

Linguistic Analysis of Emotions in Online News Comments - an Example of the Eurovision Song Contest

Ana Zwitter Vitez,^{+,*} Darja Fišer^{*,†}

⁺ Faculty of Humanities, University of Primorska, Titov trg 5, 6000 Koper, Slovenia

^{*} Department of Translation, University of Ljubljana, Aškerčeva 2, 1000 Ljubljana, Slovenia

[†] Department of Knowledge Technologies, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

E-mail: ana.zwitter@guest.arnes.si, darja.fiser@ff.uni-lj.si

Abstract

The aim of the study is to identify linguistic differences of positive and negative comments on the example of online comments on the news about the Eurovision song contest. The results show that positive comments have a typical exclamation form and simple sentence structure, and include more informal vocabulary and orthographic variation. Negative comments, on the other hand, are more likely to be formulated as statements with a complex syntax structure and with a neutral vocabulary and standard orthography. The detected differences can be explained by the communicative function of the negative comments that act as reviews and therefore call for thorough argumentation and build an individual's reflective identity.

Keywords: online news comments, linguistic analysis, sentiment analysis, Eurovision song contest

1. Introduction

Identifying emotions in language is a relevant field of research because of the strong connection between the physiological arousal of an emotion and its social display (Mygovych, 2013). If we understand how people feel, we can analyse or even predict how they will react in certain situations. This is why sentiment analysis can be used for predicting societal changes, election results, and customer satisfaction (Liu, 2015).

In online comments, analysis of emotions is particularly interesting because comments enable users to formulate their own opinion and to find their own identity independently of the official media content. Wright (2009)¹ even claims that “for many businesses, online opinion has turned into a kind of virtual currency that can make or break a product in the marketplace”.

Online news comments of the Eurovision song contest represent a specific dataset because they usually evoke polarised emotions (either users strongly support or hate Eurovision song contestants and/or their songs). The comments often even exceed the scope of the song contest itself and refer to wider political and societal issues (e.g. *Azerbejdžan podeli Rusiji 12 točk in si s tem zagotovi dostavo plina za še eno leto #Eurovision. / Azerbaijan gives Russia 12 points, thus guaranteeing their gas supply for another year #Eurovision*).

The aim of this paper is to provide a linguistic analysis of online comments in order to detect syntactic, lexical and orthographic differences between positive and negative comments.

2. Emotion Analysis in CMC

Different approaches have been developed to analyse emotions in computer-mediated communication. In data mining three basic categories are most often used: positive, negative, neutral (Smailović, 2013). These models are very useful on big datasets to study overall

trends but are not always reliable for linguistic analyses of individual texts.

In discourse analysis, much more fine-grained sentiments are typically examined: happiness (Stefanowitch 2004), shame (Retzinger, 1991), and even irony (Haverkate, 1990). These approaches are very interesting for qualitative analyses but cannot be scaled for emotion identification on bigger datasets.

In this paper we focus on a qualitative analysis of a small dataset of news comments on the Eurovision song contest in Slovenian in order to examine linguistic characteristics of opinionated texts. Once comments were manually attributed a sentiment category, they were analysed on the syntactic, lexical and orthographic level.

3. Methodology

3.1 Sample Creation

The analysis was performed on a sample extracted from the Janes corpus v0.4 (Fišer et al., 2016). The sample contains 70 comments referring to an article announcing that the Slovenian representative was selected to compete in the finals of the Eurovision song contest² published on the national television and radio online news portal RTV Slovenija. Only opinionated comments were taken into account for the study. Neutral, factual and objective comments (e.g. *Bjørn Einar Romøren*) were discarded as were off-topic comments or direct replies to a previous comment that were part of an internal debate that had nothing to do with the article they appeared under (e.g. *Kje je kolega XX? Upam, da ni zaspal! / Where is our camerad XX? I hope he hasn't fallen asleep!*).

3.2 Sentiment Annotation

First, comments were manually attributed a sentiment category (positive, negative) by two annotators. Disagreements were detected in 6 cases (8%), which were discussed in order to reach a systematic final decision. 4

¹URL: <http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html>

² URL: <http://www.rtv slo.si/zabava/glasba/evrovizija-2014/slovenija-s-tinkaro-v-evrovizijski-finale-vidimo-se-v-soboto/336356>

of the cases involved comments which consisted of two parts, expressing a different sentiment each (e.g. *Tinkari pa zaželim srečo, četudi je ta Evrovizija zadnjih 10 let z uvedbo polfinalov čisti cirkus in šov, ki ga prav tako dolgo ne jemljem več tako resno. / I wish Tinkara all the best, even though for the past 10 years the Eurovision and its semi-finals have been nothing but a circus and a show that I am not taking seriously anymore.*). In such cases, the annotators agreed to determine the prevalent sentiment in the comment. In 2 of the cases, it was not clear out of context whether the comments were meant literally, as a joke or cynical (e.g. *Pričakujem 12 točk iz Makedonije. / I am expecting 12 points from Macedonia.*). In such cases, the entire discussion thread was examined for a wider context and annotated accordingly.

3.3 Linguistic Analysis

Each sentence in the sample was analysed for sentence type (statement, exclamation, question, order), sentence structure (simple, complex), vocabulary characteristics and orthography (formal, informal). Examples of the analysis are presented in Tables 1 and 2.

SLO: Držim pesti!	
ENG: Fingers crossed!	
Sentiment	positive
Sentence form	exclamation
Sentence structure	simple
Vocabulary	/
Orthography	informal

Table 1: Linguistic analysis of a positive comment.

SLO: Kolikor slišim, se je včeraj slabo odrezala.	
ENG: As far as I heard, she did not fare well last night.	
Sentiment	negative
Sentence form	statement
Sentence structure	complex
Vocabulary	slabo
Orthography	standard

Table 2: Linguistic analysis of a negative comment.

4. Results and discussion

Comments with the same sentiment label were compared in order to detect the shared linguistic properties on the syntactic, lexical and orthographic level. The sample contains slightly more negative (53%) than positive (47%) comments.

4.1 Syntax

As can be seen from Figure 1, a large majority (86%) of the negative comments are statements (e.g. *Ne, ne bo. / No, it won't.*) with only a few examples of questions (8%) and exclamations (5%). Among the positive comments, on the other hand, there is a similar share of statements (48%) and exclamations (45%) (e.g. *Srečno! / Good luck!*). While positive comments contain no questions, there are a couple of commands (6%) (e.g. *Uživajmo in ne nergajmo kot stare babe. / Let's enjoy the show and not whine like old ladies.*).

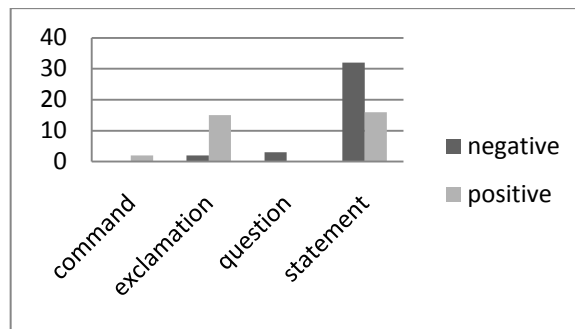


Figure 1: The distribution of different types of sentences in positive and negative comments.

The majority of the sentences in the negative comments are complex (62%) while nearly half of the sentences in the positive comments (49%) are simple. For illustration, Figure 2 contains examples of a complex negative sentence and a simple positive one.

Negative, complex:
SLO: Če zaupaš našim medijem, so še skoraj vsako leto bile kritike glede naših pesmi pozitivne, ampak rezultata pa nobenega in isto bo letos.
ENG: According to our media, despite positive reviews our songs were unsuccessful almost every year and this year won't be any different.

Positive, simple:
SLO: Imam dober občutek.
ENG: I have a good feeling about this.

Figure 2: Examples of simple and complex sentences in positive and negative comments.

4.2 Vocabulary

The vocabulary level was manually annotated following the criterion of whether a comment is characterised by a specific lexical unit carrying an opinion or not.

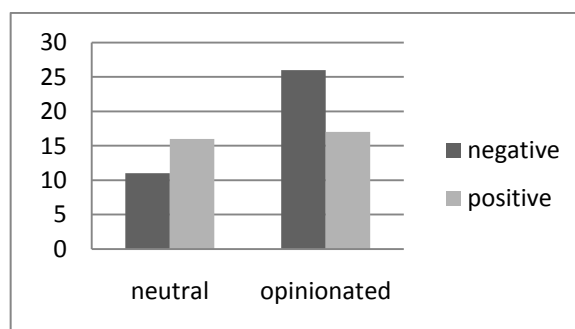


Figure 3: The distribution of neutral and opinionated vocabulary in positive and negative comments.

As Figure 3 shows, vocabulary in the negative comments is heavily opinionated (70%), e.g. *kuhna (inside deal), davkoplačevalci (taxpayers), lajna (broken record)*. About half of the positive comments are characterised by opinionated vocabulary (51%), but this vocabulary is not topic-specific and usually expresses general support, e.g. *Srečno! (Good luck!), upam (I hope), podpiramo (we support)*.

4.3 Orthography

At the orthographic level the following phenomena that are typical of CMC language were observed: informal orthography (e.g. *dejmo* instead of *dajmo*), use of all-caps (e.g. *BO*), non-standard use of punctuation (e.g. *Dajmo klobasica!!!:)* / *Do it*), and emoticons (;)). As can be seen in Figure 4, while distinctly standard orthography (78%) is used in the negative comments, nearly half of the positive comments (42%) contain non-standard orthographic features.

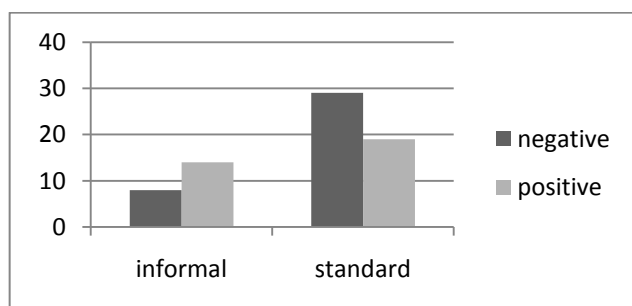


Figure 4: The distribution of non-standard and standard orthography in positive and negative comments.

For illustration, Figure 5 contains an example of a positive comment with standard orthography and an example of a positive comment with non-standard spelling.

Negative, non-standard:
SLO: *dejmo naši!!!*
ENG: *c'mon, team!!!*
Positive, standard:
SLO: *Danes imajo naši turisti še zadnji dan turistovanja na davkopláčevalske stroške.*
ENG: *Today is the last vacation day attaxpayers' expense for our tourists.*

Figure 5: Examples of standard and non-standard orthography in positive and negative comments.

5. Conclusions

The aim of the study was to identify linguistic characteristics of positive and negative comments on the example of comments on articles about the Eurovision song contest. The results show that positive comments are fewer, typically have an exclamation form and simple sentence structure, and contain more informal vocabulary and orthography. Negative comments are more numerous, are typically represented as statements with a more complex syntax structure and with distinctly general vocabulary as well as standard orthography.

While it is true that the analysed sample is small and limited to a single topic, the results are very homogenous and consistent throughout the analysis. A plausible explanation for such a discrepancy is the critical function of the negative comments which calls for thorough argumentation, not affect, and from the position of a reflective individual who acts in his own capacity, not as a member of regional or social groups adherence to which

usually shows through the use of typical vocabulary and orthography.

Our future work plan is to extend the analysis on a wider range of highly opinionated topics (sports, politics, religion, product and service reviews) and text types (blogs and blog comments, tweets, forum posts, Wikipedia talk pages). In addition, the set of sentiments will be more fine-grained in order to distinguish between different types of negative or positive sentiment such as support and cynicism that deserve special treatment.

6. 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 Non-standard Internet Slovene" (J6-6842, 2014–2017).

7. References

- Fišer, D., Erjavec, T., Ljubešič, N. (2016): JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin. *Slovenščina 2.0*, 4(2), 67–100.
- Haverkate, H. (1990). A speech act analysis of irony. *Journal of Pragmatics*, 14(1), 77–109.
- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
- Mygovitch, I. (2013). Secondary nomination in the modern English language: affective lexical units. *Visnik LNU imeni Tarasa Ševčenka*, 1(1), 206–214.
- Ritchie, G. (2004). The Linguistic Analysis of Jokes. *Journal of Literary Semantics*, 33(2), 196–197.
- Smailović, J., Grčar, M. Lavrač, N., Žnidaršič, M. (2014). Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences* 285: 181–203.
- Stefanowitsch, A. (2004). Happiness in English and German: A metaphorical-pattern analysis. In M. Achard and S. Kemmer (eds.) *Language, Culture, and Mind*, 137–149. CSLI Publications.
- Retzinger, M. S. (1998). *Violent Emotions: Shame and Rage in Marital Quarrels*. Newbury Park, CA: Sage Publications.