The Paradox of Multilingual Emotion Detection

Luna De Bruyne
LT[3], Language and Translation Technology Team
Ghent University
luna.debruyne@ugent.be

Abstract

The dominance of English is a well-known issue in NLP research. In this position paper, I turn to state-of-the-art psychological insights to explain why this problem is especially persistent in research on automatic emotion detection, and why the seemingly promising approach of using multilingual models to include lower-resourced languages might not be the desired solution. Instead, I campaign for the use of models that acknowledge linguistic and cultural differences in emotion conceptualization and verbalization. Moreover, I see much potential in NLP to better understand emotions and emotional language use across different languages.

1 Introduction

Variation and diversity are inherent to human life, not least to human language. Yet, machine learning approaches used in natural language processing (NLP) usually ignore this variation and are biased to a (consciously or subconsciously imposed) norm. Mohammad (2022) stressed that current NLP applications therefore often amplify societal inequalities and "lead to more adverse outcomes for those that are already marginalized". Indeed, we have known for some time that NLP applications show several biases, e.g., racial bias in conversational agents (Sap et al., 2019) or gender bias in machine translation (Savoldi et al., 2021). Mohammad (2022) therefore campaigns for introducing ethics sheets for AI tasks, in which diversity should be one point to be addressed, including a discussion of the design choices that impact diverse groups of people.

Besides inequality across social groups, Søgaard (2022) recently pointed at the inequality across languages as an unwanted bias in NLP (around two thirds of NLP research at top venues would be devoted exclusively to English, which has not changed over the last 10 years). However, instead of merely acknowledging these biases, he argues that it is simple to mitigate inequality amplification, for which he proposes three strategies inspired by policies for reducing carbon emissions: (i) an NLP Cap and Trade, (ii) an NLP Carbon Tax, and (iii) NLP Car-Free Sundays. As the language bias is generally towards English, this would in practice mean to (i) distribute quota for publications on English, (ii) impose a cost on researchers submitting papers on English or (iii) a one-year ban on English models. These measures should encourage groups to work on NLP systems and resources for other languages than English.

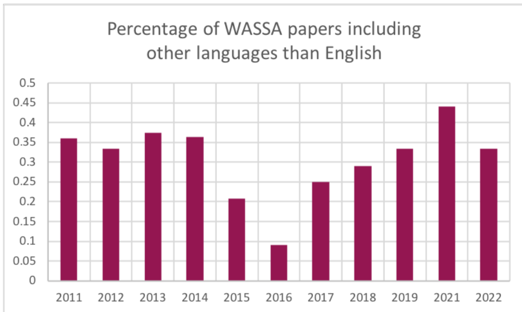


Figure 1: Proportion of papers (presented at WASSA between 2011 and 2022) including other languages than English.

For work on emotion detection (also referred to as automatic emotion recognition or AER)[1] as well, we observe a huge bias towards English resources and systems. In order to get some insight in the variety of languages addressed in research on emotion detection and related research fields, I analyzed the papers that were presented at the Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA) since 2011[2], and counted the number of papers that included other languages than English.[3] My findings,

[1] I will use these terms interchangeably.

[2] Except for the first WASSA edition in 2010, the contributions of all editions are found on https://aclanthology.org/venues/wassa/.

[3] I manually scanned all WASSA publications, except ab-

Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pages 458–466
July 14, 2023 ©2023 Association for Computational Linguistics

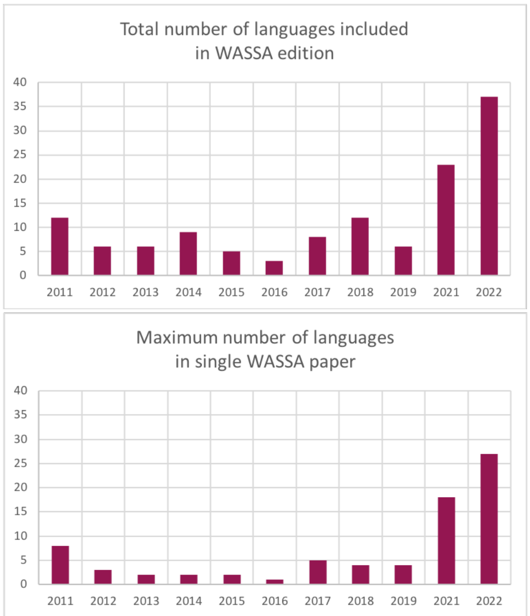


Figure 2: Number of languages that were included in the papers presented at WASSA between 2011 and 2022.

shown in Figure 1, are in line with the numbers reported by Søgaard (2022): the proportion of papers that not exclusively focused on English fluctuates around one third and did not increase over the years. However, I did observe that there was a remarkable increase in the past two years concerning the total number of languages that were included and the maximum number of languages across papers (see Figure 2).

I believe that this positive trend is fueled by large language models like multilingual BERT (Devlin et al., 2019), which lend themselves perfectly to (zero or few-shot) transfer learning. Indeed, the high number of languages at WASSA 2021 and 2022 is largely due to just three papers that include many different languages: Lamprinidis et al. (2021), Bianchi et al. (2022) and Rajda et al. (2022) respectively include 18, 19 and 27 languages, each of them performing experiments using multilingual models.

However, even though multilingual models might seem promising for tackling NLP tasks for other (and lower-resourced) languages than English, the use of multilingual models result in a paradox in the case of emotion detection: using multilingual emotion detection models inherently assumes that different languages deal with emotions the same way. This idea may be in line with classical views on emotion analysis, but goes completely against state-of-the-art evidence in psychology showing that emotions are not universal, but rather culture (and language) dependent (Barrett, 2017; Mesquita et al., 2016).

Therefore, I believe the real challenge does not lie in attracting more research on a larger number or greater variety of languages, but in studying emotion detection without falling into universalist ideas and instead acknowledging differences in emotional conceptualization and verbalization across languages. This can be achieved by creating valid datasets (original data written and annotated by native speakers), with label sets that are adjusted to the target language (using native emotion words or emotion representations that go beyond the anglocentric basic emotions).

In this position paper, I will discuss state-of-the-art psychological findings and their implications for emotion detection in NLP (Section 2). I will then expand on these implications by discussing them in the light of current papers in AER research (Section 3). Next, I will propose some research directions that can be taken in AER to better align with psychological evidence (Section 4). Finally, my viewpoints are summarized in some concluding thoughts (Section 5).

2 What psychology teaches us and what it means for AER

For a very long time, a universalist view on emotions prevailed. In such a view, it is believed that the way emotions are conceptualized and experienced is the same across different cultures and that emotions are biologically hard-wired. Especially the work on facial expressions by Ekman and his colleagues, in which participants from different cultures made similar decisions when asked to match emotion words or emotional stories with facial expressions of *anger*, *disgust*, *fear*, *joy*, *sadness* and *surprise*, consolidated the idea of universal emotions (Ekman et al., 1969; Ekman and Friesen, 1971).

However, experiments like Ekman's are biased by the Western perspective of the researcher and influenced by the used research methods (e.g., the

stracts of invited talks and submissions for shared tasks (but, for each shared task, I did include the task description paper). If the paper included other languages than English, it was usually mentioned in the paper title or in the abstract, although sometimes I had to read the dataset description to find out. For a remarkable number of papers, no language was mentioned at all. In almost all of these cases, the papers used English data. Plots for the distribution of languages in all WASSA discussions are shown in Figure 4 in the Appendix.

choice of emotion words to match with; posed instead of spontaneous emotion expressions). More recent experiments using a greater diversity in research methods and contexts, reveal diversity instead of universality (Gendron et al., 2014).

Moreover, a critical review of literature about the biological basis of emotion categories by Barrett (2006) indicates that evidence for the existence of such universal basic emotions is inconclusive. Rather, there is "cultural variation in the experience of emotion that is intrinsically driven by cultural differences in emotion categories and concepts" (Barrett, 2006, p.38). In a similar vein, Mesquita et al. (2016) claim that emotional experience is culturally constructed, which shows itself in cultural differences regarding how people communicate and talk about their emotions.

Variation in emotion conceptualization and experience on the hand, and concomitantly, emotion expression and verbalization on the other hand, both have consequences for automatic emotion recognition in NLP.

Diversity in emotion conceptualization

Not all cultures dispose of the same emotion concepts. There exist many examples of emotion concepts in specific languages that do not seem to have a translation in other languages, e.g., *toska* in Russian (described as spiritual anguish without a specific cause), *saudade* in Portuguese (described as a somewhat melancholic feeling of incompleteness), *lítost* in Czech (a state of agony and torment created by the sudden sight of one's own misery) or *fago* in Ifaluk (which has characteristics shared with the English concepts *love*, *compassion* and *sadness*). Even for concepts that are claimed to be 'basic emotions', not all languages have a word – e.g., there exists no word for *sadness* in Tahiti (Levy, 1984).[4]

Besides the untranslatability of some emotion words, there are also many differences in the connotations and meanings of emotion terms across languages (Mesquita et al., 1997; Pavlenko, 2008; Wierzbicka, 2009). The concept of *anger*, for example, is hardly the same as *gnev*, although they are usually glossed as translation equivalents in English-Russian dictionaries (Wierzbicka, 1998).

That there is variation between emotion concepts across languages, is not just because the emotion words we use to refer to them are not perfect translation equivalents (which is an inherent problem related to translation in general, and not only to the translation of emotion words), but because emotions are culturally constructed (Mesquita et al., 2016). For example, in studies comparing the emotion conceptualization between inhabitants from the United states and Japan, it was found that emotion concepts arise from the *individual* in the perspective of American respondents, while they arise from the *relationships* between individuals in Japanese respondents (Uchida et al., 2009). This reflects how in some cultures (e.g., in Japan), processes at the level of the collective are more important for constructing emotions, while in other cultures (e.g., the U.S.) individual-level processes prevail (Mesquita et al., 2016).

Implications for AER: As emotion concepts are dependent on the culture we live in and the language we speak, we should design our datasets and models accordingly. Native speakers should label texts, with emotion labels that make sense to them. We should not use the Ekman emotion taxonomy as the basis of AER without motivation, as the claim that these emotions would be universal has been disproved. Moreover, there is no reason to believe that these emotions have the same meaning as they have for speakers of English.

Diversity in emotion verbalization

The way we conceptualize and experience our emotions has of course a huge impact on how we express and verbalize them. Again, the distinction between individualistic and collectivist cultures is important. People in individualistic cultures seem to be more openly conveying emotional feelings and use a more expressive style than people from collectivist cultures, which is illustrated by the reticence of verbal and non-verbal expression of the emotion *love* by Chinese people compared to Americans (Caldwell-Harris et al., 2013). Moreover, it was found that there are several linguistic differences in the emotional expression between people from individualistic cultures – where emotion terms are related to the self and the use of nouns and adjec-

[4] The fact that a language does not have a word for specific emotion concepts, does not necessarily mean that people speaking that language cannot *conceptualize* such an emotion. However, according to Barrett (2017), conceptualization is a prerequisite for emotional *experience*. Whether Tahitians can experience *sadness* thus depends on whether they can conceptualize it. We are not sure whether having a word for a concept is necessary for having the concept, but it seems the case that having a word makes conceptualization easier (Barrett, 2017).

tives is more prominent – and collectivist cultures – where emotion terms are more often used to refer to relationships intead of the individual, and more interpersonal verbs are used (Semin et al., 2002; Mesquita et al., 2016).

However, there are even more subtle differences that have nothing to do with the individualistic-collectivist dichotomy. Languages can have very characteristic strategies for emotion verbalization, e.g., using diminutive, augmentative and pejorative suffixes in Spanish or Portuguese (Rudolph, 1990), or emotion verbalization that is focused on the human body in Russian (Wierzbicka, 1999). Also emojis, which are a common strategy to convey emotions in informal writing, show much divergence between languages and countries, sometimes even between countries that are geographically close to each other and in which the same language is spoken, like Mexico and Columbia (Kejriwal et al., 2021).

Implications for AER: As emotion verbalization is dependent on the culture we live in and the language we speak, models should be trained on texts that are written in the language for which we want to use the developed emotion detection system. Both training and evaluation data should be written and labeled by native speakers, as only native speakers might pick up on language-specific emotion verbalization strategies.

3 What we are really detecting in NLP

I will expand on the implications mentioned in the previous section by discussing recent papers dealing with multilingual AER. I will zoom in on three important aspects of automatic emotion detection: the data, the labels, and the models.

I selected three WASSA submissions from 2021 and 2022, namely those that included the highest number of languages: the papers of Rajda et al. (2022), Bianchi et al. (2022), and Lamprinidis et al. (2021). The first one includes an assessment of sentiment analysis in 27 languages (I will refer to this work as MSA, standing for multilingual sentiment analysis); the second one presents XLM-EMO, a multilingual emotion detection model evaluated on 19 languages; and the last one presents Universal Joy, an emotion detection dataset including 18 languages. Additionally, I will also discuss the work by Öhman et al. (2020), who present the multilingual emotion detection dataset XED, including 32 languages.

The data: Both in MSA (Rajda et al., 2022) and Universal Joy (Lamprinidis et al., 2021), data is used that was originally written in the target languages. While existing sentiment datasets are used in MSA, Universal Joy is created by scraping Facebook posts based on the Facebook-specific feelings tags. Also in XLM-EMO (Bianchi et al., 2022), original data from existing emotion datasets is used, although the data for some languages (French, German and Hindi) was machine-translated (from Spanish to French and German, and from English to Hindi, respectively).

I believe the use of (machine) translations is problematic, as it neglects language-specific characteristics of emotion verbalization. Moreover, as shown by Troiano et al. (2020), emotional connotations are partly lost in the machine translation process. Also for XED (Öhman et al., 2020), translated (although human-translated) data was used, namely in the form of subtitles. Although the use of translated subtitles allows for the creation of a parallel corpus – which is in itself a compelling idea – it is far from ideal to use non-original data, as such data – even if it is translated by humans – might be biased towards the source language in terms of emotion characteristics.

The labels: I will not focus on MSA here, as it uses sentiment labels instead of fine-grained emotion labels. Judging from the three other papers, there is still work to be done regarding the handling of emotion labels in multilingual datasets. It seems to be common to treat the labels across languages as one and the same category. In XLM-EMO, for example, datasets from various languages with different label sets are merged by removing instances that did not fit the labels *anger*, *fear*, *joy* or *sadness*. However, based on the literature cited in Section 2, it is hardly likely that *anger* in one language has a perfectly overlapping meaning with its translation in a different language. Moreover, this approach results in a huge loss of data and information. Universal Joy relies on the simple but nice idea of employing Facebook feelings-tags as labels. As Facebook users attach these tags themselves when posting messages, it is ensured that the labels correspond to the feelings of the writers of the posts. However, the original tags (27 different tags initially), were mapped to the five categories *anger*, *anticipation*, *fear*, *joy*, and *sadness*. The mapping happened in the same way for all languages, but again, it is not certain that these mappings make sense at the level of the individual languages. In XED, an even more risky approach is used, consisting of projecting labels that were manually annotated for the English instances to the translations of those instances in the other languages. Apart from the fact that emotion labels might not be comparable across languages, this approach assumes that utterances have the same emotional connotation, irrespective of in which language or culture it is uttered. In each of these papers, the classical view on emotion prevails, assuming that emotion categories are universal.

The models: In MSA, XLM-EMO, and Universal Joy, a pre-trained multilingual model (e.g., mBERT (Devlin et al., 2019) or XML-R (Conneau et al., 2020)) is fine-tuned on the multilingual datasets. In each of these papers, the default setting is to fine-tune on all languages at the same time. This neglects the fact that emotions are verbalized differently in different languages, and moreover enhances the classical emotion view by modeling emotion concepts as if they were one and the same category shared by all languages. Instead of acknowledging variation across languages, multilingual systems are modelling artificial universal emotion categories.

Moreover, the multilingual models in the discussed papers show a bias towards English: in each of the three datasets, there is more fine-tuning data for English than for the other languages. In fact, pre-trained models itself are already biased towards English anyway: in mBERT, for example, 21% of the training data is English.[5] This English bias is also evidenced by the zero-shot experiments described in XLM-EMO: Bianchi et al. (2022) fine-tuned mBERT on all languages except on a target language, which is respectively English, Arabic and Vietnamese. In contrast to Arabic and Vietnamese (where a language-specific model outperforms zero-shot experiments), there is almost no difference between the zero-shot performance and a language-specific model in the case of English as target language, indicating that the pre-trained model already contains information on English.

In further experiments on Universal Joy, cross-lingual fine-tuning effects are investigated: Lamprinidis et al. (2021) compare multilingual fine-tuning of mBERT (i.e., fine-tuning on all languages of the dataset) with monolingual fine-tuning (fine-tuning on only the target language). They observe positive cross-lingual effects, meaning that performance increases when fine-tuning data from other languages is added (especially when there are syntactic and typological similarities between fine-tuning and target languages). However, these positive effects were only found when the size of the target language dataset was small. For large target language datasets, including multiple languages for fine-tuning did not result in an improvement.

The experiments in the discussed papers show that multilingual models, fine-tuned on a variety of languages, can improve performance. But performance on what? The answer is: performance on classifying texts – which are sometimes not even originally written in the target language – into artificial emotion categories – that are modeled across different languages at the same time and might not make sense according to the emotion conceptualization in that target language. Moreover, the increased performance seems to be just as easy to reach by gathering more data in the target language itself.

I am not claiming that these multilingual models are by definition useless for AER. They can be a compromise in real low-resourced situations. However, we should face that, although such multilingual models are driven by a very inclusive idea, they might not be inclusive at all, and may disadvantage languages that are verbalizing and conceptualizing emotions in a different way than it is done in English.

Therefore, rather than investing in multilingual models, we should invest in better monolingual resources that are not created from an Anglocentric (or by extension: Western) perspective. Qualitative monolingual resources (respecting language-specific ways of emotion conceptualization and verbalization), are moreover needed to investigate how multilingual models really deal with the language-specificity of emotions.

4 The real challenge in AER

Similarly to Søgaard (2022), I believe we should act against the dominance of English in NLP and more specifically in emotion detection. However, nudging researchers to publish papers on other languages than English or to create multilingual datasets is not sufficient, or at least not if we do not let go of the Anglocentric perspective on emotion (Wierzbicka, 2009). Instead, we should be

[5]https://github.com/google-research/bert/blob/master/multilingual.md

aware of the Anglocentric bias that multilingual models have, and find out how language-specific emotion verbalization and conceptualization affect multilingual emotion detection. Maybe, language-specific information is or can be employed by multilingual models, but at this point, our knowledge about that is too limited. I therefore see a promising research line in using NLP to investigate how emotion verbalization exactly differs across languages, and how emotion detection models deal with such language-specific information.

Although the first question is mainly a question for psychology, computational methods can help to solve this puzzle, as illustrated by the work of Jackson et al. (2019), who performed a network analysis on emotion words in 2,474 languages to get more insight in how emotion concepts vary across languages, and Markov et al. (2018), who found evidence that emotional features depend on someone's native language by analyzing over 1,000 essays written in English by non-native speakers.

However, to investigate emotion verbalization across languages, and to create models that can deal with those differences, the prime concern is to create valid data, that is, data that is originally created in that language and annotated by native speakers. Such annotations (and thus label sets) should be adjusted to the target language, which is impossible if we keep on using English words and theories like Ekman's set of basic emotions.

Nonetheless, I do see the perks in using annotations that are comparable across languages. I therefore want to break a lance for using dimensional emotion representations like the circumplex model of affect (Russell, 1980; Barrett, 2017), instead of emotion categories. The circumplex model consists of two axes, namely *pleasure* and *arousal*, representing *core affect*. Core affect feelings lie at the heart of emotional episodes, which makes that specific emotion words are associated with specific states of core affect. The English emotion concept *anger*, for example, is prototypically associated with low valence and high arousal, but translations of emotion words might have other associations with core affect in other cultural contexts. Therefore, they can be a compelling approach in comparing emotional states (and performances of emotion detection models) across languages. In the work of Preoţiuc-Pietro et al. (2016), Buechel and Hahn (2016) and De Bruyne et al. (2021), such dimensional emotion representations were already successfully used in the context of emotion detection. In line with the ambition of making emotion detection more language inclusive, I therefore believe that combining core affect with language-specific emotion labels might be the way ahead.

5 Conclusion

In this position paper, I addressed a known issue in NLP, namely the dominance of English. I discussed this issue in the light of automatic emotion detection and argued that this dominance is not limited to the small number of papers that includes other languages than English, but is also reflected in the way current datasets and models are used. It is thus not enough to encourage research on other languages than English, but to address these languages the right way, without assuming that emotions are conceptualized and verbalized in a universal way. Therefore, it is crucial to create valid data, i.e., original data from the target language (not translated) and annotated by native speakers. Label sets should be adjusted to the target language, using native emotion words and preferably combined with labels for core affect.

Moreover, I see much potential in NLP to better understand how emotional language use differs across languages. That information can subsequently help to reveal how current multilingual models deal with such differences, or even to make them more language inclusive.

Limitations

As this is a position paper, I mainly provide thoughts here, and do not include any experiments or actions myself.

Although the goal of this paper is to combat biases in AER, it is limited to discussing the dominance of English. Other biases, like the bias towards social media texts, or the tendency to ignore neurodiversity and conditions like alexithymia and autism spectrum disorder, are not addressed in this paper.

The counting study I performed to demonstrate that the number of papers dealing with other languages than English does not increase – contrary to the number of languages that are addressed, which does show an upward trend – is only based on papers presented at the Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. Maybe other patterns could be discovered when analyzing the papers of other venues.

References

Lisa Feldman Barrett. 2006. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review*, 10(1):20–46.

Lisa Feldman Barrett. 2017. The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1):1–23.

Federico Bianchi, Debora Nozza, and Dirk Hovy. 2022. XLM-EMO: Multilingual emotion prediction in social media text. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 195–203, Dublin, Ireland. Association for Computational Linguistics.

Sven Buechel and Udo Hahn. 2016. Emotion analysis as a regression problem – Dimensional models and their implications on emotion representation and metrical evaluation. *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, pages 1114–1122.

Catherine Caldwell-Harris, Ann Kronrod, and Joyce Yang. 2013. Do more, say less: Saying "I love you" in Chinese and American cultures. *Intercultural Pragmatics*, 10(1):41–69.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Luna De Bruyne, Orphee De Clercq, and Veronique Hoste. 2021. Emotional RobBERT and insensitive BERTje: Combining transformers and affect lexica for Dutch emotion detection. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 257–263, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Paul Ekman and Wallace V Friesen. 1971. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124.

Paul Ekman, E Richard Sorenson, and Wallace V Friesen. 1969. Pan-cultural elements in facial displays of emotion. *Science*, 164(3875):86–88.

Maria Gendron, Debi Roberson, Jacoba Marietta van der Vyver, and Lisa Feldman Barrett. 2014. Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, 14(2):251.

Joshua Conrad Jackson, Joseph Watts, Teague R Henry, Johann-Mattis List, Robert Forkel, Peter J Mucha, Simon J Greenhill, Russell D Gray, and Kristen A. Lindquist. 2019. Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472):1517–1522.

Mayank Kejriwal, Qile Wang, Hongyu Li, and Lu Wang. 2021. An empirical study of emoji usage on Twitter in linguistic and national contexts. *Online Social Networks and Media*, 24:100149.

Sotiris Lamprinidis, Federico Bianchi, Daniel Hardt, and Dirk Hovy. 2021. Universal joy a data set and results for classifying emotions across languages. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 62–75, Online. Association for Computational Linguistics.

Robert I Levy. 1984. The emotions in comparative perspective. *Approaches to emotion*, pages 397–412.

Ilia Markov, Vivi Nastase, Carlo Strapparava, and Grigori Sidorov. 2018. The role of emotions in native language identification. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 123–129, Brussels, Belgium. Association for Computational Linguistics.

Batja Mesquita, Michael Boiger, and Jozefien De Leersnyder. 2016. The cultural construction of emotions. *Current Opinion in Psychology*, 8:31–36. Culture.

Batja Mesquita, Nico H Frijda, and Klaus R Scherer. 1997. Culture and emotion. *Handbook of cross-cultural psychology*, 2:255–297.

Saif Mohammad. 2022. Ethics sheets for AI tasks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8368–8379, Dublin, Ireland. Association for Computational Linguistics.

Emily Öhman, Marc Pàmies, Kaisla Kajava, and Jörg Tiedemann. 2020. XED: A multilingual dataset for sentiment analysis and emotion detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6542–6552, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Aneta Pavlenko. 2008. Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and cognition*, 11(2):147–164.

Daniel Preoţiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling valence and arousal in Facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 9–15, San Diego, California. Association for Computational Linguistics.

Krzysztof Rajda, Lukasz Augustyniak, Piotr Gramacki, Marcin Gruza, Szymon Woźniak, and Tomasz Kajdanowicz. 2022. Assessment of massively multilingual sentiment classifiers. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 125–140, Dublin, Ireland. Association for Computational Linguistics.

Elisabeth Rudolph. 1990. Portuguese diminutives as special indicators of emotions. *Grazer Linguistische Studien*.

James A Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678, Florence, Italy. Association for Computational Linguistics.

Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2021. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics*, 9:845–874.

Gün R Semin, Carien A Görts, Sharda Nandram, and Astrid Semin-Goossens. 2002. Cultural perspectives on the linguistic representation of emotion and emotion events. *Cognition & Emotion*, 16(1):11–28.

Anders Søgaard. 2022. Should we ban English NLP for a year? In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5254–5260, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Enrica Troiano, Roman Klinger, and Sebastian Padó. 2020. Lost in back-translation: Emotion preservation in neural machine translation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4340–4354, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Yukiko Uchida, Sarah S. M. Townsend, Hazel Rose Markus, and Hilary B. Bergsieker. 2009. Emotions as within or between people? Cultural variation in lay theories of emotion expression and inference. *Personality and Social Psychology Bulletin*, 35(11):1427–1439.

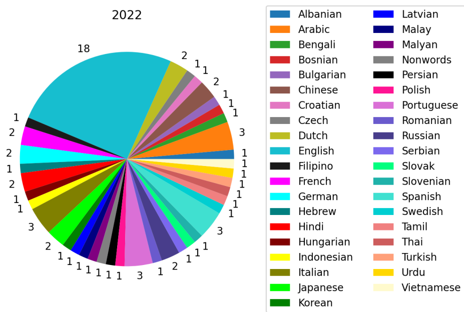
Anna Wierzbicka. 1998. "Sadness" and "anger" in Russian: The non-universality of the so-called "basic human emotions". In Angeliki Athanasiadou and Elzbieta Tabakowska, editors, *Speaking of Emotions*, pages 3–28. De Gruyter Mouton, Berlin, New York.

Anna Wierzbicka. 1999. *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press.

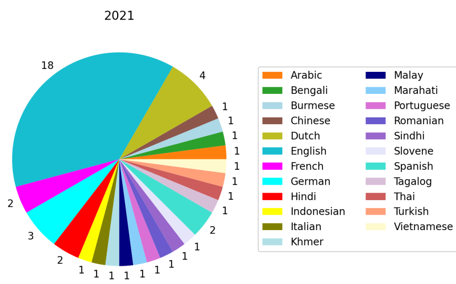
Anna Wierzbicka. 2009. Overcoming anglocentrism in emotion research. *Emotion Review*, 1(1):21–23.

Appendix

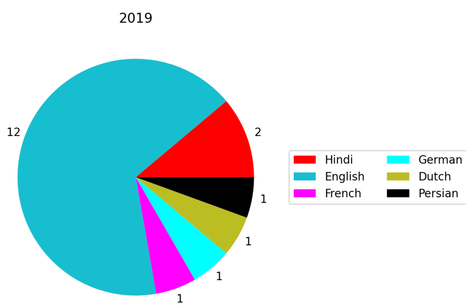
In the following pie charts, the distribution of languages treated in the WASSA editions between 2011 and 2022 are shown. The papers used for obtaining these distributions are the same as the papers used in Figure 2.



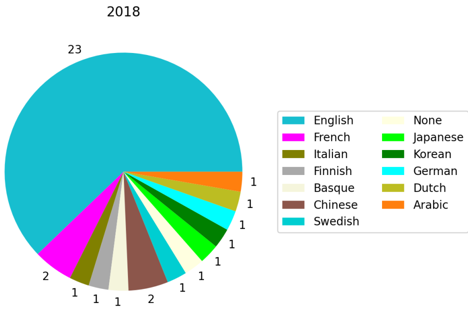
Languages included in the papers at WASSA-2022.



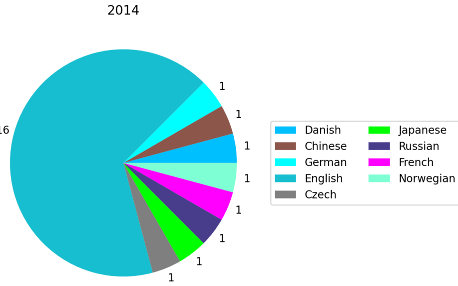
Languages included in the papers at WASSA-2021.



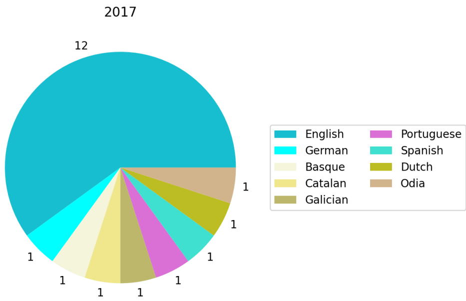
Languages included in the papers at WASSA-2019.



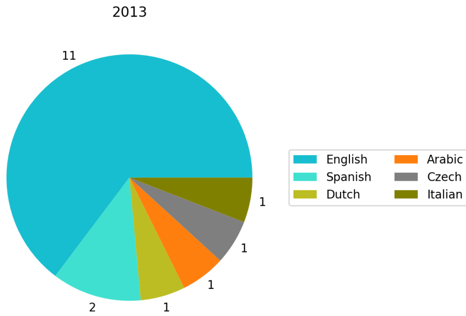
Languages included in the papers at WASSA-2018.



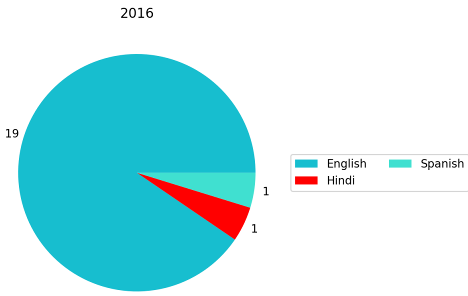
Languages included in the papers at WASSA-2014.



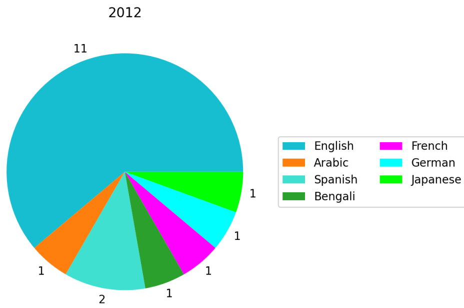
Languages included in the papers at WASSA-2017.



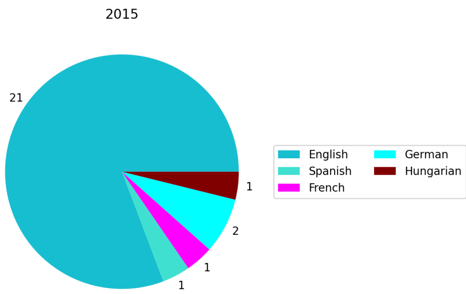
Languages included in the papers at WASSA-2013.



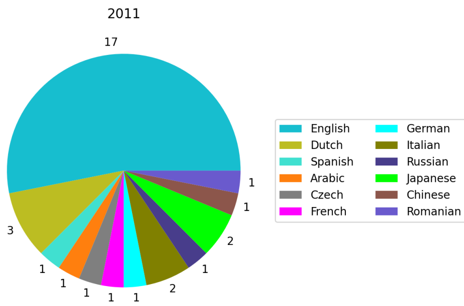
Languages included in the papers at WASSA-2016.



Languages included in the papers at WASSA-2012.



Languages included in the papers at WASSA-2015.



Languages included in the papers at WASSA-2011.

Figure 4: Distribution of languages in the WASSA contribitions between 2011 and 2022.