

Codification Within Reach: Three Clickable Layers of Information Surrounding the New Slovenian Normative Guide

Helena Dobrovoljc^{1,2}, Urška Vranjek Ošlak¹

¹ ZRC SAZU, Fran Ramovš Institute of the Slovenian Language, Novi trg 2, SI-1000 Ljubljana, Slovenia;

² University of Nova Gorica, School of Humanities, Vipavska 13, SI-5000 Nova Gorica, Slovenia

E-mail: helena.dobrovoljc@zrc-sazu.si, urska.vranjek@zrc-sazu.si

Abstract

This paper presents how language technology tools enable the integration of different types of normative data into a single language manual. The new Slovenian Normative Guide, the central normative manual consisting of normative rules and an orthographic dictionary, is based on language problems reported by language users. The normative guide consists of normative rules, and the orthographic dictionary supplements them with additional examples. The normative guide contains not only a systematic set of basic writing rules at the vowel-letter level (orthography or spelling), but also other consensual norms of the standard language. In order to effectively meet the needs of today's users of Slovenian, it was necessary to create a new concept for the orthographic dictionary so that it could effectively accompany the normative guide. In revising the normative rules, data collected on the Language Counselling Service platform were used. The normative guide is surrounded by three digitally interconnected layers of normative information; these three resources help the user navigate through the new normative view of the Slovenian language and provide arguments and explanations for the decisions made in the revision process.

Keywords: Slovenian; normative guide; orthographic dictionary; corpora research

1. Introduction

First we must point out some Slovenian peculiarities: Normative guides provide information about the acceptability of language elements for standard language use. In Slovenian, the standard language is an agreed supra-regional idiom that has been used in the written language since the middle of the 19th century. The so-called normative manuals (i.e. grammars, unabridged monolingual descriptive dictionaries of the standard Slovenian language and normative (orthographic) guides) are updated every few decades to harmonise the standard idiom with the natural language.

While grammars and descriptive dictionaries are universal concepts in linguistics, the term “normative guide”, which is the English equivalent of the Slovenian term “pravopis”, requires additional explanation. It is a manual consisting of normative rules

accompanied by an orthographic dictionary. A normative guide includes not only a systematic set of basic writing rules at the vowel-letter level (orthography or spelling), but also other consensual norms of the standard language that determine the use of lower and upper case letters, writing and syntactic use of names from other languages, writing together or apart, the status and use of loan words and proper names, punctuation, and the like (Dobrovoljc, 2016).

Until the publication of the central unabridged monolingual descriptive dictionary *Slovar slovenskega knjižnega jezika* (hereafter SSKJ), it was understandable that even simpler spelling dictionaries had to contain very concise semantic and stylistic information. After the publication of the SSKJ, the orthographic dictionary needed a new concept. Unfortunately, this did not happen with the publication of the normative guide *Slovenski pravopis 2001* (hereafter SP 2001). The codification of the Slovenian language went in two directions; discrepancies occurred not only between the normative manual as a whole and the descriptive dictionary, but also between the orthographic rules and the orthographic dictionary.

This paper presents the new *Slovenian Normative Guide* and its strategy for overcoming such inconsistencies with the help of a digital environment.

It was only after the publication of SP 2001 that it became clear that two partially overlapping dictionaries for the Slovenian language (the SSKJ and the orthographic dictionary) were not needed. A new concept for the orthographic dictionary had to be created so that it could effectively accompany the normative rules. The most typical language facts are listed in the general descriptive dictionary of the standard Slovenian language; however, the orthographic dictionary needs to include (1) the material expansion or enrichment of the normative rules (i.e. rules for the use of capital letters, borrowing, punctuation, writing together or apart) and (2) language elements that cause difficulties (i.e. atypical phrases that are difficult to use – problem-oriented approach). The starting point of this approach is the recognition that the way a word is written also depends on its meaning, which cannot be determined or represented in a dictionary without context (Moon, 2014).

2. The New Concept: Orthographic Codification of the Slovenian Language

The new approach to the elaboration of normative rules for the Slovenian language is problem-based; the problematic areas of language are identified with the help of an online language counselling service, which builds a database of Slovenian user language problems. The analysis of language use is carried out with the help of digital tools: text corpora and word sketches provide more advanced means of language processing, which allow the description of a standard language to be a more accurate representation of actual language use.

In designing the new normative guide, particular consideration will be given to the needs of language users and new linguistic facts. Organisationally, the lexical part will be produced simultaneously with the new semantic dictionary, so that its specialisation can be unambiguously normative (orthographic) and problem-oriented.

2.1 Three Layers of Information

Revision of normative rules with the help of data collected from the Language Counselling Service platform forms digitally interconnected layers of normative information.

The main source of linguistic dilemmas addressed in the new normative guide is the **Language Counselling Service** (*Jezikovna svetovalnica*) online platform, which is widely used by Slovenian language users to seek advice on linguistic choices and ambiguities; researchers use it to identify language description gaps (Dobrovoljc et al., 2020).

The platform (available at <https://svetovalnica.zrc-sazu.si>) has been active since 2012 and is based at the ZRC SAZU's Fran Ramovš Institute of the Slovenian Language. The different types of questions posted by the users of the language counselling site represent a rich and reliable source of difficulties with standard language usage which need to be taken into account in the process of revising the existing normative rules (Dobrovoljc et al., 2020).

User questions are answered by the staff of the Fran Ramovš Institute of the Slovenian Language, and each answer is approved by at least three members of the otherwise eight-member editorial board. Each answer is tagged with labels from three different levels of information:

- Language plane code (morphology, syntax, spelling, phonetics, etc.);
- Sub-area code (e.g. declension in morphology, verb-object agreement in syntax, etc.);
- Keyword (individual difficult cases and examples, e.g. COVID-19 related words).

The Language Counselling Service automatically creates a provisional online language guide with clickable codes and keywords. The platform is linked to other language resources available online (the *Fran* platform).

Language counselling points out gaps in language description and in language manuals; the creators of the new Slovenian Normative Guide (called *Pravopis 8.0*) use it as a source of language problems. *Pravopis 8.0* is an online normative manual and as such is part of the *Fran* Slovenian language portal (Fran Ramovš Institute for the Slovene Language ZRC SAZU, n.d.). The normative information surrounding it consists of three interconnected layers:

1. The **normative rules** are the theoretical part of the normative guide; they are available on the *Fran* platform under the name *Pravopis 8.0*. Each illustrative example in the rules is linked to the *ePravopis* orthographic dictionary.

2. The **Orthographic Dictionary** (*ePravopis*) is in its essence a normative dictionary; it is a growing dictionary the main purpose of which is to offer (additional) examples of the rules presented in the normative guide, a typical function of orthographic dictionaries (Verovnik, 2004). Dictionary entries contain information about spelling, pronunciation, text usage, morphological behaviour and word-formation possibilities of the included words.

3. Dictionary entries form problem-oriented groups linked to a publication called **Orthographic Categories** (*Pravopisne kategorije*), a collection of comments on how certain normative and orthographic difficulties are solved in the new orthographic dictionary, and a record of how the new orthographic dictionary differs from the current codification. Each dictionary entry is linked to its corresponding category, which contains a description of the linguistic problem and a list of all the entries included.

2.1.1 International Perspectives

As ZRC SAZU's Fran Ramovš Institute for the Slovenian Language is the only research institution dealing with the orthography of the Slovenian language, the new Normative Guide, together with the new Orthographic Dictionary and Orthographic Categories, is the central language resource of its kind for the Slovenian language. However, the idea of an online platform through which linguists can obtain data on language difficulties from a wide range of language users is not new. In 2011, a collection of language problems in the standard Slovenian language was formed, as part of the project “Sporazumevanje v slovenskem jeziku” (Communication in Slovene, available at <http://eng.slovenscina.eu>). In this project, existing online language counselling resources were used to create a manual of style. The project was based on best practices in European linguistic projects where language portals have been successfully designed using the public engagement method. There are several similar contemporary resources in other European countries, which are described below.

The “Grammis” portal of the Leibniz Institute for the German Language in Mannheim (available at <https://grammis.ids-mannheim.de>) was published in 1997. On this portal, users can find out about difficulties of the German language in the form of questions and answers as presented in the current grammar of the German language. Today, the portal has developed into a comprehensive hypermedia network information system, which is currently being expanded in terms of content and functionality.

The online Czech language handbook, created within the project “Internetová jazyková příručka” (Internet Language Reference Book, available at <https://prirucka.ujc.cas.cz>) by the Czech Language Institute of the Czech Academy of Sciences and the Faculty of Informatics of the Masaryk University (Pala & Šmerk, 2011), is currently being

expanded. The content of the reference book is based on the questions and problems posed to the linguists of the Czech language counselling centre “Jazyková poradna”. The Czech linguists still offer language advice over the phone.

A similar, partially interconnected system exists for the Croatian language (Vranjek Ošlak & Černivec, 2021). The Croatian language counselling service uses questions from speakers of Croatian as a framework for their printed language manuals. The answers are also published online (<https://jezicni-savjetnik.hr>).

The Estonian language advisory service is organised in a similar way. The Estonian “Keelenõuanded” (Language Council, available at <https://keeleabi.eki.ee>) of the Institute of Estonian Language answers language questions by phone, e-mail and mail. Estonian linguists answer questions about grammatical and orthographic difficulties in Estonian; however, they do not answer questions related to language policy or teaching. Their answers are published on a website. The same applies to the language advisory service of the Institute for the Languages of Finland (to be found at https://www.kotus.fi/en/services/telephone_counselling).

The Slovak language advisory service called “Jazyková poradňa JÚLŠ SAV” operates online; the platform is a joint project of the Sme.sk portal and the Ludovít Štúr Institute of Linguistics of the Slovak Academy of Sciences (available at <https://www.juls.savba.sk/poradna.html>). During the coronavirus pandemic, language advice played an important role in helping lay people, lecturers and teachers with their language problems.

The Language Counselling Service for the Slovene language, presented below, is at its core an advisory service; however, it is also a starting point for active public involvement in the process of updating (normative) language manuals. The main goal of the committee responsible for revising the normative rules and of the associated researchers is to produce up-to-date online manuals based on real language difficulties as pointed out by language users.

2.2 Language Counselling as a Source of Language Difficulties

In the following, we exemplify how the questions posed in the Language Counselling Service contribute to the enrichment of the new Slovenian Normative Guide, namely the normative rules. We selected two user questions that were the basis for the corresponding corpus research; they investigate the morphological behaviour of loan words:

- a) Do proper names ending in *-tz* show vowel alternation (*o* to *e*) in morphological inflexion and possessive adjective formation? Case No.1: *Fritz* – instrumental case: *Fritzem/Fritzom*, possessive adjective: *Fritzev/Fritzov*.
- b) What is the gender of loan words with atypical endings? Case No.2: *karitas* – feminine or masculine.

2.2.1 Case No.1: Morphological and word-formational variability of forms

A morphosyntactically tagged text corpus allows us to extract data on morphological duplicates. For example, if we want to know where Slovenian language users are hesitant about the syntactic use of the German name *Fritz*, the corpus provides a list of duplicate forms (Figure 1) from which we could deduce: (a) that duplicate forms *Fritzom/Fritzem* in the instrumental case are frequent in the corpus and b) that users often question the vowel alternation in the possessive adjective ending in *-ov/-ev* (*Fritzov/Fritzev*).

Fritzeva	Fritzev	8
Fritzeve	Fritzev	7
Fritzevo	Fritzev	6
Fritzovo	Fritzov	6
Fritzova	Fritzov	5
Fritzove	fritzov	5

Figure 1: A list of duplicate forms ending in *-ov/-ev* (*Gigafida 2.0* corpus)

The norm thus established shows that vowel alternation in Slovenian depends not only on pronunciation but also on the notation. If the combination of letters ⟨t⟩ and ⟨z⟩ is understood as a digraph and pronounced as [c], this triggers the realisation of vowel alternation *o* to *e* (*Fritzev*). However, if the combination is perceived only as a sequence of letters ⟨t⟩ and ⟨z⟩ and not as a digraph, the vowel alternation does not occur and the endings in written language follow the notation (*Fritzov*).

In drafting a normative rule to represent these findings in the normative guide, it is necessary to systematically examine all possible cases that could be included in the normative guide and orthographic dictionary. For this purpose, a glossary is created (Figure 2).

A review of the corpus material shows that vowel alternation *o* > *e* is merely a possibility and that such a phenomenon is systematically observed in all names ending in the final spoken [c] sound when it is written with other letters or letter combinations, e.g. Hungarian names ending in *-cz* (*Göncz*), German names ending in *-z* (*Leibniz*), or when the stem of a proper name contains the spoken [c] sound, e.g. Italian names ending in *-zza* or *-zzo* (*Tomizza*, *Campazzo*).¹

¹ This particular rule's status is currently at the proposal level and is not yet published in the new normative guide (*Pravopis 8.0*).

Osnovne oblike	Vse oblike	Osnovne oblike	Vse oblike
Osnovna oblika	Število pojavitev	Osnovna oblika	Število pojavitev
Moritz	5.025	Waltz	446
Fritz	4.227	Lubitz	434
Metz	3.052	Chemnitz	430
Auschwitz	3.020	HTZ	427
JBTZ	902	Kronplatz	365
Lutz	867	Fitz	346
Hertz	845	Getz	329
Kravitz	825	Platz	329
Blitz	784	Horowitz	316
Schultz	757	Brammertz	315
Ritz	743	Schmitz	290
Wolfowitz	721	Mateschitz	281
deutz	681	Haaretz	262
Schwartz	664	Zeit	260
Deutz	630	Berlitz	259
TZ	594	Quartz	250
Katz	485	Kaspitz	248
Stiglitz	476	Biarritz	211
Mutz	465	Leibovitz	207
Spitz	450	Strutz	207

Figure 2: Case No.1 glossary example: similar borrowed names ending in *-tz* (*Gigafida 2.0* corpus)

The normative rule must therefore include general instructions for changing the pronunciation and ending attribution in cases where a borrowed name is pronounced with the final sound [c].

The normative rules thus prepared are deductively coherent with the orthographic dictionary (*ePravopis*) on the *Fran* language portal. The realisation of this ending attribution phenomenon is expected without exceptions for all names in the dictionary referring to this rule, following the presented example – *Fritz* (Figure 3). Moreover, the background of the decisions made in the process of normative rule formation is explained in the corresponding orthographic category.

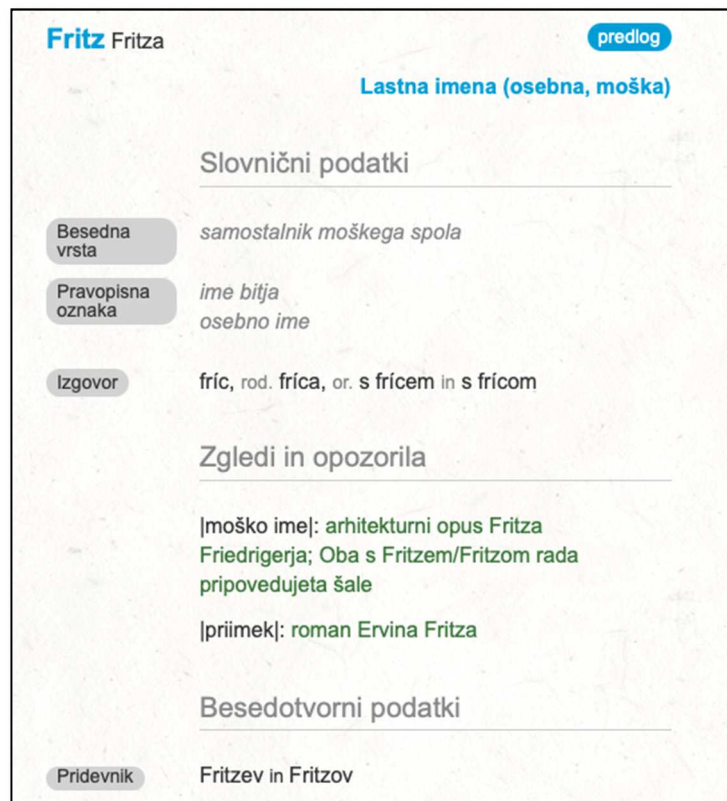


Figure 3. Orthographic dictionary entry: ending attribution phenomenon in the case of *Fritz* (see bottom line: possessive adjectives *Fritzev* and *Fritzov*)

2.2.2 Case No.2: Morphological variability of syntactic categories

Although corpus queries are in principle frequency-oriented on sets of the most common lexemes, and researchers are primarily concerned with establishing the lowest boundary of relevant hits (Holz, 2003), deviations are also important for finding out what is difficult for users of a particular language. These difficulties are not only defined by the right-or-wrong dichotomy; we are increasingly aware of standard linguistic diversity and thus language choice (Larsen-Freeman, 2000). Borrowing in Slovenian happens morphologically; we borrow both proper nouns and common nouns by adapting their grammatical categories in terms of ending to the existing system. However, often words from related languages (e.g. the Serbian and Croatian names *Užice* and *Brela*) or classical languages (*karitas* ‘charity’) indicate different possibilities, as the user experience with these nouns may be different.

The Language Counselling Service received this user question: “I have a question regarding the name of the organisation *Slovenska karitas* or in short *Karitas*. According to their website and the language use, the gender of this word is predominately feminine. The normative guide, however, determines it is masculine. That is the first problem. The second problem is capitalisation: *Škofijska karitas Koper* or *Škofijska Karitas Koper* or *škofijska Karitas Koper*. What is the correct capitalisation of the name of this organisation?”

In the case of the word *karitas*, standard language manuals recommend keeping the gender of the word as it is in the donor language (partly because of the important role of the connoisseurs of classical languages, who generally keep the gender of words: SSKJ classifies this noun as feminine and indeclinable). Lay users, however, follow the system of the Slovenian language and decline this noun like its parallel e.g. *ananas* ('pineapple', masculine and declinable). The current normative guide *Slovenski pravopis 2001* characterises both the common noun *karitas* and the proper name *Karitas* as masculine. Despite the relevant normative rule, the use of the common noun *karitas* indicates an increase in the frequency of the feminine gender (Figure 4), which is due to the related proper name *Slovenska karitas* (as the preceding adjective suggests, the noun *karitas* is feminine and indeclinable).

e je država odločila, da bo zaradi nepričakovanih naravnih nesreč **Karitasu** in Rdečemu križu podarila milijon evrov. Takšna gesta :
 zbirajo in delijo ne le po občinah, temveč tudi po Rdečem križu in **Karitasu**. Tako nihče natančno ne ve, kdaj in koliko je dobil priza
 evrov, ilirskobistriška Karitas je do zdaj za nesrečo prek škofijske **Karitas** zbrala 65.000 evrov. Vsota te, doslej znane denarne pon
 00 evrov, v kratkem pa se bo povečala še za delež, ki ga občinska **Karitas** pričakuje od slovenske. Center za socialno delo bo sklac
 lačila ter delijo hrano in mleko (občasno ju prejemajo od Škofijske **Karitas** Ljubljana ter iz intervencijskih zalog Evropske unije). Ra:
 rostovoljke, ki tako pomagajo ljudem v stiski. Še posebej v tednu **Karitas** želijo poudariti, da pomen gesla Ali me imaš rad? ni le v t
 obili ob tečajih prve pomoči. Ob večjih problemih pa sodelujemo s **Karitasom** in se dogovorimo, katere obveznosti bo prevzel kdo,
 ni je prejelo pomoč okoli 11.000 ljudi, nadaljnjih 16.000 pa je prek **Karitasa** prejelo pomoč Evropske unije. Posameznikom ali družin

Figure 4: Declination variability: *karitas* (*Gigafida 2.0*)

In normative rule-making, then, the material dictates the rule formulation, which must express the following linguistic fact (Figure 5): In short, most borrowed names in Slovenian are masculine, and in rare exceptions feminine. The noun *karitas* (originally feminine) is masculine or feminine in Slovenian.

Občna poimenovanja oblikoslovno prevzemamo glede na izrazno podobno oziroma glede na končaj. Največ prevzetih poimenovanj je moškega spola, le izjemoma so ženskega, npr. franc. **la chanson**, ž. sp. – slov. **šanson**, m. sp.; gr. **karitas**, ž. sp. – slov. **karitas**, m. ali ž. sp. Pogosto so moškega spola tudi samostalniki z neznačilnimi moškimi končnicami **-e (finale)**, **-o (avto)**, **-u (guru)**. Besede na **-a se uvrščajo v ženski spol (npr. gorgonzola, činčila)** ali v oba spola – moškega ali ženskega (**ara, panda**).

Figure 5: The new normative guide (*Pravopis 8.0*): loan words rule formulation

In Figure 5, the examples in blue contain links to dictionary entries (Figure 6) included in the orthographic dictionary (*ePravopis*). In this way, the interconnectedness of the new Normative Guide (normative rules) and the associated orthographic dictionary is ensured.

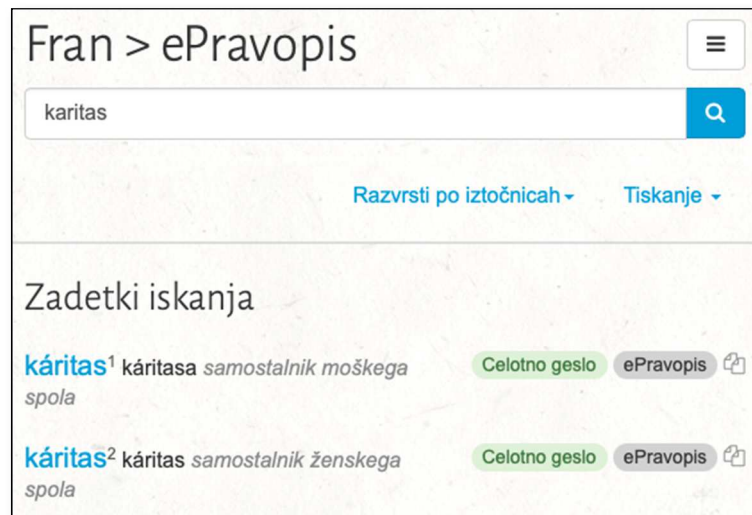


Figure 6: The new orthographic dictionary (*ePravopis*): *karitas* as headword (masculine and feminine homonyms)

3. Interconnectedness: How it Works

In the following chapter, we show how the above-mentioned three-layered interconnectedness works in the case of Slovenian temporal names, namely names of days, months, seasons, historical events and points in time such as holiday names.

3.1 Temporal Names and Capitalisation

In Slovenian, temporal names are not considered proper noun categories, as for example in English, where month names (*July, May*), names of days (*Monday, Sunday*), names of historical events (*the French Revolution*) and of points in time (*All Saints' Day*) are considered as proper names and written with a capital initial letter (Langendonck, 2007). In Slovenian, all temporal names are written without a capital initial letter regardless of syntactic position, which can be typically proper (*meseč maj* 'the month named May') or typically common-noun (*vsi trije božiči so minili mirno* 'all three Christmases passed peacefully' in the sense of 'all three Christmas holidays': Christmas Eve, New Year's Eve and the evening before Epiphany). The current fashion of writing holiday names without capital initial letters does not follow their syntactic role, but derives from a traditional agreement. In the first half of the 20th century, capitalisation was often rejected because it was characteristic of Germanic languages. Due to the nation's Austro-Hungarian past, Slovenian language speakers were constantly in a position of bilingualism, as German was the dominant language (Štih et al., 2008). Linguists therefore rejected capitalisation as something foreign and not in accordance with the history and structure of the Slovenian language (Dobrovoljc, 2004).

Certain groups in the Slovenian language community occasionally petition to change this normative rule, hoping to achieve its alternation into writing holidays with a capital initial letter. The reasons for these episodic tendencies are (1) emotional (capital initial

letters are associated with respect, especially regarding religious holidays); (2) influenced by foreign practices entering the Slovenian language, through e.g. greeting cards; or (3) are the result of certain beliefs that the lowercase initial letter reflects 45 years of enforced reduction of religious practices.

The committee responsible for revising the normative rules (*Pravopisna komisija pri SAZU in ZRC SAZU*) considered the current social situation and the requirements of various social groups when revising the chapter on upper and lower case letters in accordance with the methodology outlined above.

<i>vesel božič</i>	161
<i>Vesel božič</i>	88
<i>vesel Božič</i>	59
<i>Vesel Božič</i>	43
<i>VESEL BOŽIČ</i>	14
<i>veseli božič</i>	2
<i>veseli Božič</i>	1
<i>veselemu Božiču</i>	1
<i>veselejši božič</i>	1
<i>veselega božiča</i>	1
<i>veselega Božiča</i>	1
<i>vesele božič</i>	1
<i>vesele Božič</i>	1
<i>vesel BOŽIČ</i>	1
<i>Veselega božiča</i>	1
<i>Vesel BOŽIČ</i>	1

Table 1: Corpus query: *vesel božič* in the *slWac* corpus

In both versions of the central reference corpus for the Slovenian language *Gigafida*, proofread texts predominate, making it impossible for the corpus to reflect intuitive writing practices. In previous research (Dobrovoljc & Vranjek Ošlak, 2018), it was argued that contemporary linguistic research must also be conducted on non-standard language corpora (e.g. *Janes* or *slWac*), as they often yield different results.² The study of linguistic material was therefore also focused on various other corpora (Table 1).³ Since in Slovenian the word *božič* is homonymous with the surname *Božič*, which is relatively common, we looked for a characteristic greeting, namely *vesel božič* ‘Merry Christmas’.

² One of the reviews pointed out that “users in some user-generated contents write nonstandardly on purpose or by decision. This influences the genre itself.” This is true, of course, but it does not preclude the possibility of using non-standard language corpora as a means of comparison and for testing indicators of language change in particular. The predominant source of research is still reference and representative language corpora such as *Gigafida*.

³ All corpora used are available at: <https://www.clarin.si/noske/>.

A comparison of the corpus queries in Table 2 shows that the use of capital letters has not increased over the years (comparing the *Gigafida* 1.1 and 2.0 corpora), but capitalisation is much more noticeable in the *slWac* corpus (Slovenian websites) and in the *Janes* corpus (blog texts, online chat rooms, tweets, etc.); another obvious feature is also non-standard notations (e.g. colloquial words written according to their pronunciation). We also noticed that language users avoid having to choose upper or lower case in certain cases, namely by using only capital letters.

	Gigafida 1.1	Janes	slWac	Gigafida 2.0
<i>vesel božič</i>	71.28%	63.33%	67.64%	83.40%
<i>vesel Božič</i>	18.44%	39.20%	28.12%	13.82%
<i>vesel BOŽIČ</i>	10.17%	6.32%	4.24%	2.81%
errors		1.49%		

Table 2: Corpus query: *vesel božič* in four different corpora

In order to place the material research described above in a linguistic and social context, we carried out two further analyses. We were interested in the relationship between single-word and multi-word holiday names, so we also examined corpora material with regard to the use of the holiday name *velika noč* 'Easter' (Table 3).

	Gigafida 1.1	Janes	slWac	Gigafida 2.0
<i>velika noč</i>	54.22%	40.39%	47.69%	63.86%
<i>Velika noč</i>	38.84%	50.00%	47.31%	34.87%
<i>Velika Noč</i>	6.42%	5.87%	4.04%	1.14%
<i>VELIKA NOČ</i>	0.42%	1.78%	0.96%	0.13%

Table 3: Corpus query: *velika noč* in four different corpora

	Gigafida 1.1	Janes	slWac	Gigafida 2.0
<i>dan državnosti</i>	91.52%	67.13%	74.05%	91.52%
<i>Dan državnosti</i>	7.15%	29.98%	22.94%	7.15%
<i>dan Državnosti</i>	0.07%	0.44%	0.64%	0.07%
<i>Dan Državnosti</i>	0.02%	0.61%	0.20%	0.02%
<i>DAN DRŽAVNOSTI</i>	1.24%	1.75%	2.17%	1.24%
errors		0.09%		

Table 4: Corpus query: *dan državnosti* in four different corpora

Based on the wording in the normative rules, we checked whether the increase in the use of the capital initial was related to religious content; therefore, we conducted a

parallel study of the holiday name *dan državnosti* 'Statehood Day' (Table 4).⁴

Comparison of the results shows that capitalisation is significantly more common for religious holiday names than for others, but the predominant notation manner is still lowercase, which has been preferred for over a hundred years. The influence of multi-word holiday names is negligible.

The formulation of the normative rule is thus threefold (Figure 7). (1) Holiday names, regardless of the type of holiday (religious, national, European, etc.), continue to be written with a lowercase initial letter, as normative tradition and also prevailing usage dictate – {123}. (2) A separate admonition refers to holiday names containing proper nouns; they are capitalised, e.g. *dan Zemlje* 'Earth Day', *dan svetega Patrika* 'St. Patrick's Day' – {124}. (3) The so-called *stylistic instruction* (marked with a pencil symbol) introduces the possibility of writing holiday names in private correspondence with a capital initial, as a sign of respect, especially for religious holidays.

POIMENOVANJA PRAZNIKOV, POSEBNIH DATUMOV, DNEVOV IN MESECEV

{123} Z malo začetnico pišemo časovne opredelitve, tj. poimenovanja

- a) dnevov in mesecev: **ponedeljek, sobota; avgust, mali traven, rožnik;**
- b) državnih in verskih praznikov: **dan državnosti, dan samostojnosti in enotnosti, praznik dela; božič, pepelnica, velika noč; hanuka, ramadan/ramazan;**
- c) posebnih datumov: **ново leto, silvestrovo, martinovo;**
- č) mednarodnih, priložnostnih ali spominskih dnevov, mesecev, tednov in let: **dan spomina na žrtve holokavsta, mednarodni dan maternega jezika, svetovni dan čebel; teden razoroževanja; mednarodni mesec boja proti raku dojk; leto mladih, mednarodno leto zdravja rastlin.**

ODDAJTE KOMENTAR

{124} Kadar je v teh poimenovanjih uporabljeno lastno ime, ga pišemo z veliko začetnico: **dan Rudolfa Maistra, dan vrnitve Primorske k matični domovini, dan Zemlje.** Enako velja za pridevnike, izpeljane iz lastnega imena: **Prešernov dan; Marijino vnebovzetje (nasproti veliki šmaren ali velika maša ali velika gospojnica); Martinova sobota, Silvestrov večer (nasproti silvestrovo) in Vidov dan.**

O izlastnoimenskih pridevnikih s priponskimi obrazili *-ov/-ev* ali *-in* v stalnih besednih zvezah gl. poglavje »Pridevniki iz lastnih imen« (III. Velika in mala začetnica).


 V zasebnem pisanju se namesto male začetnice pogosto uporablja velika začetnica, še zlasti v voščilih ob verskih praznikih.

Figure 7: The new normative guide (*Pravopis 8.0*): holiday names rule formulation

Since the normative rules are illustratively and factually limited to only the most typical examples, which are of course linked to dictionary entries in the orthographic dictionary, an additional explanatory section called Orthographic Categories (*Pravopisne kategorije ePravopisa*) was conceived, which is also available online. This

⁴ Capital initial letter occurrences include those at the beginning of sentences.

section explains for each normative rule whether and how it has been changed; it also describes the possible language difficulties. For holiday names, Orthographic Categories lists all holiday names included in the orthographic dictionary, focusing not only on the initial case, but also on (a) newer holiday names in Slovenian that have not yet been included in a dictionary (e.g. *ašura* 'Ashura', *noruz* 'Novruz'), and on (b) synonymous holiday names (*judovska velika noč* 'the Jewish Easter' and *pasha* 'Passover'). Holiday identifiers were supplemented with dates where possible, otherwise with other relevant information. A link directs the language user to the relevant article of the old and new normative rules.

4. Conclusion

The new Slovenian Normative Guide draws from the knowledge of the generational, cognitive and educational diversity of language users. The creators of the new approach tried to write **the same linguistic information in different language codes and connect them**. In creating the normative manuals, an interconnected online system was designed, combining a language counselling platform, normative rules, an orthographic dictionary, and a description of normative solutions. All these levels are clickable and interlinked.

Through language-related questions asked on the Language Counselling Service platform, linguists encounter language difficulties. Through corpus material research, the working group learns about the problem and finds similar use cases. It articulates these in normative rules that concisely inform about the problem and its exceptions, and provide hints that are often stylistic in nature. All illustrative examples in the rules are included in the dictionary and linked to the corresponding dictionary entries. Normative solutions are presented in Orthographic Categories; from there, the user is referred back to the normative guide, completing the circle of interrelated linguistic information.

The same linguistic information is represented in different ways, and these different representations are interconnected so that the user can choose the most appropriate one. The choice depends on the user's prior knowledge of the language and practical experience with language use. The interconnectedness of the normative guide, orthographic dictionary, Language Counselling Service and Orthographic Categories is made possible by the digitally designed databases in the background. Language manuals designed in this way can reach a larger number of language users and be more user-friendly by making the normative information more accessible.

5. Acknowledgements

This contribution was prepared within the framework of the P6-0038 programme financed by the Slovenian Research Agency (ARRS).

6. References

- Dobrovoljc, H. (2004). *Pravopisje na Slovenskem*. Ljubljana: Založba ZRC, ZRC SAZU.
- Dobrovoljc, H. (2016). Povezljivost pravopisnih pravil in slovarja: sanje pravopiscev 20. stoletja. In Erjavec, T. & Fišer, D. (eds.) *Proceedings of the conference on Language Technologies & Digital Humanities*. Ljubljana, Slovenia. Available at: <http://www.sdjt.si/wp/wp-content/uploads/2016/09/JTDH-2016DobrovoljcPovezljivost-pravopisnih-pravil-in-slovarja.pdf>.
- Dobrovoljc, H., et al. (2020). *Kje pa vas jezik žuli?*. Ljubljana: Založba ZRC, ZRC SAZU.
- Dobrovoljc, H., Vranjek Ošlak, U. (2018). Zakaj ne z eno poizvedbo hkrati po različnih korpusih? (Troje korpusnih preverb pod primerjalnim drobnogledom). In Erjavec, T. & Fišer, D. (eds.) *Proceedings of the conference on Language Technologies & Digital Humanities*. Ljubljana, Slovenia. Available at: http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018_Dobrovoljc-H_Zakaj-ne-z-eno-poizvedbo-hkrati-po-razlicnih-korpusih.pdf.
- Fran*. Accessed at: <https://fran.si/>. (20 January 2021)
- Grammis*. Accessed at: <https://grammis.ids-mannheim.de>. (8 June 2021)
- Holz, N. (2003). Besedilni korpus Nova beseda in geslovník za Slovar novejšega besedja. *Jezikoslovni zapiski* 9, pp. 89–94.
- Institute for the Languages of Finland*. Accessed at: https://www.kotus.fi/en/services/telephone_counselling. (8 June 2021)
- Internetová jazyková příručka*. Accessed at: <https://prirucka.ujc.cas.cz>. (8 June 2021)
- Jazyková poradňa JÚLŠ SAV*. Accessed at: <https://www.juls.savba.sk/poradna.html>. (8 June 2021)
- Jezični savjetnik*. Accessed at: <https://jezicni-savjetnik.hr>. (8 June 2021)
- Jezikovna svetovalnica*. Accessed at: <https://svetovalnica.zrc-sazu.si/>. (20 January 2021)
- Keelenõuanded*. Accessed at: <https://keeleabi.eki.ee>. (8 June 2021)
- Langendonck, W. (2007). *Theory and Typology of Proper Names*. De Gruyter Mouton.
- Larsen-Freeman, D. (2009). *The Grammar of Choice*. Teaching and testing grammar. The handbook of language teaching. Blackwell.
- Moon, R. (2014). Meanings, Ideologies, and Learners' Dictionaries. In Abel, A., Vettori, C. & Ralli, N. (eds.) *Proceedings at the XVI EURALEX International Congress: The User in Focus. Part 3*. Bolzano. Available at: http://euralex2014.eurac.edu/en/callforpapers/Documents/EURALEX_Part_3.pdf.
- NoSketchEngine*. Clarin. Accessed at: <https://www.clarin.si/noske/>. (2 April 2021)
- Pala, K. & Šmerk, P. (2011). Internetová jazyková příručka. *Zpravodaj ÚVT MU*, XXI(3), pp. 14–17. Available at: <http://webserver.ics.muni.cz/bulletin/articles/666.html>.
- Pravopis 8.0*. Accessed at: <https://www.fran.si/pravopis8>. (20 January 2021)
- Pravopisne kategorije*. Accessed at: <https://fran.si/spt-kategorije>. (20 January 2021)

- Sporazumevanje v slovenskem jeziku*. Accessed at: <http://eng.slovenscina.eu>. (8 June 2021)
- Štih, P., et al. (2008). *Slovenska zgodovina: družba – politika – kultura*. Ljubljana: Inštitut za novejšo zgodovino. Available at: <http://www.sistory.si/publikacije/pdf/zgodovina/Slovenska-zgodovina-SLO.pdf>.
- Toporišič, J., et al. (eds.) (1990). *Slovenski pravopis 1 – Pravila*. Ljubljana: SAZU, Državna založba Slovenije.
- Toporišič, J., et al. (eds.) (2001). *Slovenski pravopis*. Ljubljana: SAZU, ZRC SAZU, Založba ZRC. Available at: <https://fran.si/134/slovenski-pravopis>.
- Verovnik, T. (2004). Norma knjižne slovenščine med kodifikacijo in jezikovno rabo v obdobju 1950–2001. *Družboslovne razprave* 20, pp. 241–258.
- Vranjek Ošlak, U. & Černivec, M. Slovenski in hrvaški pravopis – podobnosti in razhajanja na prehodu v digitalno dobo. In Nikolovski, G. (ed.) *Izzivi slavistike v 21. stoletju: 4. mednarodna znanstvena konferenca Slavistični znanstveni premisleki, zbornik povzetkov*. Maribor: Univerza v Mariboru, Univerzitetna založba. Available at: <https://press.um.si/index.php/ump/catalog/view/562/713/1411-1>.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

