Video Games as a Corpus: Sentiment Analysis using Fallout New Vegas Dialog

Mika Hämäläinen, Khalid Alnajjar, Thierry Poibeau

Video Games as a Corpus: Sentiment Analysis using Fallout New Vegas Dialog

Mika Hämäläinen
mika.hamalainen@helsinki.fi
University of Helsinki
Helsinki, Finland

Khalid Alnajjar
khalid.alnajjar@helsinki.fi
University of Helsinki
Helsinki, Finland

Thierry Poibeau
thierry.poibeau@ens.psl.eu
LATTICE, École Normale Supérieure
Paris, France

ABSTRACT

We present a method for extracting a multilingual sentiment annotated dialog data set from Fallout New Vegas. The game developers have preannotated every line of dialog in the game in one of the 8 different sentiments: *anger, disgust, fear, happy, neutral, pained, sad* and *surprised*. The game has been translated into English, Spanish, German, French and Italian. We conduct experiments on multilingual, multilabel sentiment analysis on the extracted data set using multilingual BERT, XLMRoBERTa and language specific BERT models. In our experiments, multilingual BERT outperformed XLMRoBERTa for most of the languages, also language specific models were slightly better than multilingual BERT for most of the languages. The best overall accuracy was 54% and it was achieved by using multilingual BERT on Spanish data. The extracted data set presents a challenging task for sentiment analysis. We have released the data, including the testing and training splits, openly on Zenodo. The data set has been shuffled for copyright reasons.

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**; **Natural language generation**; *Neural networks.*

KEYWORDS

sentiment analysis, video games as corpus, multilinguality

ACM Reference Format:
Mika Hämäläinen, Khalid Alnajjar, and Thierry Poibeau. 2022. Video Games as a Corpus: Sentiment Analysis using Fallout New Vegas Dialog. In *FDG '22: Proceedings of the 17th International Conference on the Foundations of Digital Games (FDG '22), September 5–8, 2022, Athens, Greece.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3555858.3555930

1 INTRODUCTION

Multilabel sentiment analysis is a challenging NLP task. Sentiment analysis is often conducted in a simplified positive-negative scale in the field of NLP with highly successful results [1, 2, 14, 15]. What still remains difficult is when a model needs to predict a more nuanced sentiment label such as sadness, happiness or anger. In this paper, we explore the latter case where sentiment is not reduced into one axis but rather annotated based on 8 different emotions.

FDG '22, September 5–8, 2022, Athens, Greece

ACM ISBN 978-1-4503-9795-7/22/09.
https://doi.org/10.1145/3555858.3555930

Role playing games (RPGs) contain large amounts of dialog that could be used in a variety of NLP tasks. However, thus far, the resources RPGs have have been mostly gone unnoticed in the NLP research community. The video game we base our research on in this paper, Fallout New Vegas[1], contains a very interesting annotated dataset. The Steam version of the game is available in English, Spanish, French, Italian and German and each line of dialogue in the game has been annotated for sentiment and its intensity.

Fallout New Vegas is an open-world RPG that is set in a post-apocalyptic world after a nuclear attack. The game mixes RPG and first person shooter genres. The game has several side quests and many non-playable characters (NPCs) with long prescripted dialog containing multiple dialog options. All in all, the game contains over 53,000 lines of dialog making it an ideal corpus for a variety of NLP tasks.

In the current work, we describe how we extracted the preannotated dialog from Fallout New Vegas and we present several BERT-based models for sentiment analysis. We have released the extracted dataset with train-test splits openly on Zenodo[2]. The data is of a high value because it represents a professionally annotated data source in 5 languages and in a completely new domain of text than the existing sentiment analysis corpora. We also showcase that video games are an extremely useful data source for NLP and their use as a corpus should be more widely studied. This research opens up a possibility to better understand and model emotion in dialog with computational methods.

2 RELATED WORK

In this section, we will focus on some of the research related to using video games as a resource for NLP tasks and also the latest advances in multilabel sentiment analysis.

Video games have been used as a corpus before, for example, a recent paper extracts a sentiment lexicon from Skyrim [3]. Their approach is different from us in the sense that their focus was to extract a lexicon, where as our approach extracts sentiment annotated sentences to be used as training data for a machine learning model. Sentiment lexica play a very different role in sentiment analysis than machine learning approaches [17]. Fallout 4 dialog has been used before to as data for a dialog adaptation model [6]. The paper presents a machine learning approach for adapting existing dialog in Fallout 4 to better match the condition the game character is in. Also, video game data has been used to detect persuasion [10].

Multilabel sentiment analysis is a fragmented field of research in the sense that several different approaches deal with different number of sentiment labels and the labels themselves can be higher level

[1] https://fallout.bethesda.net/en/games/fallout-new-vegas
[2] https://zenodo.org/record/6990638

file name	VDialogueA_VDialogueArcade_00163394_2
sentiment	Anger
English	I'm just saying that if it were to fall into Lake Mead and be irreparably damaged... and if you threw an EMP grenade in after it...
Spanish	Solo digo que si se cayera al lago Mead y sufriera daños irreparables... y luego tú le tiraras una granada...
French	Je dis juste que s'il devait "tomber" dans le Lac Mead et souffrir de pannes irréversibles... et qu'en plus vous lui lanciez une grenade IEM...
German	Was ich sagen will, ist, dass er ja durchaus in den Lake Meade "fallen" könnte ... und wenn Sie eine EMP-Granate hinterherwerfen würden ...
Italian	Dico solo che se dovesse cadere in Lake Mead e rimanere irreparabilmente danneggiato... e se gli gettassi dietro anche una granata IEM...

Table 1: An aligned example of one sentence in all the 5 languages

emotions (like in our case) or lower level affects. The term multilabel itself can also be understood differently, for instance there is research that calls identifying multiple sentiments (postive-negative) within a sentence multilabel sentiment analysis [11]. However, for us, multilabel means that there are more than the typical positve-negative axis to be considered in sentiment analysis.

A recent paper demonstrates multilabel sentiment analysis on 100 languages [16]. The authors use RoBERTa-XLM to extract feature vectors. These are then used in training a bi-directional LSTM based classifier model. Another line of work [7] compares several different multilabel classification methods on the task of sentiment analysis showing that RAkEL [12] gave the best performance on raw token input. They show promising results using both of the models but conclude that multilabel sentiment analysis is far from an easy task using machine learning methods.

3 EXTRACTING THE DATA

Fallout New Vegas stores data in a proprietary binary format that cannot be parsed easily without specialized tools. Fortunately, the developers have released an official modding tool called Garden of Eden Creation Kit (GECK)[3] that allows us to extract all in-game dialog. We use the version of the game that is distributed through Steam[4]. GECK outputs a TSV file out of which we save the text, file name and sentiment. We use the file name information to align sentences in different languages with each other. And example of the resulting data in the different languages can be seen in Table 1.

	Eng	Deu	Ita	Spa	Fra
Anger	3335	2407	2189	611	143
Disgust	932	679	658	206	34
Fear	1620	969	726	140	34
Happy	4029	2351	1916	375	103
Neutral	39802	29339	25664	5819	1626
Pained	994	846	780	89	12
Sad	1055	714	649	248	64
Surprised	1649	1082	930	167	59

Table 2: Number of dialog lines per language and sentiment label

The problem we ran into was that GECK is rather buggy, we tried to patch it using an unofficial tool called GECK Extender[5], but it did not patch the main bug we encountered. GECK manages to export the entire in-game dialog correctly for English, but it crashes after some time for other languages. It seems to crash consistently at the same step for a given language. For German and Italian, it extracts majority of the dialog before crashing, where as for Spanish and French, it crashes rather early in the extraction process. We also tried out some unofficial tools such as fo76utils[6] but they did not work at all with Fallout New Vegas. As a result, we have slightly different amount of data for each language as seen in Table 2.

Because the data is so small for Spanish and French, we will not use them for training. Instead, we use the sentences from the Spanish and French data for testing. We also exclude the translations of these sentences in other languages from the training and use them for testing. This way all languages share the testing data and no translation of the test sentences appears in training for any of the languages. The rest of the data for English, German and Italian is used for training. Because the Neutral label is so overwhelmingly present in the data, we limit it to 3000 samples in the training data for each language.

The data itself is relatively clean because it is dialog displayed to the end user in the video game. Some of the dialog lines include additional annotation inside of brackets: *{nervous, hiding a secret}Corporal White? I don't know where he - {obviously changing his lie mid-sentence}uh, I mean, never heard of him. Uh, I gotta go...* We removed these extra annotations with regular expressions. Apart from this, no additional preprocessing was done.

4 SENTIMENT ANALYSIS

We model the sentiment analysis task as a sequence classification task, where the model has to predict the sentiment label given a Fallout New Vegas dialog sentence. We experiment with several BERT-based [5] models and a RoBERTa-based [8] model. We experiment both in a multilingual and a monolingual scenario. We split 15% of the training data for validation for each trainable language.

4.1 Multilingual setting

In our multilingual models, the model is trained with the training data for English, Italian and German. Then the models are evaluated using the evaluation splits for English, Italian and German, and the entire data set for Spanish and French which do not have enough data for training.

[3]https://geck.bethsoft.com/index.php?title=Main_Page
[4]https://store.steampowered.com/app/22380/Fallout_New _Vegas/
[5]https://geckwiki.com/index.php/GECK_Extender
[6]https://github.com/fo76utils/fo76utils

	English						German						Italian						Spanish						French					
model	BERT			ROBERTA			BERT			ROBERTA			BERT			ROBERTA			BERT			ROBERTA			BERT			ROBERTA		
	pre	rec	f1	pre	rec	f1	pre	rec	f1	pre	rec	f1	pre	rec	f1	pre	rec	f1	pre	rec	f1	pre	rec	f1	pre	rec	f1	pre	rec	f1
Anger	0.16	0.29	0.21	0.18	0.30	0.22	0.14	0.23	0.17	0.15	0.22	0.17	0.14	0.23	0.18	0.17	0.25	0.20	0.12	0.15	0.13	0.16	0.19	0.17	0.13	0.29	0.18	0.16	0.22	0.18
Disgust	0.07	0.11	0.09	0.11	0.11	0.11	0.09	0.12	0.10	0.11	0.11	0.11	0.10	0.11	0.10	0.11	0.10	0.11	0.04	0.02	0.03	0.07	0.05	0.06	0.09	0.15	0.11	0.00	0.00	0.00
Fear	0.17	0.29	0.22	0.13	0.28	0.17	0.05	0.06	0.06	0.08	0.13	0.10	0.03	0.03	0.03	0.10	0.18	0.13	0.07	0.09	0.08	0.06	0.10	0.08	0.14	0.12	0.13	0.12	0.21	0.15
Happy	0.11	0.28	0.16	0.12	0.38	0.19	0.08	0.18	0.11	0.11	0.33	0.16	0.10	0.28	0.14	0.11	0.32	0.16	0.09	0.23	0.13	0.12	0.39	0.18	0.07	0.20	0.10	0.10	0.31	0.15
Neutral	0.82	0.59	0.69	0.82	0.57	0.67	0.80	0.63	0.70	0.81	0.61	0.69	0.81	0.62	0.70	0.81	0.62	0.70	0.79	0.67	0.72	0.80	0.63	0.70	0.83	0.58	0.68	0.82	0.64	0.72
Pained	0.08	0.08	0.08	0.08	0.07	0.07	0.04	0.03	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.03	0.04	0.04	0.03	0.04	0.03	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Sad	0.10	0.14	0.12	0.10	0.14	0.12	0.09	0.11	0.10	0.09	0.10	0.09	0.09	0.11	0.10	0.10	0.13	0.12	0.10	0.10	0.10	0.10	0.10	0.10	0.04	0.08	0.06	0.12	0.14	0.13
Surprised	0.08	0.18	0.11	0.07	0.19	0.10	0.09	0.17	0.12	0.07	0.17	0.10	0.06	0.13	0.09	0.09	0.20	0.12	0.06	0.14	0.09	0.08	0.17	0.11	0.04	0.07	0.05	0.14	0.25	0.18
Overall accuracy	**0.51**			0.50			**0.52**			0.51			**0.52**			**0.52**			**0.54**			0.52			0.49			**0.55**		

Table 3: Results of the multilingual models (precision, recall, F1-score)

We use the transformers Pyton library [13] to fine-tune the multilingual BERT [5][7] and the multilingual XLMRoBERTa [4][8] for sequence classification task. Both of the models are trained with the same data for 3 epochs with 500 warm-up steps and a weight decay of 0.01. The labels are predicted using the softmax function.

Because both of the models are multilingual they should learn to predict the sentiment even for Spanish and French which were completely held out from the training data. Because we intend to test the models on these two languages, we do not introduce any language labels to distinguish between languages during the training.

	English			German			Italian		
	pre	rec	f1	pre	rec	f1	pre	rec	f1
Anger	0.22	0.31	0.26	0.18	0.28	0.22	0.17	0.30	0.22
Disgust	0.12	0.18	0.15	0.09	0.13	0.11	0.07	0.11	0.09
Fear	0.17	0.32	0.23	0.04	0.03	0.03	0.04	0.06	0.05
Happy	0.14	0.41	0.21	0.12	0.31	0.17	0.12	0.29	0.17
Neutral	0.82	0.60	0.69	0.81	0.64	0.72	0.81	0.58	0.68
Pained	0.05	0.09	0.06	0.03	0.06	0.04	0.03	0.03	0.03
Sad	0.11	0.13	0.12	0.11	0.13	0.12	0.11	0.20	0.14
Surprised	0.10	0.25	0.14	0.08	0.17	0.11	0.08	0.18	0.11
Overall accuracy	0.52			0.54			0.50		

Table 4: Results of the language specific models

4.2 Monolingual setting

In addition to training multilingual models with the data for all available languages, we are interested in seeing whether monolingual models perform better or worse in this task than the multilingual model. Monolingual models are usually trained with more data for the particular language in question than what is present in the multilingual models. We train separate models for English, Italian and German, training and testing them only with language specific data.

For English, we use original BERT model [5][9]. For Italian, we use the Italian BERT[10] provided by the MDZ Digital Library team (dbmdz) at the Bavarian State Library. The model is trained on OPUS and OSCAR corpora. The German BERT model[11] is also provided by the same team and it has been trained on a variety of corpora such as a Wikipedia dump, EU Bookshop corpus, Open Subtitles, CommonCrawl, ParaCrawl and News Crawl. We use the same training parameters as for the multilingual models.

5 RESULTS AND EVALUATION

The results of the multilingual models can be seen in Table 3. The results are calculated using the *classification_report* method of Scikit-learn [9]. There is not a big difference between the multilingual BERT and XLMRoBERTa models. XLMRoBERTa seems to be slightly worse than the multilingual BERT, however, it has a better overall accuracy for French. All in all, the sentiments that are the most difficult ones to predict seem to be *disgust* and *pained*. Perhaps because they might depend more on the audio cues than textual ones. After the *neutral* label, *anger* and *happy* seem to be the easiest ones to predict for the model, although the overall performance is not very good.

We can see that for both of the multilingual models could learn to transfer the sentiment analysis to Spanish and French, which were not used in the training at all. The overall accuracy is comparable to that of the languages that had training data, however, the results for some labels are rather poor. For instance, for French, neither of the models predicted any *pained* sentiment sentences right.

The results for the monolingual models can be seen in Table 4. The results are slightly better for all languages than Italian when comparing to the multilingual models. However, there is no major increase in the overall accuracy either despite the models having been pretrained with more language specific data. We believe that this is due to the fact that Fallout belongs to a very different domain of text than what is represented by the training data of the BERT models.

Next, we will take a closer look at the results of the multilingual BERT model as it seemed to give better results than the XLM-RoBERTa based one. If we look at the aligned sentences that were predicted wrong for all of the languages, we can see that there are 269 such cases. On a label level, the model predicted the same sentence wrong for all languages 29 times for *surprise*, 30 for *happy*, 11 for *fear*, 46 for *anger*, 103 for *neutral*, 31 for *sad*, 8 for *pained* and 11 for *disgust* label. It is interesting that the model did not commit the same errors for all languages even though the sentences are translations of each other, and if the model was to truly capture the multilinguality of semantics, it ought to make same mistake for sentences that are each other's translations.

[7] bert-base-multilingual-cased
[8] xlm-roberta-base
[9] bert-base-uncased
[10] dbmdz/bert-base-italian-uncased
[11] bert-base-german-dbmdz-uncased

For example the following *neutral* sentence was predicted either as *sad* or *disgust* depending on the language: *Why the need for a bunch of old warhorses like us?* (*disgust*), *Woher rührt der Bedarf nach einem Haufen alter Haudegen wie uns?* (*disgust*), *Come mai il bisogno di un gruppo di vecchi veterani come noi?* (*sad*), *¿Para qué necesitas a un montón de veteranos decrépitos como nosotros?* (*sad*) and *Pourquoi faire appel à des vieux bourrins comme nous ?* (*disgust*). The difference might be explained by the translation strategy used. The Spanish and Italian sentences use the word veteran where as the other languages use horse related vocabulary *warhorse*, *Haudegen* and *bourrin*.

If we look at sentences that were predicted correctly for all of the languages, we can see 370 such cases. 363 of them are *neutral*, 4 *anger*, 2 *surprise* and 1 *happy*. This means that while the model gets the prediction right for at least one language, it seldom gets it fully right for all of the languages. For example, the following sentence was classified correctly as surprise by all of the models: *Huh? Look, man, me and Diane, we don't dig on that politics stuff, savvy? We just make the product and make it get to a good home*, *Häh? Ich und Diane, wir stehen nicht so auf Politikkram, ja? Wir stellen nur Ware her und sorgen dafür, dass sie ein ordentliches Zuhause bekommt*, *Eh? Guarda, amico, io e Diane non ci occupiamo di politica, chiaro? Realizziamo solo il prodotto e facciamo sì che arrivi a destinazione*, *¿Eh? Escucha, colega, Diane y yo no nos metemos en cosas de política, ¿entiendes? Solo fabricamos el producto y lo hacemos llegar a buen puerto* and *Hein ? Écoutez, mec, Diane et moi, on n'est pas branchés politique, d'accord ? On fabrique les produits et on essaie d'en vivre le mieux possible*. In all of the cases, the translators had retained the initial *huh?* which probably gave the model a good cue for predicting the label correctly as *surprise*.

6 CONCLUSIONS

We have presented a new multilingual data set for sentiment analysis consisting of 8 sentiment labels. The data has been extracted from Fallout New Vegas which already contained sentiment labels. In addition to the labels, the game data contained sentiment intensity scores, but we did not utilize them in this research. We have made the data publicly available on Zenodo[12]. The data has been shuffled for copyright reasons.

Multimodal classification of text remains a challenging NLP problem and our experiments on sentiment analysis are no exception. The overall accuracies of the models are rather low, but they are in line with the usual accuracies obtained in similar multilabel NLP classification tasks. Both multilingual BERT and XLMRoBERTa were able to learn to analyze sentiment in Spanish and French while they were excluded from the training.

In the future, we are interested in conducting sentiment analysis multimodally, because we believe that it would be helpful for many of the sentiments expressed in the corpus, as intonation and how the sentences are pronounced may contain better cues about the sentiment than pure text. It should be possible to get at least audio data because the game has voice acting. However, before that, we need to tackle the practical problem of actually extracting the audio files from the game. Currenlty, GECK outputs *File not found* for the audio files of each sentence.

[12]https://zenodo.org/record/6990638

ACKNOWLEDGMENTS

This work was partially financed by the Society of Swedish Literature in Finland with funding from Enhancing Conversational AI with Computational Creativity, and by the Ella and Georg Ehrnrooth Foundation for Modelling Conversational Artificial Intelligence with Intent and Creativity. This research has received mobility funding from Nokia Foundation under grant number 20220193.

REFERENCES

[1] Khalid Alnajjar. 2021. When Word Embeddings Become Endangered. *Multilingual Facilitation* (2021).
[2] Jeremy Barnes, Robin Kurtz, Stephan Oepen, Lilja Øvrelid, and Erik Velldal. 2021. Structured Sentiment Analysis as Dependency Graph Parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*. Online.
[3] Thérese Bergsma, Judith van Stegeren, and Mariët Theune. 2020. Creating a sentiment lexicon with game-specific words for analyzing NPC dialogue in the elder scrolls V: Skyrim. In *Workshop on Games and Natural Language Processing*.
[4] 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*. 8440–8451.
[5] 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*. Minneapolis, Minnesota, 4171–4186.
[6] Mika Hämäläinen and Khalid Alnajjar. 2019. Creative Contextual Dialog Adaptation in an Open World RPG. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo, California, USA) *(FDG '19)*.
[7] Shuhua Monica Liu and Jiun-Hung Chen. 2015. A multi-label classification based approach for sentiment classification. *Expert Systems with Applications* 42, 3 (2015), 1083–1093.
[8] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv* abs/1907.11692 (2019).
[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[10] Teemu Pöyhönen, Mika Hämäläinen, and Khalid Alnajjar. 2022. Multilingual Persuasion Detection: Video Games as an Invaluable Data Source for NLP. *The Proceedings of DiGRA 2022* (2022).
[11] Jie Tao and Xing Fang. 2020. Toward multi-label sentiment analysis: a transfer learning based approach. *Journal of Big Data* 7, 1 (2020), 1–26.
[12] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. 2010. Random k-labelsets for multilabel classification. *IEEE transactions on knowledge and data engineering* 23, 7 (2010), 1079–1089.
[13] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 38–45.
[14] Linyi Yang, Jiazheng Li, Padraig Cunningham, Yue Zhang, Barry Smyth, and Ruihai Dong. 2021. Exploring the Efficacy of Automatically Generated Counterfactuals for Sentiment Analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*. Online, 306–316.
[15] Xiaocui Yang, Shi Feng, Yifei Zhang, and Daling Wang. 2021. Multimodal Sentiment Detection Based on Multi-channel Graph Neural Networks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 328–339.
[16] Selim F Yilmaz, E Batuhan Kaynak, Aykut Koç, Hamdi Dibeklioğlu, and Suleyman Serdar Kozat. 2021. Multi-Label Sentiment Analysis on 100 Languages With Dynamic Weighting for Label Imbalance. *IEEE Transactions on Neural Networks and Learning Systems* (2021).
[17] Emily Öhman. 2021. The Validity of Lexicon-based Sentiment Analysis in Interdisciplinary Research. In *Proceedings of the Workshop on Natural Language Processing for Digital Humanities*.