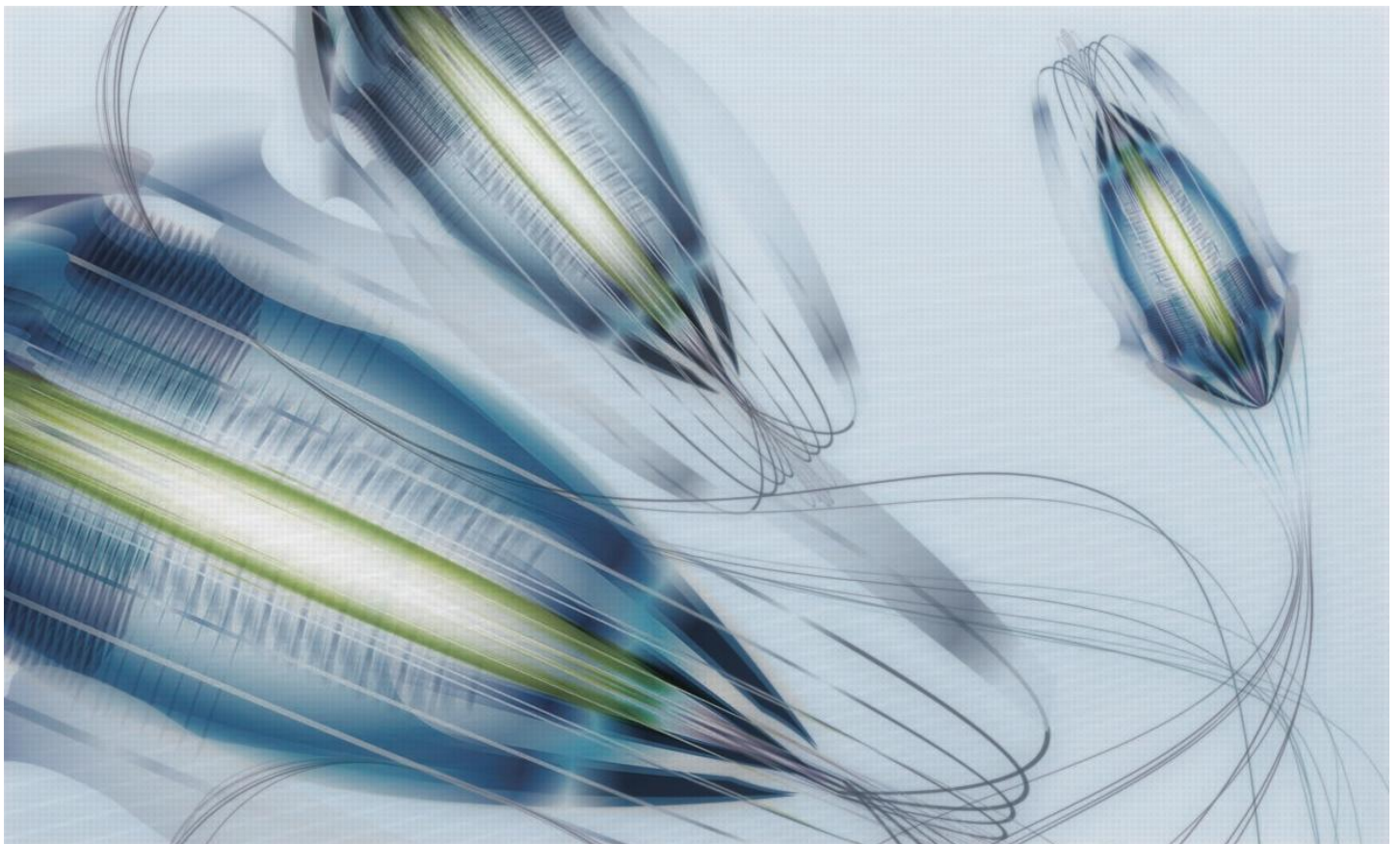# The LiLaH Emotion Lexicon of Greek, Kurdish, Turkish, Spanish, Farsi and Chinese

Jens Lemmens, Ilia Markov and Walter Daelemans

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

The LiLaH Emotion Lexicon of Greek, Kurdish, Turkish, Spanish, Farsi and Chinese

Computational Linguistics & Psycholinguistics Research Center

CLiPS Technical Report Series (CTRS)

February, 2023

CTRS-009

Authors

Jens Lemmens

CLiPS Research Center, University of Antwerp (BE)

jens.lemmens@uantwerpen.be

Ilia Markov

CLTL, Vrije Universiteit Amsterdam (NL)

i.markov@vu.nl

Walter Daelemans

CLiPS Research Center, University of Antwerp (BE)

walter.daelemans@uantwerpen.be

The LiLaH Emotion Lexicon of Greek, Kurdish, Turkish, Spanish, Farsi and Chinese

Jens Lemmens[1], Ilia Markov[2] and Walter Daelemans[1]

[1] University of Antwerp, [2] Vrije Universiteit Amsterdam

Abstract. In this technical report, we present manual translations in Greek, Farsi, Spanish, Kurdish, Turkish, simplified Chinese and traditional Chinese of the NRC word-emotion association lexicon (Mohammad and Turney, 2013). For each of the aforementioned languages, we collaborated with one native speaker who checked and - if necessary - corrected the automatically generated translations of the lexicon according to the provided annotation guidelines. The manual translations can be found here: http://hdl.handle.net/10032/tm-a2-v7

Keywords: Emotion lexicon, sentiment analysis, minority languages

1   Introduction

Emotion lexica are resources that have proven to be useful for sentiment analysis, emotion detection, hate speech detection and related tasks (Ljubešić et al., 2020; Mohammad et al., 2018). They are, however, expensive and time-consuming to create, which is why they are often scarce for minority languages (Buechel et al., 2020). In a similar fashion to the LiLaH emotion lexicon of Croatian, Dutch and Slovene (Ljubešić et al., 2020), we present the manual translations of the NRC emotion lexicon for Greek, Kurdish, Turkish, Spanish, Farsi, traditional Chinese and simplified Chinese. This resource contributes to the following already existing lexica for these languages:

- Greek: Greek Sentiment Lexicon (Tsakalidis et al., 2018)
- Kurdish: AffectVec[1] multilingual
- Turkish (Dehkharghani et al., 2015)
- Spanish: Spanish Emotion Lexicon (SEL)[2]
- Farsi: Manual translations[3] of the NRC lexicon, previously released
- Chinese: AffectVec multilingual

2   NRC Word-Emotion Association Lexicon

The NRC word-emotion association lexicon was constructed by Mohammad and Turney (2013). It consists of 14,182 word entries accompanied by binary association labels for 2

---

[1] http://emotionlexicon.org/multilingual/
[2] https://mailman.uib.no/public/corpora/2012-December/016707.html
[3] https://github.com/mhbashari/NRC-Persian-Lexicon

sentiments ("positive" and "negative") and 8 emotions ("anger", "disgust", "fear", "sadness", "joy", "trust", "anticipation" and "surprise").

The provided translations of the NRC lexicon are automatically generated with Google Translate, but an extrinsic evaluation in previous research has shown that manual corrections of these translations result in higher performance when using the lexicon for a hate speech detection task (Ljubešić et al., 2020). We therefore apply the same methodology to other understudied languages.

3   Translation Procedure

For the translations of the NRC lexicon, one native speaker (all master students) for each of the aforementioned languages was asked to evaluate and if necessary correct the automatic translations generated with Google Translate. They were instructed with the following guidelines (identical to Ljubešić et al., 2020 (p. 154)):

- We only review entries associated with at least one sentiment or emotion, as these are the only relevant entries in the emotion lexicon.
- While translating, the sentiment and emotion labels should be taken into account by the translator.
- If a source word is polysemous, we translate only those senses that relate to the given sentiment and emotion.
- We include all target synonyms of the source word that are frequent and that are not polysemous or homonymous with target words of a different sentiment and emotion.
- Given that the original lexicon was not part-of-speech encoded, we do not encode it in the translations either. We add translations in multiple parts-of-speech if this follows the guidelines above.

The manual translations resulted in 6,468 emotion-associated entries in 7 languages.

Acknowledgements

---

[4] https://www.lilah.eu

Bibliography

Sven Buechel, Susanna Rücker, and Udo Hahn. 2020. Learning and Evaluating Emotion Lexicons for 91 Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1202–1217, Online. Association for Computational Linguistics.

Rahim Dehkharghani, Yucel Saygin, Berrin Yanikoglu, and Kemal Oflazer. 2015. SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. In *Language Resources and Evaluation*, 50(3), pages 667–685. https://doi.org/10.1007/s10579-015-9307-6

Nikola Ljubešić, Ilia Markov, Darja Fišer, and Walter Daelemans. 2020. The LiLaH Emotion Lexicon of Croatian, Dutch and Slovene. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 153–157, Barcelona, Spain (Online). Association for Computational Linguistics.

Saif Mohammad and Peter Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. Computational Intelligence, 29. https://doi.org/10.1111/j.1467-8640.2012.00460.x.b

Saif Mohammad and Peter Turney. 2010. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34, Los Angeles, CA. Association for Computational Linguistics.

Adam Tsakalidis, Symeon Papadopoulos, Rania Voskaki, Kyriaki Ioannidou, Christina Boididou, Alexandra I. Cristea, Maria Liakat, and Yiannis Kompatsiaris. 2018. Building and evaluating resources for sentiment analysis in the Greek language. In *Language Resources & Evaluation*, 52, pages 1021–1044. https://doi.org/10.1007/s10579-018-9420-4