# Extracting Causes of Emotions from Text

**Alena Neviarouskaya**
Toyohashi University of Technology
Toyohashi, Japan
alena@kde.cs.tut.ac.jp

**Masaki Aono**
Toyohashi University of Technology
Toyohashi, Japan
aono@kde.cs.tut.ac.jp

## Abstract

This paper focuses on the novel task of automatic extraction of phrases related to causes of emotions. The analysis of emotional causes in sentences, where emotions are explicitly indicated through emotion keywords can provide the foundation for research on challenging task of recognition of implicit affect from text. We developed a corpus of emotion causes specific for 22 emotions. Based on the analysis of this corpus we introduce a method for the detection of the linguistic relations between an emotion and its cause and the extraction of the phrases describing the emotion causes. The method employs syntactic and dependency parser and rules for the analysis of eight types of the emotion-cause linguistic relations. The results of evaluation showed that our method performed with high level of accuracy (82%).

## 1 Introduction and Background

Emotional reactions to three salient aspects of the world, namely (1) events and their consequences, (2) agents and their actions, and (3) objects, are based on the nature of cognitive origins and can be triggered under specific conditions (Ortony et al., 1988). The cognitive model of emotions (OCC model of emotions) arranges 22 emotions in three substantially independent classes according to the aspects of the world that are in focus of evaluation.

Recently, the task of automatic recognition of distinct emotions conveyed in text has been gaining increased attention of researchers in the areas of natural language processing and computational linguistics (Alm, 2008; Aman and Szpakowicz, 2008; Boucouvalas, 2003; Chaumartin, 2007; Katz et al., 2007; Kozareva et al., 2007; Liu et al., 2003; Neviarouskaya et al., 2011; Purver and Battersby, 2012; Strapparava and Mihalcea, 2008; Suttles and Ide, 2013). To understand emotions expressed in written language, it is important to analyse the causes of emotions ("*what caused a particular emotion*") and eliciting conditions ("*under what conditions*"). The challenge of emotion cause detection in text has been recently tackled by Chen et al. (2010), who developed two sets of linguistic pattern-based features (manually generalized patterns and automatically generalized patterns) for extraction of causes for emotions in Chinese. The linguistic-pattern-based methodology described in (Chen et al., 2010) inspired the development of a method for the identification of Italian sentences that contain emotion cause phrases and the retrieval of emotion – emotion cause phrase couples (Russo et al., 2011). In their subsequent work, Caselli et al. (2012) semi-automatically assigned polarity values to Italian nouns that potentially represent nominal cause events associated with emotions.

In this work, we introduce a novel method for automatic extraction of emotion causes. The main contributions of our work are as follows: (1) development of a corpus of emotion causes and (2) deep analysis of cause events specific for 22 emotions from the OCC model. The analyses of emotional causes in sentences, where emotions are explicitly indicated through emotion keywords, and conditions that lead to emotional experience can provide the foundation for research on challenging task of recognition of implicit affect from text.

## 2 Development and Analysis of the Corpus of Emotion Causes

### 2.1 Creation of the Dataset of Sentences with Explicitly Indicated Emotions

In the text of (Ortony et al., 1988), about 130 tokens (emotion words) have been distributed between 22 emotion types. For example, '*glad*' and



*International Joint Conference on Natural Language Processing*, pages 932–936,
Nagoya, Japan, 14-18 October 2013.

'*happy*' correspond to *Joy* emotion class; '*scared*' and '*terrified*' are associated with *Fear* emotion; and '*awe*' and '*esteem*' describe *Admiration* emotion. We consider these tokens as seed terms for extraction of sentences that contain information on what caused the particular emotion.

In addition to 22 sentences provided in (Ortony et al., 1988) as examples for each emotion type, we manually collected 510 sentences with emotion tokens and explicitly mentioned emotion causes from online ABBYY Lingvo dictionary (*http://www.lingvo-online.ru/en*). 118 emotion tokens were found productive, resulting in at least one cause-containing sentence per emotion token.

The corpus consisting of 532 sentences was manually annotated. The annotation task included the following subtasks: (1) to define an agent or an experiencer of emotion specified by emotion token; (2) to delimit the phrase describing the cause of emotion; (3) to define the linguistic relation between emotion and its cause; (4) to classify the cause event as positive, negative, or neutral; and (5) to extract tokens that influence the polarity of the phrase.

### 2.2 Corpus Analysis

We performed the detailed analysis of the created corpus. The agent or experiencer of emotion specified by emotion token was defined in 495 sentences (93% from the whole corpus). In the corpus, about 46% of sentences are related to positive emotions, and about 54% of sentences express negative emotions.

The analysis of polarity of cause events from the annotated corpus showed the following distribution of the causes according to the sentiment categories: (1) positive – about 27%; (2) negative – about 29%; and (3) neutral – about 44% of the cause events. These figures emphasize the fact that the cause of emotion expressed in text is not necessarily described by sentiment words. Interesting observation is that cause events are negative in 2.9% of sentences with positive emotions, and positive cause events occur in 4.5% of sentences with negative emotions (for example, '*And people changed from diet to diet and felt guilty* [negative emotion] *because they continued to like the things they weren't supposed to*').

The important feature that was identified in each sentence was the linguistic relation between emotion and its cause. Based on the analysis of the annotated data, we distinguish eight types of such linguistic relations:

1. One-word preposition (OWP). For example, '*at*' in the sentence '*And while she gaped with disappointment <u>at</u> his lukewarmness, he got himself away, at ten*'.
2. Complex preposition (CP). For example, '*because of*' in the sentence '*He was himself a Greek, and there were many who felt offended <u>because of</u> his height*'.
3. Coordinating conjunction (CC). For example, '*for*' in the sentence '*La Cote was much depressed, <u>for</u> he had scored here the worst failure of his campaign*'.
4. Subordinating conjunction (SC). For example, '*because*' in the sentence '*And people changed from diet to diet and felt guilty <u>because</u> they continued to like the things they weren't supposed to*'.
5. Subject (SUBJ). For example, in the sentence '*<u>His tone</u> scared her more than anything she could remember*', the subject '*his tone*' represents the cause of *Fear* emotion expressed by the verb '*scared*'.
6. Verb or predicate (V). For example, the predicate '*filled with*' connects the *Joy* emotion with its cause in the sentence '*As for the captain, the presence in his room of the children, who came to cheer up Ilusha, <u>filled</u> his heart from the first <u>with</u> ecstatic joy*'.
7. Object (OBJ). For example, in the sentence '*I adore <u>poetry</u>*', the object '*poetry*' triggers *Love* emotion that is reflected through the verb '*adore*'.
8. Attributive nominal (ATT). For example, in the sentence '*It is a <u>sad tale</u>, a very <u>sad tale</u>*', emotional adjective '*sad*' describes the noun '*tale*' through attributive nominal relation (in this sentence, '*tale*' causes *Distress* emotion).

In Table 1, the specific emotion-cause linguistic relations that were found in our corpus of sentences are listed according to their frequency. One-word prepositions (including '*to*', '*for*', '*of*', '*at*', '*with*', '*by*', '*about*', '*over*' etc.) acting as linkages between emotion tokens and phrases describing the cause of emotion occur in about 68.2% of sentences. Subordinating conjunctions (examples include '*that*', '*when*', '*because*', '*as*' etc.) constitute about 21.4% of sentences. The object and subject are the next frequent relation types (about 6% and 2.3% of sentences, respectively).

## 3 Method for Extraction of Emotion Causes

Our method for automatic extraction of emotion causes is based on the analysis of syntactic and dependency information from the parser. In our

| Relation | Type | Frequency (number) | Frequency (%) |
|---|---|---|---|
| *to* | OWP | 77 | 14.47 |
| *for* | OWP / CC | 73 | 13.72 |
| *that* | SC | 63 | 11.84 |
| *of* | OWP | 48 | 9.02 |
| *at* | OWP | 42 | 7.89 |
| *with* | OWP | 37 | 6.95 |
| object | OBJ | 32 | 6.02 |
| *by* | OWP | 25 | 4.70 |
| *about* | OWP | 22 | 4.14 |
| *when* | SC | 21 | 3.95 |
| *over* | OWP | 20 | 3.76 |
| *because* | SC | 15 | 2.82 |
| subject | SUBJ | 12 | 2.26 |
| *in* | OWP | 9 | 1.69 |
| *on* | OWP | 7 | 1.32 |
| attribute | ATT | 6 | 1.13 |
| *as* | SC | 5 | 0.94 |
| *if* | SC | 5 | 0.94 |
| *as though* | SC | 4 | 0.75 |
| *filled with; fostered by; trigger* | V | 3 | 0.56 |
| *after* | OWP / SC | 1 | 0.19 |
| *as if* | SC | 1 | 0.19 |
| *because of* | CP | 1 | 0.19 |
| *from* | OWP | 1 | 0.19 |
| *under* | OWP | 1 | 0.19 |
| *without* | OWP | 1 | 0.19 |

Table 1. Emotion-cause linguistic relations and their frequency in the corpus

| Id | Token | Lemma | Dependency | Tags |
|---|---|---|---|---|
| 1 | Most | many | qn:>2 | @QN> %>N DET SUP PL |
| 2 | doctors | doctor | subj:>3 | @SUBJ %NH N NOM PL |
| 3 | are | be | v-ch:>4 | @+FAUXV %AUX V PRES |
| 4 | attracted | attract | main:>0 | @-FMAINV %VP EN |
| 5 | to | to | ha:>4 | @ADVL %EH PREP |
| 6 | medicine | medicine | pcomp:>5 | @<P %NH N NOM SG |
| 7 | because | because | pm:>9 | @CS %CS CS |
| 8 | they | they | subj:>9 | @SUBJ %NH PRON PERS NOM PL3 |
| 9 | look | look | cnt:>4 | @+FMAINV %VA V PRES |
| 10 | forward | forward | goa:>9 | @ADVL %EH ADV |
| 11 | to | to | ha:>9 | @ADVL %EH PREP |
| 12 | curing | cure | pcomp:>11 | @<P-FMAINV %VA ING |
| 13 | disease | disease | obj:>12 | @OBJ %NH N NOM SG |

Table 2. Example of parser output

work we employ Connexor Machinese Syntax (*http://www.connexor.eu/technology/machinese/*) that is applied to each sentence in order to get lemmas, dependencies, syntactic and morphological information (see example in Table 2). Using parser output, the method extracts phrases that characterize the emotion causes.

The algorithm detects and extracts cause phrases introduced by prepositions (OWP and CP) through three rules:

1. POSTMODIFIER rule: if morphological tag of the cause marker is *PREP* and this preposition is linked with the emotion token through *mod* syntactic relation, then extract all tokens related to this preposition.
2. NEXT TOKEN rule: if morphological tag of the cause marker is *PREP* and syntactic relation of this preposition is unavailable (*null* relation), then if this cause marker directly follows the emotion token, extract all tokens related to this preposition.
3. VERB-MEDIATED RELATION rule: if morphological tag of the cause marker is *PREP* and this preposition is directly connected with verb, to which emotion token is related within the clause, and the id of preposition is higher than that of emotion token, then extract all tokens related to this preposition.

The rules for extraction of phrases connected to emotion tokens through conjunctions (SC and CC) are as follows:

1. THAT rule: if morphological tag of the '*that*' cause marker is *CS* and the id of conjunction is higher than that of emotion token, then if verb of subordinate clause, to which the conjunction '*that*' is connected, is related to emotion token through chain of relations, extract all tokens related to the verb of subordinate clause.
2. DEPENDENT CLAUSE rule: if morphological tag of the cause marker is *CS* or *CC*, and the dependent verb, to which conjunction is related, is connected to the main verb, to which emotion token is related (here, the emotion token might be the verb itself), then extract all tokens related to the verb of dependent clause.

To detect verbs for the above rules, the algorithm looks for the following functional tags: *@+FMAINV* (finite main verb), *@-FMAINV* (nonfinite main verb), and *@<P-FMAINV* (nonfinite clause as preposition complement).

The extraction of emotion causes represented by either subject (SUBJ), or predicate (V), or object (OBJ), or attributive nominal (ATT) linguistic relations is based on the analysis of *subj*, *obj*, and *att* syntactic relations and the corresponding tokens.

## 4 Evaluation

Based on the emotion cause phrases extracted by human annotator from our corpus consisting of 532 sentences, we evaluated the appropriateness of the phrases extracted by our algorithm. In each pair of phrases, the number of words was calculated (namely, number of gold standard tokens and number of automatically extracted tokens). Then, the number of words correctly extracted by our algorithm was found, and we calculated precision, recall, and F-score for each automatically extracted phrase. The results averaged over all the phrases are given in Table 3 (including the results on different groups and all emotion cause linguistic relations).

| **Emotion cause linguistic relations** | **Accuracy of phrase extraction** | | |
|---|---|---|---|
| | **Precision** | **Recall** | **F-score** |
| Prepositions (OWP, CP) | 0.715 | 0.723 | 0.700 |
| Conjunctions (SC, CC) | 0.470 | 0.549 | 0.473 |
| Subject, predicate, object, and attributive nominal (SUBJ, V, OBJ, ATT) | 0.787 | 0.793 | 0.772 |
| All relations | **0.670** | **0.692** | **0.658** |
| All relations (after improving the method based on error analysis) | **0.821** | **0.852** | **0.810** |

Table 3. Evaluation of the appropriateness of automatically extracted emotion causes

As seen from the obtained results, our algorithm achieved the highest level of precision (0.787) in extracting emotion cause phrases represented by subject, predicate, object, and attributive nominal linguistic relations, while it was least precise (0.470) in case of emotion causes introduced by conjunctions. We obtained good results considering all emotion cause linguistic relations: precision in 0.670, recall in 0.692, and F-score in 0.658.

We performed an error analysis on the sentences, where our method failed to extract correct phrases. The classification and distribution of errors is given in Table 4. In most cases (about 44.8%), the method failures were due to missing rule for infinitive marker '*to*' (morphological tag *INFMARK*>, in contrast to preposition tag *PREP*). For example, '*to*' in the sentence '*In that regard, New Zealand is proud <u>to</u> work towards nuclear disarmament with the other members of the New Agenda Coalition*'. About 22.4% of errors were caused by inability of the parser to output correct tags for syntactic relations. Analysis of '*when*' as a relative adverb (*ADV* and *WH* morphological tags), in addition to it as a subordinating conjunction, would deal with about 13.4% of errors. We found that the emotion causes represented by subordinate clauses without such a marker of subordination as '*that*' pose the main challenge, as the parser outputs *null* relations for such dependent clauses (for example, clause '*I never had to lie then*' in the sentence '*I reckon I was so glad I never had to lie then*'). The analysis of errors showed the necessity to improve several rules (such as THAT, POSTMODIFIER, SUBJ, and OBJ rules). The method would also benefit from adding reference resolution. For example, using reference resolution, the method could extract '*these difficulties*' instead of '*they*' as emotion cause from the sentence '*I could not dwell upon these difficulties fully, for they made me far too uneasy*'.

| **Error type** | **Frequency (number)** | **Frequency (%)** |
|---|---|---|
| Infinitive marker '*to*' | 60 | 44.78 |
| Null or incorrect tag from parser | 30 | 22.39 |
| '*When*' as a relative adverb | 18 | 13.43 |
| Missing subordinating conjunction '*that*' | 11 | 8.21 |
| THAT rule | 4 | 2.99 |
| POSTMODIFIER rule | 3 | 2.24 |
| Emotion phrase '*look forward*' | 3 | 2.24 |
| Reference resolution | 3 | 2.24 |
| Coordinating conjunction in SUBJ and OBJ rules | 2 | 1.5 |
| **Total** | **134** | **100** |

Table 4. Classification and distribution of errors

After improving the emotion cause extraction method by adding and modifying the rules, we obtained the following evaluation results: precision in 0.821, recall in 0.852, and F-score in 0.810 (last row in Table 3). In that way, our method performed with about 15% gain in accuracy.

## 5 Conclusions

The main contributions of our work are the creation of a corpus of emotion causes specific for 22 emotions from the OCC model and the development of a novel method for extraction of emotion causes from sentences based on the analysis of syntactic and dependency information provided by the parser. In future research we plan to improve our emotion cause extraction method and incorporate the automatic detection of an experiencer of emotion specified by emotion token and the classification of causes as positive, negative, or neutral.

## References


Andrew Ortony, Gerald L. Clore, and Allan Collins. 1988. *The Cognitive Structure of Emotions*. Cambridge University Press.

Cecilia O. Alm. 2008. *Affect in Text and Speech*. PhD Dissertation, Urbana, IL: University of Illinois at Urbana-Champaign.

Saima Aman and Stan Szpakowicz. 2008. Using Roget's Thesaurus for Fine-Grained Emotion Recognition. *Proceedings of the Third International Joint Conference on Natural language Processing (IJCNLP 2008)*, pp. 296-302.

Anthony C. Boucouvalas. 2003. Real Time Text-to-Emotion Engine for Expressive Internet Communications. In *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments*, Ios Press, pp. 306-318.

Francois-Regis Chaumartin. 2007. UPAR7: A Knowledge-Based System for Headline Sentiment Tagging. *Proceedings of SemEval-2007*.

Phil Katz, Matt Singleton, and Richard Wicentowski. 2007. SWAT-MP: the SemEval-2007 Systems for Task 5 and Task 14. *Proceedings of SemEval-2007*.

Zornitsa Kozareva, Borja Navarro, Sonia Vazquez, and Andres Montoyo. 2007. UA-ZBSA: A Headline Emotion Classification through Web Information. Proceedings of SemEval-2007.

Hugo Liu, Henry Lieberman, and Ted Selker. 2003. A Model of Textual Affect Sensing Using Real-World Knowledge. *Proceedings of the International Conference on Intelligent User Interface*s, pp. 125-132.

Matthew Purver and Stuart Battersby. 2012. Experimenting with Distant Supervision for Emotion Classification. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 482-491.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to Identify Emotions in Text. *Proceedings of the 2008 ACM Symposium on Applied Computing*, pp. 1556-1560.

Jared Suttles and Nancy Ide. 2013. Distant Supervision for Emotion Classification with Discrete Binary Values. *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 121-136.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2011. Affect Analysis Model: Novel Rule-Based Approach to Affect Sensing from Text. *International Journal of Natural Language Engineering*, 17(1):95-135. Cambridge University Press.

Ying Chen, Sophia Y. M. Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion Cause Detection with Linguistic Constructions. *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 179-187.

Irene Russo, Tommaso Caselli, and Francesco Rubino. 2011. EMOCause: An Easy-Adaptable Approach to Emotion Cause Contexts. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, ACL-HLT 2011, pp. 153-160.

Tommaso Caselli, Irene Russo, and Francesco Rubino. 2012. Assigning Connotation Values to Events. *Proceedings of the Eight International Conference on Language Resources and Evaluation*, pp. 3082-3089.