Grumpiness ambivalently relates to negative and positive emotions in ironic Austrian German text data

Andreas Baumann[1,♠] and **Nicole Bausch**[1,♢] and **Juliane Benson**[1,♢] and **Sarah Bloos**[1,♢]
and **Nikoletta Jablonczay**[2,♠] and **Thomas Kirchmair**[2,♠] and **Emilie Sitter**[1,♢]
[1]University of Vienna, Faculty of Philological and Cultural Studies
[2]University of Vienna, Faculty of Social Sciences
♠{andreas.baumann,nikoletta.jablonczay, thomas.kirchmair}@univie.ac.at
♢{a01615138,a12240767,a12110417,a01648097}@unet.univie.ac.at

Abstract

We present a quantitative analysis of grumpiness as expressed in Austrian German text data. Based on a sample of annotated texts, we examine to what extent grumpiness relates to emotional properties and stylistic features. We show that grumpiness is mostly related to emotional configurations characteristic of anger but that grumpiness can alternatively signal positive emotions in ironic contexts.

1 Introduction

Grumpiness is one of the notorious characteristics of Austrian culture. With far-reaching consequences: Vienna dropped to the final position[1] in the category 'friendliness' in a recent expat city ranking.[2] The issue with grumpiness is, however, more intricate than one would think. In linguistics and cultural studies, grumpiness was shown to be vaguely related to verbal aggression (Havryliv, 2017) and even thought to be associated with positive characteristics like sense of humor. Grumpiness is seen as a kind of charm, adding to the city's unique character and identity (Creath, 1995; Chen and Wu, 2019).

Despite its socio-cultural relevance, research on the topic lacks a systematic and quantitative assessment of which emotions Austrian grumpiness actually relates to. In this contribution, we conduct a statistical analysis of emotional and stylistic associates of grumpiness. Our analysis is based on a sample of texts written in Austrian German that were annotated and enriched with respect to various emotional and stylistic properties. We demonstrate that grumpiness results from a complex interaction of emotional features and irony, and that grumpiness does not exclusively signal negative emotions.

[1]https://www.derstandard.at/story/2000141285183/
[2]https://www.internations.org/expat-insider/

2 Background

According to research in cognitive psychology, grumpiness is an emotional state that is often associated with dissatisfaction, annoyance, bad temper, and irritation (Barker et al., 2020; Brosschot et al., 2010; Dietvorst et al., 2021). As such, grumpiness can be a temporary state of mind, caused by factors such as lack of sleep, stress, or physical discomfort (Deonna and Teroni, 2009), or it can be a more persistent aspect of someone's personality.

Dimensional models of emotion allow for a characterization of emotional states along several axes, most often valence (ranging from negative to positive), arousal (ranging from calm to aroused), and dominance (ranging from submissive to dominant) (Russell, 1980; Calvo and Mac Kim, 2013), often referred to as VAD model.

Considering grumpiness from the perspective of the VAD model, the emotional state is considered likelier to be negative, because it is associated with unpleasant experiences. Grumpy people tend to focus on the negative aspects of their experiences and may have difficulties finding pleasure or enjoyment in everyday activities (Watson and Clark, 1984). In terms of arousal, the judgement is less clear. Grumpy people may feel tired or sluggish and less motivated or interested in their surroundings. In an experiment on facial expressions, grumpiness was shown to be associated with relatively low arousal (Barker et al., 2020). However, they may also experience moments of increased arousal, e.g., when they become agitated or frustrated by a particular situation (Dietvorst et al., 2021).

As far as dominance is concerned, grumpiness could be potentially associated with a sense of powerlessness or frustration, and hence submissive emotions (Leach and Weick, 2018). On the other hand, grumpiness is related to anger, which is characterized by low valence, high arousal, and high dominance (Calvo and Mac Kim, 2013). Thus, it would be interesting to see where exactly grumpiness is located in the VAD space.

How emotional states like grumpiness are intertwined with texts like poetry, literature or, more recently, the vast amount of text data produced on social media has become a field of interdisciplinary interest. For this purpose, also data science and the digital humanities are constantly working on new modelling techniques mainly using techniques from NLP like keyword detection or lexica to predictive modelling, there has been a shift to more sophisticated, state-of-the-art neural networks.

What they all share is the search for the best combination of stylistic, structural and semantic features to determine the emotions or 'tone' of interest. The solution depends mainly on the data and goal. For the detection of ironic comments for example, besides using standard features like word count or PoS distributions (Alm et al., 2005), it has proven useful to include interjections, punctuation, capitalization, use of first-person pronouns, repetitions, negations or even labelled emoticons as features (Ortega-Bueno et al., 2018). It was also indicated by Reyes et al. (2012) that special linguistic features like morphosyntactic ambiguity -– linked with lesser syntactic complexity -– are useful for inferring irony as well. This is relevant because irony and grumpiness show a distinct connection: irony is often used to soften an angry remark or criticism, with the speaker appearing to be more in control (Dews et al., 1995).

Diving deeper into the matter, Van Hee (2017) shows that lexical features like character and punctuation flooding in tweets (e.g. in words like 'Loooove') outperformed word n-grams in irony detection next to structural and sentiment features like tags, valence or polarity scores. Nonetheless, the best results were yielded when combining all three feature-sets. The author concludes that certain features suit certain 'types' of irony.

The addition of stylistic features in general does statistically improve the overall performance of emotion detection models (Malheiro et al., 2016) but they do not seem to work equally well alone, and they don't have an effect as high as semantic features. Hence, it makes sense to take stylistic features into account when investigating grumpiness manifested in text data.

3 Data

3.1 Annotation

We based our analysis on the Million Posts corpus (Schabus et al., 2017). It consists of postings taken from the user forum of the Austrian news website http://derstandard.at. Texts represent a sample of the Austrian variety of German. This user forum is a suitable resource for studying grumpiness as it accomodates a large population of users with diverse political views (mostly excluding strong right-wing attitudes) so that topics are typically discussed vividly and emotionally (note, though that the forum is moderated, hence hate-postings do not get published if they are detected). About 3500 of the texts in the corpus have been already labeled with respect to sentiment (pos/neu/neg; three categorical labels per posting). We computed average sentiment ratings for each posting and found that only 69 texts in the data set show a positive sentiment. To create a balanced sample, we sampled a roughly equal amount of neutral and negative texts and ended up with a stratified sample of 200 texts in total.

Subsequently, texts were annotated with respect to five characteristics: arousal, dominance, abstractness, irony, and grumpiness. Annotators were asked to judge the texts with respect to these characteristics based on a five-point Likert scale. All texts were labeled by three annotators each. All annotators (some of which are authors of this paper) were students speaking German as their first language (they received course credit and no monetary compensation for their labeling efforts). Annotators were provided with the parent posting (if it existed) and the title of the news article postings related to as additional context.

We computed Chronbach's α to assess inter-annotator agreement. Apart from abstractness with $\alpha = 0.31$, inter-annotator agreement was sufficiently high[3] (arousal: $\alpha = 0.67$; dominance: $\alpha = 0.69$; irony: $\alpha = 0.78$; grumpiness: $\alpha = 0.75$) and comparable with the quality of the ratings in the Million Posts Coprus (Schabus et al., 2017). Notably, the relatively high inter-annotator agreement for grumpiness was reassuring for our study (see Figure 1).

[3] Values of Chronbach's α greater than 0.8 are considered to be good, values between 0.7 and 0.8 are considered to be acceptable, and values below 0.5 are interpreted as unacceptable (Li et al., 2016; Streiner, 2003).

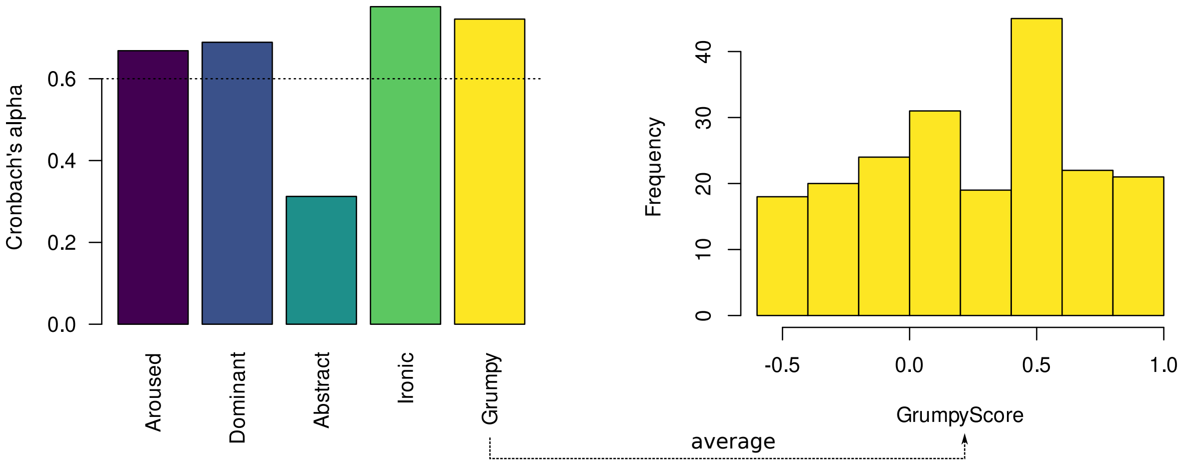

Figure 1: Left: inter-annotator agreement (Chronbach's α) for all annotated features (acceptability threshold shown by dashed line). For each text, GrumpyScore was computed as average of all grumpiness ratings. Right: distribution of GrumpyScore in the sample of 200 texts.

3.2 Emotional features

In a next step, the average of all annotator ratings was computed for each characteristic and each text to obtain overall scores (ArousedScore, DominantScore, AbstractScore, IronicScore, GrumpyScore). All scores including sentiment taken from the Million Posts Corpus (SentiScore) were subsequently scaled to the interval $[-1, 1]$ in such a way that 0 corresponds to a neutral score. The histogram in Figure 1 shows that GrumpyScore is fairly equally distributed across the interval $[-0.5, 1.0]$. That is, the texts in the sample were classified as rather grumpy on average (despite the sample being balanced with respect for sentiment).

3.3 Stylistic features

In order to capture potential stylistic correlates of grumpiness, we derived a range of linguistic variables. First, we used the Flair PoS tagger to compute the fraction of Nouns, Verbs and Adjectives for each text. Second, we counted the number of Colons, Periods, ExlamationMarks, and QuestionMarks, as well as the number of happy (:)), sad (:(or :/), and blinking (;)) emoticons (HappyEmoticon, SadEmoticon, BlinkEmoticon, respectively). Finally, we retrieved TextLength measured as the number of characters, as well as TypeTokenRatio to include a proxy for lexical diversity.

4 Analysis

4.1 Emotional and stylistic features

What is the relative impact of emotional and stylistic features on grumpiness? To shed light on this question, we first computed a linear (Gaussian) regression model in which GrumpyScore depends on all other 18 features described in the previous section. We used the per text computed reciprocal of the standard deviation of the grumpiness ratings as weights in the model, so that texts with a more accurate GrumpyScore are weighted higher in the model. The resulting model shows a reasonably high goodness of fit at $R^2 = 0.68$ (and a fairly symmetric residual distribution), indicating that grumpiness is characterized well by the emotional and stylistic features at hand.

Since much information in the data about the outcome is shared among the 18 predictors, we employed AIC-driven top-down model nesting to optimize the previously computed linear model. The resulting model (which scores the lowest AIC and $R^2 = 0.67$) features eight predictors, five of which show statistically non-trivial effects: grumpiness is associated with high arousal, high dominance, negative sentiment, and, to a lesser extent, irony. Thus, the linear model suggests grumpiness to be associated with anger (which is itself characterized by high arousal, high dominance and low valence). Interestingly, the number of verbs shows a particularly strong positive impact on the outcome variable. See Table 1 for a breakdown.

To get insights into the ranking of the predictors, we computed relative variable importance based on the AIC scores of all sub-models of the maximal model featuring 18 predictors (Burnham and Anderson, 2004). More specifically, we derived Akaike weights for all sub-models and, for each predictor, computed relative variable importance as

Predictor	Coef.	SE	t value
(Intercept)	0.32	0.33	0.95
ArousedScore	**0.36**	0.06	5.56
DominantScore	**0.45**	0.08	5.55
IronicScore	**0.11**	0.04	2.64
TypeTokenRatio	-0.62	0.35	-1.77
SentiScore	**-0.28**	0.04	-7.29
SadEmoticon	0.14	0.10	1.40
Adjectives	0.32	0.19	1.74
Verbs	**0.76**	0.21	3.56

Table 1: Effects on GrumpyScore in the optimal linear model. Bold indicates statistically non-trivial effects at a 5% significance level.

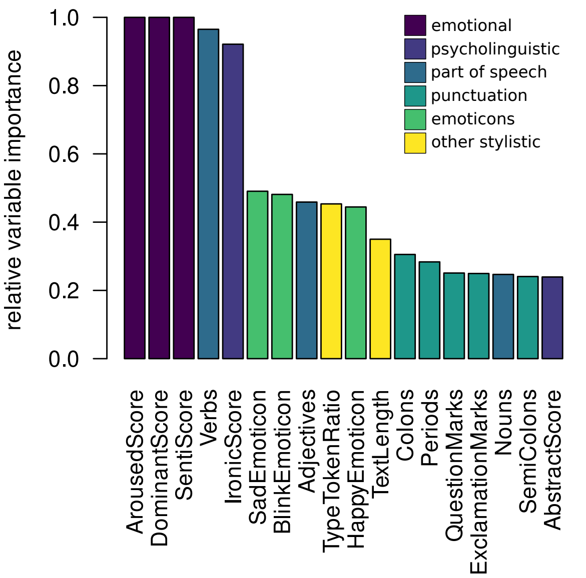

Figure 2: Relative variable importance based on multiple linear regression models.

the sum of Akaike weights of all models in which that predictor is present. The ranking is shown in Figure 2. There seem to be three different groups: emotional features and the number of verbs are most important for inferring grumpiness. The predictors in the second group are only roughly half as important. Interestingly, the group shows all emoticon counts. The remaining predictors (mostly punctuation, but also the number of nouns) display the lowest relevance for inferring grumpiness.

4.2 Grumpiness in the VAD space

The significant effects of all emotional predictors in the model make clear that grumpiness is unsurprisingly associated with specific emotional aspects. To explore the location of grumpiness in the emotional space spanned by valence, arousal, and dominance, we used generalized additive models (GAM) (Wood, 2006). Here, GrumpyScore is predicted by three interacting variables SentiScore, ArousedScore, DominantScore). The interaction was implemented as a smooth tensor-product term (number of knots $k = 5$). Due to the distribution of GrumpyScore (Figure 1, right), we used a Gaussian link function. Again, reciprocal standard deviations of GrumpyScore were used as weights like in the linear model.

The model is visualized in the upper panel of Figure 3. It displays the valence-arousal space for four different dominance bins. Light colors (yellow) indicate a stronger association with grumpiness than dark colors (purple). It can be seen that grumpiness increases with dominance (in line with the linear model), and that grumpiness is associated with high arousal and low valence, i.e., it is co-located with emotional categories like anger. This particularly holds true for submissive scenarios but is weakened as dominance increases. High dominance apparently allows for a slightly more positive association with grumpiness.

4.3 Interaction with irony

In the linear model, the significant effect of irony is particularly interesting. We computed a second GAM, but this time GrumpyScore was predicted by SentiScore, ArousedScore, and IronicScore in order to assess the effect of irony of the location of grumpiness in the valence-arousal space. The result is shown in the lower panel of Figure 3. In line with the linear model, the effect of irony is weaker than that of dominance (overall, the plotted surface does not become substantially lighter).

Interestingly, if irony is low (first plot) grumpiness is relatively strictly confined to the negative and aroused region of the emotional space. However, if irony scores high, there are relatively high associations of grumpiness with negative *and* positive regions, while (valence-wise) neutral regions show diminished grumpiness. This indicates that grumpiness is highly ambivalent in ironic settings: grumpiness could either correspond to angry contexts but also to joyful ones (but not to indifferent contexts).

5 Discussion and conclusion

In this paper, we presented a quantitative analysis of linguistically represented grumpiness based on a sample of texts that were annotated for various

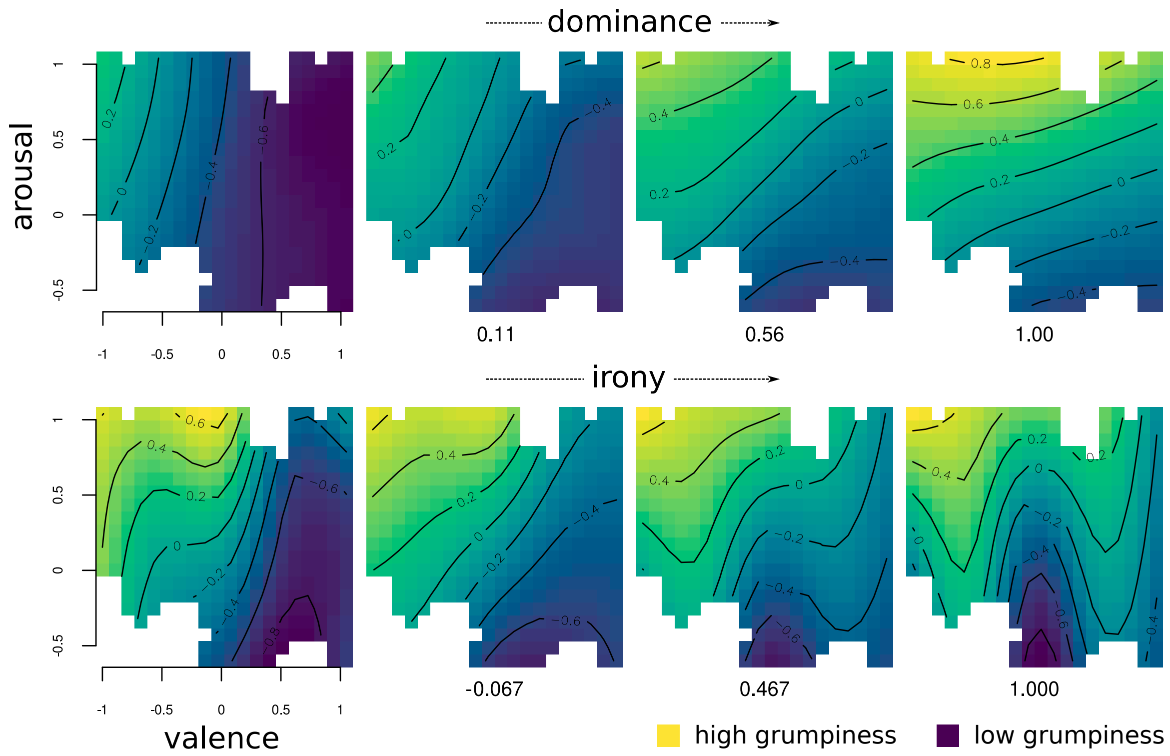

Figure 3: Valence-arousal spaces modulated by dominance (upper panel) and irony (lower panel), based on GAMs.

emotional properties and enriched with stylistic information. The main result of our analysis is that, overall, grumpiness is associated with anger. However, this clear association does not hold in ironic contexts. Here, grumpiness can relate to either negative or positive emotions. Interestingly, this means that knowledge of whether or not a text is ironic (i.e., a certain sensitivity with respect to irony) does not suffice to categorize grumpy utterances. Individuals require additional information to decode the emotional state underlying a grumpy utterance.

This result is in line with the observation that Austrian grumpiness can signal humor as well. (Creath, 1995; Chen and Wu, 2019; Havryliv, 2017). Whether or not this intricate relationship between emotion, grumpiness, and irony is responsible for the fact that Viennese people tend to be perceived as unfriendly as suggested by surveys among expats (see footnote 1 and 2), still needs to be looked at more closely.

Another result of our modeling analysis is that grumpiness seems to be associated with an extensive usage of verbs (as opposed to nouns and adjectives). Given that verbs are typically less concrete than other lexical categories, this result seems surprising at first sight. Nominal style is typical of less aroused genres like legal or scientific texts, while verbal style is generally represented more strongly in everyday speech (Radovanovic, 2001). Either way, the results point at the relevance of stylistic cues when inferring emotional states from text.

It is evident that our study is subject to limitations. For one, the number of texts as well as the number of annotations per text is not large. However, inter-annotator agreement was sufficiently high (in particular as far as grumpiness is concerned) and the fact that our models show statistically robust effects, high goodness of fit, and relatively small standard errors despite the small sample size is reassuring. In addition to a larger number of texts (and annotators), potential follow-up studies would need to take different genres into account. Clearly, considering spoken corpora would be most relevant in this regard (however, forum postings represent an already relatively informal genre).

Finally, it would be interesting to see to what extent grumpiness ratings from raters with different social, linguistic, or geographic backgrounds deviate from each other. This would help to shed light on how linguistically expressed grumpiness is perceived cross-culturally.

Supplementary materials

The analysis can be reproduced in the following project on Posit Cloud: https://posit.cloud/content/5527995. The processed data set of all aggregated scores our analysis is based on is available at https://phaidra.univie.ac.at/o:1634249. A supplementary analysis involving several emotional lexica can be found here: https://phaidra.univie.ac.at/o:1634258.

Acknowledgements

We would like to thank Jennie Atwell, Ines Konnerth, Maximilian Moser, Marja Nikolcic, Sarah Sulollari, Lale Tüver, and Sebastian Wegerer for additional help with data annotation.

References

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. In *Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 579–586.

Megan S Barker, Emma M Bidstrup, Gail A Robinson, and Nicole L Nelson. 2020. “grumpy” or “furious”? arousal of emotion labels influences judgments of facial expressions. *Plos one*, 15(7):e0235390.

Jos F Brosschot, Bart Verkuil, and Julian F Thayer. 2010. Conscious and unconscious perseverative cognition: is a large part of prolonged physiological activity due to unconscious stress? *Journal of Psychosomatic Research*, 69(4):407–416.

Kenneth P Burnham and David R Anderson. 2004. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Rafael A Calvo and Sunghwan Mac Kim. 2013. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3):527–543.

Yeong-Shyang Chen and Shou-Tsung Wu. 2019. Social networking practices of viennese coffeehouse culture and intangible heritage tourism. *Journal of Tourism and Cultural Change*, 17(2):186–207.

Richard Creath. 1995. From königsberg to vienna: Coffa on the rise of modern semantics. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie*, 34(1):113–118.

Julien A Deonna and Fabrice Teroni. 2009. Taking affective explanations to heart. *Social Science Information*, 48(3):359–377.

Shelly Dews, Joan Kaplan, and Ellen Winner. 1995. Why not say it directly? the social functions of irony. *Discourse Processes*, 19(3):347–367.

Evelien Dietvorst, Marieke Hiemstra, Dominique Maciejewski, Eeske van Roekel, Tom Ter Bogt, Manon Hillegers, and Loes Keijsers. 2021. Grumpy or depressed? disentangling typically developing adolescent mood from prodromal depression using experience sampling methods. *Journal of Adolescence*, 88:25–35.

Oksana Havryliv. 2017. Verbale Aggression: das Spektrum der Funktionen. *Linguistik Online*, 82(3).

Stefan Leach and Mario Weick. 2018. From grumpy to cheerful (and back): How power impacts mood in and across different contexts. *Journal of Experimental Social Psychology*, 79:107–114.

Junyi Jessy Li, Bridget O’Daniel, Yi Wu, Wenli Zhao, and Ani Nenkova. 2016. Improving the annotation of sentence specificity. In *Proc. of the 10th International Conference on Language Resources and Evaluation*, pages 3921–3927, Portorož, Slovenia.

Ricardo Malheiro, Renato Panda, Paulo Gomes, and Rui Pedro Paiva. 2016. Emotionally-relevant features for classification and regression of music lyrics. *IEEE Transactions on Affective Computing*, 9(2):240–254.

Reynier Ortega-Bueno, Carlos E Muniz-Cuza, José E Medina Pagola, and Paolo Rosso. 2018. Uo upv: Deep linguistic humor detection in spanish social media. In *Proc. of the 3rd Workshop on Evaluation of Human Language Technologies for Iberian languages*, pages 204–213.

Milorad Radovanovic. 2001. On nomimal and verbal style: cultures or languages in contact? *International Journal of the Society of Language*, 151(3):41–48.

Antonio Reyes, Paolo Rosso, and Davide Buscaldi. 2012. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12.

James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178.

Dietmar Schabus, Marcin Skowron, and Martin Trapp. 2017. One million posts: A data set of german online discussions. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1241–1244.

David L. Streiner. 2003. Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80:99–103.

Cynthia Van Hee. 2017. *Can machines sense irony?: exploring automatic irony detection on social media*. Ph.D. thesis, Ghent University.

David Watson and Lee A Clark. 1984. Negative affectivity: the disposition to experience aversive emotional states. *Psychological Bulletin*, 96(3):465.

Simon N Wood. 2006. *Generalized additive models: an introduction with R*. CRC.