

Fear and Loathing on Twitter: Attitudes towards Language

Damjan Popič,* Darja Fišer*†

*Faculty of Arts, University of Ljubljana
Aškerčeva 2, SI-1000 Ljubljana, Slovenia

†"Jožef Stefan" Institute

Jamova cesta 39, SI-1000 Ljubljana, Slovenia

E-mail: damjan.popic@ff.uni-lj.si, darja.fiser@ff.uni-lj.si

Abstract

The paper deals with the sociolinguistic concept of prestige imbued in the notion of standard language, and the social status connected to the inherent language skill (or lack thereof). To this end, we analyse Slovenian tweets pertaining to language use and the (in-)correctness of other users' use of language, propose a typology, especially in cases where language use is used as an argument against someone's qualifications or beliefs.

Keywords: orthography, (linguistic) prestige, computer-mediated communication

1. Introduction

The paper deals with a corpus-enhanced sociolinguistic analysis of attitudes towards language of Slovenian users on Twitter. In the Slovenian culture, the ability to produce linguistically correct texts has long been a test of a person's sophistication, education, and social class. This stance was further entrenched by the purist efforts, perpetuated by the country's constant struggle against imperialist forces, and the notion persists to this day.¹

For this reason, we analyse the attitudes of Slovenian Twitter users towards language, and the types of discourse that language-related debates feature in. By using the keywords *comma*, *orthography*, *grammar*, *Slovene*, and *language* we isolate tweets in which users comment on language from the corpus JANES-Tweet v1.0 (Fišer et al. 2017). We analyse these examples from a sociolinguistic standpoint of prestige. This kind of attitude towards language is very common in linguistically conservative communities, and is often accompanied by a shared belief that the attitude towards language is a moral issue. This is visible in language use as (supposed) improper use of language is often used as an *ad hominem* argument aiming to suppress the relevance of that person's beliefs (on any matter).

Furthermore, recent developments have shed a new light on the attitudes towards language proficiency on social media all over the world, especially with Donald Trump. His tweets have been subject to much scrutiny, linguistic and otherwise.² This shows that the language on Twitter, although supposedly informal in nature, is very much considered to be public.

¹ This is perhaps best described by the latest development in which the Slovenian Parliament voted on amending the War Grave Act (with 54 ayes and 6 nays). Following the parliamentary conclusion, a (superfluous) comma will be removed from the first stanza of Oton Župančič's poem that will be inscribed on the memorial of national reconciliation.

² See a comprehensive R-based statistical analysis of his tweets in relation to the device used for posting (Robinson 2016; <http://varianceexplained.org/r/trump-tweets/>).

2. Background

The Slovenian normative language tradition has been built around the German syntactic system and is highly prescriptive in nature. This means that the syntax in Slovene is highly structured, and this combination of a highly complex syntax and strict language rules imposed by the normative language guide (i.e. *pravopis*, from Ger. *Rechtsschreibung*) makes it notoriously difficult to write "correctly." For instance, the dreaded comma is the bane of students' lives, with many recent studies showing that Slovenes in general are underperforming severely when it comes to commas (Popič et al. 2016).

3. Analysis

In this section, we present the results of the analysis of tweets containing at least one of the selected keywords: *comma*, *orthography*, *grammar*, *Slovene*, and *language*. The analysis involved a manual content analysis of the entire extracted sub-corpus and drafting a typology that allows us to cover the attitudes towards language imbued in the extracted tweets, for all the keywords. As the following tables indicate, the attitudes differ in number and tenor, however, scathing and disparaging tweets are present in each of the categories. Table 1 gives the frequencies of the selected keywords in the corpus.

Keyword	Frequency
<i>vejica</i> (the comma)	1,978
<i>slovnica</i> (grammar)	930
<i>pravopis</i> (orthography)	432
<i>jezik</i> (language)	11,202
<i>slovenščina</i> (Slovene)	4,607
Total	19,149

Table 1: Keyword frequencies.

As Table 1 demonstrates, by far the most frequent of the five keywords is *language*, with *Slovene* being second. However, the expression *language* is relatively general and also used in other senses (as in 'tongue' in the physical sense, in phrasemes, etc.). Considering the much narrower meaning of *vejica* (it almost exclusively

pertains to the comma as a means of punctuation), it is evident that the comma plays an immensely significant role with 1978 occurrences. For this reason, we devote the most attention to this keyword and deal with the rest of the keywords in pairs, focusing above all on the differences in the attitudes between the keywords.

3.1 *Vejica* ‘the comma’

The comma holds the most prestigious place in the Slovenian linguistic tradition, and performs several functions in discourse. Considering the examples extracted from our corpus, we can categorize the extracted examples in the categories given in Table 2 (including a free and standardised translation for each).

Attitude	Example
inquisitive	@strankaDL Zakaj je v ministrov izjavi vejica? ;-) Šoltes: Aktualno politico <i>Why is there a comma in the minister's statement?</i>
informative	@drVinkoGorenak Spoštovani, manjka vam vejica za prvim ne. Prav je: "Ne, ne <i>Dear Sir, you are missing a comma following the first 'ne'. It should read: "Ne, ne</i>
lamenting	nekoc v eseju s "PS" dodal deset vejic in pripisal vstavi po potrebi <i>I once put ten commas in the postscript in an essay, with a comment: "Insert where appropriate."</i>
jocular	Če bi Krpan imel vejico v žepu, bi jo gotovo postavil <i>If Krpan [a Slovenian folk hero] had a comma in his pocket, he'd certainly use it.</i>
dismissive	@finance_si @NovaSlovenija Še vejice porihajte. Da ne boste kot Pojbič <i>Get your commas in order. So you aren't like Pojbič...</i>
defensive	čaki mal.. A zarad vejic je pa človek nepismen?? ;-) #svašta <i>Wait a second... So [the misuse of] commas make[s] you illiterate? #whatever</i>
apologetic	tudi jaz bi napisal z vejicami, če bi mi omejitev 140 znakov <i>I'd use commas if I had more than 140 characters</i>
idiomatic	Najboljši članek! Podpišem vse do vejice in pike! Desnica naj zakoplje <i>The best article ever! I agree with every comma and full stop!</i>

Table 2: Tweets relating to the comma.

As Table 2 demonstrates, we have identified eight different attitudes towards the comma in our dataset. The inquisitive attitude pertains to actual questions regarding the use of the comma. The second, i.e. the informative attitude, is complimentary to the inquisitive one as it provides explanation(s) on the use of the comma and/or answers to questions on the use of the comma. The example we provided above includes a detailed (and quite possibly condescending) explanation directed at a Slovenian MP explaining why his use of the comma in his tweet was incorrect.

The lamenting attitude covers examples of exasperation over the perceived difficulty over the use of the comma. The example above laments about the school experience of a former pupil who asked his teacher to insert commas for him if required. The jocular attitude covers examples that apply irony on the Slovenian situation and the obsession with commas, whereas dismissive tweets use (alleged) misused commas to portray a person or institution as incompetent in general. For instance, in our example in Table 2, a person directs a comment at the Twitter account of the Slovenian conservative party NSi, saying that they should get their writing in order lest they should “go full Pojbič”, referring to a Slovenian MP famous for his poor grammar and spelling in his tweets. It is of note that the Slovenian politicians (especially conservative) and official institutions are under constant scrutiny when it comes to language (in social media), especially regarding the comma.

On the other hand, the apologetic and defensive attitudes aim to explain or justify one’s “transgressions” with the comma, each in its own way. While the apologetic approach involves providing reasons for a particular example of comma misuse (either not knowing the rules or more pragmatic excuses like haste, typos, etc.), the defensive attitude conveys a stronger reaction, either against the user(s) exposing the misuse or against the relevance of the comma itself (normally in contrast to the meaning). The final category involves examples containing idiomatic references to the comma, i.e. using the comma as a metonymic expression for ‘language’ or ‘text’.

3.2 *Slovenščina* ‘Slovene’ and *jezik* ‘language’

As both keywords display very similar attitudes, we deal with them in a single section, with the obvious intention of displaying the attitudes relating to the “mother tongue”. A classification of tweets containing the two keywords is given in Table 3.

The tweets containing *Slovene* and *language* can be classified in five categories. The category General covers tweets providing general information and inquiries on both keywords. It is interesting to note that the example containing a purist tweet is again directed at the conservative MP Vinko Gorenak, who faced criticism, as recorded in Table 2, for his comma transgressions. This time he is the target of two purist comments, one lambasting him for improper grammar, and one for using a word that does not exist in Slovene on TV.

Attitude	Jezik 'language'	Slovenščina 'Slovene'
General/ inquisitive	A je slovenščina res najtežji jezik na svetu? <i>Is Slovene really the hardest language to learn in the world?</i>	Kako prevedemo "small talk" v slovenščino? <i>How do you say "small talk" in Slovene?</i>
nationalist	s slovenščino, našim maternim jezikom, delamo kot svinja z mehonom. <i>We treat Slovene, our mother tongue, like dirt.</i>	UPORABLJA ISKLJUČNO SLOVENSKE BESEDE. Slovenščina to smo SLOVENCIALI NIMAMO SVOJIH <i>Use only Slovene words, we are Slovenes after all.</i>
purist	@drVinkoGorenak Prosim za podnapise v slo jeziku: kaj pomeni "paložek"? @PlanetTV <i>Subtitles please: what does "paložek" mean in Slovene?</i>	@drVinkoGorenak "bodite točna" v slovenščini ne obstaja. Če že vikate, vikajte <i>"Bodite točna" [i.e. the mixing of plural and singular forms] doesn't exist in Slovene. If you mean to address someone formally, stick to it.</i>
jocular	casov " the day after yesterday". Jeziki ji ful laufajo <i>"The day after yesterday." She's great with languages.</i>	Ha odkar delam s kamionarji, mi slovenscina neki sepa. <i>Ever since I started working with truckers, my Slovene has deteriorated.</i>
idiomatic	vidva takorekoč govorita isti jezik. In kaj bi bilo <i>You speak the same language, as it were.</i>	vidim, da si bil zgolj v službi slovenščine. ČE ČM! <i>I see you were only in Slovene's service.</i>

Table 3: Tweets relating to language and Slovene

The tweets that employ a jocular perspective either expose the difficulties people have with Slovene or are meant as a ridicule of the Slovenian hard-line stance on language, whereas, in idiomatic tweets, the keyword *language* is very common, and *Slovene* much less so. The most significant difference in attitudes towards the comma and towards language/Slovene lies in the generality – the tweets belonging to the latter category are much more general (as well as less vicious and dismissive), but at the same time more nationalist.

3.3 Pravopis 'orthography' and slovnica 'grammar'

As in section 3.2, we combine the related keywords into a single analysis. This is especially relevant because most people confuse or substitute orthography and grammar, the latter most often being an umbrella term for all language-related matters, but in the Slovenian linguistic tradition, the two are distinct entities.

Attitude	slovnica 'grammar'	pravopis 'orthography'
purist	slovnica pa pravopis 1,vsebina 5 <i>You get an F for grammar and orthography, and an A for content.</i>	Vzgleđ? Pravopis pod vzglavnik, nekompetentnež <i>"Vzgleđ"? [an ungrammatical form of vzgleđ] Put the orthography under your pillow, you incompetent...</i>
apologetic	domače. Tle ni prostora za visoko slovnico. Če ne morem na #hodok, sem pa <i>This is no place for elevated grammar.</i>	tw je za vsebino, ne pravopis :) mene je bolj strah <i>Twitter is meant for content, not orthography.</i>
informative	slavisti. Ali pač? Po slovenski slovnici je desni prilastek pri kraticah <i>The Slovenian grammar on modifiers with acronyms says that...</i>	osebno se tudi zdi tako. Ampak pravopis je stvar konsenza. Greste na <i>It may seem like that but orthography is all about consensus.</i>
inquisitive	po ;-) Saj menda "greve" ni po slovnici? #samvprašam #slovnica <i>Surely "greve" is ungrammatical?</i>	inovatorke. Ali cestarji obvladajo pravopis? via <i>Do road workers know orthography?</i>
idiomatic	Lukšič: Ko gre za slovnico političnega delovanja, je treba <i>When it comes to the grammar of political action...</i>	le za demokracija, črn dan za pravopis! #krivopisje #domoljupi <i>Not just democracy, today is a sad day for orthography.</i>
nationalist	slabo znanje slovnice se imenuje "čefurščina" ;)) <i>Poor grammar is indicative of čefurščina [a pejorative term for the dialect of Slovene spoken by immigrants from the countries of the former Yugoslavia]</i>	slovenski narod, niti osnovnega pravopisa ne pozna. Kak veš, da si v <i>Alas, the nation of Slovenia – they don't even know the basic orthography.</i>
jocular	uvede policijska ura in odpravi slovnica. Just sat through Simeon ten <i>Let's implement curfew and abolish grammar altogether.</i>	crknu!?! Kako se reče po noven slo pravopisu beseda KLON? PODOBNIK! <i>What is a clone called in Slovene? It's PODOBNIK [a surname of two Slovenian MPs, as well as a very literary word for a lookalike in Slovene].</i>

Table 4: Tweets relating to grammar and orthography

The attitudes contained in the tweets relating to orthography and grammar are given in Table 4. The attitudes in tweets pertaining to orthography and grammar can be categorized in 7 different categories. As can be seen in our purist examples, users on Twitter employ a rather harsh tone when it comes to these keywords. The same seems to hold true for nationalist tweets, which, as in our examples, usually employ a hard-line stance or one of a exasperation over the use of language of people writing in Slovene.

With apologetic tweets, we can see an interesting trend that users themselves, upon being faced with accusations of having “poor grammar”, are trying to discredit Twitter as a means of “formal” communication and instead establish it as a means of fast, direct, and informal communication. The concepts of “orthography” and “grammar” are thus used as synonyms for “form”, whilst the significance of “meaning” is pointed out.

4. Conclusion

In this paper, we were interested in discovering whether or not the stereotypical attitudes towards written Slovene have been preserved and transformed in computer-mediated communication, in our case on Twitter. The analysis shows that, in spite of the belief that informality is an integral part of new media, all traditional notions pertaining to Slovene are still very much alive in the digital age. What is more, some forms of prejudice seem to be flourishing online due to the lack of oversight. For instance, personal attacks with dismissive tweets aiming to destroy one’s credibility are very common, especially in regard to the comma, whereas more general topics attract more general attitudes, but just as vicious (purist and nationalist).

5. Acknowledgements

The work described in this paper was funded by the Slovenian Research Agency within the national basic research project “Resources, Tools and Methods for the Research of Nonstandard Internet Slovene” (J6-6842, 2014-2017).

6. References

- Fišer, D., Erjavec, T., Ljubešić, N. 2017. The compilation, processing and analysis of the Janes corpus of Slovene user-generated content. Wigham, C.R. & Ledegen, G. (ur.). *Corpus de communication médiée par les réseaux: construction, structuration, analyse*. Collection Humanités Numériques. Paris: L’Harmattan (in print).
- Popič, D., Fišer, D., Zupan, K., Logar, P. (2016). Raba vejice v uporabniških spletnih vsebinah. *Proceedings of the Language Technologies and Digital Humanities Conference*. Ljubljana, Slovenia: 149–153.
- Robinson, D. (2016). Text analysis of Trump's tweets confirms he writes only the (angrier) Android half. (<http://varianceexplained.org/r/trump-tweet>)