GERMAN LOANWORDS IN CROATIAN NEWSPAPER CORPORA

GERMANIZMI U HRVATSKOM NOVINSKOM KORPUSU

Lidija Tepeš Golubić

Zagreb University of Applied Sciences

ABSTRACT

In this paper German loanwords that have been found automatically through detection in four widely circulated Croatian newspapers (Slobodna dalmacija, Novi list, Večernji list and Glas Slavonije) have been researched during a period of one year. The research is based on written, that is, digital sources – the newspapers, which have been automatically searched for German loanwords. The list of German loanwords itself has been compiled by their listing from all the relevant written sources – dictionaries and scientific papers.

Even though the German language historically had quite a large influence on some parts of Croatia, the research has shown that German loanwords occur less than expected in contemporary newspaper texts. Their occurrence in the Croatian language is, however, systematic and it is possible to detect them by using automatic methods.

Keywords: German loanwords, digitized newspaper corpora, automatic detection

SAŽETAK

U radu se istražuju germanizmi koji su automatsko detekcijom pronađeni u četirima visokotiražnim hrvatskim dnevnim novinama (Slobodna Dalmacija, Novi list, Večernji list i Glas Slavonije) tijekom perioda od godinu dana. Istraživanje se temelji na pisanim odnosno digitalnim izvorima – novina, kroz koje su se automatski pretraživali germanizmi. Sam popis germanizama sastavljen je popisivanjem istih iz relevantnih pisanih izvora – rječnika i znanstvenih radova.

Iako je njemački jezik povijesno imao velikog utjecaja na jezik pojedinih dijelova Hrvatske, istraživanje je pokazalo da se germanizmi u suvremenim novinskim tekstovima pojavljuju rjeđe od očekivanog. Njihovo je pojavljivanje u hrvatskom ipak sustavno te ih je moguće detektirati uporabom automatskih metoda.

Ključne riječi: *germanizmi, digitalizirani novinski korpus, automatska detekcija*

ZUSAMMENFASSUNG

Germanismen im kroatischen Zeitungskorpus In der Arbeit unter dem Titel" Germanismen im kroatischen Zeitungskorpus"werden Germanismen erforscht, welche durch eine automatische Detektion im Zeitraum von einem Jahr in vier Tageszeitungen mit hocher Verlagszahl gefunden wurden (Slobodna Dalmacija, Novi list, Večernji list i Glas Slavonije). Die Forschung basiert aus Quellen im digitalen und gedruckten Zeitungsformat bei welchen automatisch nach Germanismen gesucht wurde. Allein die Liste mit Germanismen wurde durch Auflistung von Germanismen aus Wörterbüchern und Forschungsarbeiten zusammengestellt. Obwohl die deutsche Sprache einen großen Einfluss auf manche Teile Kroatiens hatte, zeigte die Forschung, dass Germanismen in zeitgenössischen Zeitungsartikeln seltener vorkommen als erwartet. Deren Erscheinung in der kroatischen Sprache geschieht systematisch und man kann sie anhand von Nutzung verschiedenenr automatischen Methoden detektieren.

Schlüsselwörter: *Germanismen, digitalisierter Zeitungskorpus, automatische Detektierung*

1. INTRODUCTION
1. UVOD

The influence of foreign languages on Croatian territory and thus the influence of the German language in Croatia has been very strong during the past few centuries and it has even stretched to the level of the official language in the period of the establishment of the Military Frontier (1535-1871) that was a strategically important border area of the Habsburg Monarchy, and thereby a fertile ground for bilingualism and more intense use of the German language [1]. The imperial education reform (1764) has also helped in spreading of the German language in that period, as well as the introduction of the German language into the classes of public schools. Joseph II used a resolution from 1784 to unsuccessfully introduce the German language as the official language of the army, judiciary and in education for quite a short time (1786-1790) [2].

Consequently, such influence has left deep traces in both the spoken and written form of the Croatian language.
This paper investigates the traces of the German language in contemporary written Croatian language, analyzing daily newspapers that represent a mirror of the contemporary Croatian language. Researches of the German loanwords manly put an accent on the linguistic aspect of the research. Regarding German loanwords in newspaper corpora, paper from Krpan [3] presents loanwords taken from the German language analysed in specific volumes of the weekly newspaper "Narodni list", published in the period between 1955 and 2005, and adaptations of these loanwords on morphological and semantic levels.

The following paper describes the computational approach to detect German loanwords in newspaper corpora and to determine their frequency. Accent on the computational approach and not the linguistic analysis is different and new approach in detecting and researching German loanwords in Croatian.

The aim was to use the available language technologies to show the frequency and ways (context) in which German loanwords are used in the Croatian language on the basis of their appearance in regional daily newspapers.

2. THE FORMATION OF THE GERMAN LOANWORDS LIST
2. FORMIRANJE POPISA GERMANIZAMA

The first step to detect the German loanwords in Croatian regional daily newspapers using language technology was to form a list of German loanwords found in the relevant written sources. The basic material for the list of German loanwords was found in Croatian dictionaries, dictionaries of foreign words and regional dictionaries, as well as relevant scientific and professional papers that investigate German loanwords in the Croatian language.

The list of German loanwords that have been further researched using language technologies was formed from the following dictionaries and PhD thesis: *Rječnik hrvatskoga jezika* (edited by Jure Šonje, 2000), *Veliki rječnik hrvatskoga jezika* (Vladimir Anić, 2003), *Novi rječnik stranih riječi* (Bratoljub Klaić, 2012), *Rječnik stranih riječi*" (Šime Anić, Nikola Klaić and Želimir Domović, 2002), *Agramer. Rječnik njemačkih posuđenica u zagrebačkom govoru* (Zrinjka Glovacki-Bernardi, 2013), *Esekerski rječnik* (Velimir Petrović, 2008), PhD thesis by Ivo Medić *Kulturno-historijsko značenje i lingvistička analiza njemačkih pozajmljenica kod zagrebačkih obrtnika (obrada metala, drva i kože): prilog pitanju utjecaja njemačkog jezika na hrvatske specijalne jezike* (1962), PhD thesis by Teo Binder (1954) and PhD thesis by Velimir Piškorec *Germanizmi u podravskom dijalektu* (2001).

Eseker Dialect Dictionary by Velimir Petrović has been included in the list of dictionaries we have used to compile the list of German loanwords because we had expected that we would find recorded German loanwords that are usually used only in spoken Croatian language., even though we are talking about a dictionary of Osijek-German dialect. The dictionary has been published quite recently (2008) hence has become relevant research material.
These dictionaries and PhD thesis contain German loanwords that have become part of the Croatian standard language. The total number of distinct German loanwords collected from these dictionaries and PhD thesis was 17,988 types.
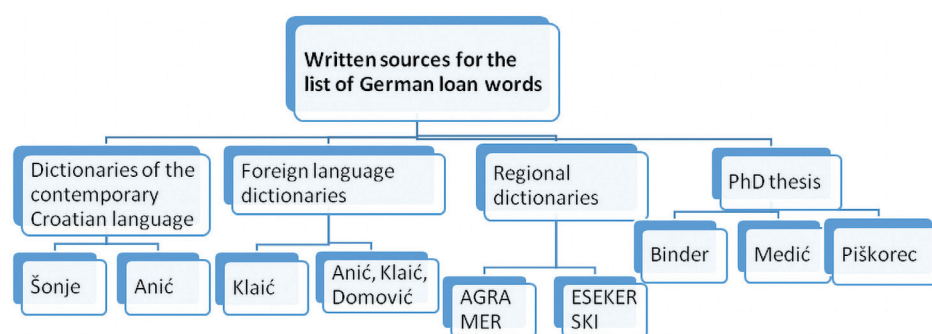
Figure 1 *Written sources for the formation of the list of German loanwords*

Slika 1 *Pisani izvori za formiranje popisa germanizama*

3. METHOD

3. METODA

Our first step in this research was to compile a list of German loanwords detected in the above mentioned sources. All variants in the writing of German loanwords in the analyzed sources were taken into account, as well as the classification of German loanwords proposed by different authors. These variants were confirmed in the Croatian Language Portal (http://hjp.znanje.hr/), which cannot be used as a source of German loanwords, since it does not allow a search by etymology, ie. it is limited to entry-based search.

Based on the list, a database was formed containing all German loanwords recorded in these sources. Apart from the loanwords, the database additionally contains the original German word that was borrowed by the Croatian language and its translation.

Furthermore, all unambiguous Croatian equivalents were added to each German loanword and, in case the Croatian equivalent did not exist, each German loanword was given a descriptive translation.

The database also contains data about the source of each German loanword and its part of speech, which proved to be important in detecting fraudulent German loanwords, ie. meanings in which the word is of Croatian and not German origin (for example: *gol* – Croatian adjective meaning 'naked' and *gol* – German loanword meaning 'touchdown', 'hit').

The collected list of German loanwords was automatically compared with a corpus of digitized daily newspapers *Večernji list*, *Novi list*, *Glas Slavonije* and *Slobodna Dalmacija*.

Previous studies of German loanwords were carried out using a literary-linguistic approach, where a computer was used only for data recording and editing. Our method automatically detects German loanwords in all their morphosyntactic forms, using existing dictionary databases of the Croatian standard language and automatically generates all possible morphosyntactic forms of each word, thus expanding the list of German loanwords collected from Croatian dictionaries and major works dealing with linguistic analysis of German loanwords.

Figure 2 *Newspaper article in Novi list with German loanword "birtija" in the headline. Source: http://www.novilist.hr/Vijesti/Rijeka/Scena/Posljednjih-100-Palach-je-i-pos-o-i-kuca-i-birtija (16.6.2015)*

Slika 2 *Novinski članak u Novom listu s germanizmom "birtija" u naslovu. Izvor: http://www.novilist.hr/Vijesti/Rijeka/Scena/Posljednjih-100-Palach-je-i-pos-o-i-kuca-i-birtija (16.6.2015.)*

We believe that the use of words borrowed from other languages, thereby also German loanwords, enrich the language and form an important element of the journalistic style, but also influence the creativity of journalists or newspaper articles. The information is conveyed by the articles in daily newspapers, but the authors also need to intrigue the reader [4]. Use of internationalisms, journalisms and loanwords are considered admissible characteristics of the style of jounalism.

The following table (Table1) shows data on the circulation of daily newspapers published as a report by the Croatian Chamber of Commerce, Industry and Technology (CCC) on May 25. 2015. The table reveals only part of the report, since the original report contains data on circulation of all 524 newspapers published in Croatia. The entire report can be found on CCC's website: *http://www.hgk.hr/javne-ovlasti-u-sektoru* (25.5.2015).

Table 1. *Information on the circulation of daily newspapers*
Tablica 1. *Informacije o izdanju dnevnih novina*

Town	*Name of the newspaper*	*The average sold (2014) p-sold; b-free*	*The average planned circulation (2015)*
Split	Slobodna Dalmacija	25.469 p. 248 b.	35.500
Rijeka	Novi list	18.515 p. 448 b.	24.794
Zagreb	Večernji list	49.500 p.	69.927
Osijek	Glas Slavonije	7.435 p.	12.000

As for daily newspapers, only a brief overview is given, related to their circulation in Croatia. The daily newspaper *Večernji list* is a political journal with evening news, that has been published in Zagreb since June 30, 1959.

Večernji list has ten editions in Croatia (Dalmatia, Istria-Primorje-Lika, Slavonia and Baranja, Slavonia, Podravina and Bilogora, Varaždin and Medjimurje, Zagorje, Sisak, Karlovac and Zagreb), one for Bosnia and Herzegovina and one international edition.

Daily newspaper *Novi list* is the oldest of the four in this research included Croatian newspaper. Launched in 1900 by Franjo Supilo, the first issue was called *Novi list*, but the name was later changed to *Riečki novi list*.
Riečki novi list that still exists today was first published in 1947 in Rijeka, and despite having its name changed several times it is the most important newspaper in that region.

The first issue of the newspaper *Glas Slavonije* was released in 1943 in the temporarily liberated territory of Western Slavonia. This newspaper is published in the city of Osijek and represents the newspaper with the largest circulation in Eastern Slavonia. Croatian daily newspaper *Slobodna Dalmacija* was founded in 1943, and represents the newspaper with the largest circulation in Dalmatia.

4. COMPUTER PROCESSING OF GERMAN LOANWORDS COLLECTED FROM DICTIONARY SOURCES
4. RAČUNALNA OBRADA GERMANIZAMA PRIKUPLJENIH IZ RJEČNIČKIH IZVORA

Croatian language belongs to inflectional languages and as such allows nouns, adjectives and verbs to have a number of different forms which indicate the case, number, tense, person and other grammatical categories. The words that belong to the open vocabulary of the Croatian language can appear in texts in different morphosyntactic forms. In order to conflate these forms to their canonical form (lemma), they are often normalized to a common root by stemming or to the basic form by the process called lemmatization.Stemming involves the removal of affixes, prefixes and suffixes from the word forms in order to obtain the root (stem) common to all forms.

The stem obtained by such a method does not necessarily correspond to the word root in linguistic terms. Simply put, the stem obtained by such a method is what is left after the prefix and suffix have been removed.

Morphological variants / word forms have different endings, but essentially describe the same word meaning. These different variants, therefore, can be conflated to a representative form called root.

The lemmatization process is similar to determining the stem of words. The difference is that lemmatization reduces a word down to its basic form, lemma, instead of the word stem.

Each German loan word can take a different form in the Croatian language, but its meaning does not change. For instance, words *bauštelom* (sg. instrumental case)*, bauštelama* (pl. instrumental case) and *baušteli* (sg. Dative case) all represent variants of the same lemma (*bauštela –* '*construction site*') and these words are called *types,* while each appearance of these words in the corpus is called *token.* Thus, a series of plural forms (nominative, genitive, dative, accusative, vocative, locative, instrumental) of the lemma *bauštela –* '*construction site*' consists of the following 7 types: *bauštele, bauštela, bauštelama, bauštele, bauštele, bauštelama i bauštelama.*

On the other hand, German loanwords *bauštela,* '*construction site*', and *bauštelac* '*construction worker*' are semantically different, because they belong to different lemmas.

Thus, lemma of the noun represents its nominative singular form, such as *bager* 'excavator', while the lemma of a verb represents its infinitive form, such as *fušati* '*bungle*'.

In some cases, as proved in our research, it is not possible to uniquely determine the lemma of a word. For example, the word *lista* belongs to two different lemmas: noun *list* 'leaf' and verb *listati* 'leaf /leaf through'.

The lemmatization process assigns the correct lemma to the word *lista*.

Morphological analyzers are tools that assign a word its lemma, enabling us to find a word in the language database or a lexicon / dictionary.

Morphological generators are tools that assign all appropriate morphosyntactic forms to a lemma (such as singular and plural forms for nouns), enabling us to find a word in a digital corpus. Consequently, morphological analyzers and generators represent the important components of a natural language processing system.

In determining the stem and suffix of German loanwords (ie. lemma and suffixes), we used the Croatian Morphological Lexicon (HML: http://hml.ffzg.hr/).

The Croatian Morphological Lexicon contains around 11.000 lemmas and more than 4.000.000 word forms. Lexicon is intended for professionals and systems for data search, text mining and various tasks of natural language processing for the Croatian language. HML is distributed under the CC-BY-NC-SA license.

The final lemmatization of all 17,988 German loanwords was performed using the reverse word dictionary for German loanwords, which sorts each entry according to its last letter and the subsequent letters up to the beginning of the word, enabling the addition of appropriate suffixes to each lemma (for example, the word *cement* in reverse dictionary has this form *tnemec*). Consequently, all the words that have the same suffix appear consecutively in such a dictionary.

The first step in the process of detection of German loanwords in the daily newspapers was to determine all possible forms resulting from declension (nouns and adjectives) and conjugation (verbs) of each lemma. After that, the set consisting of forms created by adding a suffix to each word stem was prepared.

This process, when done automatically, is usually error-prone, even for the original Croatian words.

The RegExp method (Regular Expression method) was used to search for German loanwords in newspaper articles, creating a regular expression for each lemma. The search strings were created in accordance with a model: the suffix (manually determined by existing grammar rules) was added to the manually determined word root (the shortest common string before phonetic or morphological alterations).

Example:

• stem & (suffix1 | suffix2 | suffixN ...)

• grup& (a | u | om ...)

The Regular Expression enabled the generation of grammatical rules using the RegExp matrix.

The root endings make up a large collection of fragments that connect the root of a word with its suffixes, such as, number, case, time or person. For example, the Croatian language has a different paradigm for female, male and neutral genders, as well as the animate and inanimate, which are marked by inflectional suffixes and different root endings in seven cases of singular and plural.

Example:

• lemma → grupa (eng. group)

The stem was determined as *grup*, and suffixes of the declination paradigm for feminine nouns were added to the stem (a | e | i| | o | u | om | ama). As a result, the RegExp method provided the following structure:

• \bgrup (a | e | i| o | u | om | ama) \b

This regular expression enabled us to find the following inflectional forms of lemma *grupa* in the newspaper text: *grupa*, *grupe, grupi, grupu, grupo, grupom, grupama*.

The Regular Expression method was firstly applied to a small sample set of 1360 lemmas, described in [5]. But, the automatic determination of the root has not yielded satisfactory results.

Therefore, the "rule of the last character belonging to lemma" was implemented to correct the errors of the Regular Expresion method.

Example:

• lemma → cement

The detected stem was *cement*. The suffixes of the declination paradigm for masculine nouns were added to the detected stem, respectively a | u | om| i | a | ima | e.

The RegExp method provided the following structure:

• \bcement (a | u |om | i | a | ima | e) \b

This regular expression enabled us to find the following inflectional forms of lemma *cement* in the newspaper text: *cementa, cementu, cemente, cementom, cementima* but it did not allow us to find the lemma *cement* itself.

Furthermore, in this experiment, the ending character of each lemma was singled out, lemmas were sorted in an ascending way based on that character and the character was then converted into one of the possible root suffixes.

This procedure proved to be effective for many other German loanwords, such as, *bankrot* 'bankruptcy'*, bizmut* 'bismuth'*, ballast* 'ballast'*, bruderšaft* 'brotherhood'*, recept* 'recipe', etc.

The resulting expression was:

• cement - t = cemen + (ta | tu |tom | ti | ta | tima | te)

Table 2. *German loanwords with suffix -ati and associated suffixes for this paradigm*

Tablica 2. *Germanizmi na -ati i pridruženi nastavci za paradigm*

German loanwords with suffix -ati and associated suffixes for this paradigm		
German loanword	**Type of verb**	**Paradigm**
abradirati	finite/infinite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu
abrihtati	finite/infinite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu
abriktati	finite/infinite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu
abšisati	finite/infinite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu
abšmalcati	finite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu
abšminkati	infinite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu
abštehati	infinite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu
abzecati	infinite	ti t m š 0 mo te ju o la lo li le j jmo jte jući vši h smo ste še hu

A label of the morphosyntactic category and a paradigm number was assigned to each German loan word in the reversed dictionary. These data allowed us to extract heuristic information about the equivalent stem classes.

Furthermore, all morphosyntactic forms (nouns and adjectives in all cases of singular and plural, verbs in the present tense, present perfect, past perfect and future perfect, as well as truncated infinitive, participle present participle and past participle) have been generated for all 17,988 lemmas of German loanwords collected from all available sources (dictionaries and PhD thesis).

The table below (Table 2), for example, shows a few verbs that are associated with suffixes of the verb paradigm -ati.

Computer analysis of the frequency of German loanwords in daily newspapers showed that most German loanwords appeared in *Novi list*, followed by *Slobodna Dalmacija* and *Večernji list*, with a relatively equal number of tokens. The lowest occurrence of German loanwords was detected in the daily newspaper *Glas Slavonije*.

In the cross section of all newspapers, we detected 191 distinct lemmas of German loanwords.

Table 3. *Number of tokens in computer-processed daily newspapers*

Tablica 3. *Broj pojavnica u računalno obrađenim dnevnim novinama*

Name	TOTAL
Glas Slavonije	3650
Novi list	5429
Slobodna Dalmacija	4946
Večernji list – Hrvatska	4259
TOTAL NUMBER OF GERMAN LOANWORDS	**18284**

As for parts of speech, out of 191 German loanwords, four words were adjectives, 28 were verbs, 15 were indeclinable, 101 words were masculine nouns, 42 were feminine nouns and only one was a neutral noun.

We can conclude that the most frequent German loanwords are masculine nouns, followed by feminine nouns, verbs and adjectives.

The situation regarding parts of speech of German loanwords in newspapers is very similar to the web corpus which was analyzed in another study, with the exception of adjectives that are slightly more frequent than verbs.

5. THE OCCURRENCE AND FREQUENCY OF GERMAN LOANWORDS AS PER THEIR LOCATION IN THE TEXT
5. *UČESTALOST POJAVLJIVANJA GERMANIZAMA OBZIROM NA MJESTO U TEKSTU*

Newspaper headlines are divided into three categories: nominative, informative and advertising.

Nominative headlines usually contain an assessment of the news. Therefore, such headlines mainly contain nouns and adjectives. If a verb appears in the nominative headline, it usually has a verbless form. In the smaller research conducted 2012 [5] we have found the following examples of nominative headlines in *Večernji list*:

> –*Krastavci: svježi na salatu, dinstan i uz meso* – 'Cucumbers: fresh and as a salad, stewed with meat' (3.6.2012)

> –*Do Dubrovnika i trajektom, i koridorom kroz Neum i mostom* – 'Reaching Dubrovnik by ferry, by a corridor through Neum and by bridge' (6.6.2012);

> –*Investicijski krah od 200 milijuna €, -* 'A 200 million € investment collapse', (15.6.2012).

The informative headlines convey the content and are dominated by different verb forms, exclamation and question marks and question words. We have found the following example of an informative headline in *Večernji list*:

> –*Hajduk izvisio: Križanac produljio sa Splitom pa sa Zužulom letio helikopterom* – 'Hajduk got scooped: Križanac extended contract with Split and then flew in a helicopter with Zužul' (1.6.2012)

Advertising headlines offer the content, and therefore consist of incentive and suggestive words and word forms such as imperatives, exclamation and question marks, hyphens. Following examples of advertising headlines in *Večernji list* [5]:

> – *PPS-Galeković s parketima osvaja Rusiju i Azerbajdžan* – 'PPS-Galeković conquers Russia and Azerbaijan with parquets' (18.6.2012);

> – *Pristupačna češka limuzina idealna za mlade obitelji* – 'Affordable Czech sedan ideal for young families', (26.6.2012);

> – *Slatko i hladno, savršen par* – 'Sweet and cold, a perfect couple', (30.6.2012)

The headlines often appear without a context and as general as possible, but then subheadings confirm the content. Headlines need to be provocative in order to intrigue the reader. Therefore, the authors often resort to strong, distinctive and persuasive words, circumlocutions, local dialectal words, proverbs and sayings, contrast and irony. The authors play with words to arouse the interest of their readers.

According Stolac and Vlastelić [6] commercials target the end user in order to convince them and even to persuade them. The use of German loanwords in newspaper texts could also be justified, according to the aforementioned context, by enticing the reader to read the text because they have been attracted by a German loanword that they have found familiar and that has caught their attention.

As for the frequency of the occurrence of German loanwords in newspapers, our research revealed that German loanwords usually occur in the content of the text, followed by titles, headlines and subtitles in the last place.

Table 4 shows the frequency of German loanwords given the location in the text (headline, title, subtitle and content) in all four daily newspapers.

German loanwords in the content of daily newspapers articles are, equally represented regardless of the region in Croatia. This led us to the conclusion that the writing style in the analyzed daily press is not necessarily affected by its region.

The largest number of German loanwords that were detected in daily newspapers, 67 words, originate from *Novi rječnik stranih riječi* (*New Dictionary of Foreign Words*) by Bratoljub Klaić, while 53 German loanwords originate from the regional dictionaries *Agramer. Rječnik njemačkih posuđenica u zagrebačkom govoru (Agramer. Dictionary of German loanwords in Zagreb speech)* by Glovacki-Bernardi and almost an equal number of German loanwords – 52, originate from Petrović's *Esekerski rječnik (Eseker dictionary)*.

Veliki rječnik hrvatskoga jezika, "Great Dictionary of the Croatian language" by Vladimir Anić. The source of 48 out of 191 detected German loanwords was *Veliki rječnik hrvatskoga jezika,* "Great Dictionary of the Croatian language" by *Vladimir Anić.* Furthermore, 36 German loanwords originate from the PhD thesis by Ivo Medić, while 33 words originate from the doctoral dissertation by Teo Binder.

Table 4. *Number of tokens in computer-processed daily newspapers*

Tablica 4. *Pojavnost germanizama s obzirom na mjesto u tekstu*

SOURCE	Glas Slavonije	Novi list	Slobodna Dalmacija	Večernji list	TOTAL
Headline	82	160	102	99	443
Title	116	139	111	158	524
Title2	70	71	55	105	301
Subtitle	111	152	93	156	512
Content	3271	4907	4585	3741	16504
TOTAL	3650	5429	4946	4259	18284

Also, 29 German loanwords originate from the *Rječnik stranih riječi,* "Dictionary of Foreign Words" by Anić, Klaić and Domović.

The lowest number of German loanwords originate from the *Rječnik hrvatskoga jezika*, "Dictionary of the Croatian language" edited by Šonje et al., only 28 words, and an equal number originate from the doctoral thesis of Velimir Piškorec.

6. CONCLUSION
6. ZAKLJUČAK

Computer analysis of the frequency of German loanwords in daily newspapers showed that most German loanwords appeared in *Novi list*, followed by *Slobodna Dalmacija* and *Večernji list,* with a relatively equal number of tokens. The lowest frequency of German loanword occurrence was detected in the daily newspaper *Glas Slavonije*. In the cross section of all newspapers, we detected 191 distinct lemmas of German loanwords.

As for parts of speech, 4 detected words were adjectives, 28 were verbs, 15 were indeclinable, 101 words were masculine nouns, 42 were feminine nouns and 1 was a neutral noun.

Therefore, the most frequent German loanwords are masculine nouns, followed by feminine nouns, verbs and adjectives. The situation in the newspapers regarding parts of speech of German loanwords is very similar to the web corpus which was analyzed in another experiment, with the exception of adjectives that are slightly more frequent than verbs.

As for the frequency of occurrence of German loanwords in newspapers, generally speaking, German loanwords most frequently occur in the content (body) of the text, followed by titles, headlines and subtitles in the last place.

We also discovered that German loanwords in the content of daily newspapers articles are, regardless of the Croatian region, equally represented. This led us to the conclusion that the writing style in the analyzed daily press is not necessarily region specific. "Regionally specific" refers to German loanwords whose appearance in text and speech are characteristic of a certain region for example dialects in Zagreb, Rijeka, etc.

What stands out is the large number of, so-called false German loanwords, such as *rat*, which are not found in the meaning of German loanword in the newspaper corpus (e.g. *rat* is found in the meaning of *war,* not in the meaning of *advice*), and such false German loanwords have a very high frequency of occurrence [4]. German loanwords that have been found in newspaper texts have been subsequently subjected to manual processing. A certain number of words that have been included in our list of German loanwords showed to also have their Croatian form and/or meaning (above in the text the example of the word 'rat') and in examples that we have found by computer search only the Croatian meaning of the word can be found. Therefore, we have named such German loanwords „false German loanwords" for the purpose of our study. A great number of such false German loanwords have been given by the word *grad* that in almost all the examples has the meaning 'city' and not the meaning of a German loanword 'grade'.

Language technologies have enabled us to determine how often German loanwords extracted from written sources appear in contemporary (written) texts - Croatian regional daily newspapers. We consider newspapers as contemporary corpora because they in a way mirror the everyday use of the Croatian language. We have also established a mechanism for the analysis of German loanwords in Croatian language that can be applied to other languages. We have explored the digitally prepared materials (newspapers – *Novi list, Slobodna Dalmacija, Glas Slavonije* and *Večernji list*) and revealed the frequency of occurrence of German loanwords in headlines, titles, subtitles, and the content of the daily newspapers.

The analysis has revealed that daily newspapers in a period of one year contained only 191 lemmas of German loanwords and 18,284 tokens. Thus, we can conclude that German loanwords in contemporary Croatian language appear systematically, but less than one would expect given the centuries of intensive influence of the German language in Croatia. Research has also shown that German loanwords listed in standard dictionaries of the Croatian language appear uniformly in the regional editions of daily newspapers, regardless of the region, which means that they are not necessarily region specific.

Such a statement has been proven by computational analysis of digitized daily press over a period of twelve months. Statistics have shown that German loanwords occur evenly in the newspapers from all four Croatian regions, i.e. there is no single region where German loanwords appear more frequently. As for the frequency of occurrence in newspapers, the overall conclusion is that German loanwords most frequently occur in the body of the text, followed by titles, headlines and, lastly, subtitles.

The digital processing of daily newspapers did not enable analysis of German loanwords by sections, or the determination of specific use of German loanwords in certain fields of human endeavor. An exception was, for instance, the word *box*, which belongs to the sports dictionary, but does not necessarily occur only in the sports section and can also be used as Anglicism (box = enclosed area) and not as a German loan word. In this study we revealed the frequency of occurrence of German loanwords in texts of contemporary Croatian language (daily newspapers).Although the total number of German loanwords which are part of the contemporary Croatian dictionaries and other consulted sources was 17.988, the number of distinctive lemmas found in newspaper texts was 191.

7. REFERENCES
7. REFERENCE

[1.] Žepić, S. (2002): Zur Geschichte der deutschen Sprache in Kroatien. Zagreber Germanistische Beiträge 11, University of Zagreb

[2.] Puškar, K. (2010): Tragom njemačko-mađarsko-hrvatskih jezičnih dodira: problematika jezika posrednika. Cris. Year. XII. Nr. 1/2010.P. 129-14. ISSN 1848-9621

[3.] Krpan, T. (2013): Germanizmi u Narodnom listu od 1955. do 2005. godine. Liburnia: međunarodni znanstveni časopis za kulturu, turizam i komuniciranje. Vol. 2 No. 1, 2013. ISSN 1848-5332

[4.] Tepeš Golubić, L. (2016): Germanizmi u digitalnim novinskim korpusima hrvatskoga jezika. Thesis. Zagreb.

[5.] Tepeš Golubić, L.; Mikelić Preradović, N.; Boras, D. (2013): Semi-automatic detection of germanisms in Croatian newspaper texts // Human Language Technologies as a Challenge for Computer Science and Linguistics / Vetulani, Zygmunt ;Uszkoreit, Hans (ur.). Poznan, Poland: Fundacja Uniwersytetu im. A. Mickiewicza. P. 173–177

[6.] Stolac, D.; Vlastelić, A. (2014): Jezik reklama. Hrvatska sveučilišna naklada – Filozofski fakultet Sveučilišta u Rijeci. Zagreb – Rijeka. ISSN 0586-0296

SOURCES: DICTIONARYS AND DISSERTATIONS
IZVORI: RJEČNICI I DISERTACIJE

Anić, V. (2003): Veliki rječnik hrvatskog jezika, Novi liber, Zagreb. ISBN 978-953-0-40040-5

Anić, Š.; Klaić N.; Domović Ž. (2002): Rječnik stranih riječi. Sani-Plus. Zagreb.

Binder, T. (1954): Die deutschen Lehnwörter in der kroatischen Essegger Mundart. Thesis. Vienna.

Glovacki-Bernardi, Z. (2013): Agramer. Rječnik posuđenica u zagrebačkom govoru. Novi liber. Zagreb. ISBN 978-953-325-035-9

Klaić, B. (2012): Novi rječnik stranih riječi. Školska knjiga. Zagreb. ISBN 978-953-0-40939-2

Medić, I. (1962): Kulturno-historijsko značenje i lingvistička analiza njemačkih pozajmljenica kod zagrebačkih obrtnika (obrada metala, drva i kože). Thesis. Zagreb.

Petrović, V. (2008): Esekerski rječnik. Filozofski fakultet. Zagreb. ISBN 978-953-175-283-1

Piškorec, V. (2001): Germanizmi u podravskomdijalektu. Disertacija. Filozofskifakultet Sveučilišta u Zagrebu.

Šonje, J. (2000): Rječnik hrvatskoga jezika. Leksikografski zavod Miroslav Krleža i Školska knjiga, Zagreb.

AUTHOR · *AUTOR*

Lidija Tepeš Golubić - biograpgy can be found in the Polytechnic & Design Vol. 5, No. 2, 2017.

Correspondence: ltepes2@tvz.hr