

Emotion Role Labeling and Stimulus Detection*

Roman Klinger
roman.klinger@ims.uni-stuttgart.de

2023-04-24

Emotion analysis and the Need for Structured Predictions

Emotion analysis, as I described it in my [previous blog post which focused on appraisal theories](#), typically focuses on predictions based on a predefined textual unit. This could be a sentence, a Tweet, or a paragraph. For instance, given the text

I am very happy to be able to meet Donald Trump.

one could task an automatic emotion analysis system to output:

1. the emotion that the person writing the text felt at writing time or wants to express (author emotion),
2. the emotion that a person might develop based on the text (reader emotion),
3. the emotion that is explicitly mentioned in the text (text level emotion).

For (1), it's pretty clear that the author of the text "I" feels joy. The text expresses that quite clearly, which also fits (3). To understand what emotion the reader might feel (2) depends on various aspects. If they like the author and Donald Trump, for instance, there might be some increased chance that sharing the joy of the author is more likely than a negative emotion. I also think that (3) is not really a task - it's more an underspecified setup, in which the decision whose emotion is to be detected is left open.

There has been some work on the question of perspective - the reader vs. the writer ([Buechel and Hahn 2017](#)). What we cannot do, however, with such text classification approach is to extract the emotion of other mentioned entities, here, "Donald Trump". The difference between the tasks is that there is always an author and a reader, but the entities are flexible parts of the text. To assign an emotion to an entity, we first need to know which entities in the text are mentioned.

If text classification is sufficient depends on the actual task. For social media analysis, extracting the emotion of authors of full tweets makes a lot of sense. For literature, the

*Published as blog post at <https://www.romanklinger.de/blog/2023-04-24-roles/>

author's emotion obviously is not that relevant, similarly for the journalist's emotion when writing a news headline. For some domains, it is much more intuitive to look at the emotions of entities that are mentioned. Given, for instance, the following news headline (from [Bostan, Kim, and Klinger \(2020\)](#))

A couple infuriated officials by landing
their helicopter in the middle of a nature reserve.

the question which emotion the author felt is probably irrelevant – it's a journalist, they don't feel anything about it, and if they do, it probably doesn't matter. We might be interested to understand what emotion is caused in a reader, for instance to improve recommendation systems (to only read good news; or to find headlines which are suitable for clickbait). Still, arguably, in such tasks the emotions that are felt by interacting entities are more relevant for analysis of news. Here, we would like to know that "officials" are described to be angry. We could also try to infer an emotion of "A couple" - perhaps they were pretty happy (anyway, they have a helicopter).

Coming back to the example of "[The sorrows of the young Werther](#)" that have been mentioned [earlier](#), finding out which emotions are ascribed to entities in a novel clearly requires more than just text-level classification, to not just be a straight oversimplification.

Finally, we might also want to know which event (object, or other person) is described to cause the emotion. Being able to do that would allow us to automatically extract social network representations ([Barth et al. 2018](#)) and understand which stimuli are often described to cause a specific emotion.

Structured Sentiment Analysis

This is all pretty related to another more popular task that you might have heard of: aspect-based sentiment analysis. Trying to understand not only if a text is positive, but what aspect is described to be positive, who the opinion holder is, and which words express this opinion. This is now an established task in sentiment analysis. As a recent example, the [SemEval Shared Task on Structured Sentiment Analysis \(Barnes et al. 2022\)](#) aimed at detecting parts of the text corresponding to the *opinion holder*, the *expression*, and the *target*, as the organizers illustrate in [Figure 1](#).

The setup of structured sentiment analysis or aspect-based sentiment analysis is older and more established than structured emotion analysis. However, transferring sentiment analysis to emotion analysis is not entirely straight-forward. One reason is that tasks do not perfectly align:

- Detecting an **opinion holder** in sentiment analysis totally makes sense. Such thing like an "emotion holder" does, however not really exist. It would be the person experiencing or feeling an emotion, to whom we could refer as the **emoter** (we could also say feeler, but that word is more ambiguous). This also shows one difference between emotion analysis and sentiment analysis in the sense of opinion analysis - expressing an emotion is often not a voluntary process (sometimes not even conscious), while this is more often the case for an opinion. Also, opinions could develop more out of a conscious cognitive process.

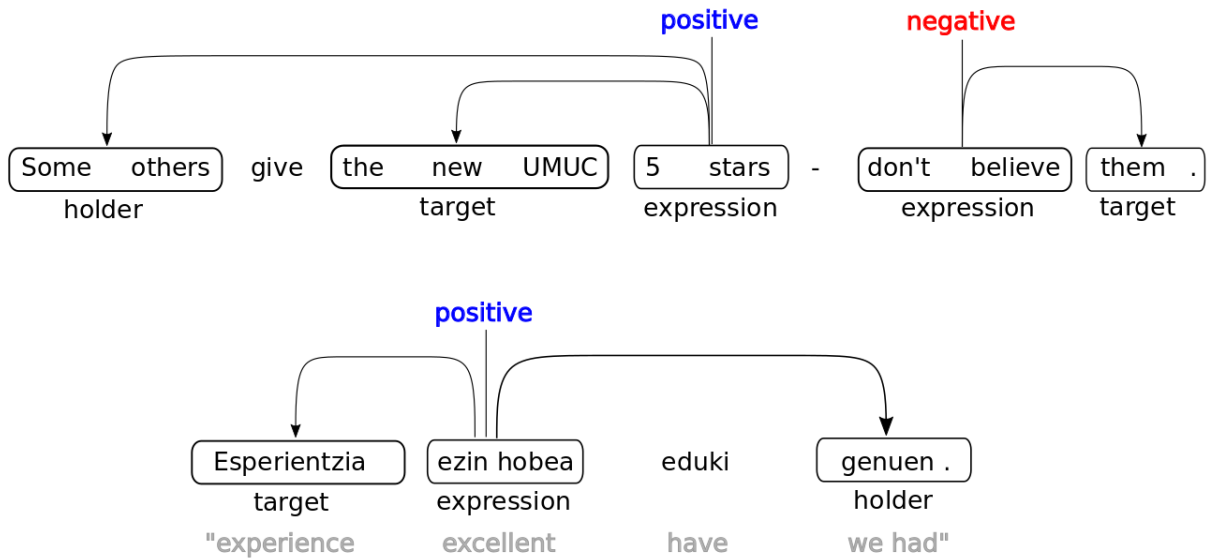


Figure 1: Visualization of the SemEval Task on Structured Sentiment Analysis in 2022

- The **aspect/target** in sentiment analysis might correspond to two things in emotion analysis. It can be a **target**, I can be angry *at* something or someone, that is not necessarily the *cause* of that emotion. I can be angry at a friend, because she did eat my emergency supply of chocolate. But I cannot be *sad at* somebody. In emotion analysis, we care more about the **stimulus** or **cause** of an emotion.
- The **evaluative phrase** in sentiment analysis (something is *good* or *bad*) pretty clearly corresponds to emotion words (something makes someone *sad* or *happy*).

To understand the relations between these tasks, we conducted the project [SEAT](#) (Structured Multi-Domain Emotion Analysis from Text) between 2017 and 2021, to which [CEAT](#) is the successor (which started in 2021).

Data Sets and Methods for Full Structured Emotion Analysis

There are now a couple of data sets available to develop systems that detect emoters and causes. Recently, the project [SRL4E](#) aggregated several of them into a common format ([Campagnano, Conia, and Navigli 2022](#)), including the ones by Gao et al. (2017), [Liew, Turtle, and Liddy \(2016\)](#), [Mohammad, Zhu, and Martin \(2014\)](#), and Aman and Szpakowicz (2007). I will focus in this blog post on our own work, namely [Kim and Klinger \(2018\)](#), [Kim and Klinger \(2019a\)](#), and [Bostan, Kim, and Klinger \(2020\)](#). Not part of SRL4E is x-enVENT ([Troiano et al. 2022](#)), because it has been published more recently, but we will also talk about this.

The two corpora by Kim and Klinger (2018) and Bostan, Kim, and Klinger (2020) aimed at developing resources that enable the development of models that recognize emotion labels for all potential emoters mentioned in the text and the relations between them (that one entity is part of a target or a cause):

There are two main differences in these data:

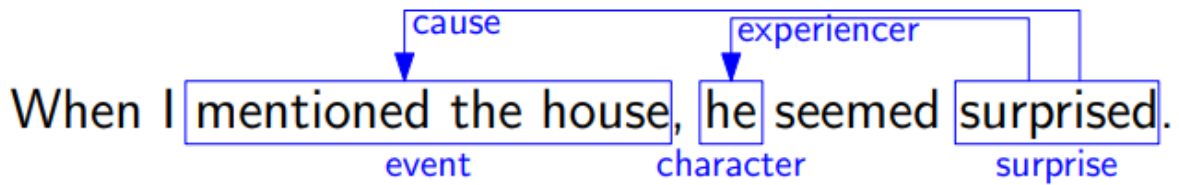


Figure 2: Visualization of one annotation in the REMAN corpus

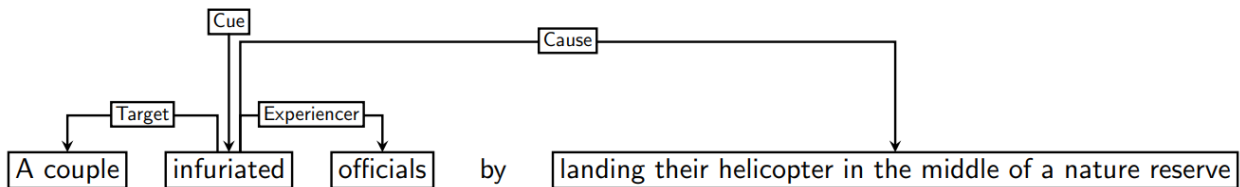


Figure 3: Visualization of one annotation in the GoodNewsEveryone corpus

1. Annotation Procedure (crowdsourcing vs. carefully trained annotators)
2. Domain (Literature vs. News headlines)

The REMAN corpus

When we started with the annotation of the REMAN corpus, we were involved in the [CRETA Project](#), a platform that combined multiple projects from the digital humanities. There was some focus on literary studies, and therefore we decided to annotate literature. We chose Project Gutenberg, because of its relative diversity and accessibility. However, literature comes with challenges - it's not exactly written to communicate facts concisely and clearly. Emotion causes and the associated roles can be distributed across longer text passages, and we expected the annotation to be difficult, because of the artistic style. This led to some decisions:

- Each data instance consists of three sentences, in which the one in the middle is selected to contain the emotion expression. The sentences around would only be annotated for roles.
- We performed the annotation with students from our institute with whom we could interactively discuss the task (before fixing the annotation guidelines and letting them annotate independently).

These decisions led to quite some ok inter-annotator agreement, but was still clearly below tasks that are more factual. Particularly detecting the cause spans was challenging. We attributed this to the subjective nature of emotions and the domain being quite challenging.

The GoodNewsEveryone corpus

For the GoodNewsEveryone corpus, we decided therefore to move to crowdsourcing, to be able to retrieve multiple subjective annotations which were then aggregated. Emotion role labeling is a structured task, and this required a multi-step annotation procedure. To not make the data more difficult than necessary, we chose a domain that is characterized by short instances: news headlines. This came, however, with another set of difficulties, namely the missing context. We did not anticipate that it would be so hard for annotators to interpret specific events. That was particularly the case when annotators from the US were tasked to annotate UK headlines (or the other way around).

Simplifying Role Labeling to Stimulus Detection or Entity-Specific Predictions

To our knowledge, up until today, there is no work on fully extracting emotion role graphs automatically. The most popular subtask is arguably **emotion cause/stimulus detection**, in which the part of the text is to be detected that describes what caused an emotion. In Mandarin, it is common to formulate the task as clause classification. It seems that in English, stimuli are often described with non-consecutive text passages or cannot be mapped clearly to clauses. Therefore, in English, it is more common to detect emotion stimuli on the token level (L. A. M. Oberländer and Klinger 2020). We also worked on stimulus detection quite a bit, as part of the corpus papers mentioned above, and additionally in German (Doan Dang, Oberländer, and Klinger 2021). We also wanted to understand if knowledge about the roles can improve emotion classification (L. Oberländer, Reich, and Klinger 2020) (yes), and how emotions are actually ascribed to a character in literature (Kim and Klinger 2019b) (depends on the emotion category).

We decided to additionally follow another research direction. While, clearly, the emotion stimulus plays an important role as the trigger to the affective sensation, there is no emotion without the person experiencing it. If we believe that emotions help us in surviving in a social world, we also need to put the entity that feels something on the spot. Our first attempt was Kim and Klinger (2019a), in which we annotated the data based on **entity relations**.

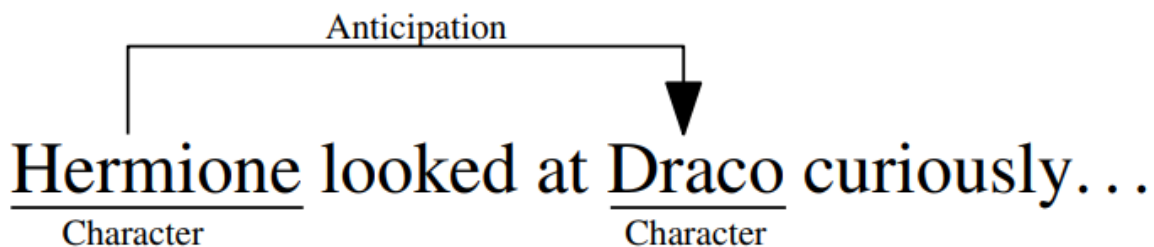


Figure 4: Visualization of one annotation in the Emotional Relations corpus

We left it to the automatic model to figure out which parts of the text are important to decide which emotion somebody feels and took the stance that the relation between

characters is important to be analyzed. We did that with a pipeline model, which detects entities, assigns emotions, and aggregates them in a graph.

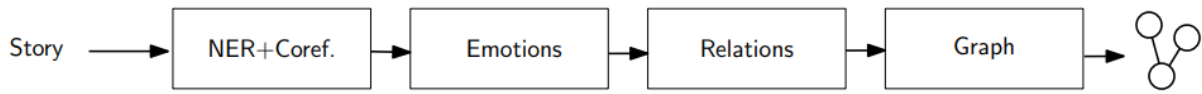


Figure 5: Visualization of the Emotional Relations Pipeline

This work follows the motivation to analyze social networks most directly, because in the evaluation of the model, we evaluated on the network level - it was fine to miss some entity relation, as long as we find it somewhere in the text.

The second and more recent paper acted as an aggregating element between our work of appraisal-based emotion analysis and emotion role labeling (Troiano et al. 2022). We went back to in-house annotations based on trained experts, because we wanted to acquire **entity-specific emotion and appraisal annotations** which we needed to first develop together with annotators. Therefore, this paper also acted as a preliminary study to Troiano, Oberländer, and Klinger (2023) which we already discussed in a previous blog post. Here, we reannotated a corpus of event reports (based on Troiano, Padó, and Klinger (2019) and Hofmann et al. (2020)), but only from the perspective of the author (the person who lived through the event and told us about that), but in addition from the perspective of every person participating in the event.

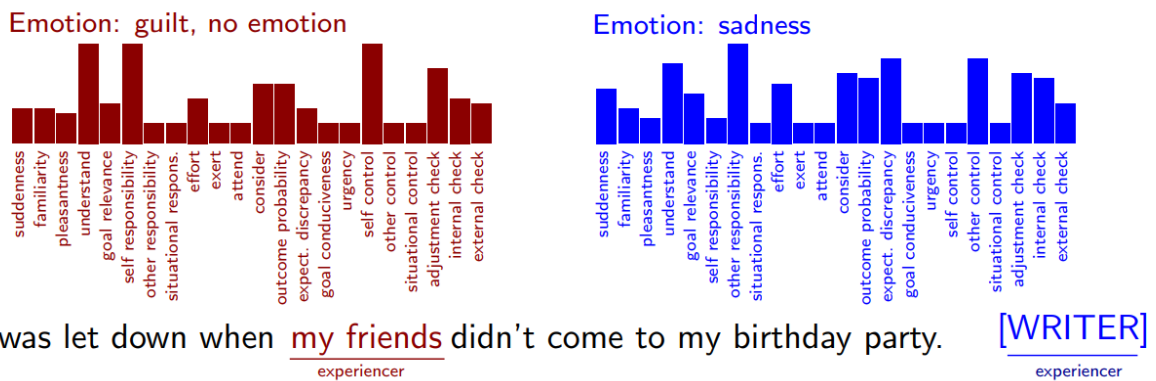


Figure 6: Visualization of x-enVENT Example

We left the relation between entities underspecified, but in the analysis of the data, that can be quite clearly observed, on the emotion and the appraisal-level. For instance, when one person is annotated to feel responsible, that decreases the probability that the other person is also responsible. As self/other-responsibility is an appraisal dimension known to be relevant for the development of guilt, shame, and pride, this also influences the emotion. We also did perform automatic modeling experiments, which very clearly showed that simple text classification does not entirely capture the emotional content of text - it conflates multiple emotional dimensions into one (Wegge et al. 2022).

Summary

In this blog post, I summarized the work we performed on emotion role labeling which is a way to represent emotions described in text. In contrast to text classification, it is more powerful to accurately represent what's in the text, but the modeling is also more challenging. Because of that, various subtasks have been defined, including experiencer specific emotion detection and stimulus detection; which both focus just on one role.

Why is this important? Most of what I wrote about is about resources, and only a bit about modeling and automatic systems. Before automatic systems can be developed, we need corpora, not only to train models, but also as a process to understand the phenomenon. I think that the emotion role labeling formalism is powerful enough to represent all relevant aspects of emotions as they are expressed in text, but it is challenging to create high-quality corpora. Further, it is challenging because sometimes, a stimulus cannot be exactly located in text. Emotions do not develop just based on one single event that can be referred to with a name or a short text. That might be ok in news data, but in literature, an event can be described with many more words, perhaps stretching over pages or even a whole book.

What comes next? Some data sets exist now, and we have a good understanding of the challenges in annotation. For each subtask, there also exist various modeling approaches. We have also seen that emotion classification and role detection influence each other ([L. A. M. Oberländer and Klinger 2020](#)). While emotion stimulus detection and emotion classification is very commonly addressed as a joint modeling task now in Mandarin ([Xia and Ding 2019](#)), we do not yet have joint models that find all roles and emotion categories together. Such structured prediction models might not only provide a better understanding of what's expressed in text than single predictions, the quality of models on each subtask might also improve, because the various variables interact. *In my opinion, developing such models is still one of the most important tasks in emotion analysis from text. This will not only help to develop better performing natural language understanding systems. It can also contribute to develop a better understanding of the realization of emotions in text.*

Acknowledgements

I would like to thank all my collaborators with whom I worked on role labeling and emotions. These are (in no specific order) Laura Oberländer, Evgeny Kim, Bao Minh Doan Dang, Kevin Reich, Max Wegge, Valentino Sabbatino, Amelie Heindl, Jeremy Barnes, Tornike Tsereteli, and Enrica Troiano.

This work has been funded by the German Research Council (DFG) in the project [SEAT, KL 2869/1-1](#).

Bibliography

- Aman, Saima, and Stan Szpakowicz. 2007. "Identifying Expressions of Emotion in Text." In *Text, Speech and Dialogue*, edited by Václav Matoušek and Pavel Mautner, 196–205. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Barnes, Jeremy, Laura Oberlaender, Enrica Troiano, Andrey Kutuzov, Jan Buchmann, Rodrigo Agerri, Lilja Øvrelid, and Erik Velldal. 2022. "SemEval 2022 Task 10: Structured Sentiment Analysis." In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, 1280–95. Seattle, United States: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.semeval-1.180>.
- Barth, Florian, Evgeny Kim, Sandra Murr, and Roman Klinger. 2018. "A Reporting Tool for Relational Visualization and Analysis of Character Mentions in Literature." In *Book of Abstracts - Digital Humanities Im Deutschsprachigen Raum*. Cologne, Germany. <http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf>.
- Bostan, Laura Ana Maria, Evgeny Kim, and Roman Klinger. 2020. "GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception." In *Proceedings of the 12th Language Resources and Evaluation Conference*, 1554–66. Marseille, France: European Language Resources Association. <https://www.aclanthology.org/2020.lrec-1.194>.
- Buechel, Sven, and Udo Hahn. 2017. "EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis." In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 578–85. Valencia, Spain: Association for Computational Linguistics. <https://aclanthology.org/E17-2092>.
- Campagnano, Cesare, Simone Conia, and Roberto Navigli. 2022. "SRL4E - Semantic Role Labeling for Emotions: A Unified Evaluation Framework." In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 4586–4601. Dublin, Ireland: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.314>.
- Doan Dang, Bao Minh, Laura Oberländer, and Roman Klinger. 2021. "Emotion Stimulus Detection in German News Headlines." In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, 73–85. Düsseldorf, Germany: KONVENS 2021 Organizers. <https://aclanthology.org/2021.konvens-1.7>.
- Gao, Qinghong, Jiannan Hu, Ruifeng Xu, Gui Lin, Yulan He, Qin Lu, and Kam-Fai Wong. 2017. "Overview of NTCIR-13 ECA Task," 6.
- Hofmann, Jan, Enrica Troiano, Kai Sassenberg, and Roman Klinger. 2020. "Appraisal Theories for Emotion Classification in Text." In *Proceedings of the 28th International Conference on Computational Linguistics*, 125–38. Barcelona, Spain (Online): International Committee on Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.11>.
- Kim, Evgeny, and Roman Klinger. 2018. "Who Feels What and Why? Annotation of a Literature Corpus with Semantic Roles of Emotions." In *Proceedings of the 27th International Conference on Computational Linguistics*, 1345–59. Santa Fe, New Mexico, USA: Association for Computational Linguistics. <https://aclanthology.org/C18-1114>.
- . 2019a. "Frowning Frodo, Wincing Leia, and a Seriously Great Friendship: Learning to Classify Emotional Relationships of Fictional Characters." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Com-*

- putational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 647–53. Minneapolis, Minnesota: Association for Computational Linguistics. <https://www.aclanthology.org/N19-1067>.
- . 2019b. “An Analysis of Emotion Communication Channels in Fan-Fiction: Towards Emotional Storytelling.” In *Proceedings of the Second Workshop on Storytelling*, 56–64. Florence, Italy: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-3406>.
- Liew, Jasy Suet Yan, Howard R. Turtle, and Elizabeth D. Liddy. 2016. “EmoTweet-28: A Fine-Grained Emotion Corpus for Sentiment Analysis.” In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, 1149–56. Portorož, Slovenia: European Language Resources Association (ELRA). <https://aclanthology.org/L16-1183>.
- Mohammad, Saif, Xiaodan Zhu, and Joel Martin. 2014. “Semantic Role Labeling of Emotions in Tweets.” In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 32–41. Baltimore, Maryland: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-2607>.
- Oberländer, Laura Ana Maria, and Roman Klinger. 2020. “Token Sequence Labeling Vs. Clause Classification for English Emotion Stimulus Detection.” In *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*, 58–70. Barcelona, Spain (Online): Association for Computational Linguistics. <http://www.romanklinger.de/publications/OberlaenderKlingerSTARSEM2020.pdf>.
- Oberländer, Laura, Kevin Reich, and Roman Klinger. 2020. “Experiencers, Stimuli, or Targets: Which Semantic Roles Enable Machine Learning to Infer the Emotions?” In *Proceedings of the Third Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media*. Barcelona, Spain: Association for Computational Linguistics. <https://www.aclanthology.org/2020.peoples-1.12/>.
- Troiano, Enrica, Laura Ana Maria Oberlaender, Maximilian Wegge, and Roman Klinger. 2022. “X-enVENT: A Corpus of Event Descriptions with Experiencer-Specific Emotion and Appraisal Annotations.” In *Proceedings of the Language Resources and Evaluation Conference*, 1365–75. Marseille, France: European Language Resources Association. <https://aclanthology.org/2022.lrec-1.146>.
- Troiano, Enrica, Laura Oberländer, and Roman Klinger. 2023. “Dimensional Modeling of Emotions in Text with Appraisal Theories: Corpus Creation, Annotation Reliability, and Prediction.” *Computational Linguistics* 49 (1). https://doi.org/10.1162/coli_a_00461.
- Troiano, Enrica, Sebastian Padó, and Roman Klinger. 2019. “Crowdsourcing and Validating Event-Focused Emotion Corpora for German and English.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4005–11. Florence, Italy: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1391>.
- Wegge, Maximilian, Enrica Troiano, Laura Ana Maria Oberlaender, and Roman Klinger. 2022. “Experiencer-Specific Emotion and Appraisal Prediction.” In *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)*, 25–32. Abu Dhabi, UAE: Association for Computational Linguistics. <https://aclanthology.org/2022.nlpccs-1.3>.
- Xia, Rui, and Zixiang Ding. 2019. “Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1003–12. Florence, Italy: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1096>.

[\[Download this post as PDF\]](#)