



**HAL**  
open science

## Annotating Emotions in Acquired Brain Injury Patients' Narratives

Salomé Klein, Amalia Todirascu, Hélène Vassiliadou, Marie Kuppelin, Joffrey Becart, Thalassio Briand, Clara Coridon, Francine Gerhard-Krait, Joé Laroche, Jean Ulrich, et al.

### ► To cite this version:

Salomé Klein, Amalia Todirascu, Hélène Vassiliadou, Marie Kuppelin, Joffrey Becart, et al.. Annotating Emotions in Acquired Brain Injury Patients' Narratives. Patient-Oriented Language Processing, Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health)@ LREC-COLING 2024, May 2024, Torino, Italy. pp.26-36. hal-04636082

**HAL Id: hal-04636082**

**<https://hal.science/hal-04636082>**

Submitted on 8 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Annotating Emotions in Acquired Brain Injury Patients' Narratives

Salomé Klein<sup>1</sup>, Amalia Todirascu<sup>1</sup>, Hélène Vassiliadou<sup>1</sup>, Marie Kuppelin<sup>2</sup>,  
Joffrey Becart<sup>1</sup>, Thalassio Briand<sup>1</sup>, Clara Coridon<sup>1</sup>, Francine Gerhard-Krait<sup>1</sup>, Joé  
Laroche<sup>1</sup>, Jean Ulrich<sup>1</sup>, Agata Krasny-Pacini<sup>2</sup>

University of Strasbourg, <sup>1</sup>UR 1339/LiLPa & FRLC / <sup>2</sup>INSERM UMR-S 1329

Le Portique, 14, rue René Descartes, 67084 Strasbourg Cedex (BP 80010)

{salklein, todiras, vassili, joffrey.becart, briandt, gerhard, laroche}@unistra.fr, {marie.kuppelin,  
clara.coridon, jean.ulrich}@etu.unistra.fr, agata.krasny-pacini@upecam.assurance-maladie.fr

## Abstract

In this article, we aim to measure the patients' progress in recognizing and naming emotions by capturing a variety of phenomena that express emotion in discourse. To do so, we introduce an emotion annotation scheme adapted for Acquired Brain Injury (ABI) patients' narratives. We draw on recent research outcomes in line with linguistic and psychological theories of emotion in the development of French resources for Natural Language Processing (NLP). From this perspective and following Battistelli et al. (2022) guidelines, our protocol considers several means of expressing emotions, including prototypical expressions as well as implicit means. Its originality lies on the methodology adopted for its creation, as we combined, adapted, and tested several previous annotation schemes to create a tool tailored to our spoken clinical French corpus and its unique characteristics and challenges.

**Keywords:** Emotion, Annotation, Acquired Brain Injury, Spoken French

## 1. Introduction

Corpora enriched with emotional information become increasingly important, especially in clinical linguistics. In the field of Natural Language Processing (NLP), finding a consensus on emotional categories and their expression as well as developing an automatic emotion annotation system is a widely recognized challenge (see for instance EmotionX challenge in Hsu and Ku, 2018 or, for French, the DEFT challenges 2015; 2018). In automatic emotion detection, most of the existing corpora are written. Spoken French and particularly patients' narratives have been little explored, due to the lack of available data (Amblard et al., 2020). Patients' narratives represent valuable data for doctors, linguists and NLP researchers, as they can be used for diagnosis purposes, to evaluate the effectiveness of a therapy or to detect imminent signs of a crisis. These corpora represent real challenges for NLP systems due to phenomena related to the oral channel such as disfluencies, repetitions, hesitations. Various expressions of the emotions can be found in these narratives and NLP systems fail to identify them, because of the large variability of their expressions: simple words, multi-word expressions, entire sentences.

Our corpus is composed of patients' narratives suffering from Acquired Brain Injury (ABI), who present difficulties naming their emotions. In this article, we aim to measure the patients' progress in recognizing and naming emotions by capturing a variety of phenomena that express emotions in discourse. To do so, we introduce an emotion annotation scheme to improve understanding of ABI patients' narratives. We exploit recent research outcomes in the development of French NLP resources for emotion annotation (Etienne

et al., 2022; Troiano et al., 2022; Cortal et al., 2023; Etienne, 2023), in line with linguistic and psychological theories of emotions. Following Battistelli's et al. (2022) guidelines, our protocol (§ 4) considers several means of expressing emotions, including prototypical expressions (e.g. 'I'm angry') and implicit means. Its originality lies on the methodology adopted for its creation. We combined, adapted, and tested several previous annotation schemes (§ 3.2) to create a new one adapted to our spoken clinical French corpus and its unique characteristics and challenges. The narrative corpus shows the patients' difficulties, and the need to encode the subtle and non-standard ways they use to convey them. Our scheme enables a lightweight and flexible annotation that codes the specific features of emotional expressions in spoken language, such as lexical repetitions, hesitations, non-standard idioms, and emotion expressions diluted over several sentences.

The paper is organized as follows. First, we present the context of our research (§ 2), followed by the psycholinguistic and NLP research projects used to create our annotation scheme (§ 3). We then describe our corpus, our annotation scheme, and our methodology (§ 4). Finally, we discuss the annotation campaign, the results obtained, and the drawbacks of our method (§ 5). We conclude with the future perspectives, including some solutions to the obstacles encountered (§ 6).

## 2. Context of the Study

Our study takes place within the participatory research project GRoupeRegulationEMOtion for people with acquired brain injury (GREMO; NCT 05 39 34 92 *Regulating Emotions and Behaviors After Brain Injury*) involving clinical psychology,

linguistics and NLP. GREMO is an intensive emotion regulation intervention comprising group sessions and individual psychotherapy based on Dialectical Behavior Therapy (DBT) developed by Linehan (2015). Patients with an ABI may suffer from emotional dysregulation for which evidence lacks to establish adapted treatments. Moreover, patients' alexithymia (affect disorder) and lack of insight biases the usual measures of patient-reported outcomes. One of the aims of the GREMO group is to help patients (re)learn emotional regulation skills and to associate physiological and behavioral changes, such as increased heart rate, screaming, and sweating, with specific emotional states like anger and fear. They also learn to link these changes with stereotypical situations, such as being lied to, yelled at, or abandoned.

The overall goal of the project is to explore a new objective measure of emotional regulation through linguistic and pragmatic markers, that could be used to test the effectiveness of DBT in ABI (Kuppelin and Krasny-Pacini, 2023). In addition to the other standard outcome measures, 45 patients with chronic ABI were voice-recorded for 40-90 minutes semi-directive interviews. The interviewer asks the patient to recount emotionally-charged memories or to describe the emotions they feel when shown an emotionally-inducing or a neutral image, at 3 time points 5-month apart: at the beginning of the baseline (T0, 5 months before the therapy), immediately before therapy (T1, to explore retest effects and stability of response) and immediately after therapy (T2, to explore for gains due to a 5-month intensive DBT). All patients are fluent but have pathological scores on the clinical evaluation scales: they have impaired cognitive functions and suffer from emotional regulation difficulties.

### 3. Emotion in Psycholinguistics and in NLP: A Brief Overview

We do not aim to provide an exhaustive overview of the literature on emotions and the distinctions between *emotion*, *feeling*, *affect* or *sensation* and *opinion* that vary greatly from one author to another. In the psychiatric field, Sifneos (1996: 138) recommends "for clarifications purposes, that in future studies 'affect' should be used as a general term to include 'emotion' with its somatic components and 'feeling' with its subjective experiential aspects." Some linguists suggest, on the contrary, to avoid this too conventional and technical term which does not reflect the difference between inner feelings (i.e. contingents) and behaviors or attitudes that are subject to external judgement (Polguère, 2013). In NLP, the label of *emotion* seems to prevail and is used as a generic term including opinions and attitudes (Suttles and Ide, 2013; Bostan and Klinger, 2018).

It is crucial to consider these theoretical aspects which inevitably influence annotators' decisions. Speaker's intuition regarding the difference between an attitude and an emotion strictly speaking highlights the difficulty of defining semantically vague concepts in everyday words. While we aspire obtaining a fine-grained robust description model of emotions applicable to various corpora, the paradox is that the greater the number of categories added, the lower the inter-annotator agreement (see Öhman, 2020).

#### 3.1 Emotion in Psycholinguistics

It is well known that there is no widely accepted and satisfactorily used classification of emotions. The psychologist Ekman (1992) considers that emotions are discrete, categorizable units which can be reduced to a finite number of primary or primitive emotions (fear, sadness, disgust, joy, surprise and anger). Other classifications reckon with the dimensional viewpoints, advocated by Wund (1903) and Russel (1980) in particular, who see emotions as belonging to positive/negative axes, and of high or low intensity (see Galati and Sini, 1998). This also ties in with the Appraisal Theory (Lazarus, 1991; Frijda, 2007; Ellsworth, 2013) where emotions involve an evaluation that has caused a given reaction.

Linguistically speaking, emotions have been studied in terms of their explicit expression, using idioms and terms that directly describe emotional states (e.g. 'I feel bad', 'I'm happy'; Anscombe, 1995; Flaux and Van de Velde, 2000). Emotion is understood in the sense of a physio-psychological manifestation in an affected 'place'. Nonetheless, statements conveying emotional information are not limited to those that explicitly name an affect. Micheli (2014) for instance distinguishes between explicit and implicit modes of expressing emotions. Wharton and de Saussure (2022: 670) also insist on considering "expressions that are irreducible to purely conceptual or propositional meanings" such as interjections. In the same vein, Etienne and Battistelli (2021) note two modes of expressions which do not rely on emotional labels: those arising from situations typically associated with a particular emotion, such as a funeral, a party or get slapped, and those being cues or consequences of the emotion expressed in the discourse, such as interjections, exclamatory statements and descriptions of behaviors associated with emotions. To sum up, emotions can be expressed explicitly through emotion labels, while implicit modes of expressions are inferential.

In our study, we consider that emotions stand out from other affective phenomena such as sensations, feelings, moods because of their *praxis* which underlies their categorization in language (see also the primitive "WANT to do" in Wierzbicka, 1992 or "the urge to act" in the GREMO program). The concept of *emotion* in the

repertoire of everyday language is defined by typical situations to which we react with certain affective manifestations that may be followed by typical behaviors: emotions are in a sense a type of behavior, or at least are strongly linked to behaviors/attitudes that they motivate in response to practical scenarios.

### 3.2 Emotion Annotation in NLP: Important Findings

Due to the difficulties of classifying emotions, authors of automatic emotion detection systems were from the very beginning aware of the fact that emotion annotation is not a simple emotion tag assignment. In other words, it seems impossible to just apply any annotation guidelines randomly. Existing annotation schemes consider different categories (private states, beliefs, thoughts, opinions, emotions), different emotion taxonomies, and procedures in one or two steps (word level and/or sentence level annotations) that may include or not intensity and polarity. And, most importantly, they may consider or not negation and modality. Finally, it is not always clear if context elements play a role in annotators' decisions nor if implicit means are processed.

Dimensional viewpoints and primary emotions are the most common for the modeling of lexical resources such as "Affects Lexicon" (Augustyn and Tutin, 2009), "Polarimots" (Gala and Brun, 2012) or "Diko" (Lafourcade et al., 2015). These resources generally use the Ekman's classification, the most widely accepted in NLP, which requires clear and simple categories as stated above. The use of a categorial classification is here the easiest way to implement a model for automatic emotion detection tools (Bhaumik et al., 2023; Cortal et al., 2023). Other systems create corpora taking into account the dimensional issues of the emotions proposed by the appraisal theory (Troiano et al., 2023). Besides Ekman's primary emotions, some annotation schemes add "complex emotions" (combinations of primary ones) following Turner's sociological taxonomy (2007): e.g. 'pride' as the combination of 'happiness' and 'fear' (Aman and Szpakowicz, 2007; Chen et al., 2009; Etienne and Battistelli, 2021). Plutchik's (1980) taxonomy identifies eight basic emotions: joy, fear, disgust, anger, sadness, surprise, confidence, and anticipation. This taxonomy is included in some annotation guidelines, such as in Giouli et al. (2014). But it is important to note that the use of Ekman's six primary emotions or Plutchik's taxonomy does not limit annotations to six or eight labels. Most annotation schemes associate each of them with relevant keywords (e.g. 'anger' can be associated with words such as *hate*, *dislike*, and *disgust*). Other models include several tags from the outset: Vidrascu (2007), based on Kappas et al. (1991) work, considers 20 labels of discrete emotions, Augustyn (2015) opts for 41

preestablished emotional categories, Abdul-Mageed and Ungar (2017) or Demszky et al. (2020) built a large dataset for the 24 labels of the Plutchik's wheel.

Another crucial point in emotion annotation tasks is polarity and intensity. Most of annotation schemes distinguish between positive, negative, none for polarity and low, medium, high (sometimes extreme) intensity or measured on a scale from 1 (low) to 5 (extreme).

Emotion annotation might sometimes be divided into 2 subtasks: a) emotion detection, which distinguishes between emotional and neutral content, generally at the utterance level and b) emotion classification, which assigns an emotion tag to a word (Aman and Szpakowicz, 2007; Chen et al., 2009; Giouli et al., 2014). Most projects opt though for a word-level, chunk or clause-level annotation (Vidrascu, 2007; Wiebe et al., 2005; Augustyn, 2015; Roman et al., 2015). Finally, sentences containing negation or modality are usually filtered out. As far as the distinction implicit/explicit is concerned, explicit mentions are preferred, pragmatic considerations being context dependent and thus difficult to apply by the NLP systems.

To sum up, there is no unified emotion classification in the existing NLP annotation projects, resources, or tools. Units' delimitation is also variable: word-, chunk- or sentence-level are used to annotate emotions and the adjacent properties. Some of them consider intensity or polarity, but usually negation, modality or pragmatic aspects are neglected (except for Grabar and Dumonet, 2015).

## 4. Annotation Scheme

We opted for creating our own guidelines by combining and adapting existing ones (§ 3.2 and 5.2) to focus on specific aspects of our study. The aim is not to annotate the emotions felt by the patient, nor the interviewer's discourse. Our goal is to capture the variety of ways in which a patient expresses emotions, including when they are attributed to others, denied, or modalized. The annotation is made, beyond the sentence, by trying to limit the influence of the general context and has two layers. The first is an utterance-level annotation, which allows annotators to distinguish between sentences that convey emotion from those that do not (Yes/no answers). This level also encodes information on polarity, intensity, and emotional categories, which are detailed below:

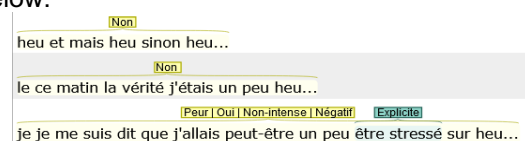


Figure 1: Sentence Layer (yellow) in INCEPTION (Klie et al., 2018)

The last sentence in Figure 1 is annotated for emotion (i.e., as conveying an emotion or not, ‘Oui/Yes’), intensity (‘non-intense’), and polarity (‘Negative’). The emotion category (‘Peur’/‘Fear’) is also tagged.

The second layer identifies the salient emotional expression in the labeled sentences (see Fig. 1 ‘être stressé’ ‘to be nervous’). Emotions are encoded as either explicit or implicit, depending on whether they are directly mentioned or suggested by cues in the discourse. Implicit marks of emotional expression can be further specified as either ‘suggested’ or ‘manifested’ (see below). The diagram in Figure 2 illustrates the different levels and types of annotation:

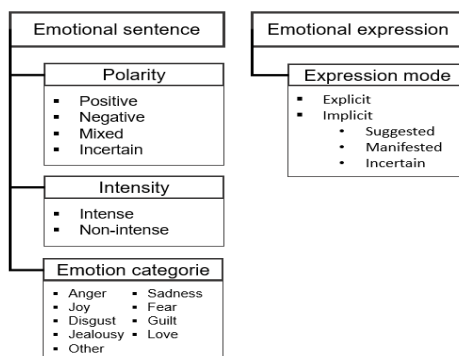


Figure 2: Annotation scheme

When a sentence expresses an emotion, it is then tagged with more specific modalities: polarity (positive, negative, mixed, uncertain), intensity (intense or non-intense), and emotional category. Sentence-level annotation of polarity considers the overall evaluation of the sentence and negation markers invert it. This layer also includes intensity, which can be intrinsic to the linguistic expression used (e.g. ‘terrified’ signals a high level of fear, ‘enraged’ a high level of anger), conveyed with modifiers (‘very angry’) or syntactic phenomena such as repetition and accumulation. The polarity and intensity features are encoded at the sentence-level to capture a range of intensifying phenomena without annotating them explicitly, thus reducing the cognitive cost for the annotator.

The properties of emotional category correspond to a fine-grained annotation of the emotion expressed by the tagged sentence. Eight emotional categories were chosen based on the DBT therapeutic manual given to patients: *anger*, *sadness*, *joy*, *fear*, *disgust*, *guilt*, *jealousy*, *love*. These emotions, their characteristics, triggers, and consequences are discussed in detail during therapy sessions. Five of them (the first listed) correspond to Ekman’s primary emotions. *Surprise* was not kept because it is considered too fleeting. The remaining three are complex, social emotions (§ 3.2). An ‘Other’ choice was also included, permitting the annotator to add new

categories, bringing the total number of categories to nine.

Moreover, the second layer distinguishes between the explicit and implicit way of expressing emotion. Explicit expressions directly designate the psychological states involved using nouns, verbs, adjectives, locutions, and multi-word expressions that refer directly to an emotion. The annotator can use a list of keywords corresponding to the emotional category based on EMOTAIX lexicon (Piolat and Bannour, 2009). To determine whether an expression falls under the explicit category of a designated emotion, we apply the lexicographical criteria cited in Etienne and Battistelli (2021): the expression should be found into a list of keywords associated with an emotion.

Although annotators were encouraged to choose from the list of emotion labels, an open-ended field was also available in case they felt it necessary to add a label (Devillers et al., 2002). We knew that there was a risk of irrelevant answers, but we thought that it was worth trying to account for the possible diversity of emotions based on speakers’ intuition. We will briefly discuss this point in section 6.2.

If the emotional triggers do not meet the above conditions, they are annotated as implicit expressions. There are two types of implicit expressions. First, ‘suggested emotions’ (Etienne and Battistelli, 2021; see also Micheli’s 2014 ‘supported’ emotions) refer to a situation, an event, or a detail which is a socio-culturally accepted reason for the source of an emotion (for example, a funeral is seen as a situation generating sadness). Then, as discussed in Etienne and Battistelli (2021), ‘displayed’ and ‘behavioral’ emotions, are expressed implicitly through actions or descriptions of actions resulting from an emotion. This kind of emotion is called in our scheme ‘manifested’ (Fig. 2). In this case, the inference is made in the opposite direction from the suggested emotion. This allows for the annotation of phenomena such as laughing, crying, or clenching one’s teeth in response to an emotional state. If a fine-grained inferential label of the implied emotion cannot be attached to either category, the annotator can tag the expression as ‘Uncertain’, allowing for the identification of more borderline expressions.

To avoid contextual influences, annotated sentences were taken out of context. Annotators were asked to start annotating transcripts from the end of the transcription to the beginning to limit subjective biases. Additionally, the number of labels for all properties has been reduced to facilitate annotators’ decisions and homogenize annotations. Decision trees were also provided:



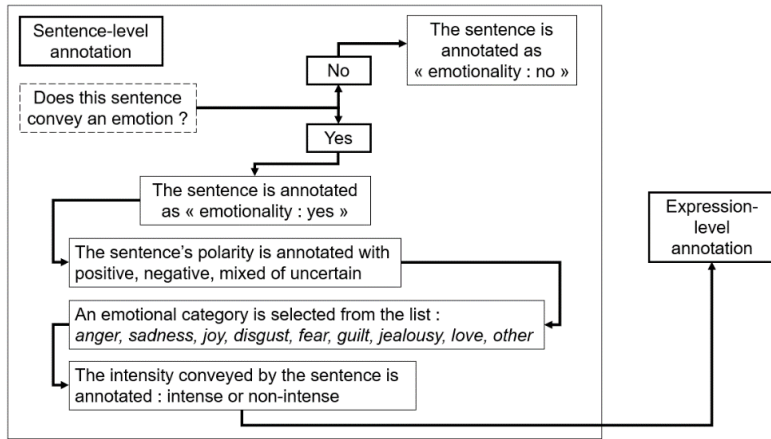


Figure 3: Decision tree for the sentence-level annotation

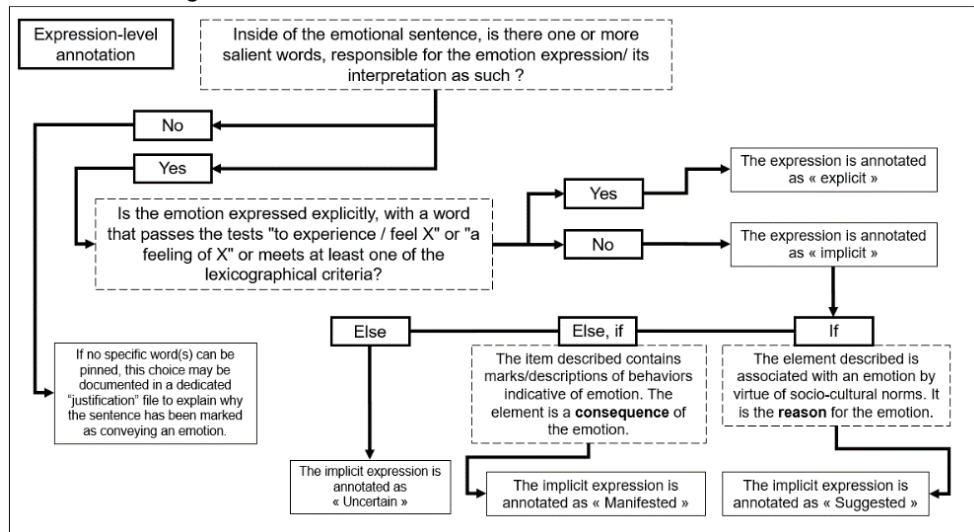


Figure 4: Decision tree for the expression-level annotation

## 5. Application of the Annotation Guidelines

The corpus was annotated by nine expert linguists. All annotators were trained to use the tool INCePTION (Klie et al., 2018) with which the campaign was conducted. Nine of the ten transcripts were annotated by three different annotators, making it easier to resolve any discrepancies in the annotations. The remaining transcript was annotated by all nine annotators and served as a control annotation to assess more accurately inter-annotators' variability, their understanding and proper application of the guidelines. The corpus was annotated in a blind manner, without knowledge of the patient or the recording time. The annotation campaign lasted 3 weeks. In general, the Emotionality task (Yes/no) was easier to perform than deciding on the categories of emotions (§ 5.2).

### 5.1 Corpus Description

The corpus contains transcriptions of interview recordings conducted with patients at the three stages of the protocol (§ 2). The recordings were transcribed following specific XML transcription guidelines that included tags for turns of speech

and disfluency phenomena such as hesitation and repetition. The corpus was then segmented into sentence units using the Whisper automatic voice recognition tool developed by OpenAI (Radford et al., 2022). This tool segments recognized sentences based on prosodic and syntactic parameters. Whisper was solely used for the segmentation step, as it did not perform well on the transcription task due to the unique features related to ABI patients' speech, such as stuttering, long pauses, and non-standard speech rate. The corpus was segmented into sentences by comparing Whisper's output with our manual transcriptions and by incorporating them where the tool added full stops. It was manually corrected when necessary.

To begin the annotation test phase, we first trained the annotators on a transcript extract. We then selected 10 transcripts, totaling 7 hours, 41 minutes and 1 second of recording time. The patients' speech consists of 58,625 tokens. The interviewers' speech is not meant to be annotated as already mentioned. The 10 transcripts include 5 pairs of recordings produced at two different times by the same person. Of these 5 pairs, 2 are control pairs, allowing us to compare

speeches at T0 and T1. The remaining 3 pairs are GREMO pairs of T1 and T2 recordings, i.e., pre- and post-therapy.

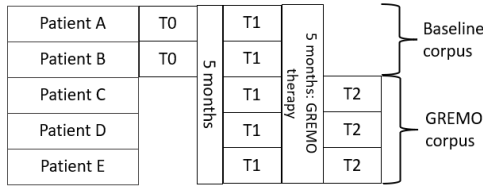


Figure 5: Corpus composition

This distribution of the transcriptions enables us to compare the annotations before and after therapy (§ 2).

## 5.2 Annotation Results

Cohen's Kappa inter-annotator agreement (Cohen, 1960) for the 'Emotionality' (Yes/No) property, which indicates whether a sentence includes emotional information or not, is ranged from 0.4 to 0.74 among all the pairs of annotators, with an average of 0.60. Thus, most annotators agreed on the emotional expression of a segment. At this stage of the study, our primary focus was to compare the ability to consistently recognize an 'emotional' sentence. Nevertheless, we can mention the following scores for the other properties annotated. The 'Polarity' feature ranges from 0.38 to 0.68 among annotator pairs, while the 'Emotional Category' ranges from 0.22 to 0.62. By way of comparison, we can mention Kim and Klingers's (2018) findings who reported agreement ranging from 0.06 to 0.40 for the annotation of 8 emotions out of 1115 sentences.

### Emotionality.

patient code	recording time	sentence number	emotional sentence	proportion of emotional
A	T1	670	260	38,81%
A	T2	477	141	29,56%
B	T1	406	179	44,09%
B	T2	324	158	48,77%
C	T1	291	131	45,02%
C	T2	396	150	37,88%
D	T0	120	27	22,50%
D	T1	291	55	18,90%
E	T0	434	180	41,47%
E	T1	473	208	43,97%
Total		3882	1489	

Figure 6: Overall results for the Emotionality feature

More than 33% of sentences uttered by the patient were annotated as containing an emotional expression. This was expected as the narration of emotional memories naturally leads to emotion expressions. Variability in the proportion of emotional sentence (ranging from 18,9% to 48,8%) reflects differences in narrative strategies among individuals. We hypothesize that this could also serve as a reliable marker for alexithymia, which will be explored in future work.

*Polarity.* Out of the 1489 annotated sentences, 445 (29.9%) were positive, 938 (62.9%) were negative, 65 (4.4%) were of mixed polarity, and 41 (2.8%) uncertain. The sentences marked as 'Uncertain' were mostly related to contexts of incomprehension (e.g. 'I may have felt a lot of emotions, but I can't remember them now') or general discussions about emotions.

*Intensity.* A total of 332 sentences (22%) of the 1489 annotated sentences in the corpus were considered as intense. It is worth noting that, for 4 out of the 5 patients, the number of sentences judged as carrying intense emotional information increased between the 1<sup>st</sup> and 2<sup>nd</sup> corpus recordings (T0-T1 or T1-T2; see section 2).

*Emotion categories.* The emotional categories assigned to the sentences in our corpus expressing emotions are distributed as follows:

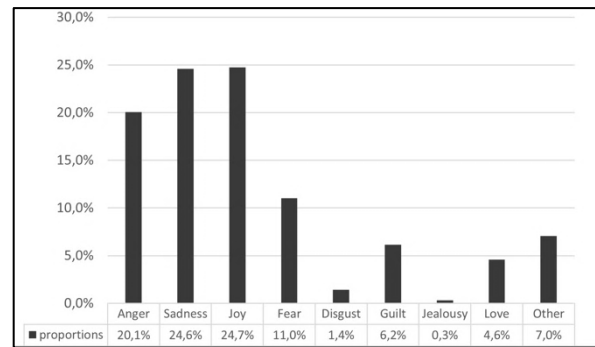


Figure 7: Distribution of the emotional categories

The categories of *disgust* and *jealousy* were rarely used for annotation, while the 'Other' category accounts for more than one sentence out of every 20. This indicates an imbalance between the emotional categories taught to patients during therapy and the concepts expressed during narration, which will be discussed in section 6.2.

*Emotion expression.* Regarding the second level of annotation for the trigger expression of emotional interpretation, we annotated 2054 expressions within the 1489 emotional sentences in our corpus. On average, there were 1.36 emotional expressions per sentence. Four out of five patients showed an increase in the proportion of emotional expressions per sentence between the first and second recording. This might suggest a higher concentration of emotional terms in their narrative, indicating an improvement in the structuring and delivery of the emotional message after the therapy.

*Expression modes.* Out of the 2054 annotations, 804 were explicit expressions of emotions, which accounts for approximately 40%. The remaining 1250 expressions were categorized as 70% suggested (886) and 30% manifested (364).

Although they learn to articulate their emotions more, as their speech is not artificial, it is governed by general pragmatic principles: for instance, it is redundant to say that a funeral is sad.

### 5.3 Related Work: Comparison with other Emotion Annotation Guidelines

Etienne's and Battistelli's guidelines (2021) suggest annotating several elements such as the experiencer, the cause and the consequence of the emotion, the affect relationship or emotional passage schema. Additionally, the guidelines offer to annotate explicit and implicit emotions (and their subtypes).

These guidelines have been selected as the basis of our annotation scheme due to the distinction explicit/implicit emotions and the variety of subtypes of implicit emotions. However, unlike Etienne et Battistelli (2021), we do not annotate the experiencer, the cause, the consequence or the affect relations, due to the spoken clinical nature of our corpus and the specific purpose of our annotation scheme. Besides, contrary to Etienne and Battistelli (2021), polarity and intensity were added at the sentence level to make it easier to take into account the negation and the intensifiers at a higher-level annotation.

Thus, three labels were added for polarity: positive, negative and uncertain. The latter was used to identify contexts in which emotion is mentioned but the narrative is not axiologically marked. Polarity tagging was based on the results reported by Wiebe et al. (2005), who include positive, negative, other and none, and by Vidrascu (2007) who tags polarity as positive, negative and unknown (see Bostan and Klinger, 2018 for an overview). Valence ("the pleasantness of the stimulus"; Warriner et al., 2013) is important when we deal with more complex emotional expressions (e.g. 'conforter'/'to comfort' suggests sadness but a positive polarity). Additionally, annotating polarity helped us capture all cases of denied or modalized emotion. By doing so, we proceeded in the opposite way to Etienne and Battistelli (2021): for a sentence like "Paul n'est pas heureux/Paul is not happy", the authors annotate 'happy' in the 'joy' category, whilst we tag the whole sentence in negative polarity (negation of a positive emotion) and annotate it as 'sadness'. Along with polarity, the intensity of the emotion was tagged in a simpler way than other guides, that is intense or not intense (Augustyn, 2015 and Roman et al., 2015 distinguish two levels of intensity: medium/high and low/non-low). This approach, which does not just rely on the lexicon, allowed for strong charged expressions to be retrieved: e.g. 'Je n'y arrive pas'/'I can't manage it', 'ce n'est pas facile'/'it's not easy', 'c'est trop pour moi'/'it's too much for me', 'Je ne peux plus le faire'/'I can't do

it anymore', 'ce n'est pas la peine'/'there is no point (in doing...)'.

Our two-step annotation is inspired by Giouli et al. (2014). The authors annotate in emotion both at the utterance-level (for an entire sentence, emotion: yes/no) and at the word or the multi-word expression-level (emotion tags). The aim is to obtain a corpus annotated in several levels, with the wider context of what we called the 'sentence\_emo', and the finer context of the expression (word or multi-word) conveying an emotion (Aman and Szpakowicz, 2007). The double level makes it possible to create a corpus of negative examples, with sentences containing no expression of emotion. In the context of a spoken corpus, automatic segmentation is based on pauses. Annotation at the utterance-level allowed us thus to remove truncated and incomplete segments from the annotation.

Finally, contrary to Etienne and Battistelli (2021), our annotation is intended to be context-free. A major contribution of our work relies on the guidelines given to annotators: as mentioned before, annotators were asked to consider each sentence separately, preferably in disorder, without taking the context into account for the analysis. As the patients' transcripts narrate memories of their lives at the first person, which are sometimes difficult and emotionally charged, it is easy for the annotator to fall into empathy and annotate contexts that are too broad because of a possible identification with the patient.

Our approach can also increase our knowledge of many lexicalized turns of phrase specific to oral expression.

## 6. Discussion

### 6.1 Results

This pilot annotation was conducted on a reduced corpus and will undergo further iterations to measure the effects of the therapy. Despite the small size of the corpus, many overall trends can be discussed.

*Emotion categories.* Joy, sadness, and anger are the dominant emotional categories (70% of the annotations). The categories of anger, sadness, fear, disgust, guilt, and jealousy represent 63.6% of the emotional labels used, which is consistent with the negative polarity distribution of our corpus at 62.9%. The remaining discrepancy is due to words like 'conforter/to comfort' or phrases like 'Je ne me suis pas fâché/I didn't get angry', in which the polarity is reversed whilst the emotional tag is not. This distribution of categories was expected, given the nature of our corpus. The themes of accident, disability, and difficulty are central in patients' discourse and usually generate negative emotions.



*Expression modes.* Approximately 60% of the emotional triggers are implicit. More than two-thirds represent suggested emotions, which are prototypical situations associated with the feeling of an emotion. Unlike Etienne et al. (2022), our categories are not equally distributed. The 'Manifested' label combines Etienne and Battistelli's 'Displayed' and 'Behavioral' categories, but it is our lowest frequent mode of expression (17%). Patients mainly use narration of situations to justify an emotion (suggested emotions at 43%) and explicitly express their feelings (40%). This may be an attempt to gain empathy from the listener. The patient may tend to focus on emotional triggers and feelings rather than actions taken in response to those emotions, such as yelling, kicking, or storming out of a room. These manifested emotions may cause the patient to feel exposed, guilty, and out of control.

## 6.2 Difficulties and Adjustments

This first annotation campaign was discussed during a feedback meeting intended to talk about the difficulties encountered by the annotators and to improve the scheme.

One of the main problems was the segmentation of units expressing emotions, especially due to the properties of spoken data, where an emotional expression can be interrupted by hesitations, repetitions, and revisions. Stricter criteria were therefore introduced, including the annotation of the light verb together with the noun of the emotion ('avoir peur/'be afraid', 'me faire peur/'scare me'), the non-annotation of verb auxiliaries and of specifiers (determiners).

The issue of annotating pragmatic markers like phatic elements has also arisen as a specific problem in our spoken speech corpora. The question is whether conversation rituals, such as 'pardon/sorry' or 'pas de problème/no problem' produced in dialogue contexts should be annotated or not as emotion expressions. We decided that if the segment is explicitly a phatic marker and not in a sentence containing an emotional expression, it should not be annotated. Indeed, words with positive or negative connotations, such as 'problem', can be found in non-emotional sentences. In the sentence 'moi qui aime bien le vélo j'ai été faire un tour' ('I like cycling, so I went for a ride'), 'aime bien' ('to like') is a positive subjective predicate, but the sentence expresses an opinion rather than an emotion and therefore it is not annotated.

One of the issues we anticipated (§ 4) was the presence of a free field for adding emotional expressions. The annotators added 32 labels, some of which had interesting aspects, but many of them were also redundant. The categories *jealousy* and *disgust* were largely underused due to misunderstanding of the labels. We changed them to 'covetousness' and 'lassitude' to make

their specificities more comprehensible to annotators. The inter-annotator agreement on Emotion Categories is quite low, due to the large number of emotions and the difficulty of choosing only one label for the segment. One annotator also used the 'Other' category to make double emotion annotations. This modification allows for the specification of 'Mixed' polarities, and it was decided to keep the possibility of double annotation, as in Etienne and Battistelli (2021). This will enable the coding of more complex and nuanced emotions, such as 'dismay', and 'anguish'. The average inter-coder agreement shows that the polarity is a demanding task, mainly because of the difficulty to capture it when negation appears.

Finally, annotators frequently encountered difficulties extracting annotations from the context due to their empathy overtaking them and despite following the instructions to annotate backwards from the text. To address this issue, a unique identifier will be added to each sentence, allowing them to be presented in a random order and limiting the contextual effect.

## 7. Conclusion and Further Work

The annotation scheme presented in this paper is a valuable resource for annotating emotions in French patient narratives, which are currently scarce. It aims at annotating emotion at sentence-level and expression-level. We provide a lightweight and flexible scheme suited for annotating non-standard language constructions. Our approach synthesizes concepts from various previous works, in line with Etienne and Battistelli (2021) and Troiano et al. (2022), creating a cohesive scheme. It is theoretically motivated, as it considers both the linguistic studies of pragmatic and semantic expression of emotion (Plantin, 2011; Micheli, 2014; Wharton and de Saussure, 2022) and the psycholinguistic aspects of it, in the context of the DBT applied to ABI patients.

Since manual annotation is a tedious and time-consuming task, we intend to use the final annotated corpora to fine-tune a pre-trained language model for automatic annotation of our corpus. These annotations will be tested as linguistic markers of the success of the DBT therapy among the recorded patients, hopefully contributing to the relevance of linguistics in the clinical context. To the best of our knowledge, this work is one of the first of its kind in annotating transcripts of patients' narratives, and as such is a pilot contribution to annotating emotion expression in spoken French.

Finally, our corpus will be used in future work for deep learning approaches or distant supervision tasks as training and validation data.

## 8. Acknowledgements

This study was supported by LiLPa (UR 1339) and the FRLC (Fédération de Recherche en Langage et Communication) and uses data from the GREMO-LCA project supported by UGECAM Alsace and by the French Eastern Interregional Group of Clinical Research and Innovation (GIRCI Est; 2021). We wish to thank EMOI-TC patients for their participation in the protocol and for sharing their emotional and behavioral difficulties in the recording.

## 9. Bibliographical References

- Abdul-Mageed, M. and Ungar, L. (2017). EmoNet: Fine-grained emotion detection with gated recurrent neural networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1) 718-728.
- Aman, S., and Szpakowicz, S. (2007). Identifying Expressions of Emotion in Text. In *Proceedings of the 10th International Conference Text, Speech and Dialogue*, p. 196-205.
- Amblard, M., Braud, C., Li, C., Demily, C., Franck, N., and Musiol, M. (2020). Investigation par méthodes d'apprentissage des spécificités langagières propres aux personnes avec schizophrénie (Investigating Learning Methods Applied to Language Specificity of Persons with Schizophrenia). In C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla, and S. Schneider (Eds.), *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2: Traitement Automatique des Langues Naturelles* (p. 12-26). ATALA et AFCP.
- Anscombre, J.-C. (1995). Morphologie et représentation événementielle: Le cas des noms de sentiment et d'attitude. *Langue française*, 105(1):40-54.
- Augustyn, M. (2015). *Annotations des marques de la subjectivité langagière: Discours rapporté, passages entre guillemets et lexique des affects. Manuel de codage. Consulté à l'adresse <https://lidilem.univ-grenoble-alpes.fr/ressources/corpus/marques-subjectivite>*
- Augustyn, M., and Tutin, A. (2009). Constitution d'un corpus annoté autour du lexique des émotions: Collocations et fonctions lexicales. In D. Beck, J. Milićević, and A. Polguère (Eds.), *Actes de la Quatrième conférence internationale sur la Théorie Sens-Texte*. p.25-35.
- Battistelli, D., Étienne, A., and Lecorvé, G. (2022). L'émotion à un niveau textuel: La fonction structurante des émotions observée à partir d'annotations. *Discours*, 30.
- Bhaumik, A., Bernhardt, A., Katsios, G., Sa, N., and Strzalkowski, T. (2023). Adapting Emotion Detection to Analyze Influence Campaigns on Social Media. In J. Barnes, O. De Clercq, and R. Klinger (Eds.), *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis* (p. 441-451). Association for Computational Linguistics.
- Bostan, L.-A.-M., and Klinger, R. (2018). An Analysis of Annotated Corpora for Emotion Classification in Text. In E. M. Bender, L. Derczynski, and P. Isabelle (Eds.), *Proceedings of the 27th International Conference on Computational Linguistics* (p. 2104-2119). Association for Computational Linguistics.
- Chen, Y., Lee, S. Y. M., and Huang, C.-R. (2009). A cognitive-based annotation system for emotion computing. *Proceedings of the Third Linguistic Annotation Workshop on ACL-IJCNLP '09*, 1-9.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Cortal, G., Finkel, A., Paroubek, P., and Ye, L. (2023). *Emotion Recognition based on Psychological Components in Guided Narratives for Emotion Regulation*. In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL2023)* (p. 72-81), May 5, 2023, Association for Computational Linguistics
- Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., and Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (p. 4040-4054).
- Devillers, L., Rosset, S., Bonneau-Maynard, H., and Lamel, L. (2002). Annotations for Dynamic Diagnosis of the Dialog State. In M. González Rodríguez and C.P. Suarez Araujo (Eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. European Language Resources Association (ELRA).
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4), 169-200.
- Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, 5(2), 125-131.

- Etienne, A. (2023). *Analyse automatique des émotions dans les textes : Contributions théoriques et applicatives dans le cadre de l'étude de la complexité des textes pour enfants* [Thèse de doctorat, Paris Nanterre].
- Etienne, A., and Battistelli, D. (2021). *Annotation manuelle des émotions dans des textes écrits avec la plateforme Glozz*. [Research Report]. MoDyCo, Université Paris Nanterre.
- Etienne, A., Battistelli, D., and Lecorvé, G. (2022). A (Psycho-)Linguistically Motivated Scheme for Annotating and Exploring Emotions in a Genre-Diverse Corpus. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (p. 603-612). European Language Resources Association.
- Flaux, N., and Van de Velde, D. (2000). *Les noms en français : Esquisse de classement*. Ophrys.
- Frijda, N. H. (2007). *The laws of emotion*. Lawrence Erlbaum Associates Publishers.
- Gala, N., and Brun, C. (2012). *Propagation de polarités dans des familles de mots : Impact de la morphologie dans la construction d'un lexique pour l'analyse d'opinions*. Actes de la conférence TALN (Traitement Automatique des Langues, Grenoble).
- Galati, D., and Sini, B. (1998). Les mots pour dire les émotions : Recherche sur la structure du lexique émotionnel italien. *Revue de Sémantique et Pragmatique*, 4, 139-161.
- Giouli, V., Fotopoulou, A., and Mouka, E. (2014). Annotating sentiment expressions for lexical resources. In P. Blumenthal, I. Novakova, and D. Siepmann (Eds.), *Les émotions dans le discours / Emotions in Discourse* (p. 281-296). Peter Lang D.
- Grabar, N., & Dumonet, L. (2015). Automatic computing of global emotional polarity in French health forum messages. In *Lecture Notes in Computer Science* (p. 243-248).
- Hsu, C.-C., and Ku, L.-W. (2018). SocialNLP 2018 EmotionX Challenge Overview: Recognizing Emotions in Dialogues. *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media* (p. 27-31).
- Kappas, A., Hess, U., and Scherer, K. R. (1991). Voice and emotion. In *Fundamentals of nonverbal behavior* (p. 200-238). Cambridge University Press.
- Kim, E., and Klinger, R. (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*, (p.1345-1359). Association for Computational Linguistics.
- Klie, J.-C., Bugert, M., Boulosa, B., Eckart de Castilho, R., and Gurevych, I. (2018). The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In D. Zhao (Ed.), *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations* (p. 5-9). Association for Computational Linguistics.
- Kuppelin, M., and Krasny-Pacini, A. (2023). La thérapie comportementale dialectique et la régulation émotionnelle. In M. Kuppelin, I. Tavares, C. Jourdan, and A. Krasny-Pacini (Eds.), *Gérer les émotions en rééducation* (Sauramps medical).
- Lafourcade, M., Le Brun, N., and Joubert, A. (2015). Vous aimez ?... ou pas ? Likelt, un jeu pour construire une ressource lexicale de polarité. In J.-M. Lecarpentier and N. Lucas (Eds.), *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles. Articles courts*.
- Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, 46(4), 352-367.
- Linehan, M. (2015). *DBT skills training manual* (Second edition). The Guilford Press.
- Micheli, R. (2014). *Les émotions dans les discours : Modèle d'analyse et perspectives empiriques*. De Boeck Duculot.
- Öhman, E. (2020). Emotion annotation: Rethinking emotion categorization. *CEUR Workshop Proceedings*, 2865 (p. 134-144).
- Piolat, A., and Bannour, R. (2009). EMOTAIX : un scénario de Tropes pour l'identification automatisée du lexique émotionnel et affectif. *L'année psychologique*, 109, 655-698.
- Plantin, C. (2011). *Les bonnes raisons des émotions*. Peter Lang CH.
- Plutchik, R., (1980). A General Psychoevolutionary Theory of Emotion. In R. Plutchik and H. Kellerman (Eds.), *Emotion, theory, research, and experience: Theories of emotions* (p. 3-33). Academic Press.
- Polguère, A. (2013). Les petits soucis ne poussent plus dans le champ lexical des sentiments. In F.H. Baider and G. Cislaru, *Cartographie des émotions : Propositions*

- linguistiques et sociolinguistiques* (p. 21-42). Presses Sorbonne Nouvelle.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2022). *Robust Speech Recognition via Large-Scale Weak Supervision*.
- Roman, N. T., Piwek, P., Carvalho, A.M.B.R., and Alvares, A.R. (2015). Sentiment and Behaviour Annotation in a Corpus of Dialogue Summaries. *Journal of Universal Computer Science*, 21(4), 561-586.
- Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.
- Sifneos, P. E. (1996). Alexithymia: Past and present. *The American Journal of Psychiatry*, 153(Suppl.), 137-142.
- Suttles, J., and Ide, N. (2013). Distant supervision for emotion classification with discrete binary values. *International Conference on Intelligent Text Processing and Computational Linguistics*.
- Troiano, E., Oberlaender, L.A.M., Wegge, M., and Klinger, R. (2022). x-enVENT: A Corpus of Event Descriptions with Experiencer-specific Emotion and Appraisal Annotations. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (p. 1365-1375). European Language Resources Association.
- Troiano, E., Oberländer, L. and Klinger, R. (2023). Dimensional Modeling of Emotions in Text with Appraisal Theories: Corpus Creation, Annotation Reliability, and Prediction. *Computational Linguistics*, 49(1):1-72.
- Turner, J. H. (2007). *Human Emotions: A Sociological Theory*. Routledge.
- Vidrascu, L. (2007). *Analyse et détection des émotions verbales dans les interactions orales*. [Thèse de doctorat, Université Paris Sud].
- Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191-1207.
- Wharton, T., and Saussure, L. de. (2022). The pragmatics of emotion, argument and conflict. In G.L. Schiewer, J. Altarriba, and B.C. Ng (Éds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science [HSK] 46/1* (p. 664-680). De Gruyter.
- Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2-3), 165-210.
- Wierzbicka, A. (1992). Defining emotion concepts. *Cognitive Science*, 16, 53-581.
- Wundt, W. (1903). Naturwissenschaft und Psychologie. In *Ethik: Eine Untersuchung der Thatsachen und Gesetze des sittlichen Lebens* (W. Engelmann).