistic phenomenon; a considerable number of English words continue to appear in use despite the fact that acceptable Croatian equivalents

are readily available to language users. This can be taken as evidence corroborating the prestigious status of English among speakers of other languages, as well as proof of its overarching influence on all domains of human activity, especially ICT and popular culture.

The Database, in its current size and scope, presents a valuable addition to language resources in view of open-science policy. Both the Database and the *ENGRI* corpus, created primarily for the purposes of database compilation, are freely available as tools for other researchers whose topics of interest include, but are not limited to, language contact, borrowing process, language prestige, corpus linguistics, or even cognitive processing of foreign words in a recipient language. It serves as a unique tool for the Croatian language, offering a systematic representation of unadapted English words, while also providing insight into the frequency of their use among Croatian language speakers. Furthermore, the model of data representation in the Database provides a foundation for all types of contrastive linguistic research on borrowed lexis, where various factors such as word length, type, or frequency are in focus. Since our data are time-sensitive in nature, our intention is to repeat the compilation process and gather data from texts published after 2020, which would allow us to conduct diachronic studies into the status of English words in Croatian.

ACKNOWLEDGEMENTS

The study outlined in this paper has been supported in part by the Croatian Science Foundation (HRZZ) under project number UIP-2019-04-1576.

References

[Dataset] Bogunović, I., Jelčić Čolakovac, J. & Borucinsky, M. (2022). The database of English words and their Croatian equivalents. figshare. DOI: https://doi.org/10.6084/m9.figshare.20014712.v1

[Dataset] Bogunović, I. & Kučić, M. (2021). *Korpus hrvatskih novinskih portala ENGRI* [Corpus of Croatian news portals ENGRI]. https://urn.nsk.hr/urn:nbn:hr:187:920822.

[Dataset] Bogunović, I., Kučić, M., Ljubešić, N. & Erjavec, T. (2021). Corpus of Croatian news portals ENGRI. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1416

[Dataset] Bogunović, I. & Kučić, M. (2022). The database of English words in Croatian.xlsx. figshare. DOI: https://doi.org/10.6084/m9.figshare.20014364.v1

Brdar, I. (2010). *Engleske riječi u jeziku hrvatskih medija* [English words in the language of Croatian media]. *Lahor 10*, 174–189.

Alex, B. (2005). An unsupervised system for identifying English inclusions in German text. In C. Callison-Burch & S. Wan (Eds.), *43. Proceedings of the Annual Meeting of the Association for Computational Linguistics* (pp. 133–138). The University of Michigan. https://dl.acm.org/doi/10.5555/1628960.1628985

Alvarez-Mellado, E. (2020). An annotated corpus of emerging Anglicisms in Spanish newspaper headlines. In *Proceedings of The 4th Workshop on Computational Approaches to Code Switching* (pp. 1–8). European Language Resources Association. https://arxiv.org/abs/2004.02929

Andersen, G. (2012). Semi-automatic approaches to Anglicism detection in Norwegian corpus dana. In C. Furiassi, V. Pulcini & F. R. González (Eds.), *The anglicization of European lexis* (pp. 111–130). John Benjamins. https://doi.org/10.1075/z.174.09

Bogunović, I. & Kučić M. The database of English words in Croatian. Under review.

Bujas, Ž. (2019). *Novi englesko-hrvatski rječnik* [The new English-Croatian dictionary]. Zagreb: Nakladni zavod Globus.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511486999

Čepon, S. (2017). Anglicizmi v poslovni nomenklaturi turistinih podjetij v Sloveniji. *Revija za ekonomske in poslovne vede* 2, 35–49.

Ćoso, B. & Bogunović, I. (2017). Person perception and language: A case of English words in Croatian. *Language & Communication*, *53*, 25–34. https://doi.org/10.1016/j.langcom.2016.11.001

Drljača, B. (2006). *Anglizmi u ekonomskome nazivlju hrvatskoga jezika i standardnojezična norma* [Anglicisms in the economic terminology of the Croatian language and the standard language norm]. *Fluminensia*, 18(1), 65–85.

Drljača Margić, B. (2014). Contemporary English influence on Croatian: A university students' perspective. In A. Koll-Stobbe & S. Knospe (Eds.), *Language Contact Around the Globe* (Proceedings of the LCTG3 Conference, pp. 73–92). Peter Lang.

Entlová, G. & Mala, E. (2020). The occurrence of anglicisms in the Czech and Slovak lexicons. *Xlinguae, 13*(2), 140–148. https://doi.org/10.18355/XL.2020.13.02.11

Filipović, R. (1990). *Anglicisms in Croatian or Serbian: Origin – development – meaning*. Školska knjiga.

Furiassi, C. & Hofland, K. (2007). The retrieval of false anglicisms in newspaper texts. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years On* (pp. 347–363). Brill/Rodopi. https://doi.org/10.1163/9789401204347_020

Görlach, M. (Ed.). (2002). *An Annotated Bibliography of European Anglicisms*. Oxford University Press. https://doi.org/10.1515/9783484431027.15

Godwin-Jones, R. (2019). Contributing, creating, curating: Digital literacies for language learners, *language learning & technology*, *19*(3), 8–20. https://www.lltjournal.org/item/10125-44427/

Greenall, A. K. (2005). To translate or not to translate: Attitudes to English loanwords in Norwegian. In B. Preisler, A. Fabricius, H. Haberland, S. Kjærbeck & K. Risager (Eds), *The consequences of mobility* (pp. 212–226). Roskilde University.

Halonja, A. & Hudeček, L. (2014). Pokloni mi svoj *selfie* [Give me your *selfie*]. *Hrvatski jezik*, *2*, 26–27.

Hudeček, L. & Mihaljević, M. (2005). *Nacrt za višerazinsku kontrastivnu englesko-hrvatsku analizu* [An outline of a multilevel contrastive Croatian-English analysis]. *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, 31, 107–151. https://hrcak.srce.hr/9381

Jelčić Čolakovac, J. & Borucinsky, M. (2023). In the melting pot of web-crawled texts: The challenges of extracting English words and phrases from Croatian corpora. *International Journal of Applied Linguistics*, *34*(1), 166–182. https://doi.org/10.1111/ijal.12485

Kavgić, A. (2013). Intended communicative effects of using borrowed English vocabulary from the point of view of the addressor: Corpus-based pragmatic analysis of a magazine column. *Jezikoslovlje*, *14*(2–3), 487–499. https://hrcak.srce.hr/112204

Kay, G. (1995). English loanwords in Japanese. *World Englishes*, *14*(1), 67–76. https://doi.org/10.1111/j.1467-971X.1995.tb00340.x

Kilgarriff, A., Rychlý, P., Smrž, P. & Tugwell, D. (2004). Itri-04-08 The Sketch Engine. *Information Technology*, pp. 105–116.

Kučić, M. (2021). Creating a web corpus using GO. In M. Koričić et al. (Eds.), *2021 44th International Convention on Information, Communication and Electronic Technology* (MIPRO) (pp.1676–1678). Croatian Society for Information, Communication and Electronic Technology - MIPRO: Rijeka. DOI: https://doi.org/10.23919/MIPRO52101.2021.9597093

Luján García, C. (2017). Analysis of the presence of Anglicisms in a Spanish internet forum: some terms from the fields of fashion, beauty, and leisure. *Alicante Journal of English Studies*, *30*, 281–305. https://doi.org/10.14198/raei.2017.30.10

Ljubešić, N. & Erjavec, T. (2011). HrWaC and slWac: compiling web corpora for Croatian and Slovene. In I. Habernal & V. Matoušek (Eds.), *Text, speech and dialogue, lecture notes in computer science* (pp. 395–402). Springer.

Ljubešić, N. & Klubička, F. (2016). {bs, hr, sr} wac-web corpora of Bosnian, Croatian and Serbian. In F. Bildhauer & R. Schäfer (Eds.), *Proceedings of the 9th web as corpus workshop (WaC-9)* (pp. 29–35). Association for Computational Linguistics. http://dx.doi.org/10.3115/v1/W14-0405

McKenzie, R. M. (2010). *The social psychology of English as a global language: Attitudes, awareness and identity in the Japanese context*. Springer. https://doi.org/10.1007/978-90-481-8566-5

Međeral, K. (2016). *Jezične bakterije – pomagači ili* štetočine *u jezičnome organizmu?* [Language bacteria – helpers or foes in the language organism?]. *Hrvatski jezik*, 3, 1–10. https://hrcak.srce.hr/171398

Mihaljević Djigunović, J. & Geld, R. (2003). English in Croatia today: Opportunities for incidental vocabulary acquisition. *Studia Romanica et Anglica Zagrabiensia*, *43*, 335–352. https://hrcak.srce.hr/21021

Muhvić-Dimanovski, V. & Skelin Horvat, A. (2006). *O riječima stranoga podrijetla i njihovu nazivlju* [On words of foreign origin and their terminology]. *Filologija*, *44-47*, 203–215. https://hrcak.srce.hr/22242

Muhvić-Dimanovski, V. & Skelin Horvat, A. (2008). Contests and nominations for new words-why are they interesting and what do they show. *Suvremena lingvistika*, *65*(1), 1–26. https://hrcak.srce.hr/25183

Muhvić-Dimanovski, V., Skelin Horvat, A. & Hriberski, D. (2016). *Rječnik neologizama u hrvatskome jeziku* [The dictionary of neologisms in Croatian]. www.rjecnik.neologizam.ffzg.unizg.hr

Nikolić-Hoyt, A. (2005). *Englesko-hrvatski jezično-kulturni dodiri* [English and Croatian in language and cultural contacts]. In D. Stolac, N. Ivanetić & B. Pritchard (Eds.), *Jezik u društvenoj*

interakciji (Zbornik radova sa savjetovanja održanoga 16. i 17. svibnja u Opatiji) (pp. 353–358). Zagreb: Hrvatsko društvo za primijenjenu lingvistiku.

Patekar, J. (2019). *Prihvatljivost prevedenica kao zamjena za anglizme* [The acceptability of loan translations as substitutes for anglicisms]. *Fluminensia*, *31*(2), 143–179. https://doi.org/10.31820/f.31.2.17

Pulcini, V., Furiassi, C. & Gonzales, F. R. (2012). The lexical influence of English on European languages: From words to phraseology. In V. Pulcini, C. Furiassi & F. R. Rodrigues (Eds.), *Anglicization of European lexis* (pp. 1–27). John Benjamins. https://doi.org/10.1075/z.174.03pul

Rüdiger, S. (2018). Mixed feelings: Attitudes towards English loanwords and their use in South Korea. *Open Linguistics*, *4*, 184–198. https://doi.org/10.1515/opli-2018-0010

Serigos, J. R. L. (2017). *Applying corpus and computational methods to loanword research: new approaches to Anglicisms in Spanish*. [Unpublished doctoral thesis]. University of Texas at Austin.

Tadić, M. (2022). *European language equality: Report on the Croatian language*. European Language Equality (ELE): Berlin. https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_7__Language_Report_Croatian_.pdf

Tadić, M., D. Brozović-Rončević & Kapetanović, A. (2012). *Hrvatski jezik u digitalnom dobu* [The Croatian language in the digital age]. Springer. https://doi.org/10.1007/978-3-642-30882-6_9

Zourou, K. (2012). On the attractiveness of social media for language learning: a look at the state of the art. *Alsic. Apprentissage Des Langues et Systèmes d'Information et de Communication*, *15*(1). https://doi.org/10.4000/alsic.2436

***

Jasmina Jelčić Čolakovac received her MA degree in English language and History in 2011 at the Faculty of Arts and Sciences in Rijeka. She obtained her PhD degree in Applied Linguistics in 2017 at the University of Ljubljana. Her research interests include English loanwords in Croatian and the processing of metaphoric expressions in bilingual speakers. She has been part of the research team in the newly established Laboratory for Language, Cognition & Neuroscience (LaconLab) since 2020.

Irena Bogunović received her MA degree in English and Croatian languages in 2008 at the Faculty of Arts and Sciences in Rijeka. She obtained her PhD degree in Cognitive Sciences in 2017 at the University of Zagreb. Her research interests include English loanwords in Croatian and their neurocognitive processing by bilingual speakers. She has been acting as the head of the newly established Laboratory for Language, Cognition & Neuroscience (LaconLab) since 2020.