Emotion Carrier Recognition from Personal Narratives

Aniruddha Tammewar, Alessandra Cervone, Giuseppe Riccardi
Signals and Interactive Systems Lab, University of Trento
{aniruddha.tammewar, alessandra.cervone, giuseppe.riccardi}@unitn.it

Abstract

Personal Narratives (PN) - recollections of facts, events, and thoughts from ones own experience - are often used in everyday conversations. So far, PNs have mainly been explored for tasks such as valence prediction or emotion classification (i.e. *happy, sad*). However, these tasks might overlook more fine-grained information that could nevertheless prove relevant for understanding PNs. In this work, we propose a novel task for Narrative Understanding: Emotion Carrier Recognition (ECR). We argue that automatic recognition of emotion carriers, the text fragments that carry the emotions of the narrator (i.e. 'loss of a grandpa', 'high school reunion'), from PNs, provides a deeper level of emotion analysis needed, for instance, in the mental healthcare domain. In this work, we explore the task of ECR using a corpus of PNs manually annotated with emotion carriers and investigate different baseline models for the task. Furthermore, we propose several evaluation strategies for the task. Based on the inter-annotator agreement, the task in itself was found to be complex and subjective for humans. Nevertheless, we discuss evaluation metrics that could be suitable for applications based on ECR.

1 Introduction

A Personal Narrative (PN) is a recollection of an event, facts or thoughts felt or experienced by the narrator. People tell PNs in the form of stories to themselves and to others to place daily experiences in context and make meaning of them (Lysaker et al., 2001). PNs are generally complex and rich in information, involving multiple sub-events, participants, emotions of the narrator, relations of the narrator with the participants (Bruner, 2009) and with other important involved entities (Tammewar et al., 2020). The rich information provided through PNs can help better understand the mental state of the narrator, thus PNs are often used in psychotherapy (Angus and McLeod, 2004). Often, in psychotherapy sessions the client gets an invitation from the therapist to tell his story/PN (Howard, 1991). Through PNs, the client provides the therapist with a rough idea of his orientation toward life and the events and pressures surrounding the particular presenting problem (Howard, 1991). Nowadays, mental well-being applications are widely used, which allows the collection of PNs in the form of a journal/diary from clients easier and in digital form.

Automatic Narrative Understanding (ANU) is a growing field of research in the Natural Language Processing community (Bamman et al., 2019). ANU aims at extracting different important and useful information from Narratives, according to the requirements of the target application. Examples of tasks include reading comprehension (Chen, 2018), summarisation(Nenkova et al., 2011), and narrative chains extraction(Chambers and Jurafsky, 2008). Although PNs are widely used in psychotherapy, very few ANU tasks have been proposed to analyze PNs from the perspective of mental well-being.

An interesting ANU task, useful in the mental well-being domain, is valence (emotional value associated with a stimulus) prediction from spoken PNs. The Self-Assessed Affect Sub-challenge, part of the Interspeech 2018 Computational Paralinguistics Challenge (ComParE) (Schuller et al., 2018), focused on predicting the emotional state of the narrator in terms of valence, after narrating an event. Another task from emotion analysis that can be borrowed to ANU is Emotion Cause Extraction (ECE), aimed at identifying what might have triggered a particular emotion in a character from a narrative or the narrator himself (Lee et al., 2010). In ECE, the cause of an emotion is usually a single clause (Chen et al., 2010) connected by a discourse relation (Mann and Thompson, 1987) to another clause explicitly expressing an emotion (Cheng et al., 2017), as in this example from (Gui et al., 2016): “$< cause >$ *Talking about his honours,* $< /cause >$ *Mr. Zhu is so* $< emotion >$ *proud* $< /emotion >$.’ The focus of ECE, however, has been mainly on text genres such as news and microblogs so far. Applying ECE on the genre of PNs is complicated, given the complex structures of PNs, involving multiple sub-events and their attributes, such as the characters involved, time, and place. In PNs the cause as well the emotion may not be explicitly expressed in a keyword/clause or multiple keywords might be required to express the same, thus making it difficult to define a discourse relation.

The analysis of PNs from personal journals can provide deeper emotion analysis such as the trends of emotion states over a period. For example, Tammewar et al. (2019) worked on predicting valence from the PN and found how different text fragments, including not only sentiment words but also events (’high school exam’) or people (’grandpa’) proved to be particularly useful to predict the valence score of a narrative across both neural and classic machine learning models. Following up on this intuition, (Tammewar et al., 2020) propose *emotion carriers* as the concepts from a PN that explain and carry the emotional state of the narrator. Emotion carriers include thus not only explicitly emotionally charged words, such as “happy”, but also mentions of people (friend, grandparents), places (school, swimming pool), objects (guitar), and events (internship) that carry an emotional weight within a given narrative. For example, in the narrative in Table 1 the emotion carriers recognize the word “family” (orig. “Familie”) as carrying a positive emotion for the narrator. Tammewar et al. (2020) propose an annotation schema for emotion carriers and perform an annotation experiment of German PNs from the Ulm State-of-Mind in Speech (USoMs) (Rathner et al., 2018) corpus. Based on the analysis of the inter-annotator agreement, they conclude that the task is complex and subjective. Nevertheless, the task has potential benefits in many applications in the medical domain and in particular in the mental well-being.

Automatically recognizing emotion carriers from PNs could be useful to identify text fragments that closely relate to the narrator’s emotional state and possibly find trends in them. In the scenario of a mental well-being application, which collects digital PNs in the form of a journal, emotion carriers could be useful for automatic summarization and analytics of the user’s state over time. Additionally, a conversational agent could, for example, ask follow-up questions based on the recognized emotion carriers to elicit more information useful to better understand the emotional relation between the carrier and narrator over time.

In this work, we propose Emotion Carriers Recognition (ECR), a novel task of narrative understanding. We present different baseline models for ECR based on Conditional Random Fields (CRF) using the USoMs corpus annotated with emotion carriers information following Tammewar et al. (2020) to analyze the feasibility of the task. We also propose different evaluation strategies for ECR, to compare the model’s performance to the inter-annotator agreement obtained from humans and to evaluate the results from the perspective of its usefulness for a particular target application for mental well-being, where the task is to begin a conversation with the user about a recognized emotion carrier.

2 Data

In this work, we use a corpus of Spoken Personal Narratives manually annotated with text spans/fragments that best explain the emotional state of the narrator (emotion carriers) following the annotation schema proposed in (Tammewar et al., 2020). The PNs are in German and taken from the Ulm State-of-Mind in Speech (USoMs) corpus (Rathner et al., 2018), which was used and released in the Self-Assessed Affect Subchallenge, a part of the Interspeech 2018 Computational Paralinguistics Challenge (ComParE) (Schuller et al., 2018). The USoMs dataset consists of PNs along with self-assessed valence and arousal scores collected from 100 participants. The participants were asked to recollect two positive and two negative PNs, each with a duration of approximately five minutes. The narratives are automatically transcribed using the Amazon Transcribe Service[1]. Later, narratives from 66 participants, consisting of 239 PNs, were annotated with the emotion carriers by four annotators (25 narratives from the USoMs data were removed because

[1] https://aws.amazon.com/transcribe

PN fragment:	Und	hm	die	Gefhle	dabei	waren,	dass	man	sich	hm
Gloss:	And	um	the	feelings	there	were,	that	you	yourself	um
Annotation:	O|O|O|O	O|O|O|O	O|O|O|O	O|O|O|O	O|O|O|O	O|O|O|O	O|O|O|O	O|O|O|O	O|O|O|O	O|O|O|O
PN fragment:	einfach	**freut**	**und**	**glcklich**	**ist** ,	dass	man	eine	**Familie**	
Gloss:	easy	pleased	and	happy	is,	that	you	a	family	
Annotation:	O|O|O|O	I|O|O|O	I|O|O|O	I|O|O|I	O|O|O|I	O|O|O|O	O|O|O|O	O|O|O|O	I|I|I|O	
Translation:	*And uh, the feelings were that you are uh just pleased and happy that you have a family ...*									

Table 1: A small text fragment from a PN annotated with emotion carriers. The first row reports the original German words from the PN, the second row shows the corresponding English translation, while the third row shows the annotations. The annotation is performed by 4 annotators, thus for each token, there are 4 IO labels. For the token "Familie" the annotation is I|I|I|O, which means that the first three annotators classified it as *I* while the forth as *O*. The intensity of the red color in the background for the PN fragment also highlights the number of annotators who annotated the token (from lightest for 1 annotator to the darkest for all annotators).

of issues like noise). The annotation task involved recognizing and marking the emotion carrying text spans as perceived by the annotators from the PNs. They were asked to select sequences of adjacent words (one or more) in the text, that explain why the narrative is positive or negative for the narrator, focusing more on the words playing an important role in the event such as people, locations, objects.

The data can be represented in the *IO* encoding, as shown in the example of a small fragment of a PN from Table 1. We consider the document as a sequence of tokens, where each token is associated with the label *I* if it is a part of an emotion carrier (text span) or the label *O* if it is not. In this way, a continuous sequence of tokens with label *I* represents one emotion carrier. In the third row *Annotation*, we show the manual annotations done by the four annotators. Here, we have four annotators and thus four labels for each token. It can be observed how the annotators perceive emotion carriers differently, making the task subjective.

The number of annotations (text-spans) annotated by the annotators per narrative varies from 3 to 14 with an average of 4.6. When punctuation is excluded, on average, the number of tokens per annotation consists of 1.1 tokens for three annotators, while for the forth annotator annotates longer segments consisting of 2.3 tokens (avg.). On average, a narrative consists of 820 tokens, while a sentence consists of 34 tokens. The Tokenization is performed using the spaCy toolkit[2], while the sentence splitting is performed using the punctuation provided in the transcriptions.

[2] https://spacy.io/

3 The Emotion Carriers Recognition Task

We propose a novel narrative understanding task of automatic Emotion Carrier Recognition (ECR) from PNs. PNs are collected and used in psychotherapy in different ways. Diary is a popular life-logging tool for storing memories and daily experiences in the form of PNs, aiding recollection (Machajdik et al., 2011; Sellen et al., 2007). Diaries have shown to improve adherence by increasing the consciousness of the clients about their condition and have proven to be effective in gaining deep insights into a clients well-being. The diary can be used by a therapist to learn about the clients behavior and routines (Gjengedal et al., 2010). The emotion carriers from the PNs may provide deeper insights about how events, relations, and other elements from the clients' life affect their mental state. Our task focuses on the automatic ECR from PNs in the scenario of a mental well-being mobile application with the broader aim of building a conversational agent that can start a conversation with the client based on the PNs shared, to elicit more information about important carriers.

We pose the recognition of emotion carriers from a given PN as a sequence labeling problem. We consider the document as a sequence of tokens. Hence, the target of the sequence-labeling problem is to assign a label *I* or *O* to each token in the PN.

4 Baseline Models

Conditional Random Fields (CRF) (Lafferty et al., 2001) is a widely used machine learning algorithm for sequence-labeling problems in NLP, such as

Part of Speech tagging, Named Entity Recognition. As CRF works on sequences of inputs and output labels, it is a natural fit for our work.

As CRF by default cannot use the probability distribution of the classes, only one class must be provided in the output. In all following experiments and evaluations, for each token, we provide the label *I* if at least one of the four annotators annotated the token, otherwise, it is tagged as *O*.

We train and test the sequence-labelingmodels at narrative and sentence levels. In the narrative level, we consider the entire narrative as one sequence (**Narrative**), while in the sentence level, we consider one sentence as a sequence. In this way, we can analyze how the length of a sequence affects the performance of the CRF model. Also note that at the sentence level, the model does not have access to the other parts of the narrative. The limited access to context may affect performance. The sentence-level sequences are further considered in two ways : 1) **SentAll:** all the sentences are considered 2) **SentCarr:** only sentences containing at least one emotion carrier are considered. In the real scenario, we would have to extract carriers from the entire narrative or all the sentences, as we do not know beforehand which sentences contain the carriers. Thus, our focus is on better prediction on the *SentAll* or *Narrative*.

For the CRF model, we use the context window of ± 3 with features such as the token, its suffixes, POS tag, a prefix of POS tag, sentiment polarity. We also experiment with pre-trained word-embeddings as features.

Data			**Embs**	**Results (F1)**	
Data level	**#Train**	**#Test**		**class-I**	**micro**
SentWith	1705	402	Yes	0.8	0.76
			No	0.8	0.78
train: SentCarr test: SentAll	1674	8154	Yes	0.35	0.96
			No	**0.35**	**0.96**
SentAll	6771	1382	Yes	0.25	0.96
			No	0.28	0.96
Narrative	192	47	Yes	0.28	0.96
			No	0.25	0.96

Table 2: Results of CRF based models with different combinations of features and data segmentation. The highlighted model is used in further experiments and evaluations.

5 Evaluation

In this section, we propose different evaluation strategies for the ECR task.

5.1 Token Level

The token level evaluation measures the performance of predicting *I* or *O* class for each token in a sequence. We use this metric to evaluate the baseline CRF based models with different combinations of features and data segmentation. We are concerned more about the prediction of the class *I*, as our aim is to be able to find one or more important carriers to start a conversation with the narrator.

As discussed earlier, considering a real-world scenario, we need the model to perform well on the *SentAll* or *Narrative* data. We find that the model trained on *SentCarr* performs best on the *SentAll* data, as can be observed in the Table 2. For further evaluation, we use the model trained on *SentCarr*, thus recognition would be done on the sentences and not the entire narrative as one sequence.

Using the *SentCarr* model, we extract the continuous sequences of tokens that are tagged as *I*. These text spans are considered as the emotion carriers recognized by the model. In the next metrics, we evaluate the model by comparing this set of carriers with the set of manually annotated reference carriers.

Carriers counts: The model is trained on 191 narratives while evaluated on 48 narratives. In Table 3, we show the statistics of the number of carriers present per narrative in the reference annotation and the prediction of the model.

	Min	**Max**	**Mean**	**Std**
Reference	7	16	11.62	2.24
Predicted	4	35	13.29	5.05

Table 3: Counts of carriers annotated by the the human annotators (reference) and the automated system (predicted)

5.2 Agreement Metrics

Here we use the metrics that were used to evaluate the inter-annotator agreement between the four annotators (pair-wise) in (Tammewar et al., 2020), to evaluate the performance of our models. The metric used is *positive (specific) agreement*. This evaluation is important as it compares the performance of the system with the inter-annotator agreement, which can loosely be considered as human performance.

We also explore the different criteria to decide whether two annotations match or not, as used in the original metrics. The evaluations are based on *Exact Match*, where the two carriers match if they are exactly the same and *Partial/soft Match*, where the token overlap of the two carriers wrt the reference carrier is considered. Other parameters in the matching criteria include *position of the carrier in the narrative*, which has two possibilities *position agnostic* and *position matters/considered*. Another criterion is based on the matching of *tokens* vs *lemmas*. Note that in all matching strategies, we remove the stopwords and punctuation from the annotation as we are interested in the content words.

Results: As expected, with the loosening of the matching criteria, the results improve. A similar trend is observed in the inter-annotator agreement. When we move from *a* to *b*, we are removing the stopwords and punctuation from the predicted and reference carriers. This improves the results significantly. The reason behind this is the fact that the reference annotations, which were also used for training the CRF model, as mentioned earlier, contain all the tokens that are tagged by at least one annotator. As noticed by Tammewar et al. (2020), in the annotations, one of the annotators usually annotates longer spans than others. They also observed that many annotations also contain punctuation and stopwords. To understand this issue, let us take an example of concept annotation. For a concept like a printer, the annotators could select the spans with the printer,, the printer, or just printer,. With our strategy for creating reference annotations, we end up selecting the longest span "with the printer," which contains stopwords like *with, the* and a punctuation mark (,). However, this might not be the case in the model's output (as the training data also contain concepts marked by only one annotator). To reduce this effect, one way is to remove the stopwords and punctuation (strategy b) and another is to use the partial match (strategy c). Both strategies worked out for us, resulting in very similar scores with the two strategies. We notice a significantly large jump in the model's performance from *c* to *d* compared to the inter-annotator agreement. Although confirmation is needed, we think this could be because the model is trained and tested at the sentence level, thus the position in the narrative is not taken into consideration. Additionally, the performance further improves slightly when we match lemmas instead of tokens (from *d* to *e*).

sr	Parameters	Prec	Recall	F1	IAA
a	Exact-match position agnostic token level (w/ stopwords and punctuation)	0.236	0.270	0.252	0.252
b	Exact-match position agnostic token level	0.353	0.397	0.374	NA
c	Partial match position considered token level	0.356	0.381	0.368	0.320
d	Partial match position agnostic token level	0.533	0.603	0.566	0.399
e	Partial match position agnostic lemma level	0.549	0.620	0.583	0.403

Table 4: Evaluation based on the agreement metrics (positive agreement) with different parameter configurations. For each configuration, the corresponding inter-annotator agreement (IAA) score is in the last column.

5.3 Recognized at least *k* carriers

As mentioned earlier, we are interested in starting a conversation with the narrator about a particular carrier. To start a conversation, we would need at least one emotion carrier to talk about. In this evaluation metric, for each narrative, we measure if at least *k* carriers from the reference are recognized by the model. A carrier is considered recognized if it is an exact match. When matching, we remove stop-words and punctuation. We perform two evaluations, considering and not considering the position of the carrier in the narrative. The results are described in Table 5. For our goal of starting a conversation about a particular carrier, the results seem overwhelmingly good. However, an important question remains, how many of the recognized carriers are useful for a conversation?

5.3.1 Sentiment vs Content Carriers

The annotations include sentiment words as well as content words. To study if the model is bi-

Type of carriers	Position	at least k recognized		
		k=1	k=2	k=3
all carriers	considered	97.9	83.3	58.3
	agnostic	100	89.6	72.9
content carriers	considered	81.3	41.7	18.8
	agnostic	87.5	50	29.2
sentiment carriers	considered	83.3	50	18.6
	agnostic	89.6	56.3	22.9

Table 5: Evaluation based on the fraction of narratives in which at least k carriers are recognized correctly by the model. The evaluation is performed separately for all the carriers, only content carriers and only sentiment carriers.

ased towards the recognition of carriers with sentiment words (angry, joy) versus content words (internship, parents) in emotion carriers, we further divide the annotations (reference and predicted) into sentiment and content carriers and perform the similar evaluation on them separately. For this analysis, we calculate the sentiment polarity of each annotation using the textblob-de[3] library following (Tammewar et al., 2020), which uses the polarity scores of the words from senti-wordnet for German (with simple heuristics), similar to the English senti-wordnet. If the score is 0 the carrier is considered a content carrier, otherwise a sentiment carrier.

In Table 6, for each narrative we study the fraction of sentiment and content annotations in reference and predictions. We see that on average half of the annotations are classified as content carriers. The manual analysis of the annotations shows that the classification using textblob-de is not perfect. While it can recognize the content carriers properly, we see some examples of sentiment-carriers are also being classified into the content-carriers. Some examples of correctly recognized content carriers include *Bachelorarbeit (bachelor thesis), Magenprobleme (stomach problems), Durchhaltevermgen (stamina)* while an example of sentiment-carriers that are classified as content-carriers include *unzufrieden (unsatisfied)*

[3] https://textblob-de.readthedocs.io/en/latest/

	Min	Max	Mean	Std
Reference	0.08	0.70	0.41	0.14
Predicted	0.18	0.73	0.45	0.15

Table 6: The fraction of carriers in each narrative that carry sentiment polarity

Next, we do the evaluation based on the *at least k recognized* metric for each group independently. In Table 5, we compare the results for the content and sentiment carriers. We observe a decline in the performance compared to the evaluation of all carriers. Nevertheless, we find that in 87.5 % of the narratives (position agnostic), we can predict at least 1 emotion carrier, which is a requirement for starting a conversation.

6 Conclusions

In this work, we proposed Emotion Carriers Recognition (ECR), a novel narrative understanding task where the model is asked to recognize the text spans that best explain the emotional state of the narrator from a personal narrative. We investigated various baselines for the task and explored several evaluation metrics. The proposed task allows having a deep analysis of the emotional state of the narrator, recognizing all text fragments, including mentions of events, people, or locations, which carry an emotional charge. We argue that ECR could be particularly useful in the context of a mental well-being application, for analysis, summarization, and particularly conversational applications to better understand users' mental health.

We also evaluated our system from different perspectives. The token level performance indicates the performance of the CRF baseline model on the tokens, which is further used to recognize carriers. Whereas the *agreement metrics* and *at least k recognized* metrics evaluate the model's performance in recognizing the text spans (one or more continuous tokens). With the agreement metrics, we could loosely compare the model's performance with respect to the human performance (inter-annotator agreement) and found that the model performance is better or close to the human one. Finally, with the *at least k recognized* metric, we verify the feasibility of the integration of the model in a real-world application.

References

Lynne E Angus and John McLeod. 2004. *The handbook of narrative and psychotherapy: Practice, theory and research*. Sage.

David Bamman, Snigdha Chaturvedi, Elizabeth Clark, Madalina Fiterau, and Mohit Iyyer. 2019. Proceedings of the first workshop on narrative understanding. In *Proceedings of the First Workshop on Narrative Understanding*.

Jerome S Bruner. 2009. *Actual minds, possible worlds*. Harvard University Press.

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*, pages 789–797.

Danqi Chen. 2018. *Neural reading comprehension and beyond*. Ph.D. thesis, Stanford University.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 179–187. Association for Computational Linguistics.

Xiyao Cheng, Ying Chen, Bixiao Cheng, Shoushan Li, and Guodong Zhou. 2017. An emotion cause corpus for chinese microblogs with multiple-user structures. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(1):6.

Eva Gjengedal, Sissel Lisa Storli, Anny Norlemann Holme, and Ragne Sannes Eskerud. 2010. An act of caring–patient diaries in norwegian intensive care units. *Nursing in Critical Care*, 15(4):176–184.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-driven emotion cause extraction with corpus construction. In *EMNLP*, pages 1639–1649. World Scientific.

George S Howard. 1991. Culture tales: A narrative approach to thinking, cross-cultural psychology, and psychotherapy. *American psychologist*, 46(3):187.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. 2010. A text-driven rule-based system for emotion cause detection. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 45–53. Association for Computational Linguistics.

Paul Henry Lysaker, John Timothy Lysaker, and Judith Thompson Lysaker. 2001. Schizophrenia and the collapse of the dialogical self: Recovery, narrative and psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, 38(3):252.

Jana Machajdik, Allan Hanbury, Angelika Garz, and Robert Sablatnig. 2011. Affective computing for wearable diary and lifelogging systems: An overview. In *Machine Vision-Research for High Quality Processes and Products-35th Workshop of the Austrian Association for Pattern Recognition. Austrian Computer Society*, pages 2447–2456.

William C Mann and Sandra A Thompson. 1987. Rhetorical structure theory: A theory of text organization. *Technical Report*, pages 87–190.

Ani Nenkova, Kathleen McKeown, et al. 2011. Automatic summarization. *Foundations and Trends® in Information Retrieval*, 5(2–3):103–233.

Eva-Maria Rathner, Yannik Terhorst, Nicholas Cummins, Björn Schuller, and Harald Baumeister. 2018. State of mind: Classification through self-reported affect and word use in speech. *Proc. Interspeech 2018*, pages 267–271.

Björn W Schuller, Stefan Steidl, Anton Batliner, Peter B Marschik, Harald Baumeister, Fengquan Dong, Simone Hantke, Florian B Pokorny, Eva-Maria Rathner, Katrin D Bartl-Pokorny, et al. 2018. The interspeech 2018 computational paralinguistics challenge: Atypical & self-assessed affect, crying & heart beats. In *Interspeech*, pages 122–126.

Abigail J Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, and Ken Wood. 2007. Do life-logging technologies support memory for the past? an experimental study using sensecam. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 81–90.

Aniruddha Tammewar, Alessandra Cervone, Eva-Maria Messner, and Giuseppe Riccardi. 2019. Modeling user context for valence prediction from narratives. *Proc. Interspeech 2019*, pages 3252–3256.

Aniruddha Tammewar, Alessandra Cervone, Eva-Maria Messner, and Giuseppe Riccardi. 2020. Annotation of emotion carriers in personal narratives. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1510–1518, Marseille, France. European Language Resources Association.