On the Relationship between Frames and Emotionality in Text

Enrica Troiano*, Vrije Universiteit Amsterdam, The Netherlands e.troiano@vu.nl
Roman Klinger, IMS, University of Stuttgart, Germany klinger@ims.uni-stuttgart.de
Sebastian Padó, IMS, University of Stuttgart, Germany pado@ims.uni-stuttgart.de

Abstract Emotions, which are responses to salient events, can be realised in text implicitly, for instance with mere references to facts (e.g., "*That was the beginning of a long war*"). Interpreting emotions thus relies on the readers' background knowledge, but that is hardly modeled in computational emotion analysis. Much work in the field is focused on the word level and treats individual lexical units as the fundamental emotion cues in written communication. We shift our attention to the event knowledge they evoke. We leverage frame semantics, a prominent theory for the description of event meanings, and show it is well-suited for the study of emotions: frames build on a "semantics of understanding" whose assumptions rely precisely on people's world knowledge. Our overarching question is if the events that are represented by frames possess an emotion dimension. We hypothesise that they do, and that such a dimension can be distinguished qualitatively for different groups of frames.
To carry out a large corpus-based correspondence analysis, we automatically annotate texts with emotions as well as with FrameNet frames and roles, and we analyse the correlations between them. Our main finding is that substantial groups of frames have an emotional import. With an extensive qualitative analysis, we show that they capture several properties of emotions that are purported by theories from psychology. These observations contribute to advancing the two strands of research that we combine: emotion analysis can profit from the event-based perspective of frame semantics; in return, frame semantics gains a better grip of its position vis-à-vis emotions, an integral part of word meanings.

1 Introduction

Human life is interwoven with emotions. They echo in our brain, body, behaviors, and attract for this reason a diverse range of disciplines (Barrett et al., 2016, Part I). Psychology, among others, has entered a century-long endeavor to explain how emotions arise, with appraisal theories (Smith and Ellsworth, 1985, i.a.) providing a viewpoint that is widely accepted today: emotions are *responses to (internal or external) events*, specifically to circumstances *evaluated* as salient by their experiencers (Scarantino, 2016). Understanding how humans evaluate events is thus fundamental to discuss this affective phenomenon, and appraisal theories offer many fertile insights on the matter. They spell out, for instance, some human reactions to events, like neurophysiological changes, motor expressions and motivational tendencies (Scherer, 1989). From the perspective of an observer, these hint at what other people feel: the blushing on one's cheeks might reveal an episode of shame, the raising of a brow could indicate disappointment.

Emotions also pervade the sphere of verbal communication, where an observer infers the mental state of others by interpreting their utterances. Decoding emotions from words is key to successful communication, since emotions represent an important aspect of the meaning that speakers and writers intend to convey (Scheff, 1973). This is the idea that fuels (computational) emotion analysis in natural language processing (Canales and Martínez-Barco, 2014), a research field geared towards the creation of systems that sense emotions like humans do. Emotion analysis mainly approaches its task as text classification. It models the import of verbal expressions either as discrete categories, like anger and joy, or through scalar features such as valence and arousal (Nandwani and Verma, 2021). A central challenge in this regard is that emotions are expressed in language in a myriad of ways. At times they emerge explicitly, with words that point to an emotion state by definition (e.g., "*I'm happy*"). Other times, however, emotions can be expressed without unequivocal cues, mental states or evaluative attitudes: writers can describe a stimulus event (e.g., "*my granddad died*", "*my team won the match*", which likely spark sad-

* The work was carried out while the first author was affiliated with IMS, University of Stuttgart, Germany.

ness and joy), or their reaction to it (e.g., "*I cried*", "*I smiled*"), trusting that the correct emotional interpretation of their production will be drawn by the readers via pragmatic inference (Grice, 1975).

How can emotions be associated with such factual statements? Psychology explains the link via empathy and affective role taking (Mehrabian and Epstein, 1972; Eisenberg and Miller, 1987; Omdahl, 1995), and natural language processing connects emotion decoding more directly to world knowledge. Its starting point is that words possess specific connotations in the collective imagination (Clore et al., 1987) – e.g., *die*: sadness, *win*: joy, *ghost*: fear. Accordingly, it stores such connotations as dictionaries of word-to-emotion associations (Strapparava and Valitutti, 2004; Mohammad and Turney, 2013).

Word-level dictionaries leverage the assumption that individual words are the crucial, emotion-revealing linguistic units. This view is practically useful, but it neglects an important point, namely the impact of the context in which words occur, and thus the paradigmatic and syntagmatic information that allow people to infer emotion meanings. For instance, the surrounding verbal context of "*boiling*" helps disambiguate if this predicate refers to a heat reaction with a nonemotional tone ("*the water is boiling*") or to an emotional turmoil ("*she is boiling with anger*"). Much work in emotion analysis disregards this type of background knowledge. Approaches that embed emotion meanings into latent vector spaces (Felbo et al., 2017; Li et al., 2017, i.a.) capture contextual information, but they are less transparent to investigation than lexical methods.

In this article, we consider *frame semantics* (Fillmore, 1982) as a source of lexical abstractions that is appropriate for specifying emotions in a dictionary. Frame semantics proposes a formalism (viz., frames) and a practical resource (Berkeley FrameNet, (Baker et al., 1998)) to describe linguistic meanings with a combination of predicates (i.e., frames) and arguments. This "semantics of understanding" or U-semantics (Fillmore, 1985) explains the difference in meaning between "*the water is boiling*" and "*she was boiling with anger*" in terms of reference to two different frames that are evoked by the sentences, respectively. This frame-level disambiguation arguably makes use of knowledge about how the world is organised that is necessary to recognise which of the two sentences is emotional. It also suggests that frames bear a potential value for studies in emotion analysis, even though they are usually dismissed in the computational study of emotions.

We believe that there are many affinities between emotions and frames. Not only does FrameNet dedicate multiple frames to emotions (e.g., EMOTION_DIRECTED and EMOTION_OF_MENTAL_ACTIVITY), but it pays attention to events, similar to appraisal theories. Figure 1 illustrates this point: frame semantics focuses on abstractions of real-life situations (frames) determined by the structural properties of an event portrayed in text; appraisal theories study emotions as responses to events, whose properties are evaluated by the event participants. The primacy of events in both domains implies that verbal descriptions of emotion-triggering events (e.g., "*my team won*") can be represented by frames. Other emotion expressions can report (frame-evoking) events as well, from the assessment of the stimuli ("*that's great*"), to the occurrence of affective experiences and related reactions ("*I'm happy*", "*I'm all steamed up!*").

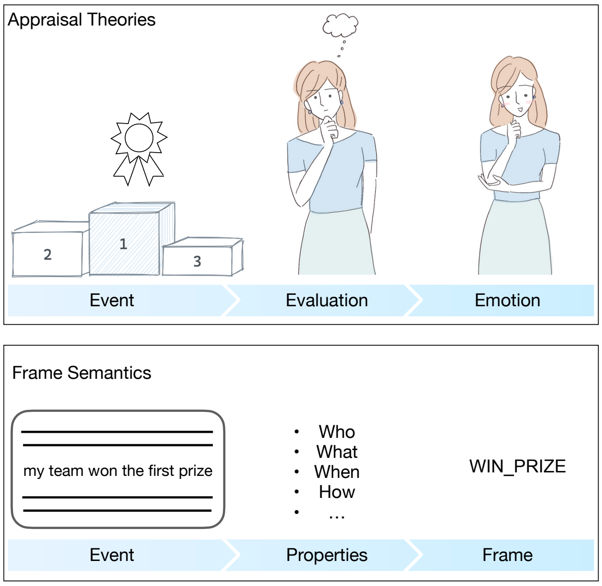

Figure 1: A comparison between the two fields we tap on. Frame semantics studies texts by focusing on events and their characteristics. Appraisal theories, interested in how emotions emerge in humans, also start from the consideration of events; they pay further attention to how event characteristics, as evaluated by individuals, lead to specific emotion reactions.

Based on this parallel between the two blocks in Figure 1, this article investigates the relationship between frames and emotions. As a first step, we refrain from analysing different emotions, and concentrate our attention on *emotionality*[1], i.e., whether a text has an emotion content, irrespective of what it is. We ask: are FrameNet frames associated with emotionality? Our expectation is that emotionality represents an integral part of more frames than those indicated in FrameNet as emotion-related ones. By borrowing the definition of emotional experiences from appraisal theories (i.e., emotions as processes engaging events, event evaluations, personal reactions), and assuming that all such

[1]We will use "emotion" and "emotionality" interchangeably.

diagnostic features can be communicated via language (e.g., events: "*my team won the match*", evaluations: "*victory was well deserved*", reactions: "*I'm happy*", "*I'm all steamed up!*"), one can conjecture that many frames that are apparently affect-less correspond in fact to the conceptualization of some emotion components. Verifying our conjecture is relevant from two complementary perspectives. For researchers in emotion analysis, we put FrameNet up to scrutiny as a suitable tool to tackle the emotional import of sentences. This could provide insights into the linguistic level at which an affective meaning comes to actualise (e.g., in the relation between words rather than words in isolation), and guide the field towards better automatic text interpretations. For frame semanticists, on the other hand, we inspect whether emotions are an underlying component of the meaning of frames.

At the methodological level, we avoid making assumptions as to which frames are emotional, but exploit an automatic procedure to identify them at scale. We start from a large unlabelled corpus of contemporary American English, on which we add two independent layers of automatic annotation, to label sentences both with binary emotion categories and the frames that they evoke. Then, by investigating the mutual information between the two, we provide evidence that emotionality is an important aspect of frames (by association, not by definition). Besides frames with no emotional import (e.g., STORING) and frames that are associated with some degree of emotion (e.g., CURE), FrameNet includes a substantial group of strongly emotional frames. Among these are instances evoked by unambiguously emotional predicates (e.g., EMOTION_DIRECTED, FEAR), and others expressing strongly emotionally loaded events (e.g., DYING), bolstering our perspective on the affective dimension of language as described by appraisal theories. As a concrete result of our analysis, we release a resource[2] with frames-to-emotion associations that can be employed in alternative to typical word-to-emotion lexicons.

The paper starts with an overview of relevant fields. Section 2 introduces emotions, with a focus on how they are studied in text, and Section 3 describes FrameNet, in relation to emotions and the task of semantic role labelling. Section 4 presents the experimental setting used to address our research question. Our main contribution is presented in Section 5, which also elaborates on a possible grouping of the FrameNet emotion vocabulary with a qualitative analysis grounded in appraisal theories, followed by an extensive discussion of its implications in Section 6. We conclude with a summary of the present work and indicate viable ventures for future research.

[2] Available at `http://www.ims.uni-stuttgart.de/data/FrameEmotionalityMapping`.

2 Emotion in Language

Emotions in Psychology. The body of psychological literature on emotions is extensive and controversial. The field has long established that these states can be investigated systematically (cf. Dixon, 2012, p. 338), but it has reached little consensus on the details, specifically concerning what emotions are, and whether (and which) can be considered cross-cultural universals. Several theories focus indeed on diverse sets of emotions, motivated by specific views on their evolutionary relevance (Ekman, 1992; Plutchik, 2001), or on their underlying dimensions (Russell and Mehrabian, 1977). Ultimately, however, different research lines agree on one point. There exists a handful of "diagnostic features" which indicate that an emotion is taking place (Scarantino, 2016): typically, a starting cause is there (e.g., an event happens); it is evaluated by its experiencers; and it sparks in them some concrete effects, like changes in their voice and posture.

To organise these observations, appraisal theories study emotions in terms of sets of evaluations (Moors et al., 2013; Scherer, 1984). When a stimulus presents itself to an individual, it is evaluated (i.e., appraised) in relation to the individual's goals, beliefs and desires. For this reason, an appraisal corresponds to specific effects – if I win the competition, I might smile and feel a pleasant sensation because winning supports my well-being; my opponent likely does not have the same reaction. Such effects involve various subsystems, all of which are engaged in an emotion process together with the cognitive appraisal. They consist of a neurophysiological component (i.e., bodily symptoms, like heart beating faster), a motor component (i.e., facial and vocal expressions), a motivational component (i.e., action dispositions), and a subjective feeling component (e.g., winning the competition feels good) (Scherer, 2005).

From psychological research, we retain the idea that an emotion episode involves at least three aspects that can mirror in language: emotion stimuli (i.e., what happens), evaluations (how that is assessed in the light, e.g., of who initiates or is affected by the stimulus), and reactions (e.g., bodily manifestations of emotions).

Emotions in Linguistics. Since emotions are not a primarily linguistic phenomenon, they have remained outside the scope of much work in theoretical linguistics (Kiefer, 1988). Searle's pragmatic framework (1976), for instance, touches upon expressive acts that convey feelings and attitudes, but it lumps emotions together with multiple other aspects of social interaction.

A more direct account of this phenomenon is given by Martin and White (2003). Tapping into the framework of Systemic Functional Linguistics, they analyse emotion expressions in language, and conclude that evaluations play a central role. Such evaluations emerge from descriptions of qualities of entities, through modal adjuncts that reflect the position of writers towards an event (e.g., "*sadly, ...*"), through communication of behavioural processes (e.g., "*he smiled at him*"), as well as mental (e.g., "*he liked him*") and relational ones (e.g., "*he felt angry at him*"). Hence, theories of appraisal, both in psychology and in language, converge on the consideration of embodied manifestations of emotions – either in real life or through language.

Emotions in NLP. The examples above illustrate the data of interest for computational emotion analysis, whose chief task is to classify emotions from text. Works in the field face the choice of following one psychological theory. The selection is usually based on both the textual domain under consideration, as well as its match to the emotions documented by the considered theory. Some opt for dimensional models. Accordingly, they map linguistic data into a continuous space (Preoţiuc-Pietro et al., 2016; Yu et al., 2016; Buechel and Hahn, 2017), like the space comprising the dimensions of valence, arousal and dominance (Russell and Mehrabian, 1977). Others rely on discrete emotion models (e.g., Ekman, 1992; Plutchik, 2001). They associate text to categories like *anger*, *disgust*, *sadness*, either at the sentence-level (Felbo et al., 2017; Li et al., 2017; Schuff et al., 2017) or at the word-level (Mohammad and Turney, 2013; Strapparava and Valitutti, 2004). The latter strand of research leverages the idea that part of a language vocabulary can be described in terms of its emotional meaning (Clore et al., 1987; Hobbs and Gordon, 2011) in order to create affect-oriented lexicons, i.e., resources that formalise the link between emotions and a specific language (Buechel et al., 2020; Chen and Skiena, 2014), encompassing words with an emotion denotation (e.g., the noun *joy*) as well as words with an emotion connotation (e.g., *party*⟶*joy*).

Only a few works have brought psychological concepts to bear on NLP on a more fundamental level than the acquisition of sets of labels that should be looked for in text (Balahur and Tanev, 2016; Shaikh et al., 2009; Udochukwu and He, 2015, i.a.), and they have rarely relied on a concept of emotions as processes involving complex evaluations (exceptions are Hofmann et al., 2020; Stranisci et al., 2022; Troiano et al., 2023).

Our work differs from previous studies in emotion analysis (Abdul-Mageed and Ungar, 2017; Felbo et al., 2017; Demszky et al., 2020, e.g.,) in various respects. We study emotionality instead of a fine-grained set of emotions; we analyse if the emotion information is contained in a well-established resource for semantic role labelling; and we bring together for the first time in the field a theory of emotions (appraisals) with a theory of semantics (frames).

3 Frame Semantics

FrameNet. The theory of frame semantics fundamentally assumes that utterances are understood via frames (Fillmore, 1982). A frame represents a situation fragment that serves to match a word (or a group thereof) to the bundle of knowledge it presupposes (Ruppenhofer et al., 2016). For instance, the term "*abandon*" evokes a conceptual category instantiated by different events (e.g., leaving a membership group, or metaphorically, quitting a bad habit) which comprise a series of participants (e.g., the group being left, the person dropping out of it). The corresponding frame, ABANDONMENT, binds together these bits of knowledge.

For English, the Berkeley FrameNet project (Baker et al., 1998) has been curating the lexical resource FrameNet. It provides an inventory of predicates (lexical units), roles (arguments), and frames. Its latest release (FrameNet 1.7) counts over 13k lexical units and 1.2k frames, which connect to one another via specific frame-to-frame (f2f) relations such as INHERITANCE, SUBFRAME, or USING (Fillmore et al., 2004).

An example for the frame ABANDONMENT from the database[3] is in Table 1. ABANDONMENT can be evoked by verbs (boldfaced in the example sentences (1), (2) and (3)) but also by other lexical units such as adjectives and nouns. It has the roles of AGENT and THEME representing the "frame elements" that participate in the situation, where the former expresses the entity leaving the latter. Moreover, this frame links to INTENTIONALLY_AFFECT via an INHERITANCE relation. That is, it inherits properties from this broader conceptual class, and can thus be considered a specific kind of INTENTIONALLY_AFFECT situations.

Frame Identification and Emotion Analysis. In addition to the frame database, FrameNet comprises sentence annotations, like the examples (1), (2), and (3) in Table 1. Such annotations have been used for semantic role labelling (SRL), a task aimed at identifying and labelling the semantic roles that the arguments of a predicate (operationalised as word spans) fill with respect to the event expressed by the predicate (Gildea and Jurafsky, 2002; Màrquez et al., 2008). The specific set of roles depends on the adopted model. Other than FrameNet, PropBank (Palmer et al., 2005) and Abstract Meaning Representation (Banarescu et al., 2013) are commonly used options.

A number of such systems for FrameNet-based SRL have been made available as off-the-shelf tools. Among them are the role labeller that leverages sentence and discourse context by Roth and Lapata (2015), the probabilistic models of Das et al. (2010) which use latent

[3]Frame definitions can be found at: `https://framenet.icsi.berkeley.edu/fndrupal/frameIndex`.

Frame: ABANDONMENT	
Definition	An Agent leaves behind a Theme effectively rendering it no longer within their control or of the normal security as one's property.
Lexical Units	abandon.v, abandoned.a, abandonment.n, forget.v, leave.v
Elements	Agent, Theme
F2F relations	Inherits from: INTENTIONALLY_AFFECT
Example Sentences	(1) Perhaps [he $_{\text{Agent}}$] **left** [the key $_{\text{Theme}}$] in the ignition. (2) [She $_{\text{Agent}}$] **left** [her old ways $_{\text{Theme}}$] behind. (3) **Abandonment** [of a child $_{\text{Theme}}$] is considered to be a serious crime in many jurisdictions.

Table 1: Example of a FrameNet frame. In the three example sentences, boldfaced words are frame-evoking predicates, bracketed words are arguments.

variables of lexical-semantic features to facilitate frame predictions for unknown predicates, and the labeller of Swayamdipta et al. (2017) that detects FrameNet frames and frame-elements.

Frame-based semantic parsers have proven useful in applications like text-to-scene generation (Coyne et al., 2012) and question answering (Shen and Lapata, 2007). Yet, they have never been fully leveraged to address emotions. For example, Ghazi et al. (2015) annotated 820 FrameNet sentences with emotions, but these were sampled based on their link to only one emotional frame (i.e., EMOTION_DIRECTED). On the other hand, the research line in emotion analysis centered on semantic roles (Mohammad et al., 2014; Oberländer and Klinger, 2020; Oberländer et al., 2020) identifies the portions of texts corresponding to emotion causes, emotion holders, and eventually, the targets towards which an emotion is directed, but it disregards frames.

Being the first study that links frame semantics and emotion analysis, we concentrate on frames and leave roles aside. These have an important function which we use implicitly as means that help identify frames in context. For example, they provide a cue that the conceptual situation evoked by, e.g., the predicate "*treats*" in "*the doctor treats the patient with aspirin*" can be distinguished from that in "*the bully treats the student with disdain*", but we leave the specific analysis of the relationship between roles and emotions for future work.

Emotions in FrameNet. Frames appear to be a valuable formalism to study emotions because FrameNet has an affective core: a small part of the database is ostentatiously concerned with emotions (e.g., FEAR), and some of the others can be traced back to a relevant emotion frame through the relations present in the database – for instance, FLEEING can be related to the FEAR frame via the USE relation (Ruppenhofer, 2018). Past research has indeed provided qualitative evidence of the emotional quality of various frames (Ruppenhofer et al., 2016), but it has done so by focusing on a limited and pre-defined vocabulary of items. In fact, the exact set belonging to the emotion domain is not spelled out, partly because FrameNet is a database under constant development, and partly because emotional meanings are only one type of the world knowledge inferences that can be made from frames – representing all of them would be unfeasible for the FrameNet curators. Our approach can identify them automatically and at large.

In his manual analysis of the emotion domain in FrameNet, Ruppenhofer (2018) discusses the criteria that guided the allocation of lexical units under specific frames. Some of them are the constraint that the lexical units in a frame should accept the same types and number of syntactic dependents, and the idea that specific frames are differentiated by the role of subject/object that is filled in by an emotion participant (EXPERIENCER_SUBJ/EXPERIENCER_OBJ). According to such criteria, words that indicate different emotions can fall within the same group of predicates. Conversely, words with the same lexical root are allowed to be part of different frames (e.g., the verb "*anger*" belongs to EXPERIENCE_OBJ together with the verb "*please*", but the adjective *angry* does not). There is in this sense a crucial difference between a dictionary-based approach to emotion analysis and a frame semantics one: The latter organises emotion words by reflecting similarities between their linguistic realisations, more than to account for their glossary characterisation.

While we employ frames as a way of grouping words, one could opt for other semantic word organisations to study the affective dimension of meaning. For example, WordNet (Fellbaum, 1998) arranges words into a large network of relations potentially useful for our goal. However, FrameNet has an important advantage over other lexical databases. Its construction principle is not focused on words per se but on the frames that these evoke, as (interrelated) classes of events (Baker and Fellbaum, 2009). This allows to capture the emotional closeness between words that might be far apart in regards to their grammatical classes and meaning (e.g., the noun "*pleasure*" and the verb "*abhor*"), but which belong to the same event class in FrameNet (e.g., both "*pleasure*" and "*abhor*" are lexical units of EXPERIENCER_FOCUSED_EMOTION).

4 Methods

Our goal is to study (a) to what extent (i.e., quantitatively) the emotionality of texts is mirrored in the frames that the texts evoke; (b) if there is a qualitative difference between the emotionality that frames carry; and whether (c) these aspect can help in starting a discussion of emotions in FrameNet. FrameNet contains a narrow emotion nucleus, but for most of the frames their 'emotionality status' (whether or not the situation is emotional) is not specified. This constitutes the core of our investigation.

Accessing data with the two types of information that we need is not straightforward. No resource for emotion analysis is labelled with frame semantics information, except for the dataset by Ghazi et al. (2015), which is limited in size and only includes emotion-bearing texts. Likewise, corpora for frame-semantic parsing do not contain emotion annotations – at least, not for the vast majority of frames. As a solution, we devise a method that combines the use of neural technologies and prior knowledge about language as contained in FrameNet: we correlate the categorical variable of emotionality (obtained through an emotion classifier) with that of frame membership (grasped by a frame identification tool).

We use this correlation to find categories of frames (inherently emotional, inherently nonemotional, and others) and to explain their belonging to one category or another in quantitative terms. Focussing on the emotional frames, we conduct a qualitative discussion based on Scherer's theory (1984), which explains emotions as processes involving the subsystems of an organism (cognitive, motivational, motor, etc.), and has a theoretical counterpart in linguistics.[4]

Data. We base our study on an unlabelled corpus, the 2020 version of COCA[5] (Davies, 2015), which is much larger than any existing resource for emotion analysis.[6] Its texts were collected from 1990 to 2020 in different domains, namely blogs, magazines, newspapers, academic texts, spoken interactions, fiction, TV and movie subtitles, and webpages. Except for academic texts, which have an arguably impartial language, we consider all other domains, split their paragraphs into sentences, exclude sentences containing words that are masked for copyright reasons and those with less than 3 tokens (tokenization performed with the python library *nltk*[7]). The preprocessed data that we use comprises ~44M sentences and ~536M tokens.

Bridging Data-driven Learning and Semantic Resources. To obtain frames and emotion information, we bypass the use of human annotation which would be prohibitively expensive. We resort instead to an automatic procedure, adopting a two-step methodology illustrated in Figure 2. First, texts are associated with emotion labels (through an emotion classifier) and frames (via a tool for frame identification); second, we carry out a corpus-based correlation analysis where the association between the two annotation sides is quantified and interpreted.

Because this approach exposes us to the risk of mistakes made by the emotion classifier and the frame identifier, we adopt experimental design strategies that boost the robustness of our empirical observations.[8] One is to employ a corpus with a considerable number of datapoints, which showcase a variety of linguistic realizations of emotions, and evoke frames across both emotion-bearing and nonemotional expressions. Second, we carry out the emotion annotation with classifiers learnt on multigenre data, a strategy that promotes the generalization ability of emotion detection models (Tafreshi and Diab, 2018); for frame labelling, we use an artificial neural network-based technology that has shown to generalise well over unseen sentences and predicates (Swayamdipta et al., 2017). Third, we evaluate the emotion classifier against a manually-annotated sample of our texts as an additional check of

[4]There exist also other appraisal-based theories, like the OCC model (Ortony et al., 1988) which describes the eliciting conditions of emotions (i.e., consequences of events, agents' actions and aspects of objects), how these are appraised along binary criteria (e.g., desirability–undesirability), and how specific evaluations cause emotions deterministically (e.g., if a condition holds, a certain reaction follows). Yet, the OCC model does not fit our goal. It sees emotions as descriptive structures of prototypical situations, and its binary evaluations, which are purely conceptual constructs, have little to do with the linguistic expression of events. By contrast, the tool that we use for event representation, frame semantics, is primarily linguistic and might not match the conceptual considerations of the OCC.

[5]`https://www.english-corpora.org/coca/`

[6]An overview of existing resources in computational emotion analysis can be found in Bostan and Klinger (2018).

[7]`https://www.nltk.org`

[8]We discuss the limitations of our approach in Appendix A.

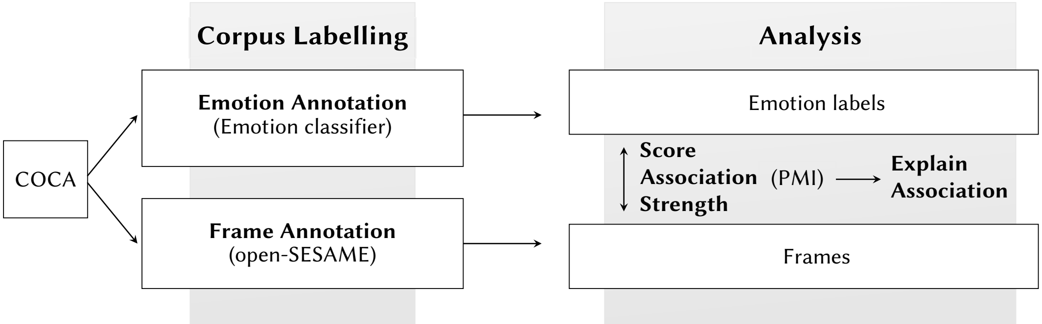

Figure 2: Our two-step experimental setting. Corpus Labelling: automatic annotation of sentences extracted from the corpus of contemporary American English with emotions and frames, separately, with the emotion classifier being evaluated on a subset of the corpus previously annotated by human judges, and the tool for frame identification evaluated on a subset of MASC as out-of-domain data. Analysis: the two strands of annotations are brought together via PMI, to first score and then explain the association between frames and emotionality.

its reliability, and we do the same for the frame identifier using out-of-domain data. Lastly, we conduct statistical analyses to limit the role of chance in positing frame-emotion associations, and we explain them qualitatively as a safeguard of the quality of our findings.

We now proceed to describe the individual components shown in Figure 2.

4.1 Corpus Labelling

As a first step, we label texts with emotion- and frame-related information. The systems used here are trained separately on different corpora. It is thus necessary to assess their domain independence and get insight into how well they apply to COCA.

Emotion Classification. We start by gathering various resources for emotion analysis that span textual domains similar to those in COCA, from webpages to literary texts: GoEmotion (Demszky et al., 2020), Grounded-Emotions (Liu et al., 2007), EmoInt (Mohammad and Bravo-Marquez, 2017), TEC (Mohammad, 2012), SSEC (Schuff et al., 2017), enISEAR (Troiano et al., 2019), ISEAR (Scherer and Wallbott, 1997), Tales (Alm et al., 2005), DailyDialogs (Li et al., 2017), and Emotion-Stimulus (Ghazi et al., 2015). These datasets feature diverse emotion schemata; we make them consistent to our binary setup by mapping their original labels into the *nonemotional* and *emotional* classes, depending on whether a text was marked as having no emotion, or as having one out of a rich set of alternatives (e.g., *joy*, *fear*, *disgust*, *hope*, *surprise*, *guilt*).

Instead of extracting our test set from this data, we use a portion of COCA. Made available by Troiano et al. (2021), the sample contains 700 texts labelled at the sentence level by three in-lab raters.[9] They are balanced across the domains that we consider, and their annotation encompasses the same binary categories of our concern. The nonemotional label corresponds to the absence of any emotion content, the emotional class represents sentences that display either of two qualities: (1) having an emotion as a central component of their meaning, thanks to the presence of an emotion word (as in "*I am so happy to see you*") or the description of an internal state of an entity ("*And there she was, desperate for her family*"); (2) describing an event, a concept or a state of affairs to which the annotators would personally associate an emotion ("*She was being pretty arrogant to me*", "*I saw my best friend*"). The annotators were tasked to judge the texts by giving their own emotion reaction, and not to try and reconstruct that of the text authors. Thus, they were allowed to associate similar events to different labels. For instance, the passing away of an unknown entity could be linked to a nonemotional judgment, while that of a person resonating with their own experience (e.g., the mention of a pet) could receive the opposite label.

Next, we train multiple models on the concatenation of the selected (training) resources: we fine-tune BERT (Devlin et al., 2019) models[10], adding a classification layer that outputs the labels *emotion* or *nonemotional*. Different models are obtained by varying the data on which they learn the classification task: the rationale is to identify a subset of training resources that yields a classifier capable of reliably judging out-of-domain data (i.e., COCA). Hence, we evaluate each model on the manually annotated COCA sample, with the majority vote determining the ground truth.[11] We

[9] `https://www.ims.uni-stuttgart.de/data/emotion-confidence`

[10] `https://huggingface.co/docs/transformers/`

[11] Associating the 700 sentences to the majority vote resulted in 474 emotional and 226 nonemotional data points. Cohen's κ (1960) agreement between this ground truth and the three annotators was .6, .8, .6, respectively. The annotators' decisions were unanimous for 304 emotional and 88 nonemotional instances.

	Frame Id		
	P	R	F1
FrameNet 1.7	.85	.85	.85
MASC	.78	.78	.78

Table 2: Evaluation of the frame identifier provided by Swayamdipta et al. (2017) against FrameNet data and MASC frame-annotated data.

pick the model that performs best on this test set to annotate the rest of the corpus. It reaches a performance of .67 F1 score[12]) Details on model selection are in Appendix B.

Frame Identifier. Models and corpora for semantic role labelling are scarcer than emotion-centered ones. Here, we require a system which, given a sentence, identifies the set of FrameNet frames that are evoked by each of the predicates, as well as the corresponding predicate arguments. To this end, we use open-SESAME[13]. Developed by Swayamdipta et al. (2017), it is a freely available interpreter for SRL with state-of-the-art performance, based on segmental recurrent neural networks (Kong et al., 2016). We re-train the provided implementation[14] using the sentences from the FrameNet release 1.7 (7340 for training, 387 for dev, and 2420 for testing).

We evaluate it on the FrameNet test set as in-domain data, as well as on external data. For that, we use 695 sentences (516 of which are frame-evoking) coming from MASC[15] (Ide et al., 2010), a subset of the Open American National Corpus that provides useful annotations for frame identification. MASC's texts include emails, essays, fiction, spoken transcripts, and hence, using it as a benchmark illustrates how the frame identifier performs on linguistic expressions similar to those found in COCA.

Precision, recall, and micro-averaged F1 for this frame identification task (Frame Id) are reported across both test sets in Table 2. We obtain these results using the script by Swayamdipta et al. (2017) on the full-text FrameNet annotation. When moving to out-of-domain data, we see a drop in performance (from F1=.85 to F1=.78), which might be partially due to an increase in the sentence length (avg. for the FrameNet test = 16.5 tokens, for the MASC test = 23.4 tokens) and in the average number of frames per sentence (2.8 for FrameNet, 6.5 for MASC). Still, we take these numbers to be sufficiently high that the frame identification system can be used to proceed with the annotation.

4.2 Analysis: Investigating Emotionality in Frames

Once COCA is labelled with emotionality and frames, we can finally proceed to our research question: are FrameNet frames associated with emotionality? Estimating the degree of this association requires an appropriate alignment strategy, as the labels we obtained differ in granularity: emotions refer to entire sentences, while the output of the frame parser relates to tokens. We choose the most straightforward alignment strategy: considering each frame in a sentence as having a separate and full-fledged alignment with the sentence-level emotionality label. This choice is a simplification, because the frame parser could identify multiple frames for an input sentence, and emotionality might be attributed to their inter-relation rather than their individual contribution. However, this is a transparent approach, comparable to related work such as aspect-based summarization in sentiment analysis (Hu and Liu, 2004), where multiple aspects identified at the sub-sentence level are grouped under the same sentiment label. The literature offers various weighting schemes to refine such alignments, but not all weighting schemes work equally well for all tasks (Buckley, 1993; Pekar et al., 2004; Ushio et al., 2021).

To identify patterns of frames occurring with emotionality status (emotional/nonemotional), we compute pointwise mutual information (PMI) (Church and Hanks, 1990). This information-theoretic measure quantifies the dependence between the values that two discrete random variables can take, and accounts for their chance co-occurrence. More specifically, PMI compares the probability of observing two variables together, against that of observing them independently, or by chance. In our case, the variables are the output labels of the automatic annotation procedure from the corpus labelling step. For each pair (f, e) consisting of a frame and an emotionality label, we estimate PMI as the number of times that such frame and emotionality label co-occur in the entire corpus, divided by the product of their individual frequencies. Formally, for each f and e, we compute

$$\mathrm{PMI(f;e)} = \log_2 \frac{\mathrm{p(e, f)}}{\mathrm{p(e)p(f)}} = \log_2 \frac{\mathrm{p(e \mid f)}}{\mathrm{p(e)}}.$$

As already mentioned, the number of extracted pairs (frame f, emotionality e) varies from sentence to sen-

[12]This performance is not state of the art in emotion classification. However, systems for emotion detection that work well on existing labelled resources might not perform equally well on COCA. We varied the model architecture and noticed that a model that achieved better results on in-domain data suffered from major performance loss when evaluated on the manually-annotated subsample of COCA. See Appendix B for a discussion of these classification results.

[13]`https://github.com/Noahs-ARK/open-sesame`

[14]Training hyperparameters as in Swayamdipta et al. (2017).

[15]Downloadable at: `https://www.anc.org/MASC/download/MASC-1.0.3.tgz`

	Emotional	Nonemotional
Sent. with frames	19.717.813	16.092.214
Sent. w/o frames	4.194.783	2.141.299
Number of frames	75.889.290	57.517.465

Table 3: Outcome of Corpus Labelling: number of sentences associated with the emotion and nonemotional labels, both with frames and evoking no frames, and number of frames.

tence, depending on whether one or many frames are evoked.

PMI does not have predefined bounds. Positive values indicate that a frame and an emotion connotation are semantically associated: they appear together more than one could expect by considering the two events independently. A PMI=0 indicates that there is no dependency between the two variables (i.e., emotionality and frames). Lastly, negative values indicate that f co-occurs with the considered e with less than chance expectancy and therefore is associated more with the opposite emotion label.

5 Emotionality-Frame Associations

The processing steps described in Section 4.1 result in two independent layers of annotation for the same texts, for which Table 3 shows statistics: the emotion classification module results in ≈23M sentences labelled as emotional and ≈18M as nonemotional. From this total, ≈6M sentences (i.e., ≈4M emotional and ≈2M nonemotional, row "Sents. w/o Frames") are not associated with any frame by the frame identifier. In our analysis, we do not consider these frameless sentences, which typically consist of short texts like "*That's what it was*" and "*-No, it's not a guy*". For all others (row "Sents. with Frames"), the role labeller identified 133M frames, specifically in the 76M emotional sentences and 57M in the nonemotional counterparts, with an average of 3.7 frames per sentence.

Given these numbers, we focus on the 758 unique frames which appear at least 50 times in either textual domain of COCA and analyse the PMI between those and emotionality[16], as reported in Figure 3. One might argue that emotionality, when expressed in language,

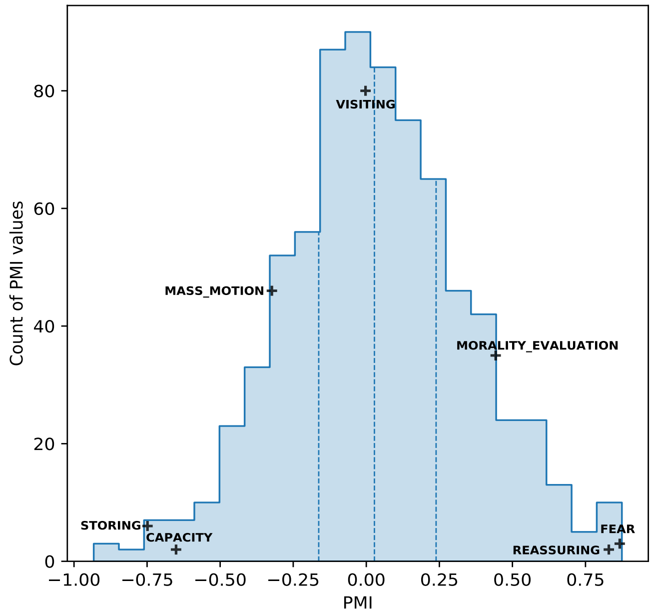

Figure 3: Histogram of PMI(f;emotional). Dashed lines: beginning of the second, third and fourth quartiles.

falls on a spectrum. Different frames can convey varying degrees of emotion, depending on factors such as context, cultural nuances, and more. But to navigate and discern patterns within this continuum, we leverage the simplifying assumption that frames comprise: a predominantly emotional vocabulary (larger than the one openly designated as emotional in FrameNet); vice versa, a set of frames that count as nonemotional; frames that can be either emotional or not, whose status is determined by the context in which they appear – they basically mirror words that dictionary-based emotion models in computational emotion analysis associate with different emotions (e.g., "*abundance*" in the lexicon of Mohammad and Turney (2013) is mapped to anticipation, disgust, joy, and trust).

We need to find lists of frames belonging to these three groups in order to evaluate our assumption. The distribution of PMI values in Figure 3 does not naturally provide such a tripartition. We could define it in many ways, for instance using PMI=0 to decide on what counts as emotional and what not. However, we adopt the quartiles of the PMI distribution because they represent a good balance between the precision and recall of our findings: as opposed to a binary separation, they shield us from considering as emotional some frames with a minimally positive PMI (due, e.g., to bias in the data or mistakes of either automatic labeller); compared to more restrictive cuts (e.g., taking the top 10% of frames as emotional), they facilitate our analysis of what frames other than those already known to be emotional are so.[17]

[16] In this binary classification setup, the distributions given by PMI(f; emotional) and PMI(f; nonemotional) are essentially symmetric: Frames which are positively correlated to one label are negatively correlated to the other. E.g., the frame MORALITY_EVALUATION illustrated in Figure 3: PMI(f; emotional)=.44, while PMI(f; nonemotional)=-.8. For this reason, we only report the emotional distribution.

[17] We rely on the thresholds for a structured and clear analysis. This

Hence, we consider the top quartile of the distribution (PMI≥.24) to correspond to frames that are consistently emotional across various contexts, as it identifies the highest 25% of PMI values positively associated with the label *emotional*. Frames in the bottom quartile (PMI≤-.16) will henceforth be treated as *nonemotional*. Both the emotional and nonemotional quartiles encompass 190 frames.

All other frames fall within the second and third quartiles of the distribution: the fact that the PMI values in Figure 3 are approximately normally distributed around 0 does not indicate the absence of a correlation between emotions and frames; it rather tells us that a large group of items are neither strongly associated with nonemotionality nor with emotionality. These 378 frames will be referred to as *contextually determined* for reasons discussed in Section 5.3.

In the following subsections, we characterise the emotional, nonemotional, and contextually-determined frames, validating the findings of the PMI procedure with a detailed qualitative analysis.

5.1 Emotional Frames

The procedure from the previous subsection has provided us with a set of frames which are purportedly emotional. In order to better understand *how* PMI values relate to the emotional aspects of frames, we ask two questions: (a), how do PMI values vary within frames? (b), how can we characterise emotional frames – can we find a clustering that is coherent according to both qualitative and quantitative criteria?

5.1.1 PMI Values across Lexical Units

In order for the notion of an "emotional frame" to have substance, we need to show that emotionality is not just the result of a small number of frequent, highly emotional lexical units in the frame, but that rather (almost) all of the frame's lexical units are emotional.

To assess whether this is the case, we compute the PMI between the label *emotional* and the lexical units of the 35 most emotional frames. We observe indeed that the frames' PMI values remain consistent across units, with minor variations. Examples are "*frightened*", "*afraid*" and "*terror*", having a PMI score of .86, .85 and .81, all close to the .86 of the corresponding frame FEAR. This tendency holds mainly for the units of frames overtly defined in terms of emotions (like FEAR with a standard deviation across lexical units of .03), but also for others, like the adjectives "*sickening*" and "*troubling*" which have a statistical association to emotionality comparable to that of STIMULUS_FOCUS, that they evoke (.70, .68 and .68, respectively; standard deviation across all units of the frame: .28), as well as "*fiasco*" (PMI= .65) and "*ruin*" (.69), headed by BUNGLING (PMI=.66, standard deviation: .31). Exceptions are lexical units that appear to have little subjective connotation. For example, for STIMULUS_FOCUS, the noun "*relaxation*", which has a less prominent evaluative undertone than the above-mentioned adjectives, deviates noticeably from the frame's PMI (.31). These numbers show that emotional frames display a fair degree of internal consistency concerning their emotionality.

5.1.2 Characterising Emotional Frames

Figure 4 (a) illustrates the 35 highest PMI-valued frames – some COCA sentences in which they appear are shown in Table 4. This small subset hints already at the diversity of the 190 emotional frames, which capture situations ranging from circumstances of interpersonal communication (e.g., OPINION, REVEAL_SECRET, WARNING) to actions (e.g., RUN_RISK), from internal motives (e.g., WILLINGNESS, RENUNCIATION) to social circumstances (e.g., HOSTILE_ENCOUNTER, PREVARICATION). A handful of these frames, like FEAR or EMOTION_ACTIVE, has a clear emotional quality. They are treated in FrameNet itself as such. However, for almost all of them (e.g., FAIRNESS_EVALUATION), an emotion content is more opaque and warrants investigation.

This diversity suggests that we need to corroborate the emotionality of the instances in the top quartile of the PMI distribution. We do that by conducting a qualitative analysis to define a few frame clusters that share emotion-related characteristics, followed by a quantitative discussion to validate our findings.

Qualitative Evidence. We build upon a discussion of the emotion vocabulary initiated by Ruppenhofer (2018). Its core idea is that it is instructive to "examine to what extent the notions FrameNet uses for its analysis do match ones found in psychological theories" (p. 96), as a way of relating the experts' understanding of emotions (formalised in theories and definitions) to the folk's understanding of their experiences (captured, e.g., by FrameNet), and linking psychological views on emotions to linguistic analyses.

We put this view into practice by manually clustering the 190 frames into different groups that map either to the emotion vocabulary in FrameNet, or to theoretically-motivated emotion properties. The clusters are:

heuristic may appear as predefining the three groups ad-hoc. But our goal is not to propose a conclusive categorization of frames in three classes that we assumed to find. Instead, we aim at understanding what brings together frames fallen under one category (see following sections). The sizeable presence of contextually-determined frames, for example, could be dismissed as an influence of the textual genre in which they appear. Our inquiry asks: Is there anything else that makes them more emotionally variable than the others? Future work could explore, e.g., clustering methods that provide a categorization of frames without the quartile-based division.

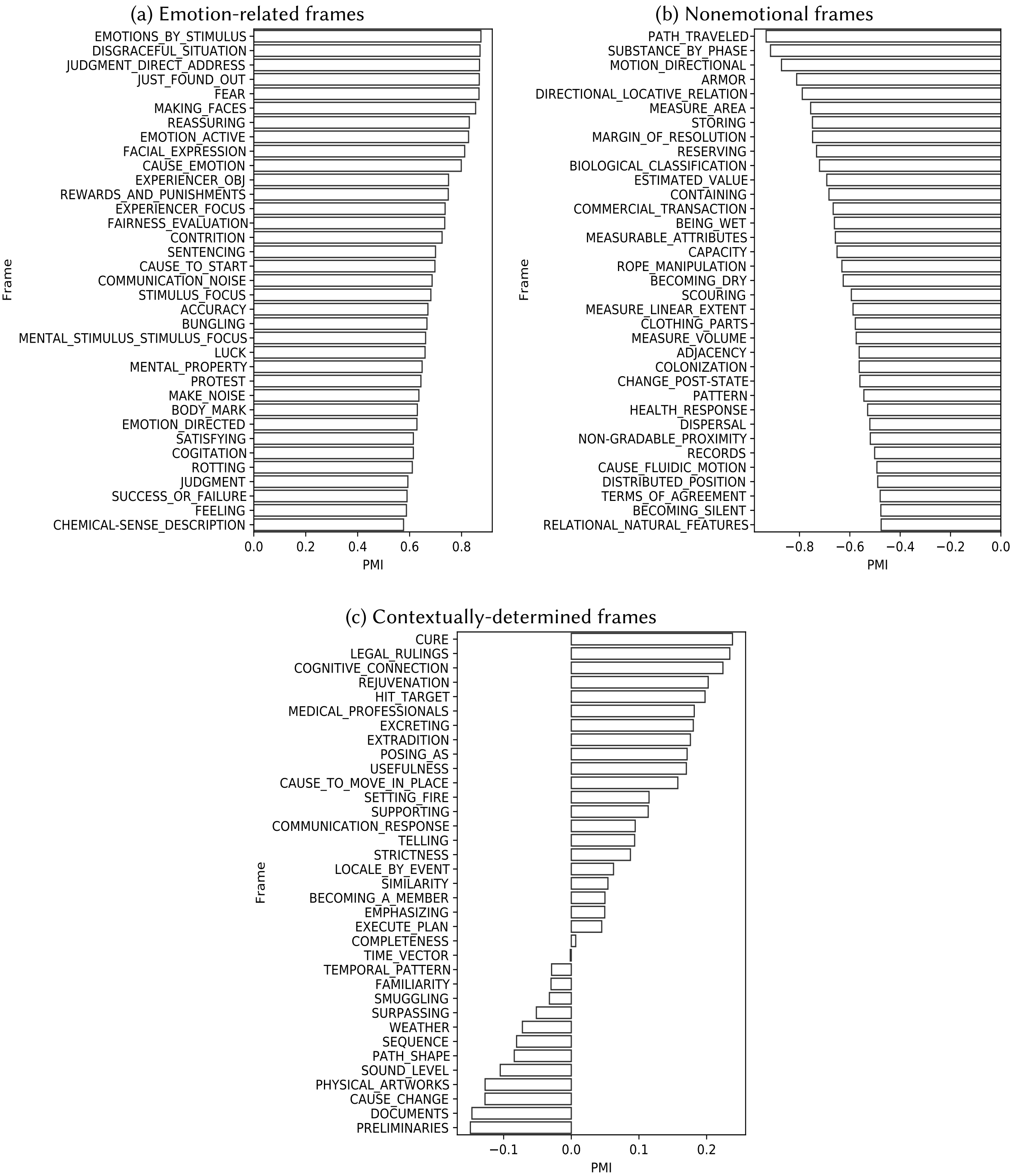

Figure 4: (a) The 35 frames with the highest PMI values in the emotional distribution, in comparison to the frames with the lowest values (b) and in-between these two extremes (c). See Table 4, 10 and 11 for example sentences in which these frames are evoked.

Frame	Text
Judgment_direct_address	Oh, thank God, thank God you're not mad at me for pushing you that day.
Emotions_by_stimulus	So glad we're friends .
Disgraceful_situation	This is outright, outrageous, disgraceful, disgusting.
Reassuring	He spoke with a dentist's tone of calm reassurance.
Cause emotion	The whole thing was quite pathetic, really, and insulting to boot.
Experiencer obj	I am surprised the judges bought it.
Communication_noise	For the first week I cried.
Stimulus_focus	The silence of the candidates is amazing.
Luck	Fortunately, adventure found him in college.
Protest	He marched, he organized, he protested, he was gassed, he was beaten, he was jailed.
Contrition	Blinking furiously, looking furiously guilty, Jimmy Lowe says, "All's I did – Ziggefoos cuts him off."
Fairness_evaluation	To the guy who is whining about how this would be so unfair if it were applied to any other social or racial group ..... God, get over yourself.
Emotion_directed	And – and she just made you happy.

Table 4: Examples of emotional frames with sentences in which they appear.

	Mean	St.dev
Overtly Emotional	.68	.15
Emotion Stimuli	.42	.14
Appraisal-based	.43	.15
Incidentally Emotional	.32	.07
All Emotional Frames	.43	.16

Table 5: PMI mean and standard deviations over lexical units within each cluster and across all 190 emotional frames.

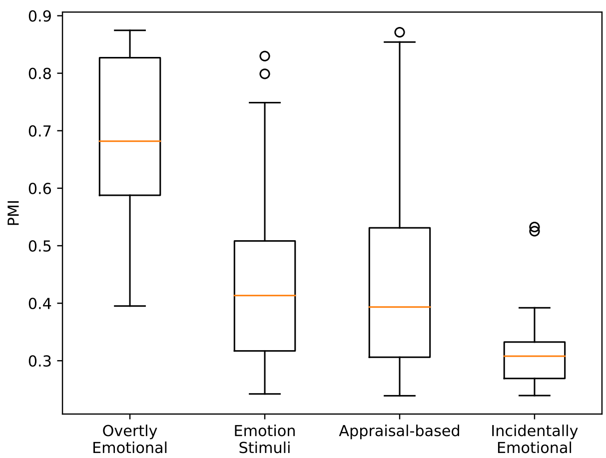

Figure 5: PMI values for each cluster.

(1) **Overtly Emotional** frames;

(2) frames that express events or concepts that might cause an emotion (i.e., **Emotion Stimuli**);

(3) **Appraisal-based** frames, which capture diagnostic features of emotions (e.g., emotion manifestations) or cognitive evaluations of situations (i.e., the factor that appraisal theories see as fundamental for an emotion to occur).

(4) **Incidentally Emotional** – the remaining frames in the top quartile which cannot be given a straightforward interpretation in terms of the first three clusters.

The list of frames classified as either cluster is given below in Tables 6 through 9. We will show that frames in each group share a common affect-laden ground, despite their variety. Before we dive into a qualitative analysis, however, we inspect some quantitative evidence.

Quantitative Evidence. While the guiding principle of our annotation is theoretically driven, the frames' membership in either cluster is our empirical decision. Actually, some items could fit into multiple clusters: hit_or_miss and attempt, which have to do with the goals and concerns of an experiencer (much in an appraisal-oriented fashion), could also be arranged among the Emotion Stimuli; desirability, that we annotated as (1), expresses a positive stance towards a circumstance and could belong to (3). Indeed, there is a large number of frames from separate clusters that are directly related to one another (e.g., a using relation holds between MISDEED, which we placed in the Emotion Stimuli, and MORALITY_EVALUATION).

Therefore, we look for quantitative validation of our annotation: Table 5 contains mean and standard deviation for each cluster across lexical units (cf. Section 5.1.1), and Figure 5 reports the per-cluster distribution of PMI values. Both corroborate the observation that items from cluster (1), Overtly Emotional, are clearly separate from the others, and cover the highest PMI values overall. This is likely due to directly emotional frames being less prone to be contextualized in text in a nonemotional manner, because they inherently signify emotion concepts. By contrast, frames in cluster (2), Emotion Stimuli, have the potential to elicit an emotional response but can be more easily contextualised without an emotional tone. For example, FEAR, in (1), *denotes* an emotion concept, while DEATH, in (2), arguably has an emotional *connotation* that could or could not be manifest in text. We perform a Mann Whitney U test between pairs of groups, as a way of controlling if the difference between the PMI values of the corresponding frames is statistically significant. This is (partially) the case: p-value $< .05$ for each pairwise comparison, except for the difference between clusters (2) and (3). Considering the conceptual overlap of these two categories[18], as well as the fact that their distinction does not reflect linguistic or semantic properties but constructs from psychology, we take this outcome as a confirmation of our initial assumption: some frames are straightforwardly emotional, while the emotionality of many others can be made sense of thanks to appraisal-grounded concepts.

5.1.3 Four Clusters of Emotional Frames

We now discuss the outcome of our manual classification in more detail, and visualise how it identifies coherent clusters in FrameNet.[19]

Overtly Emotional. This cluster encompasses 17 frames that are direct children of the node EMOTIONS (or children of its children), and can thus be considered to have an emotional status in FrameNet. Examples are JUDGMENT, EMOTION_DIRECTED and STIMULUS_FOCUS, FEELING and CONTRITION which express the internal state caused by an emotion episode. The whole list of members is in Table 6, together with the definition of this group that we used as a guideline for the task. Figure 6 illustrates them. Circled grey frames are frames, such as EMOTIONS, that are not part of the cluster but are necessary to connect the individual frames. [20] The figure demonstrates how our PMI-based analysis aligns with the FrameNet database and in particular the frame-to-frame relations. The fact that these frames form an almost connected component in FrameNet corroborates the intrinsic emotionality of its affective vocabulary.

Emotion Stimuli. 72 frames express emotion-inducing circumstances. They are shown in Table 7 and visualised in Figure 7. The frame EVENT is included in the visualisation despite belonging to the Incidentally Emotional cluster, because it delineates a generic super-category from which all other specific events branch out.

Recall that in the view of appraisal theories, events are causes of emotions: they make emotions different from other affective states, such as mood, which are more independent from the environment. Our second group of frames captures precisely this notion. It comprises items that revolve around emotion-stimulating circumstances, like ROTTING and DESTROYING, and therefore, can account for the emotionality assigned to texts that convey an affective content via purely factual descriptions. In this light, this cluster is also close to the idea underlying emotion lexicons, namely, that some words evoke mental representations that have a prototypical affective substrate, somewhat established in the collective knowledge.

For some of them, an emotional attachment might result weak at first glance, but it is clarified by looking at the texts in which they appear. MAKE_COMPROMISE, for instance, is typically evoked by sentences that bring up people sacrificing self principles; CAUSE_TO_FRAGMENT is evoked by texts depicting an entity being "broken" (e.g., being hurt by a breakup). There are also instances that do not indicate events strictly speaking, but kin concepts. Two examples are VIOLENCE and HOSPITALITY, recognised by the frame identifier in sentences that manifest appreciation for conviviality.

Appraisal-based Frames. The third cluster of 76 frames is reported in Table 8, some of which are displayed in Figure 8. This cluster formalises implicitly emotional cases like Emotion Stimuli, but it captures either properties of events, as evaluated emotion experiencers, or other emotion components that manifest in the experiencers' reactions. Similar to events, these are given a prominent role by appraisal theories: the emotion mechanism involves an experiencer who assesses

[18] Cluster (3) includes qualities of stimulus events (e.g., OPPORTUNITY) and following reactions (AGREE_OR_REFUSE_TO_ACT), which can also be considered as events themselves.

[19] For simplicity, Figure 6, 7 and 7 include only frames among the 100 with the highest positive emotional associations and do not show relations between all frames. Note that the grey nodes are not among the top 100 frames. They are illustrated to reproduce the FrameNet structure and account for how the frames under consideration (text in black) relate to one another through relations (represented by the coloured arrows, each corresponding to a specific type of relation).

[20] EMOTIONS has only a single lexical unit, the noun *emotion*, which is generally used to refer to, rather than express, emotions.

Definition	These frames are direct children of the node emotions. They must be its immediate derivation, or a derivation of one of its children nodes.
Frames	1. emotions_by_stimulus, 3. judgment_direct_address, 4. just_found_out, 5. fear, 8. emotion_active, 11. experiencer_obj, 13. experiencer_focus, 15. contrition, 19. stimulus_focus, 22. mental_stimulus_stimulus_focus, 28. emotion_directed, 32. judgment, 34. feeling, 36. desirability, 40. aesthetics, 92. predicament, 94. desiring

Table 6: Overtly Emotional frames. Each frame is numbered according to its PMI rank.

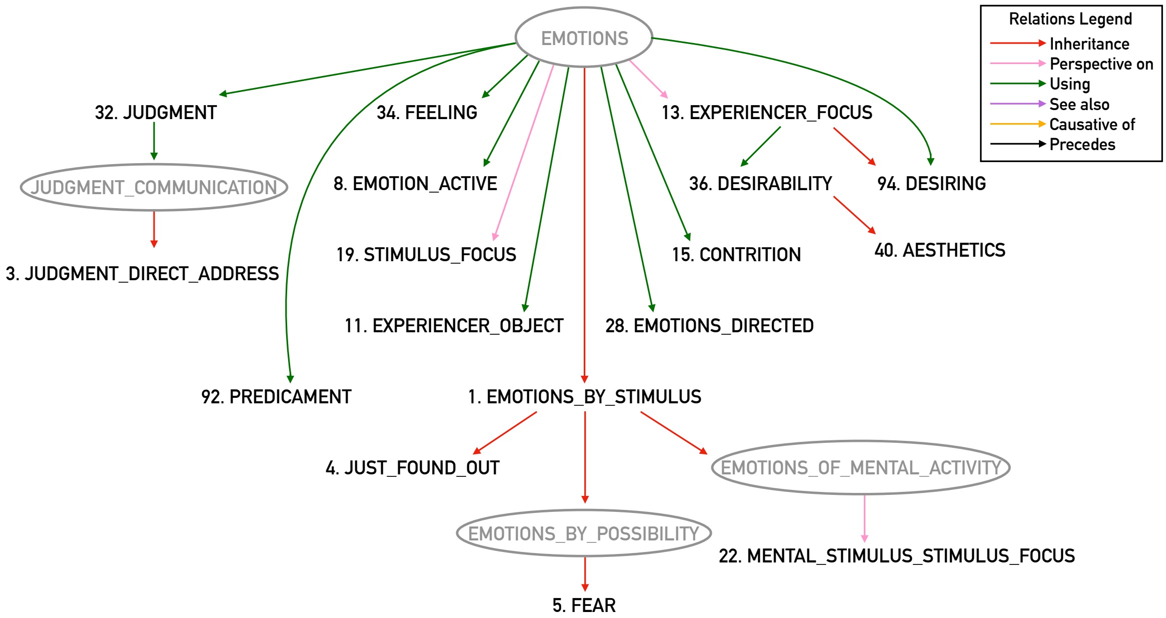

Figure 6: Emotional frames (text in black), which are children of the node Emotions, corresponding to Table 6.

Definition	These frames express circumstances that can cause an emotion.
Frames	7. reassuring, 10. cause_emotion, 12. rewards_and_punishments, 16. sentencing, 17. cause_to_start, 21. bungling, 25. protest, 31. rotting, 39. killing, 41. beat_opponent, 42. firing, 43. destroying, 45. terrorism, 46. daring, 47. verdict, 48. finish_competition, 50. offenses, 55. death, 56. recovery, 57. suasion, 60. kidnapping, 62. cause_to_experience, 66. cause_harm, 67. revenge, 69. catastrophe, 70. misdeed, 71. arrest, 72. prevent_or_allow_possession, 75. imprisonment, 80. accomplishment, 81. violence, 83. successful_action, 84. render_nonfunctional, 87. unemployment_rate, 88. warning, 89. forging, 90. renunciation, 93. assistance, 100. entering_of_plea, 101. rebellion, 106. attack, 107. repel, 108. hostile_encounter, 110. endangering, 111. cause_to_fragment, 113. rescuing, 116. prevarication, 119. subversion, 121. resolve_problem, 122. experience_bodily_harm, 124. arson, 129. medical_conditions, 134. examination, 138. infecting, 143. run_risk, 152. endeavor_failure, 153. invading, 155. theft, 158. hospitality, 159. quarreling, 162. medical_intervention, 163. bearing_arms, 166. reveal_secret, 169. escaping, 172. damaging, 173. prison, 174. make_compromise, 177. trial, 178. committing_crime, 180. surviving, 183. surrendering, 186. execution

Table 7: Emotional frames annotated as Emotion Stimuli. Each frame is numbered according to its PMI rank.

the circumstance and engages in a series of changes – i.e., subjective feelings, neurophysiological, motor and motivational alterations.

Frames concerning evaluations are, e.g., satisfying and fairness_evaluation. The latter frame, whose link to emotions seemed hazy at first, now appears as an emotional exemplar in its own right: the notion of assessment that it brings into play is central to the elicitation of emotions. In this group are also items that qualify events as endangering for the organism (e.g., dif-

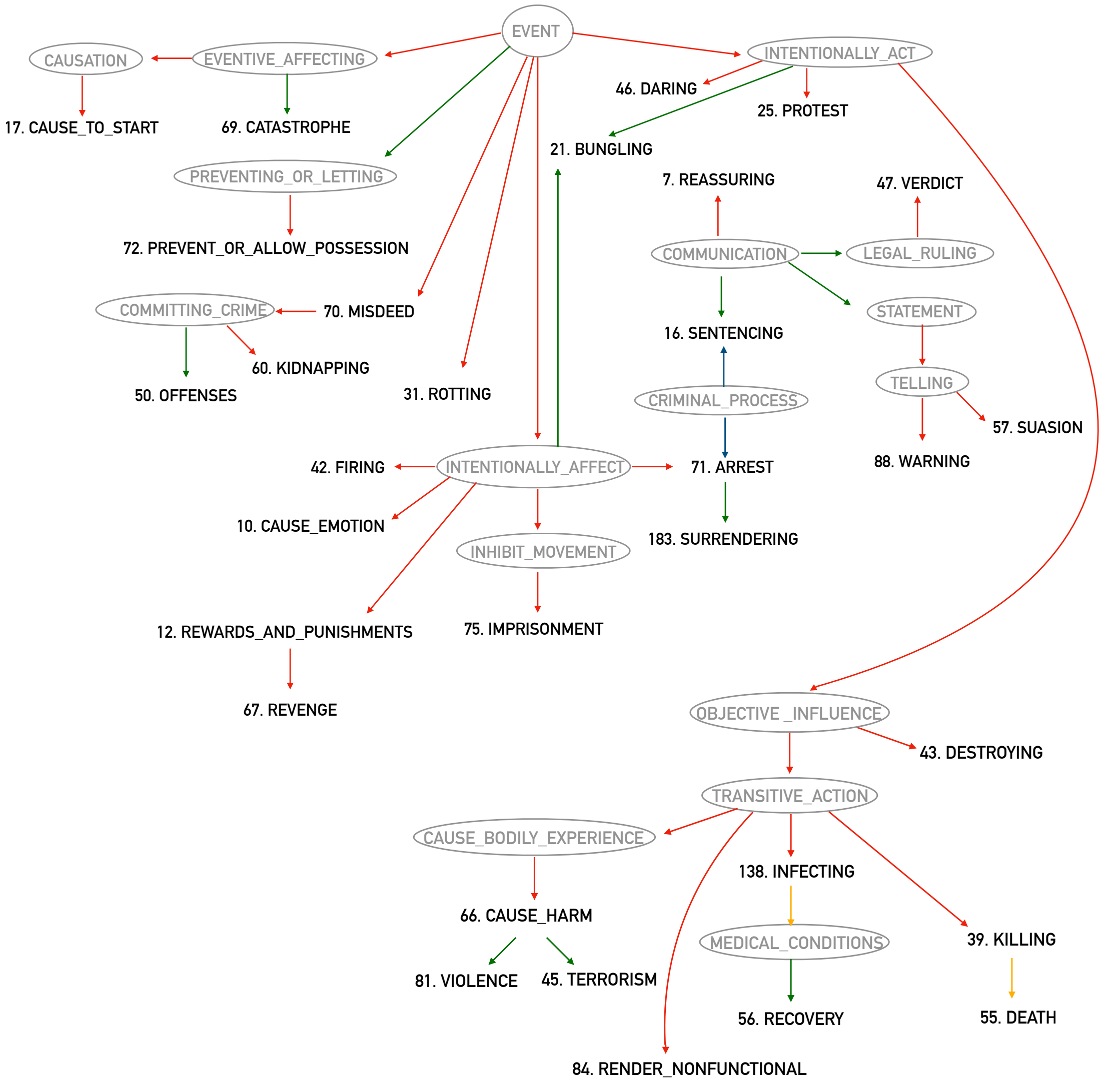

Figure 7: Emotional frames (text in black), deriving from the node EVENTS and expressing factual Emotion Stimuli, extracted from Table 7. The arrow legend is in Figure 6.

FICULTY, RISKY_SITUATION), or as fostering its well-being (e.g., LUCK, WEALTHINESS).

Some of these frames recall the *criteria* that individuals use to evaluate an environment. In the appraisal framework, they are described with a finite number of dimensions (Scherer et al., 2010). One is the coherence of the event with the personal ideals of the experiencer and with societal norms. Frames like FAIRNESS_EVALUATION and MORALITY_EVALUATION convey precisely this type of evaluation. Similarly, GRASP reflects the criterion by which events are appraised in relation to their implications – e.g., Are they relevant to the experiencer's goals? Can their consequences be estimated? It is indeed evoked by textual chunks that involve a cognizer who acquires knowledge about the significance of a given phenomenon and becomes informed to make predictions about it. Events can also be evaluated for the degree to which the experiencers are certain about what is going on (e.g., How well does the experiencer understand what is happening in the emotional situation? (Smith and Ellsworth, 1985)), which is echoed by the frame CERTAINTY, and with respect to the urgency of a reaction (REQUIRED_EVENT).

Focusing on such evaluation criteria, appraisal theories claim that specific assessments of events lead to specific emotion experiences. For instance, a lack of certainty likely results in an episode of fear or hope (Smith and Ellsworth, 1985). To an extent, this is accounted for by the relations between frames. CERTAINTY, as an example, is inherited by the node TRUST. Therefore, FrameNet relations seem to explain the affective charge of some of these frames that do not stem from EMOTIONS, but are linked to the EMOTIONS-deriving nodes all the same.

Definition	Frames capturing the link between emotions and events, namely, the saliency of the circumstance for the well-being of the experiencer, evaluations, actions, motives and responses that the experiencer takes in reaction to the event.
Frames	2. DISGRACEFUL_SITUATION, 6. MAKING_FACES, 9. FACIAL_EXPRESSION, 14. FAIRNESS_EVALUATION, 18. COMMUNICATION_NOISE, 20. ACCURACY, 23. LUCK, 24. MENTAL_PROPERTY, 26. MAKE_NOISE, 27. BODY_MARK, 29. SATISFYING, 30. COGITATION, 33. SUCCESS_OR_FAILURE, 35. CHEMICAL-SENSE_DESCRIPTION, 37. FRUGALITY, 38. AGREE_OR_REFUSE_TO_ACT, 44. CHAOS, 49. SOCIABILITY, 51. DESERVING, 53. CERTAINTY, 58. OMEN, 59. RISKY_SITUATION, 61. GUILT_OR_INNOCENCE, 63. SUBJECTIVE_INFLUENCE, 64. BEING_QUESTIONABLE, 65. PROMINENCE, 68. VOCALIZATIONS, 73. BIOLOGICAL_URGE, 74. GRASP, 76. DIFFICULTY, 77. MORALITY_EVALUATION, 78. COMING_TO_BELIEVE, 79. STINGINESS, 82. SOCIAL_INTERACTION_EVALUATION, 85. ARTIFICIALITY, 86. FLEEING, 91. HIT_OR_MISS, 95. IMPROVEMENT_OR_DECLINE, 96. WEALTHINESS, 97. CORRECTNESS, 98. COMMITMENT, 102. LEVEL_OF_FORCE_EXERTION, 104. COMPLAINING, 105. REASONING, 109. PEOPLE_BY_MORALITY, 112. SOCIAL_DESIRABILITY, 115. JUSTIFYING, 117. JUDGMENT_COMMUNICATION, 118. WILLINGNESS, 120. SENSATION, 123. INCLINATION, 125. EXPRESSING_PUBLICLY, 130. TRIGGERING, 135. EXPECTATION, 136. EXPEND_RESOURCE, 137. JUDGMENT_OF_INTENSITY, 142. TRUST, 146. OPPORTUNITY, 147. BEING_RELEVANT, 148. DEAD_OR_ALIVE, 150. AWARENESS_STATUS, 151. DYNAMISM, 154. BEING_OPERATIONAL, 157. FAME, 160. BEING_AT_RISK, 161. OPINION, 164. REQUIRED_EVENT, 170. CAUSE_IMPACT, 175. PRECARIOUSNESS, 176. MEET_SPECIFICATIONS, 179. MOTION_NOISE, 181. ATTEMPT, 185. BREATHING, 187. CONFRONTING_PROBLEM, 188. EVENTIVE_AFFECTING, 190. ATTITUDE_DESCRIPTION

Table 8: Emotional frames that capture appraisal-related properties. Each frame is numbered according to its PMI rank.

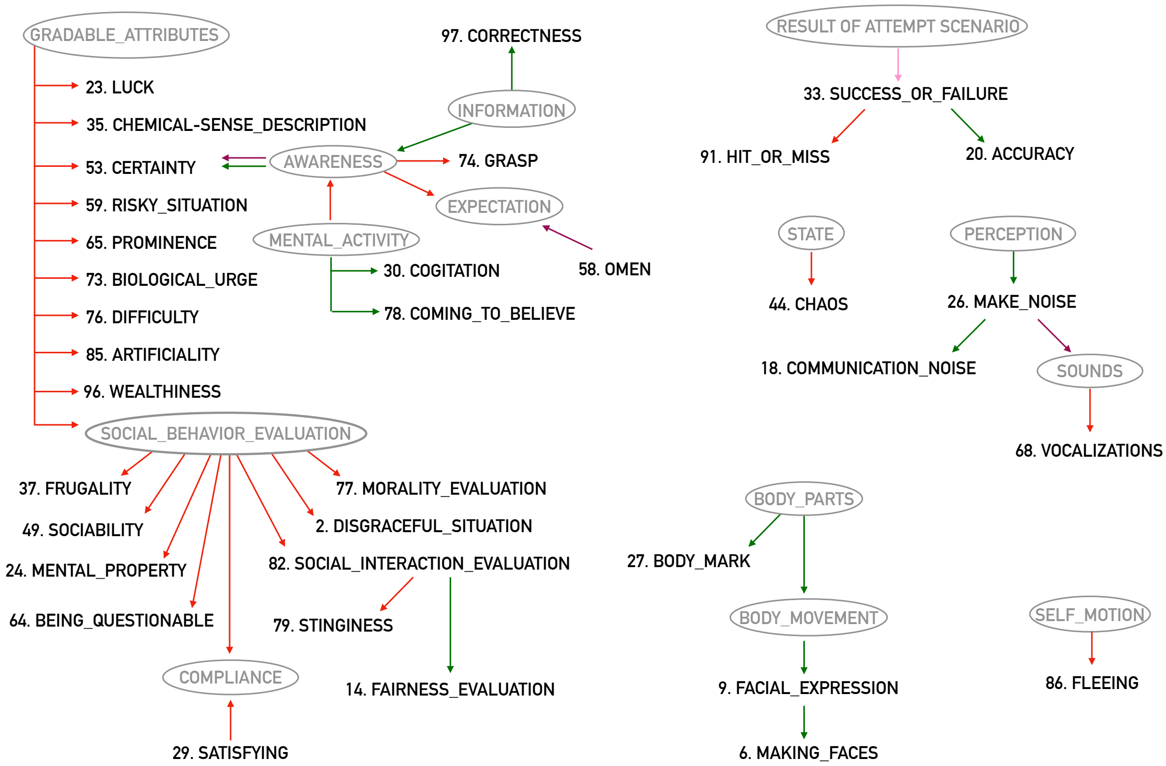

Figure 8: Emotional frames (text in black), expressing appraisal-related concepts (cf. Table 8). The arrow legend is in Figure 6.

We further observe frames that relate to the effects that emotions have on the organism (BIOLOGICAL_URGE exemplifies the involvement of internal, physiological states that can motivate action in response to an event), and frames that correspond to more observable manifestations of the emotion mechanism, such as vocal verbalizations, facial movements, and other diagnostic features that allow people to understand what their interlocutors feel. MAKING_FACES, FACIAL_EXPRESSION, COMMUNICATION_NOISE (evoked by texts like "*For the first week I cried.*") and MAKE_NOISE seize these components. Other frames, for instance REASSURING and COGITATION (a child node of WORRYING), capture external actions or internal attitudes that can occur in emotional situations.

Additional analyses of frames whose membership to the Appraisal-based cluster is not self-explanatory can be found in Appendix C.

Incidentally Emotional Frames. These 25 frames (see Table 9) rank among the lowest values in the top quartile of the PMI distribution, closer to the cutoff point than the clusters discussed so far. They hardly capture an emotion property or an emotion-inducing event; in fact, they can be argued more affine to the contextually-determined cluster of Section 5.3, to which their PMI values are close. In this analysis, they appear as emotional due to two primary factors. The first one is narrative context, the second is processing errors. We support this analysis by investigating the sentences in which these frames appear.

Regarding narrative context, recall that most COCA sentences contain multiple frames. Therefore, frames can assume emotionality from others in the same sentence, which are often narratively related. BOARD_VEHICLE and RIDE_VEHICLE, for instance, are evoked in texts that have to do with embarking on adventures and journeys: these tend to be emotionally qualified as they often mention personal stances towards such journeys (e.g., if it was pleasant). Instead, REFORMING_A_SYSTEM and CAUSE_TO_RESUME characterise texts that express an idea of personal change, of beginning (e.g., "*We may have reformed, but our enemies have not.*", "*I felt revived*"). MANIPULATE_INTO_DOING is ascribed to descriptions of bullying episodes; IRREGULAR_COMBATANTS has to do with fighters and hence a notion of brutality (comparable to KILLING and BEARING_ARMS from cluster (2)). MEDICAL_SPECIALTIES is evoked by (potentially stirring) circumstances that are related to healthcare and therapy, and RITE appears in the context of intimate meditations and expressed hopes.

Other cases seem to result directly from mistakes made by the frame identifier. With TEMPERATURE, the automatic role labeller does not understand the metaphoric use of the word "*cool*", for which that frame is usually predicted. LINGUISTIC_MEANING is a similar case. It is identified in phrases that are related to meanings and to the "making sense" of a situation, rather than in the context of a discussion about linguistic meaning.

5.2 Nonemotional Frames

Examples of nonemotional frames are in Figure 4 (b), with some corresponding texts in Table 10. We ask the same two questions about nonemotional frames that we asked about emotional frames above.

5.2.1 PMI Values across Lexical Units

To understand the difference between this group and the emotional one, we look at the PMI scores of the frames' lexical units. Mirroring what we did for the top 35 frames, we focus on the 35 most nonemotional instances at the bottom of the PMI distribution. Here, frames show a much lower internal consistency, and suggest that they act as emotionally coherent units of abstraction only above a certain PMI threshold. Indeed, the scores of lexical units instantiating a nonemotional frame spread away from that of the latter considerably, as exemplified by RELATIONAL_NATURAL_FEATURES (PMI = -.47) whose lexical instantiations encompass a vast range of values, from .01 (for the noun "*summit*") to -1 ("*shoreline*"), DISTRIBUTED_POSITION (-.48), spanning from the -.07 PMI score of "*envelop*" to -.70 of "*wreathe*", and BECOMING_SILENT (-.47), where "*quiet*" has a PMI value of -.15 as a noun and -.83 as an adjective.

This outcome is different from what we found for emotional frames, where emotionality is stable for the lexical units within frames (cf. Section 5.1). For nonemotional frames, the picture is not symmetric: the PMI variance of lexical units can be attributed to their presence (mostly) in textual contexts without emotionality, but also in some with an emotion gradation.

5.2.2 Characterising Nonemotional Frames

Compared to the emotional frames, this cluster depends much less on people's subjective involvement in the state of affairs mentioned in the texts. It includes frames expressing features of objects (e.g., BIOLOGICAL_CLASSIFICATION, ESTIMATED_VALUE, SUBSTANCE_BY_PHASE, MEASURABLE_ATTRIBUTES) or of events which have less relevance for human actors in terms of appraisals (e.g., CHANGE_OF_PHASE, BECOMING_DRY).

5.3 Contextually-determined Frames

Contextually-determined frames are those with PMI values falling in the 2nd or 3rd quartiles of the emotional

Definition	Frames that do not belong to any of the other three groups.
Frames	52. respond_to_proposal, 54. institutionalization, 99. rite, 103. linguistic_meaning, 114. board_vehicle, 126. manipulate_into_doing, 127. medical_specialties, 128. reforming_a_system, 131. economy, 132. temperature, 133. co-association, 139. affirm_or_deny, 140. behind_the_scenes, 141. appellations, 144. ride_vehicle, 145. event, 149. irregular_combatants, 156. change_of_leadership, 165. people_by_religion, 167. medical_interaction_scenario, 168. education_teaching, 171. cause_to_resume, 182. make_agreement_on_action, 184. representative, 189. touring

Table 9: Incidentally Emotional frames. Each frame is numbered according to its PMI rank.

Frame	**Text**
Path_traveled	They occur when the orbits of the moons turn edge-on to the Sun and Earth, which happens twice during Jupiter's 12-year circuit of the Sun.
Directional_locative_relation	It was known he lived across the immense valley below me.
Storing	Mark your packages with the date they were placed in the freezer so you can keep track of storage times.
Measure_area	They burned 665,000 acres; roughly 40% of the statewide total of 1.7 million acres.
Relational_natural_features	The shore is crumbling.
Becoming_silent	A silence descends on the tiny room.

Table 10: Examples of nonemotional frames with sentences in which they appear.

Frame	**Text**
Communication_response	(N) The **answer** is, you don't, or at least not with career backups. (E) The **answer** would be NO!
Give_impression	(N) Neither candidate **seemed** to have any awareness of virality . (E) You really **seem** to be exploding with creativity!
Point_of_dispute	(N) The **question**, crude as it was, hung in the air . (E) The **issue** is not whether I was a perfect pastor; I was not .

Table 11: Example sentences evoking contextually-determined frames. (E)/(N): emotional/neutral sentences. Words in boldface correspond to predicates.

distribution reported in Figure 3 (-.16≤PMI≤.24). A few examples are provided in Table 11. These items have an ambiguous emotional status, in that they present no clear association with emotionality, nor its absence.

What makes frames contextually-determined? Our hypothesis is that it is possible to set apart these cases from frames that carry an emotional (or nonemotional) load in two, non-mutually exclusive ways. First, by looking at the lexical units internal to frames, once more, to explain the sense in which these frames are different from the most external quartiles in the distribution. Second, by looking at how their emotionality changes as they co-occur with other frames. We explore these two levels separately below.

5.3.1 PMI Values across Lexical Units

The way PMI values distribute across lexical units is more similar to nonemotional frames (Section 5.2) than to emotional frames (Section 5.1.2): Values differ from each other, in such a way that contextually-determined frames, contrary to emotional ones, do not function as emotion-preserving types of units. Cases in point are the verbs "*tell*" and "*assure*", both evoking the frame TELLING, and whose emotionality association corresponds to the values .07 and .34 (i.e., "*assure*" is most often emotional than not, while "*tell*" is at times emotional); "*disparity*" and "*distinction*" are apart from one another by .51 PMI points (the first of them is the most emotional), despite being units of the same frame (SIMILARITY); likewise, CURE's lexical units "*rehabilitation*" and "*remedy*" have values falling in different quartiles of the PMI distribution (.25 and -.11, respectively). The distinguishing factor between them and the nonemotional group lies in the fact that lexical units here are more versatile. Those belonging to the nonemotional counterpart are specific to domains without emotionality (cf. MEASUREMENT_AREA, CLOTHING_COMPONENTS), and thus their occurrence in emotional contexts is not only rarer, but an artifact of either the emotion classifier (producing random errors) or of our alignment strategy (which permits nonemotional frames to inherit the emotionality of others, with which they occur). Instead, lexical units of contextually-determined frames lend themselves to assume a wide range of emotional connotations.

5.3.2 Frame Co-occurrence Patterns

Above, when characterising Incidentally Emotional frames, we already alluded to the fact that frames typically co-occur in sentences. We also see this effect for the contextually-determined frames. Figure 9 shows the frequency of these frames in three different scenarios, normalised by the total number of sentences in each of them. The two leftmost columns ($\text{Frm}_{cont.}$) report the

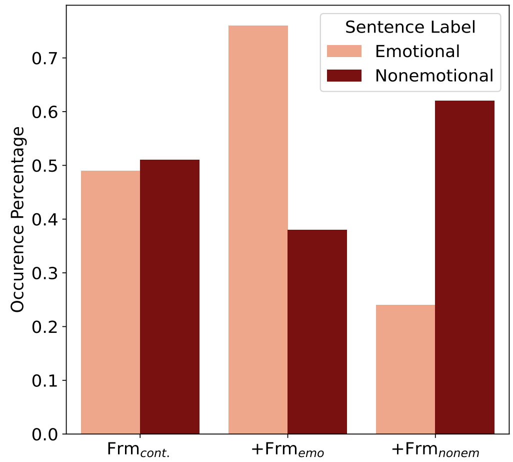

Figure 9: Distribution of emotional and nonemotional sentences evoking contextually-determined frames in isolation ($\text{Frm}_{cont.}$) and accompanied by an emotional frame ($+\text{Frm}_{emo}$) or a nonemotional one ($+\text{Frm}_{nonemo}$).

frequency of frames appearing alone in a text across the two emotion labels, corresponding to >2M emotional and nonemotional sentences. Devoid of frames interactions, these sentences help to clarify what it means for frames to be underspecified with respect to emotionality: based on a manual investigation of such sentences, contextually-determined frames appear to have less to do with properties of things or situations, compared to the nonemotionally-connotated kins. They rather represent such things (FOOD, VEHICLE, BUILDINGS) or processes (CAUSE_EXPANSION, CAUSE_TO_PERCEIVE). We notice that when these frames appear in emotional texts, they do so as side information to the main affective meaning, and do not correspond to the predicate that triggers such emotion content. For instance, CONTINUED_STATE_OF_AFFAIRS in the text "*Glad she's still on the show.*" is unrelated to the mental state of the subject. The figure also reports the count of sentences with a contextually-determined frame and one that is emotional ($+\text{Frm}_{emo}$), or one that is nonemotional ($+\text{Frm}_{neu}$). From the figure, we see that texts that contain both a contextual frame and one with a positive emotion PMI tend to be emotional; vice versa for the co-presence with a nonemotional frame, found more often in sentences labelled as nonemotional by the classifier.

Overall, the fact that these 378 frames are determined contextually shows an important aspect of the phenomenon under consideration. At times, the relationship that frames hold to their emotion content is underspecified: it is not fixed and bounded to the type of event that they formalise (i.e., it does not necessarily lie at the predicate level), but rather depends on the overall context in which the frame-evoking predicate appears. Emotion meanings make no exception in the lexical semantics panorama, where also other phenomena are to be accounted for *in context* (Cruse, 1986) – e.g., word meanings.

A manual inspection of the data also suggests that compositionality is key in the making of an emotion for those sentences corresponding to +Frm_{emo} and +Frm_{neu} in Figure 9. More precisely, we see two compositional processes. One is a "within frames compositionality", in which the predicate is (emotionally) underspecified, but its co-presence with certain arguments can turn out emotional or nonemotional. Illustrative in this regard are sentences like "*I remember this point distinctly.*" and "*I remember the magical thinking of my greatest depression.*", both associated to the frame MEMORY but with different arguments (the first sentence is recognised as nonemotional, the other as emotional). Like in the above examples, many frames are evoked by predicates that serve to introduce topical information, or subordinate sentences. The overall emotionality varies together with the content that they introduce. For instance COMMUNICATION_RESPONSE, TELLING, POINT_OF_DISPUTE, GIVING and GIVE_IMPRESSION have to do with communicative situations that could be loaded with emotionality based on how they are instantiated – what is responded, what is told, what is given (e.g., GIVING in the emotional example "*Cruella gave a gesture of resignation.*"). Similarly, UNDERGO_CHANGES describes a transformation which could be either emotional or nonemotional.

The second compositional process that we notice is an "across frames compositionality". Frames that appear in combination with a contextually-determined one contribute more to the emotional load of the sentence: the text "*[...] an old girlfriend of mine wrote me this very beautiful letter.*", which is recognised by the classifier as emotional, evokes MEMORY and the emotional AESTHETICS, while "*The words 'property value' are ones I remember.*", annotated as nonemotional by the classifier, evokes MEMORY and POSSESSION.

6 Discussion

We conducted a PMI-based analysis guided by the research question "are FrameNet frames associated with emotionality?" as well as two leading hypotheses: first, emotional frames constitute a large part of FrameNet, and second, it makes sense to talk about "emotional" frames in the sense that the lexical units within the frame behave coherently. Both assumptions proved correct. Frames that carry emotionality extend beyond the current organization of the database, as many are emotional while having a factual denotation; further, they pass this affective trait on to their lexical units.

Our manual analysis explains what frames have in common from the perspective of emotions, confirming that there are many levels of an emotion mechanism captured by frame semantics. Some frames depict concepts that seem more descriptive than affective, but it is precisely in this manner that they pick up on some important components of emotions. They correspond to some of the factors that elicit, underlie or manifest an emotion, like events, event evaluations, and emotion effects. The effects components, in particular, not only correspond to phenomena that happen in response to emotion-eliciting events (e.g., FACIAL_EXPRESSION). They can be considered events per se, and consequently, they can evoke specific frames.

We manually group these characteristics in four clusters, motivated by the original structure of FrameNet, and the fact that appraisal models and frames are grounded on a notion of event. Ruppenhofer (2018) already pointed out that appraisal theories can inform an investigation of the emotion vocabulary in FrameNet. We bolster that observation by indicating the frames to which it extends, but one could also identify other emotion properties and other links to theories different from appraisals, and organise the emotional frames accordingly. Take, for instance, the Appraisal-based cluster. In our proposal, it includes both items that contribute to eliciting an emotion (e.g., DIFFICULTY), and items that result from it (e.g., FLEEING). This is a fruitful distinction that can be made to find more fine-grained theoretical coherence in the obtained statistical associations.

Further, we empirically show that there are frames somewhat transparent to emotions: contextually-determined frames reiterate the need to think about emotionality in terms of relations between words, and raise the question of if and how frames influence the emotionality of a text, as well as its automatic classification. To what extent do predicates or arguments contribute to the decisions of an emotion classifier? Is compositionality at play?

We have previously pointed out that emotionality is a continuum, while our study approaches it through categorical lenses. From a practical standpoint, this categorization has the value of generating clear insights. But this choice introduces limitations, not least of which is a certain degree of arbitrariness in such divisions: frames do not necessarily fit into the three nonemotional, contextually-determined, and emotional "boxes" identified with the help of quartiles, and the line between contextually-determined and emotional frames (in particular the Incidentally Emotional ones) is blurred. In fact, the compelling case that emotionality is always a matter of context could be made also for many emotional frames, with the Emotion Stimuli being evident cases (events could stir an emotion or not, depending on who experiences them and how they are rendered in language). Our results prove however that a separation holds, at least in COCA, between frames for which exhibiting the emotional association is *invariably* contextual, whereas others *maintain a certain level of emotionality* – e.g., Emotion Stimuli have the tendency to denote events with potentially dramatic consequences for their experiencers, see for instance VIOLENCE or CATASTROPHE.

In sum, our analysis reveals that the relationship between the emotionality of a sentence and that of frames is not straightforward. Frames that have a strong positive or negative association to emotionality can be found in texts that express the opposite affective content overall.[21] Even the frames that FrameNet explicitly associates to the emotion domain are evoked by nonemotional sentences. EMOTIONS_BY_STIMULUS, as an example, is found by the frame identifier in the nonemotional "*I had every right to descend this stair, to walk among the glad company [...]*", because of the lexical unit "*glad*". Rather than putting the automatic annotation into question, this outcome sheds light on an important fact. Namely, sentence-level emotionality classifiers can disregard emotional subtleties. A verbal expression might have a predominant connotation to convey (e.g., a nonemotional one, in the example above), and which might be correctly identified by the automatic system; yet, by considering entities besides the subject, different emotion nuances emerge (e.g., the company is glad). Classifiers might fail to account for those, and in such cases the performance of frame identification tools can complement theirs. In line with previous work (Faruqui et al., 2015, i.a.), we thus found that approaches based on embeddings and on human-curated resources help one another also in emotion analysis.

7 Conclusion

The phenomenon of "emotions" is psychological in nature but pervades language. There, the presence of overt markers (the adjectives "*sad*", "*happy*", for instance) is not necessary for an emotion to be conveyed. These "untold" emotions spurred much attention in the field of computational emotion analysis (Balahur and Tanev, 2016; Klinger et al., 2018), which strives to automatise the ability to infer them.

Within such a context, we left traditional, lexical-based approaches of emotion analysis, because interpreting emotions can require a great deal of extra-linguistic knowledge. We considered the role that background information plays in emotions understanding, moving our attention to the meeting point between syntax and the U-semantics of Fillmore, which presupposes an acknowledgement of the physical and social world, and therefore accounts for the structural components of real-life events that stimulate emotional responses. This way, our work combined methods for computational linguistics with theories from psychology and linguistics, and it showed how these fields can influence (in fact, fertilize) one another. Below, we summarise the relevance of our findings in this interdisciplinary perspective, and point out promising next steps to take.

Summary of Findings. The observation that frames can be evoked by varied lexical units (thus capturing paradigmatic phenomena) allowed us to disregard the specific terms that instantiate them. We rather asked how frames, as conceptual abstractions that encode world knowledge, are linked to emotionality. We automatically annotated COCA with binary emotion labels and with frames, we investigated the relationship between them, and to answer our research question, we used PMI.

Our results show that there are frames with a prominent emotion import in FrameNet: be they direct children of EMOTION or not, they reflect components of emotions spelled out in the psychological literature. In other words, emotionality is a dimension of meaning that frames possess even though it is not a piece of information directly provided by the database. In addition, our qualitative analysis emphasise that individual predicates do not always carry the same type of emotion load. On the contrary, their import can depend on the context in which the predicate is situated, namely, on syntagmatic facts.

Future Work. We revealed some salient features of frames that open up possible ventures for frame semanticists. Future FrameNet developments could specify what frames carry emotionality with the use of *semantic types* (Fillmore et al., 2004). Semantic types mark general properties of frames and semantic roles, such as variations in the speech use of different lexical units, which could not otherwise be understood from the resource. In FrameNet there already exists a semantic type that is close in spirit to emotions. It indicates the polarity of lexical units like "*compliment*" and "*reprimand*", both of which instantiate JUDGMENT_DIRECT_ADDRESS and whose valence is indicated

[21] Note that there are signs of domain dependence: frames are more emotional in certain domains of COCA than in others. For example, RUN_RISK has a PMI value of .13 in textual blogs, which raises to .4 in the domain of fiction; and the frame PROTEST turns out considerably more emotional if evoked by fiction- (PMI=.77) than by TV-related texts (.47). Consistent with this observation, a Wilcoxon signed-rank test reveals a significant difference between the general PMI values of the emotional frames reported in Figure 3 and the values of the same frames in the various domains (for all of them, except for TV, p-value < .05). Therefore, emotionality is only partly consistent across genres, and this finding is in line with existing literature on the genre dependence of fine-grained emotions (e.g., Bostan and Klinger, 2018). At the same time, PMI differences are rarely as extreme as to have frames that are emotional in Figure 3 turn into nonemotional in a specific domain (that only happens for ATTITUDE_DESCRIPTION, PRECARIOUSNESS, and TEMPERATURE).

by the semantic types "Positive_judgment" and "Negative_judgment". It would be possible to adopt the same idea for the semantic clusters proposed in this paper, or for similar partitions. We refrain from modelling this information into FrameNet ourselves – an endeavor that would require careful and lexicographically motivated annotation, which exceeds the scope of our work.

Our insights can also inform computational emotion analysis. Studies in the field could aim at building systems that are simultaneously emotion- and frame-aware. The frames-to-PMI association scores that we make publicly available come handy for that purpose. Upcoming work could deepen the contribution of different parts of texts (e.g., frames, arguments, other words) on automatic emotion predictions – e.g., Do classifiers attend predicates to the same extent when judging a text that evokes an emotional frame and a text that evokes a contextually-determined frame? Lastly, research in the field that follows appraisal theories could concentrate on the intersection between frame semantics, psychology, and emotion analysis: among other events, frames proved able to model the verbal expressions of emotion components, thus capturing the multiple and nuanced realizations through which embodied emotions and the cognitive evaluations underlying them surface in language. In this regard, we have proposed an empirical mapping from frames to appraisals, but it would be important to take the reverse direction as well. Understanding to what extent frames cover the cognitive dimensions documented by appraisal theories could tell us if frame analysis can be applied as an identification strategy of such dimensions, namely, of the criteria that humans use to evaluate events, that lead to an emotion episode, and that also emerge from text: frames could thus be used as input to computational emotion analysis pipelines, making our systems more theoretically grounded.

Acknowledgements

This work has been supported by the German Research Council (DFG), project "Computational Event Analysis based on Appraisal Theories for Emotion Analysis" (CEAT, project number KL 2869/1-2).

References

Abdul-Mageed, Muhammad and Lyle Ungar. 2017. EmoNet: Fine-grained emotion detection with gated recurrent neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 718–728, Vancouver, Canada. Association for Computational Linguistics.

Alm, Cecilia Ovesdotter, Dan Roth, and Richard Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 579–586, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Baker, Collin F. and Christiane Fellbaum. 2009. WordNet and FrameNet as complementary resources for annotation. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 125–129, Suntec, Singapore. Association for Computational Linguistics.

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *COLING-ACL '98: Proceedings of the Conference*, pages 86–90, Montreal, Canada.

Balahur, Alexandra and Hristo Tanev. 2016. Detecting implicit expressions of affect from text using semantic knowledge on common concept properties. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*. European Language Resources Association (ELRA).

Banarescu, Laura, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Barrett, Lisa Feldman, Michael Lewis, and Jeannette M Haviland-Jones. 2016. *Handbook of emotions*. Guilford Publications.

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*, pages 1554–1566, Marseille, France. European Language Resources Association.

Bostan, Laura-Ana-Maria and Roman Klinger. 2018. An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Buckley, Chris. 1993. The importance of proper weighting methods. In *Human Language Technology: Pro-*

ceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993.

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*, pages 578–585, Valencia, Spain. Association for Computational Linguistics.

Buechel, Sven, Susanna Rücker, and Udo Hahn. 2020. Learning and evaluating emotion lexicons for 91 languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1202–1217, Online. Association for Computational Linguistics.

Canales, Lea and Patricio Martínez-Barco. 2014. Emotion detection from text: A survey. In *Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC)*, pages 37–43, Quito, Ecuador. Association for Computational Linguistics.

Chen, Yanqing and Steven Skiena. 2014. Building sentiment lexicons for all major languages. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 383–389.

Church, Kenneth and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29.

Clore, Gerald L., Andrew Ortony, and Mark A. Foss. 1987. The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, 53(4):751–766.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Coyne, Bob, Alex Klapheke, Masoud Rouhizadeh, Richard Sproat, and Daniel Bauer. 2012. Annotation tools and knowledge representation for a text-to-scene system. In *Proceedings of COLING 2012*, pages 679–694, Mumbai, India. The COLING 2012 Organizing Committee.

Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge university press.

Das, Dipanjan, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic frame-semantic parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 948–956, Los Angeles, California. Association for Computational Linguistics.

Davies, Mark. 2015. Corpus of Contemporary American English (COCA).

Demszky, Dorottya, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A Dataset of Fine-Grained Emotions. In *58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dixon, Thomas. 2012. "Emotion": The history of a keyword in crisis. *Emotion Review*, 4(4):338–344.

Eisenberg, Nancy and Paul A. Miller. 1987. The relation of empathy to prosocial and related behaviors. *Psychological bulletin*, 101(1):91.

Ekman, Paul. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Etzioni, Oren, Michele Banko, Stephen Soderland, and Daniel S. Weld. 2008. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74.

Faruqui, Manaal, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615, Denver, Colorado. Association for Computational Linguistics.

Felbo, Bjarke, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1615–1625, Copenhagen, Denmark. Association for Computational Linguistics.

Fellbaum, Christiane, editor. 1998. *WordNet: an electronic lexical database.* MIT Press, Cambridge, Mass.

Fillmore, Charles J. 1982. Frame semantics. In *Linguistics in the Morning Calm*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di semantica*, 6(2):222–254.

Fillmore, Charles J., Collin F. Baker, and Hiroaki Sato. 2004. FrameNet as a "net". In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Ghazi, Diman, Diana Inkpen, and Stan Szpakowicz. 2015. Detecting emotion stimuli in emotion-bearing sentences. In *Computational Linguistics and Intelligent Text Processing*, pages 152–165. Springer International Publishing.

Gildea, Daniel and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.

Grice, Paul. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax And Semantics*. Academic Press, New York.

Hartmann, Silvana, Ilia Kuznetsov, Teresa Martin, and Iryna Gurevych. 2017. Out-of-domain FrameNet semantic role labeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 471–482, Valencia, Spain. Association for Computational Linguistics.

Hobbs, Jerry R. and Andrew S. Gordon. 2011. The deep lexical semantics of emotions. In *Affective Computing and Sentiment Analysis*, pages 27–34. Springer.

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*, pages 125–138, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Hu, Minqing and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.

Ide, Nancy, Collin F. Baker, Christiane Fellbaum, and Rebecca J. Passonneau. 2010. The manually annotated sub-corpus: A community resource for and by the people. In *Proceedings of the ACL 2010 conference short papers*, pages 68–73.

Jänicke, Stefan, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. 2015. On close and distant reading in digital humanities: A survey and future challenges. In *EuroVis (STARs)*, pages 83–103.

Kiefer, Ferenc. 1988. Linguistic, conceptual and encyclopedic knowledge: Some implications for lexicography. In *Proceedings of EURALEX*, Budapest, Hungary.

Klinger, Roman, Orphée de Clercq, Saif Mohammad, and Alexandra Balahur. 2018. IEST: WASSA-2018 implicit emotions shared task. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Brussels, Belgium. Association for Computational Linguistics.

Kong, Lingpeng, Chris Dyer, and Noah A. Smith. 2016. Segmental recurrent neural networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Li, Yanran, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Liu, Vicki, Carmen Banea, and Rada Mihalcea. 2007. Grounded emotions. In *Seventh International Conference on Affective Computing and Intelligent Interaction (ACII 2017)*, pages 477–483. IEEE Computer Society.

Màrquez, Lluís, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. 2008. Special issue introduction: Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.

Martin, James R. and Peter R. White. 2003. *The language of evaluation*, volume 2. Springer.

Mehrabian, Albert and Norman Epstein. 1972. A measure of emotional empathy. *Journal of personality*, 40(4):525–543.

Mohammad, Saif. 2012. #emotional tweets. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 246–255, Montréal, Canada. Association for Computational Linguistics.

Mohammad, Saif and Felipe Bravo-Marquez. 2017. WASSA-2017 shared task on emotion intensity. In

Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 34–49, Copenhagen, Denmark. Association for Computational Linguistics.

Mohammad, Saif and Peter D. Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465.

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*, pages 32–41, Baltimore, Maryland. Association for Computational Linguistics.

Moors, Agnes, Phoebe C. Ellsworth, Klaus R. Scherer, and Nico H. Frijda. 2013. Appraisal theories of emotion: State of the art and future development. *Emotion Review*, 5(2):119–124.

Nandwani, Pansy and Rupali Verma. 2021. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1):1–19.

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.

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*, pages 58–70, Barcelona, Spain (Online). Association for Computational Linguistics.

Omdahl, Becky L. 1995. *Cognitive Appraisal, Emotion, and Empathy*. Mahwah, NJ: Lawrence Erlbaum.

Ortony, Andrew, Gerald L. Clore, and Allan Collins. 1988. *The cognitive structure of emotions*. Cambridge university press.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Pekar, Viktor, Michael Krkoska, and Steffen Staab. 2004. Feature weighting for co-occurrence-based classification of words. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 799–805, Geneva, Switzerland. COLING.

Plutchik, Robert. 2001. The nature of emotions. *American Scientist*, 89(4):344–350.

Preoţiuc-Pietro, Daniel, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling valence and arousal in Facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 9–15, San Diego, California. Association for Computational Linguistics.

Roth, Michael and Mirella Lapata. 2015. Context-aware frame-semantic role labeling. *Transactions of the Association for Computational Linguistics*, 3:449–460.

Ruppenhofer, Josef. 2018. The treatment of emotion vocabulary in framenet: Past, present and future developments. In Alexander Ziem, Lars Inderelst, and Detmer Wulf, editors, *Frames interdisziplinär: Modelle, Anwendungsfelder, Methoden*, pages 95–122. Düsseldorf University Press.

Ruppenhofer, Josef, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. Framenet ii: Extended theory and practice. Technical report, International Computer Science Institute.

Russell, James A. and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294.

Scarantino, Andrea. 2016. The philosophy of emotions and its impact on affective science. *Handbook of emotions*, 4:3–48.

Scheff, Thomas J. 1973. Intersubjectivity and emotion. *American Behavioral Scientist*, 16(4):501–511.

Scherer, Klaus R. 1984. Emotion as a multicomponent process: A model and some cross-cultural data. *Review of personality & social psychology*, pages 37–63.

Scherer, Klaus R. 1989. Appraisal theory. *Handbook of Cognition and Emotion*.

Scherer, Klaus R. 2005. What are emotions? And how can they be measured? *Social Science Information*, 44(4):695–729.

Scherer, Klaus R., Tanja Bänziger, and Etienne Roesch. 2010. *A Blueprint for Affective Computing: A sourcebook and manual*. Oxford University Press.

Scherer, Klaus R. and Harald G. Wallbott. 1997. The ISEAR questionnaire and codebook. Geneva Emotion Research Group.

Schuff, Hendrik, Jeremy Barnes, Julian Mohme, Sebastian Padó, and Roman Klinger. 2017. Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 13–23, Copenhagen, Denmark. Association for Computational Linguistics.

Searle, John R. 1976. A classification of illocutionary acts. *Language in society*, 5(1):1–23.

Shaikh, Mostafa Al Masum, Helmut Prendinger, and Mitsuru Ishizuka. 2009. A linguistic interpretation of the OCC emotion model for affect sensing from text. *Affective Information Processing*, pages 45–73.

Shen, Dan and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 12–21, Prague, Czech Republic. Association for Computational Linguistics.

Smith, Craig A. and Phoebe C. Ellsworth. 1985. Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, 48(4):186–209.

Stranisci, Marco Antonio, Simona Frenda, Eleonora Ceccaldi, Valerio Basile, Rossana Damiano, and Viviana Patti. 2022. APPReddit: a corpus of Reddit posts annotated for appraisal. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3809–3818, Marseille, France. European Language Resources Association.

Strapparava, Carlo and Alessandro Valitutti. 2004. WordNet affect: an affective extension of WordNet. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Swayamdipta, Swabha, Sam Thomson, Chris Dyer, and Noah A. Smith. 2017. Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold. *arXiv preprint arXiv:1706.09528*.

Tafreshi, Shabnam and Mona Diab. 2018. Emotion detection and classification in a multigenre corpus with joint multi-task deep learning. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2905–2913, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

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).

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*, pages 4005–4011, Florence, Italy. Association for Computational Linguistics.

Troiano, Enrica, Sebastian Padó, and Roman Klinger. 2021. Emotion ratings: How intensity, annotation confidence and agreements are entangled. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 40–49, Online. Association for Computational Linguistics.

Udochukwu, Orizu and Yulan He. 2015. A rule-based approach to implicit emotion detection in text. In *Natural Language Processing and Information Systems*, pages 197–203, Cham. Springer International Publishing.

Ushio, Asahi, Federico Liberatore, and Jose Camacho-Collados. 2021. Back to the basics: A quantitative analysis of statistical and graph-based term weighting schemes for keyword extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8089–8103, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Wauthier, Fabian L. and Michael Jordan. 2011. Bayesian bias mitigation for crowdsourcing. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.

Yu, Liang-Chih, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016. Building Chinese affective resources in valence-arousal dimensions. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 540–545, San Diego, California. Association for Computational Linguistics.

Zhuang, Liu, Lin Wayne, Shi Ya, and Zhao Jun. 2021. A robustly optimized BERT pre-training approach with post-training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.

A Limitations

The approach we described in Section 4 is common to data-driven information extraction lines of research which require no human intervention, such as the task of open information extraction (Etzioni et al., 2008), as well as to distant reading, i.e., the application of computational and statistical techniques in the field of digital humanities, aimed at uncovering global patterns in texts (Jänicke et al., 2015). Still, it incurs the risk of mistakes by both the emotion classifier and the frame identifier. The data we study was not collected for the sake of computational emotion analysis nor to study frames, and might differ in tone, topics and linguistic structures from the resources on which our automatic annotators were trained. As a matter of fact, the generalization capabilities of FrameNet-based parsers have been put into question by Hartmann et al. (2017), who found that a state-of-the-art system for SRL loses 16 percentage F1 points when evaluated against out-of-domain data. This issue also applies to emotions. Bostan and Klinger (2018) showed that systems for emotion detection tested out of domain suffer from performance drops as heavy as ≈.70 in F1 score. Overall, our findings are limited by the quality of the systems that we employ, but we believe that they provide evidence to learn something about the bond between frames and emotionality.

Some of our design choices could also be instantiated differently. For one thing, our annotation looks at emotions as a binary matter. Follow-up studies could observe if different frames carry specific emotions (anger, joy, etc.). Second, we benefit from word relations in the sense that these give context to identify frames, but we do not leverage roles, leaving this endeavor as our next research step. Third, to measure their association with emotionality, we treat all frames equally and as separate entities. While transparent, this choice does not account for within-sentence frames interactions.

B Corpus Labelling (Emotions)

We associate sentences in COCA to emotions automatically. Using a resource already labelled for emotions by humans could be a safer approach: people's judgments are arguably more reliable than those of a classifier, and this would have allowed us to only perform the frame-based strand of labelling. Yet, existing resources for affective computing have magnitudes of data points less than we need, and they typically focus on a specific type of texts, such as tweets (Mohammad, 2012), tales (Alm et al., 2005) or news headlines (Bostan et al., 2020). Employing a state-of-the-art classifier specialised in only one domain (i.e., trained on a single resource for emotion analysis) would give no guarantee that the obtained annotations are valid for our data. Moreover, we aim at observing frames as elicited by different emotion expressions, likely to be found in a mixture of textual domains.

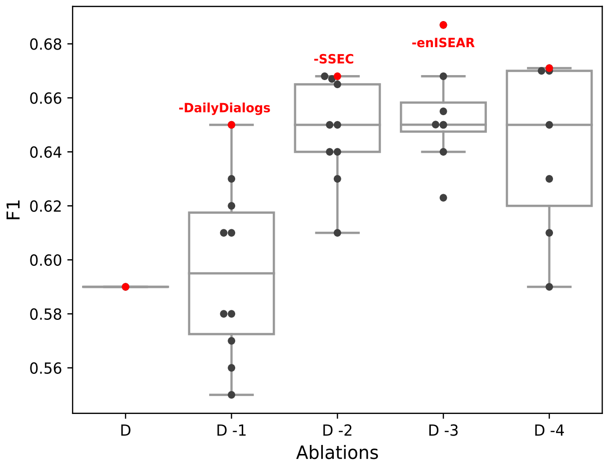

Figure 10: Model Selection: the y axis reports the F1 scores (weighted by the number of examples of each class) of the models evaluated against the annotated COCA sample. We recursively ablate datasets from the training set that yields the best model at the previous step (x axis). Dots are classifiers obtained with an ablation; the red ones indicate the best performing model: from all datasets (D), we remove each separately ("D −1"); from the set on which we obtained the best model (red dot "−DailyDialogs"), we again we remove each dataset, one at a time, thus training the next models on a collection with two datasets less than D (i.e., "D −2"); and so on.

Our model selection procedure is shown in Figure 10. Classifiers are plotted as dots in the figure, numbers on the x axis correspond to how many datasets are removed at each successive step. We kept all training parameters constant for the 35 models described in Section 4.1. They were fine-tuned for 10 epochs, setting a learning rate of $2{*}10^{-5}$ a dropout rate of 0.2, and a batch size of 32. We used AdamW as optimizer.

Recursive data elimination proceeds as a backward search. Initially, we train a classifier on all gathered corpora described in Section 4.1 ("D" in the figure, F1=.59); from these resources, we pull out each dataset separately ("D −1"), and observe that the ablation of DailyDialogs is the most beneficial (F1 increases to .65); we move on to the next ablation step and keep using the data that yielded the best performance. From that, we ablate each remaining dataset (i.e., "D −2"): now, the results reached upon removal of SSEC surpass the previously best classifier. We repeat this procedure and reach an upper bound F1 score. From the total of 35 trained models, the most competitive one

	BERT-based	RoBERTa-based
D	F1 = .83	F1 = .86
COCA sample	F1 = .69	F1 = .55

Table 12: BERT- and RoBERTa-based classifiers performance when trained and tested in domain (D) vs. trained on D and tested on the COCA sample with the majority vote treated as ground truth.

is obtained when removing DailyDialogs, SSEC, and enISEAR (F1=.69 with "D −3", which outperforms the best model in "D −4", F1=.67). We use that to annotate COCA. Note that this classifier does not correspond to the one used to select the texts for the test set.

The performances displayed in Figure 10 could be expected. First, it is hard to find classifiers that are seamlessly portable across domains. Bostan and Klinger (2018) conducted multiple experiments showing that classifiers generalise poorly across domains. They report losses as drastic as .82 F1 score when testing on out-of-domain data. For us the loss is less severe (14 points, see Table 12, column BERT-based). Second, our models are learnt on datasets whose original annotation schemata differ from one another.

For a comparison to our BERT-based model selection, we experimented with a RoBERTa-based (Zhuang et al., 2021) emotion annotator trained on the whole concatenation of corpora (D). While the latter yielded superior results when evaluated on the in-domain data, it deteriorated on the manually annotated sample of COCA as out-of-domain data. Results are reported in Table 12.

Two viable alternatives for the automatic emotion annotation step could have been: (1) to use two classifiers, having high precision for either of the considered labels – i.e., one dedicated to the labelling of the emotional category and one for nonemotional category, which could arguably be more trustworthy, and (2) to accept texts as emotional or nonemotional if the probability with which the classifier assigns a label exceeds a given threshold. However, the first case would pose the problem of deciding how to treat texts for which the two models are in disagreement with one another. In the other case, we would lose substantial data. Our decision to adopt an individual emotion labeller, with a reasonable F1, bypasses both issues.

Adopting an annotation approach entirely based on human judgments would not be unproblematic either: large data sources compiled via crowdsourcing are noisy since they are labelled by naïve judges (Wauthier and Jordan, 2011); on the other hand, annotations conducted by expert coders are more reliable, but they typically cover smaller data, and this makes empirical observations difficult to draw. We forgo the latter. Indeed, when it comes to judging emotions, the noisiness problem characterises all human-based annotations, because the task is extremely subjective and therefore can lead to extreme disagreements, irrespective of how trained the coders are. Therefore, should the results of our analysis be due to systematic misclassifications of the automatic annotator, we could assume that similar "errors" are to be found among humans.

C Appraisal-based Frames

While discussing our partition of frames, we have highlighted that many items annotated as Appraisal-based frames tap on evaluations and cognitive processes. They are more than appear at first brush. Some singular examples are:

- REASONING, which often accompanies texts where an evaluation is expressed by means of a dispute described in the text;
- FAME, appearing in sentences with assessments that are either hyperbolic, like "*Believe me it was epic.*", or that concern one's reputation and beliefs, like "*To besmirch her reputation is outrageous*".

Likewise, the placement of BREATHING, CAUSE_IMPACT and LEVEL_OF_FORCE_EXERTION in this cluster of frames might not be self-explanatory. The first two indicate an emotional reaction, (e.g., sighing and slamming a door). The last usually portrays a property of people or events (e.g., feeling fearless and strong, feeling weak). So do also the following frames:

- DYNAMISM, evoked by texts that express the intensity of an experience;
- MEET_SPECIFICATIONS, coupled in text with mentions of personal achievements, or with expressed sensations of fulfilment.