

Guidelines for the Fine-Grained Analysis of Polar Expressions

version 2.0

LT3 Technical Report – LT3 14-02

Marjan Van de Kauter, Bart Desmet and Véronique Hoste

LT3, Language and Translation Technology Team
Department of Translation, Interpreting and Communication – Ghent University
marjan.vandekauter, bart.desmet, veronique.hoste@ugent.be

URL: <http://www.lt3.ugent.be>¹

January 8, 2015

¹The reports of the LT3 Technical Report Series (ISSN 2032-9717) are available from <http://www.lt3.ugent.be/en/publications/> All rights reserved. LT3, Department of Translation, Interpreting and Communication – Ghent University, Belgium.

Contents

1	Introduction	1
2	Types of Polar Expressions	3
2.1	Private State Expressions	3
2.2	Polar Fact Expressions	4
2.3	Categorizing Polar Expressions	4
2.4	Polar Expressions with Unknown or Ambiguous Polarity	6
3	Annotation of Polar Expressions	8
3.1	Elements	8
3.2	Annotation Procedure	9
3.2.1	Polar Expression	9
3.2.2	Modifiers	13
3.2.3	Source and Source Expression	15
3.2.4	Target and Sentiment Polarity	18
3.2.5	Causativity	20
3.3	Identification of Embedded Polar Expressions	21
4	Annotation of Coreferential and Feature Relations	23
4.1	Coreferential Relations	23
4.2	Feature Relations	23
5	Annotation of Non-Polar Causal Relations	25
5.1	Elements	25
5.2	Annotation Procedure	25

Chapter 1

Introduction

In recent years, research on **sentiment analysis** (also known as **opinion mining**) has drawn increasing interest. This text mining task is aimed at the computational extraction and analysis of **evaluative text** (sentiment), i.e. opinions, feelings, emotions, etc. An important subtask of sentiment analysis is **polarity classification**, which labels the identified sentiment as being either positive or negative (Pang and Lee, 2008).

Systems that can automatically extract positive and negative sentiment from text can be used for different types of text from a wide range of domains, but up until now the field's main focus has been on the detection of opinions in user reviews and blogs (Turney, 2002; Hu and Liu, 2004; Ding, Liu, and Yu, 2008; Ounis et al., 2006)¹. In this kind of **user-generated content**, sentiment is often expressed in an explicit way: a person communicates his/her personal attitude towards a certain product or topic. An example of this is sentence (1), taken from the sentiment annotation guidelines of (Toprak, Jakob, and Gurevych, 2010). Here, the writer utters a positive opinion about *the restaurant*.

(1) *The restaurant was terrific: the best chocolate ice cream ever and platters upon platters of breakfast choices.*

As a result of the current focus on this type of data, most of the existing sentiment annotation schemes solely cover **subjective** statements which explicitly express a positive or negative evaluation (Wiebe, Wilson, and Cardie, 2005; Seki et al., 2007). However, **factual** sentences can also result in a positive or negative impression of a certain entity (Toprak, Jakob, and Gurevych, 2010). See for example sentence (2) below (taken from (Toprak, Jakob, and Gurevych, 2010)). Although the writer does not utter any personal opinion about the subject, a negative evaluation of *the computer* can be inferred from the factual content of the sentence.

(2) *The computer crashed every day.*

Such implicit expressions of sentiment are particularly common in more general text such as newswire (Musat and Trausan-Matu, 2010). We find that most existing annotation schemes are insufficient to capture all occurrences of sentiment in data like this. Therefore we propose a new scheme for the fine-grained analysis of **explicit as well as implicit expressions of positive and negative sentiment**. These so-called **polar expressions** can be divided into two categories:

¹Possible applications here are product benchmarking, understanding what voters are thinking about a certain politician or political party, etc. (Pang and Lee, 2008).

private state expressions and polar fact expressions. In Chapter 2, we explain how these types of polar expression can be identified and distinguished from each other. Chapter 3 describes the annotation procedure for polar expressions. In addition to polar expressions, our annotation scheme also covers **coreferential and feature relations** (see Chapter 4) and **causal relations** (see Chapter 5). All annotation is performed using the brat rapid annotation tool (Stenetorp et al., 2012).

The example sentences in the following chapters are taken from a non-parallel corpus of Dutch and English news articles. Note that often, not every single expression or relation in these sentences is discussed or completely analyzed. Only Chapter 6 contains the full annotation for a few of the example sentences.

Chapter 2

Types of Polar Expressions

A polar expression is a **linguistic expression of positive or negative sentiment**. Sentence (1) for example contains the polar expression *was terrific*, which expresses positive sentiment towards *the restaurant*.

(1) *The restaurant was terrific: the best chocolate ice cream ever and platters upon platters of breakfast choices.*

In our annotation scheme, two types of polar expressions are identified:

- **Private state expressions:** explicit expressions of positive or negative sentiment towards a target entity or entities (see Section 2.1)
- **Polar fact expressions:** implicit expressions of positive or negative sentiment (see Section 2.2)

In Section 2.3, we describe how these two categories can be distinguished from one another. Section 2.4 elaborates on polar expressions of which the polarity is unknown or ambiguous.

2.1 Private State Expressions

Private states are **internal states** that cannot be directly observed by others, e.g. opinions, beliefs, thoughts, feelings, emotions, judgments, etc. (Quirk et al., 1985; Wiebe, Wilson, and Cardie, 2005). They represent a person's **personal attitude towards a target entity**. Expressions of private states can be found in sentences (3) and (4). The author of sentence (3) expresses a positive evaluation of *de liquiditeit en solvabiliteit van KBC* by means of the private state expression *zijn gewoon goed*. In example (4), the private state expression *criticized* tells us that *China* holds a negative attitude towards *Mitt Romney's Jerusalem comments*.

(3) *De liquiditeit en solvabiliteit van KBC zijn gewoon goed.*

(4) *China criticized Mitt Romney's Jerusalem comments.*

Sometimes private states convey neither positive nor negative sentiment (Wilson, 2008). An example of an emotion that is **not always positive or negative** is ‘surprise’ (e.g. the astronomers’ surprise in sentence (5)). Our annotation scheme also covers expressions of non-polar private states like this; they receive the **polarity label ‘other’** (see Section 3.2.4).

(5) *Astronomers are surprised by the quantity and the variety of planets that Kepler has spotted in its first year.*

2.2 Polar Fact Expressions

Sometimes, the author of a sentence does **not explicitly** utter his/her personal attitude towards a certain topic, but an evaluation of it can nonetheless be inferred from the factual content of the sentence using common sense, world knowledge and context (Toprak, Jakob, and Gurevych, 2010). An expression conveying a piece of **factual information that, when relating to a certain target entity or entities, results in a positive or negative impression of that target/those targets** is called a polar fact expression¹. In sentence (6) for instance, *produces profit of \$2.2bn* is a polar fact expression with *BP* as its target entity, because the fact that Morgan Stanley produces profit results in a positive impression of *Morgan Stanley*. From example (7), we deduce a negative evaluation of *Kim Clijsters* on the basis of the polar fact expression *heeft de kwartfinale van het olympisch tennistornooi verloren*, which has *Kim Clijsters* as its target. Finally, sentence (8) implies positive sentiment towards *het aandeel van Monster Beverage*.

(6) *Morgan Stanley produces profit of 2.2bn.*

(7) *Kim Clijsters heeft de kwartfinale van het olympisch tennistornooi verloren.*

(8) *Het aandeel van Monster Beverage is het meest gestegen.*

2.3 Categorizing Polar Expressions

Categorizing a polar expression as either a private state expression or a polar fact expression is not always straightforward. The following describes the difference between the two types of polar expressions and points out a few pitfalls annotators should be aware of when categorizing polar expressions.

The main difference between private states and polar facts is that the former contain **subjective information**. A private state is a person’s personal point of view about something; it is an internal state that cannot be directly observed by others. Polar facts on the other hand contain **factual information**, which can be objectively verified. We illustrate the distinction between private state and polar fact expressions by analyzing examples (3) to (8).

(3) *De liquiditeit en solvabiliteit van KBC zijn gewoon goed.*

⇨ *zijn gewoon goed* = private state expression [this is a subjective evaluation; whether or not this is true cannot be objectively measured here]

¹In other sources, polar facts are referred to as objective polar utterances (Wilson, 2008) or evaluative factuais (Nigam and Hurst, 2004).

(4) *China criticized Mitt Romney's Jerusalem comments.*

⇨ *criticized* = private state expression [this is a personal attitude]

(5) *Astronomers are surprised by the quantity and the variety of planets that Kepler has spotted in its first year.*

⇨ *are surprised by* = private state expression [surprise is an inner state that cannot be directly observed by others]

(6) *Morgan Stanley produces profit of \$2.2bn.*

⇨ *produces profit of \$2.2bn* = polar fact expression [this is a fact; the bank's profit can be objectively verified]

(7) *Kim Clijsters heeft in twee sets de kwartfinale van het olympisch tennistornooi verloren.*

⇨ *heeft in twee sets de kwartfinale van het olympisch tennistornooi verloren* = polar fact expression [this is a fact; the outcome of the match can be objectively verified]

(8) *Het aandeel van Monster Beverage is het meest gestegen.*

⇨ *is het meest gestegen* = polar fact expression [this is a fact; whether or not this is true can be objectively verified by comparing the share price to other shares on the market]

(Toprak, Jakob, and Gurevych, 2010) lists a few **criteria** which can help annotators choose between private states and polar facts. When the answer to both of these questions is 'yes', the polar expression in question contains objective information and can thus be labeled as a polar fact expression.

- Does the sentence only report factual information?
- Is all information given in the sentence verifiable / falsifiable (for example by measuring something)?

Despite criteria like this, it is sometimes difficult to label a polar expression as being either a private state or a polar fact expression because the distinction between subjectivity and objectivity is not always so clear. In our annotation scheme, polar expressions can be located on a continuum from objective to subjective; the more subjective an expression, the further it is situated on the continuum. Private state expressions are located on the far right side of this **subjectivity continuum**, whereas polar fact expressions are placed on the far left side. Polar expressions that cannot unambiguously be classified as a private state or a polar fact expression, are situated somewhere between the two extremes of the continuum. Take for example sentences (9) to (14). They all express positive sentiment towards *Djokovic* (and negative sentiment towards *Roger Federer*), but the polar expressions differ in degree of subjectivity. While most annotators would probably label the first 2 or 3 as polar fact expressions, they will be likely to disagree on the categorization of the polar expressions in the other sentences. In our annotation scheme, polar expressions like this can be located somewhere between polar fact and private state expression on the subjectivity continuum (see Section 3.2.1). Annotators should however label them as private state or polar fact expressions as often as possible.

(9) *Djokovic won het van Roger Federer in 3 sets.*

⇨ *won het van* = polar expression

(10) *Djokovic versloeg Roger Federer in 3 sets.*

⇨ *versloeg* = polar expression

(11) *Djokovic klopte Roger Federer in 3 sets.*

⇨ *klopte* = polar expression

(12) *Djokovic stuurde Roger Federer in 3 sets naar huis.*

⇨ *stuurde ... naar huis* = polar expression

(13) *Djokovic maakte Roger Federer in 3 sets in.*

⇨ *maakten ... in* = polar expression

(14) *Djokovic droogde Roger Federer af in 3 sets.*

⇨ *droogde ... af* = polar expression

2.4 Polar Expressions with Unknown or Ambiguous Polarity

Unlike private states, polar facts do **not explicitly** express positive or negative sentiment. They contain factual information from which, depending on the target entity or entities, a positive or negative evaluation of that target/those targets can be inferred using common sense or world knowledge. In other words, in order to determine the polarity of the expressed sentiment, **interpretation is needed**. This poses a problem in certain text types because

- sometimes interpretation requires **domain-specific knowledge**
- sometimes a polar fact can be interpreted from **different perspectives**

These complications occur in interpreting sentence (15), taken from a financial news article. Here, *lowers* is a polar fact expression with *interest rates* as its target, resulting in a positive or negative impression of the *interest rates*. Annotators with a limited knowledge of finance might have difficulties determining whether the implicit sentiment expressed about the interest rates is positive or negative. Even for domain experts, this is not a straightforward task, since this polar fact can be viewed from different angles. A lower interest rate is positive for people who want to take out a loan, but can also be considered negative in that it may cause inflation. In our annotation scheme, polar fact expressions with **unknown or ambiguous polarity** receive the **polarity label 'unknown'** (see Section 3.2.4). However, if the author's interpretation of the expression can be deduced from the text, annotators should opt for the polarity of this interpretation instead of using the polarity label 'unknown'.

(15) *Fed lowers interest rates.*

Annotators are encouraged to **annotate any expression which they think to be polar**, even if they are not entirely sure whether the expressed sentiment is positive or negative. Using the 'unknown' label for polar expressions with unknown or ambiguous polarity enables us to determine the polarity of these expressions with the help of domain experts in a later annotation stage. Of course, the polarity labels 'positive', 'negative' and 'other' should be used as often as possible.

Chapter 3

Annotation of Polar Expressions

As described in Chapter 2, we identify two categories of polar expressions: private state expressions and polar fact expressions. Since these types of polar expressions have a similar structure, the following chapter describes one annotation procedure which can be applied to all two of them.

For each polar expression, annotators should mark several elements, which are listed in Section 3.1. In our annotation scheme, all polar expressions and their respective elements are detected and analyzed on a **sub-sentential level**. This means that a sentence can contain **multiple expressions**, and that a **mixture** of private state expressions and polar fact expressions is possible. Each of these expressions is to be analyzed individually. Furthermore, one polar expression **can be embedded** in another. Annotators should **start by detecting and analyzing polar expressions on the lowest level** using the procedure described in Section 3.2 and then work their way up in the sentence. Section 3.3 explains how to determine whether a polar expression is enclosed by another polar expression on a higher sentence level. In Chapter 6 a few example sentences can be found with their full annotation, among which a couple of sentences that contain multiple polar expressions.

3.1 Elements

A **polar expression** is a linguistic expression of positive or negative sentiment. Besides the polar expression itself, annotators should also mark the following related elements:

- **Target(s)**: the entity (or entities) sentiment is being expressed towards
- **Source**: the person or entity expressing the sentiment [not always explicated]
- **Source expression**: the linguistic expression that links the source to the polar expression [only if the source is explicated]
- **Modifier(s)**: the lexical item(s) causing a shift in the sentiment being expressed [not always present]
- **Cause**: the entity causing the positive or negative sentiment being expressed towards the target [only present in the case of a causative polar expression]

These elements and their respective attributes are annotated using the procedure in Section 3.2. Marking the polar expression is the starting point of the annotation process, since this is the core element of the annotation, to which all other elements will be linked.

The elements listed above are to be identified for both types of polar expressions. In the description of the annotation steps below, we will use sentences (4) and (16) as examples to demonstrate how the different elements and their corresponding attributes need to be annotated for private state expressions and polar fact expressions respectively.

(4) *China criticized Mitt Romney's Jerusalem comments.*

(16) *The Perth-based company announced profits had rocketed by 25 per cent.*

Further illustration will be provided through the analysis of additional example sentences. For each annotation step, it is explained under 'BRAT HOWTO' how the respective element and its attributes are to be annotated using the brat tool.

3.2 Annotation Procedure

The annotation procedure for polar expressions can be summarized as follows:

1. **Detect the polar expression**
2. **Mark all modifiers affecting this polar expression (not always possible)**
3. **Annotate the source and source expression (if the source is explicated)**
4. **Identify the target(s) of the polar expression and determine the polarity and intensity of the sentiment being expressed towards it (them)**
5. **Mark the cause of the sentiment being expressed (in the case of a causative polar expression)**

Each of these annotation steps is described in the following subsections.

3.2.1 Polar Expression

This is the linguistic expression that (explicitly or implicitly) expresses positive or negative sentiment. Besides marking the polar expression itself, annotators should also specify its type (private state expression or polar fact expression).

(4) *China criticized Mitt Romney's Jerusalem comments.*

◇ *criticized* = polar expression - private state expression

(16) *The Perth-based company announced profits had rocketed by 25 per cent.*

⇨ *had rocketed by 25 per cent* = polar expression - polar fact expression

The polar expression is the **minimum span of words that conveys the sentiment being expressed**. Furthermore, the polar expression is on the **same syntactic level as its target** (if this target can be found in the same sentence). Note that targets, sources and causes are separate elements from the polar expression and are never included in its span. Modifiers however are often included in the polar expression span because they are part of the same syntactic phrase. It is also possible for a text span to be the polar expression and its source expression at the same time (see Section 3.2.3).

There are no restrictions as to what form polar expressions can take: they can be adjectives, noun phrases, verbal constructions, etc. The only condition is that they **cannot span more than one sentence or less than one word**. Here are a few examples of grammatical phrases and constructions that can be marked as polar expressions (note that this list is not exhaustive)¹:

- **adjective or adjective phrase**

as a **pre-noun adjective (the noun being the target)**

(17) *De banken maken winst, ondanks de wispelturige economie.*

⇨ *wispelturige* = polar expression - private state expression (*economie* = target)

as a **secondary predicate (the object being the target)**

(18) *Van Leemput vindt het aandeel prijzig in absolute termen.*

⇨ *prijzig in absolute termen* = polar expression - private state expression (*het aandeel* = target)

- **noun phrase (+ preposition + target)**

(19) *Rising stars such as Avicii, David Guetta and Skrillex have recently begun collaborating with pop and hip-hop artists.*

⇨ *rising stars* = polar expression - private state expression (*Avicii, David Guetta and Skrillex* = target)

(20) *Dit betekent een stijging van de winst.*

⇨ *stijging* = polar expression - polar fact expression (*de winst* = target)

- **verb construction**

intransitive verb (the subject being the target)

(16) *The Perth-based company announced profits had rocketed by 25 per cent.*

⇨ *had rocketed by 25 per cent* = polar expression - polar fact expression (*profits* = target)

transitive verb + object (the subject being the target)

(21) *BNP Paribas boekte dan weer een winst van 6 miljard in 2011.*

⇨ *boekte ... een winst van 6 miljard* = polar expression - polar fact expression (*BNP Paribas* = target)

transitive verb (the object being the target and the subject being the source)

(22) *U.S. fears nightmarish conflict with Iran.*

⇨ *fears* = polar expression - private state expression (*nightmarish conflict with Iran* = target)

¹Some of these examples contain multiple polar expressions, but for each sentence only one expression is discussed.

(23) *Amazon wants more patents.*

⇨ *wants* = polar expression - private state expression (*more patents* = target)

transitive verb (the subject and object being the two targets)

(10) *Djokovic versloeg Roger Federer in 3 sets.*

⇨ *versloeg* = polar expression - polar fact expression (*Djokovic and Roger Federer* = targets)

phrasal verb

(24) *Het gaat Agfa-Gevaert niet voor de wind.*

⇨ *het gaat ... niet voor de wind* = polar expression - private state expression (*Agfa-Gevaert* = target)

copulative verb + subject complement (the subject being the target)

(25) *'Energiedranken zijn relatief duur,' verklaart Bart Van Leemput, hoofdanalist bij KBC Asset Management.*

⇨ *zijn relatief duur* = polar expression - private state expression (*energiedranken* = target)

resultative causative (the affected entity being the target)

(26) *Global warming melts the polar ice caps.*

⇨ *melts* = polar expression - polar fact expression (*the polar ice caps* = target)

• **a whole sentence**

(27) *Nasdaq Composite heeft dinsdag na ruim twaalf jaar de kaap van 3.000 punten heroverd. Party like it's 1999 is echter nog niet aan de orde.*

⇨ *Party like it's 1999 is echter nog niet aan de orde.* = polar expression - private state expression (*Nasdaq Composite* = target)²

As can be deduced from the examples above, polar expressions **can consist of non-consecutive tokens**.

When annotators have marked the expression span, they have to **determine the type** of the polar expression (private state expression or polar fact expression). As discussed in Section 2.3, it is sometimes difficult to label a polar expression as being either a private state or a polar fact expression, because the distinction between subjectivity and objectivity is not always clear. We therefore define a continuum ranging from objective to subjective, on which polar expressions can be located: the further an expression is situated on the schale, the more subjective it is. Polar expressions take 1 of 4 **positions on the subjectivity continuum**, denoted by the following labels: '0', '1', '2', '3'. Polar fact expressions are objective and consequently receive the label '0'. Private state expressions are situated on the other side of the schale and are labeled '3'. If it is not clear whether a polar expression should be categorized as a polar fact or a private state expression, the labels '1' and '2' can be used to locate the polar expression somewhere between the two extremes of the subjectivity continuum. '1' is chosen when annotators have a slight preference for the category 'polar fact expression', whereas the use of '2' indicates a preference for 'private state expression'. For example, in sentence (12), annotators could assign the polar expression *stuurde ... naar huis* the subjectivity label '1'. The polar expression *droogde ... af* in sentence (14) is situated a bit further on the subjectivity continuum and could therefore receive the label '2'.³

²Note that although a polar expression cannot span more than one sentence, it is possible to identify the source or target in one of the previous or next sentences. This is further discussed in Step 3 and 4 (see Section 3.2.3 and 3.2.4).

³Note however that determining the position of a polar expression on this continuum is a highly subjective task.

(12) *Djokovic stuurde Roger Federer in 3 sets naar huis.*

(14) *Djokovic droogde Roger Federer af in 3 sets.*

Nevertheless, annotators should use the labels ‘0’ and ‘3’ as much as possible.

Expressions of sentiment are sometimes **insubstantial**: this means that the sentiment being expressed is not real (Wiebe, Wilson, and Cardie, 2005). For instance in sentence (28), *fears* is a private state expression which indicates that investors hold a negative attitude towards *that Greek deposits and Greek debt will be converted...*. However, this private state only exists if Greece exits the euro and establishes a new currency, in other words the expressed sentiment is not real. By marking *if Greece does exit the euro and establish a new currency* as a modifier of conditionality (see Section 3.2.2), annotators can indicate the insubstantiality of the private state expression.

(28) *If Greece does exit the euro and establishes a new currency, investors fear that Greek deposits and Greek debt will be converted into a new currency which will sell at a steep discount to the euro.*

However, insubstantiality is **not always expressed by conditional or other modifiers**. In sentence (29) for example (taken from (Wiebe, Wilson, and Cardie, 2005)), *were illegal* is a polar expression of negative sentiment towards *his acts*. But since this evaluation is not actually uttered (the declaration of it is only being sought), the expressed sentiment is not real. Because there are no modifiers referring to this insubstantiality, the attribute ‘Insubstantial’ needs to be set on the polar expression level.

(29) *We are seeking a declaration that his acts were illegal.*

BRAT HOWTO

To mark a polar expression, highlight the corresponding text span and select ‘**PolarExpr**’. If the polar expression consists of non-consecutive tokens, mark all parts of the expression as separate polar expressions, then link them to each other by dragging one of the expressions on top of the other and select ‘**in span with**’. In the case of such a discontinuous polar expression, choose only one part for which attributes will be set and to which other elements will be linked.

Sometimes, the same text span will be part of different polar expressions. In sentence (30) for example, two polar fact expressions are targeted towards *een zeventienjarige medeleerling: heeft ... verkracht* and *heeft ... vermoord*. This means *heeft* needs to be marked as (part of) a polar expression twice, but the brat rapid annotation tool does not support this kind of annotation. However, annotating *heeft* a second time is possible if the annotator marks it as another element first (e.g. ‘Entity’ or ‘Modifier’), then double-clicks the annotation made and changes the element type to ‘PolarExpr’.

(30) *Een zeventienjarige medeleerling heeft haar verkracht en vermoord.*

For each polar expression, the following attributes need to be set:

- **Type**: Determine the type of the polar expression by locating it on the subjectivity continuum (0 for polar fact expressions, 1 for ambiguous cases where the annotator has a slight preference for the category ‘polar fact expression’, 2 for ambiguous cases where the annotator has a slight preference for the category ‘private state expression’ and 3 for private state expressions).

- **ConfidenceType:** Indicate how confident you are that the ‘Type’ attribute is set correctly (**low**, **medium** or **high**).
- **Insubstantial:** Check this attribute if a polar expression is insubstantial, but this insubstantiality is not explicited by conditional or other modifiers.
- **Causative:** Polar expressions sometimes express causality in that they refer to a positive or negative effect being exerted on a target entity or entities by another entity. Polar expressions like this are marked using the attribute ‘Is_causative’, which is discussed in detail in Section 3.2.5.

The other attributes (‘Implicit_source’, ‘Author_is_source’, ‘Impersonal_source’, ‘Is_also_source_expression’ and ‘Is_also_opinionated_source_expression’) can be left unchanged until the source and source expression of the polar expression are identified in Step 3 of the annotation procedure (see Section 3.2.3).

3.2.2 Modifiers

A modifier is a lexical item causing a shift in the polarity, strength or temporal position of the sentiment expressed by the polar expression. Modifiers are **often included in the polar expression span** because they are part of the same syntactic phrase (see Section 3.2.1). However, they need to be **analyzed separately** as well. Each modifier is assigned to a category of modification and linked to the polar expression it affects (even if it is already in the polar expression span).

(4) *China criticized Mitt Romney’s Jerusalem comments.*

⇨ the sentiment expressed by the polar expression *criticized* is affected by a past modifier (*criticized*)

(16) *The Perth-based company announced profits had rocketed by 25 per cent.*

⇨ the sentiment expressed by the polar expression *had rocketed by 25 per cent* is affected by a past modifier (*had rocketed*) and a specifying modifier (*by 25 per cent*)

We define 11 **categories** of modifiers:

- **negation** modifiers

(24) *Het gaat Agfa-Gevaert niet voor de wind.*

⇨ *niet* - affects sentiment expressed by polar expression *het gaat ... niet voor de wind*

- **intensifiers**

(31) *The pet health care industry is very profitable.*

⇨ *very* - affects sentiment expressed by polar expression *is very profitable*

(32) *Het rendement van Monster Beverage ligt op 20.747 procent!*

⇨ *!* - affects sentiment expressed by polar expression *ligt op 20.747 procent*

- **diminishers**

(25) *'Energiedranken zijn relatief duur,'* verklaart Bart Van Leemput, hoofdanalist bij KBC Asset Management.
⇨ *relatief* - affects sentiment expressed by polar expression *zijn relatief duur*

- **modality** modifiers

(33) *De eerste kwartaalresultaten van de bedrijven zullen waarschijnlijk niet slecht zijn.*
⇨ *waarschijnlijk* - affects sentiment expressed by polar expression *zullen waarschijnlijk niet slecht zijn*

- **question** modifiers

(34) *Doet eurocrisis de woonkredieten stijgen?*
⇨ *?* - affects sentiment expressed by polar expression *doet ... stijgen*

- **conditional** modifiers

(28) *If Greece does exit the euro and establish a new currency, investors fear that Greek deposits and Greek debt will be converted into a new currency which will sell at a steep discount to the euro.*
⇨ *if Greece does exit the euro and establish a new currency* - affects sentiment expressed by polar expression *fear*

(35) *Brook has to improve if he wants a world title fight.*
⇨ *if* - affects sentiment expressed by polar expression *wants*

- **future** modifiers

(33) *De eerste kwartaalresultaten van de bedrijven zullen waarschijnlijk niet slecht zijn.*
⇨ *zullen ... zijn* - affects sentiment expressed by polar expression *zullen waarschijnlijk niet slecht zijn*⁴

- **past** modifiers

(4) *China criticized Mitt Romney's Jerusalem comments.*
⇨ *criticized* - affects sentiment expressed by polar expression *criticized*

(36) *Vorig jaar maakte de groep 23 miljoen dollar verlies.*
⇨ *vorig jaar* and *maakte* - affect sentiment expressed by polar expression *maakte ... 23 miljoen dollar verlies*

- **perspective** modifiers

(37) *Lage beurskoersen zijn heel goed voor beleggers die in willen stappen.*
⇨ *voor beleggers die in willen stappen* - affects sentiment expressed by polar expression *zijn heel goed*

(38) *Complexity is bad for your health.*
⇨ *for your health* - affects sentiment expressed by polar expression *is bad*

- **specifying** modifiers

(36) *Vorig jaar maakte de groep 23 miljoen dollar verlies.*
⇨ *23 miljoen dollar* - affects sentiment expressed by polar expression *maakte ... 23 miljoen dollar verlies*

- **other** modifiers

⁴Future/past modifiers can be verbal constructions which indicate something will take place in the future/has taken place in the past.

(39) *Natuurlijk is het er bij Lady Gaga, zoals gewoonlijk, wat over.*

⇨ *zoals gewoonlijk* - affects sentiment expressed by polar expression *is ... er ... wat over* (this text span contributes to the sentiment expressed here, but cannot be assigned to a specific category)

As can be deduced from example (37), modifiers **can consist of non-consecutive tokens**.

Note that all modifiers need to be **marked and analyzed separately**. In sentence (37) for instance we find three modifiers: *zullen ... zijn* (future modifier), *waarschijnlijk* (modifier of modality) and *niet* (modifier of negation).

BRAT HOWTO

To mark a modifier, highlight the corresponding text span and select '**Modifier**'. If the modifier consists of non-consecutive tokens, mark all parts of the modifier as separate modifiers, then link them to each other by dragging one of the modifiers on top of the other and select '**in_span_with**'. In the case of such a discontinuous modifier, choose only one part to set the 'Modification' attribute and to link to other elements.

For each modifier, you will be asked to set the following attribute:

- **Modification:** Determine which category of modification the modifier belongs to (**negation, intensifier, diminisher, modality, question, conditional, future, past, perspective, specifying** or **other**).

When finished, **link the modifier to the polar expression it affects** with drag and drop (even if the modifier is already included in the text span of the polar expression).

3.2.3 Source and Source Expression

Polar expressions always have a source, but it is **not always explicitly mentioned** in the text.

In sentence (4), *China* is the source of the private state expression *criticized*. In example (25), *Bart Van Leemput, hoofdanalist bij KBC Asset Management* is the source of the polar expression *zijn relatief duur*. Note here that the full phrase and not just *Bart Van Leemput* is marked. Furthermore, a sentence (or text) can contain multiple references to the polar expression's source; we always annotate the **source mention syntactically linked to or closest to the polar expression** (e.g. *die* is the source of the polar expression *tot gevaar kon leiden* in example (40), not *de judobond*). Finally, sources are **never included in the polar expression text span** (see Section 3.2.1).

(4) *China criticized Mitt Romney's Jerusalem comments.*

(25) *'Energiedranken zijn relatief duur,' verklaart Bart Van Leemput, hoofdanalist bij KBC Asset Management.*

(40) *Eigenlijk vond de judobond het niet kunnen. Die vond dat de hoofddoek tot gevaar kon leiden.*

The source in sentence (25) is introduced by a source expression, namely *verklaart*. This text span indicates that *Bart Van Leemput, hoofdanalist bij KBC Asset Management* is the person uttering the

polar expression. If present, source expressions are marked and linked to the source they introduce. In sentence (41), source expression 's denotes that *Cher* is the source of the polar expression *criticism of*.

(41) *Cher 's criticism of Romney 's Response to Colorado Shooting a Misfire.*

If the sentence contains an explicit mention of the source, but no source expression can be identified, it is the same as the polar expression. In sentence (4), *criticized* is the polar as well as the source expression. In this case, the polar expression attribute 'Is_also_source_expression' is set and the text span does not need to be marked as 'SourceExpression' as well.

Sometimes, **source expressions can convey sentiment on an additional level: not expressed by the source of the polar expression, but by the author of the text.** Through the choice of a certain source expression, the author can express sentiment towards the polar expression itself. In sentence (42), the source *De Beweging door Democratische Verandering (MDC)* expresses positive sentiment towards *oppositieleider Tsvangirai* by means of the polar expression *al in de eerste ronde de absolute meerderheid behaalde*. This polar expression is introduced by the source expression *houdt hardnekkig vol*. On a second level, the author of the sentence questions the truthfulness of the information conveyed by this polar expression by using the words *houdt hardnekkig vol*. For source expressions like this, which express sentiment towards the polar expression they introduce, the attribute 'Is_also_opinionated' is set. If an opinionated source expression is the same as the polar expression, the polar expression attribute 'Is_also_opinionated_source_expression' is set. For instance in sentence (43), the source *the U.S.* expresses negative sentiment towards *China*, which we can deduce from the polar expression *was slandering*. On a second level, the author states he/she does not agree with the U.S.'s slandering of China by using the expression *was slandering* instead of for example *was criticizing*. For this polar expression (which is also the source expression), the attribute 'Is_also_opinionated_source_expression' is set.

(42) *De Beweging voor Democratische Verandering (MDC) houdt hardnekkig vol dat oppositieleider Tsvangirai al in de eerste ronde de absolute meerderheid behaalde.*

(43) *The U.S. was slandering China again.*

Polar expressions do not always have a source that is explicated in the same sentence. The source can either be in one of the previous or following sentences, it can be the author of the text or it can be an impersonal source.

When the source of a polar expression **only occurs in one of the previous or following sentences**, it is considered **implicit**. In the second sentence of example (44), negative sentiment is expressed towards *de cijfers* through the polar expression *zijn bijna dramatisch*, but the source (*Jan De Kerpel*) can be found in the previous sentence. Here, we mark the source in the previous sentence and set the 'Implicit_source' attribute of the polar expression.

(44) *'Het gaat snel, heel snel,' vindt Jan De Kerpel. 'De cijfers zijn bijna dramatisch.'*

In some cases, the source of a polar expression **occurs nowhere in the text (not even in one of the previous or following sentences)**, because it is **impersonal**. In sentence (45) for example, the polar expression *is onterecht door het slijk gehaald* is a passive construction. Here, the source is the impersonal **'the people'** or ('men' in Dutch), but it is not explicated in the text. In cases like this, we set the polar expression attribute 'Impersonal_source'.

(45) *Mijn naam is onterecht door het slijk gehaald.*

Sometimes the source of a polar expression is the **author of the sentence**. If this is the case, a source annotation can only be made if the text contains a **lexicalisation of the author** (possibly not in the same sentence). In sentence (46) for instance, *ik* refers to the author of the text and is marked as the source of the polar expression *ben ... virulent tegen*.

(46) *Als econoom ben ik virulent tegen de Tobin-taks.*

If a lexicalisation of the author cannot be found in the text, the **'Author_is_source'** attribute of the polar expression is set, as would be done for the polar expression *zijn gewoon* in sentence (3).

(3) *De liquiditeit en solvabiliteit van KBC zijn gewoon goed.*

In conclusion, we analyze the sources and sources expressions in example (16).

(16) *The Perth-based company announced profits had rocketed by 25 per cent.*

⇨ *the Perth-based company* = source, *announced* = source expression - introducing the polar expression *had rocketed by 25 per cent*

BRAT HOWTO

To annotate the source of a polar expression, highlight the corresponding text span and select **'Entity'**⁵. If there is no source expression, directly link the polar expression to the marked entity with drag and drop and select **'has source'**.

If there is a source expression, highlight its text span and select **'SourceExpr'**⁶, link it to the source with drag and drop, and select **'has source'**. Finally, link the polar expression to the source expression (not to the source), and select **'has source expression'**. Note that you do not have to link the polar expression to the source when there is also a source expression.

Sources are also marked and linked to the polar expression when they occur in one of the previous or next sentences. In this case, the polar expression's attribute **'Implicit_source'** is used. If the author is impersonal, use the attribute **'Impersonal_source'**. If the author is the source and the text does not contain a lexicalisation of the author, the polar expression's attribute **'Author_is_source'** is used.

If the source expression is represented by the same text span as the polar expression, it does not need to be marked as **'SourceExpr'**. Just use the polar expression's attribute **Is_also_source_expression**.

If the source expression utters sentiment towards the polar expression it introduced, check the source expression's attribute **'Is_also_opinionated'** (or the polar expression's attribute **'Is_also_opinionated_source_expression'** if the source expression equals the polar expression).

⁵If the source consists of non-consecutive tokens, mark all parts of the source as separate entities, then link them to each other by dragging one of the parts on top of the other and select **'in_span_with'**.

⁶If the source expression consists of non-consecutive tokens, mark all parts as separate source expressions, then link them to each other by dragging one of the expressions on top of the other and select **'in_span_with'**

3.2.4 Target and Sentiment Polarity

The target of the polar expression is the entity sentiment is being expressed towards. It is **never included in the polar expression text span** (see Section 3.2.1). Annotators should mark the target (or targets) of a polar expression and determine the polarity and intensity of the sentiment being expressed towards it (them).

(4) *China criticized Mitt Romney's Jerusalem comments.*

↻ private state expression *criticized* expresses negative sentiment towards the target *Mitt Romney's Jerusalem comments*.

(16) *The Perth-based company announced profits had rocketed by 25 per cent.*

↻ polar fact expression *has rocketed by 25 per cent* with target *profits* results in a positive impression of this target

Every polar expression has an explicit target: if it does not occur in the same sentence, it is identified earlier or later on in the text. In example (27), the polar expression *Party like it's 1999 is echter nog niet aan de orde.* refers to the target *Nasdaq Composite* in the previous sentence.

(27) *Nasdaq Composite heeft dinsdag na ruim twaalf jaar de kaap van 3.000 punten heroverd. Party like it's 1999 is echter nog niet aan de orde.*

If a sentence (or text) contains multiple references to the polar expression's target, we always annotate the **target mention syntactically linked to or closest to the polar expression**. In sentence (47), not *CD&V-minister Joke Schauvliege besliste een milieuvergunning toe te kennen* but *wat is* marked as the target of the polar expression "*onbegrijpelijk*".

(47) *CD&V-minister Joke Schauvliege besliste een milieuvergunning toe te kennen, wat sp.a-collega Ingrid Lieten kort nadien "onbegrijpelijk" noemde.*

When determining the sentiment being expressed towards the target, two attributes are set: **the polarity and intensity of the attitude**. The possible values for these attributes are the following:

- the **polarity** of the sentiment

positive

(37) *Lage beurskoersen zijn heel goed voor beleggers die in willen stappen.*

↻ polar expression *zijn heel goed* expresses positive sentiment towards target *lage beurskoersen*

negative

(25) *'Energiedranken zijn relatief duur,' verklaart Bart Van Leemput, hoofdanalist bij KBC Asset Management.*

↻ polar expression *zijn relatief duur* expresses negative sentiment towards target *energiedranken*

other

(5) *Astronomers are surprised by the quantity and the variety of planets that Kepler has spotted in its first year.*

⇨ polar expression *are surprised by* expresses neither positive nor negative sentiment towards target *the quantity and the variety of planets that Kepler has spotted in its first year*

- the **intensity** of the sentiment

low

(33) *De eerstekwartaalresultaten van de bedrijven zullen waarschijnlijk niet slecht zijn.*

⇨ the positive sentiment expressed by *zullen waarschijnlijk niet slecht zijn* towards *de eerstekwartaalresultaten van de bedrijven* has a low intensity

medium

(25) *'Energiedranken zijn relatief duur,' verklaart Bart Van Leemput, hoofdanalist bij KBC Asset Management.*

⇨ the negative sentiment expressed by *zijn relatief duur* towards *energiedranken* has a medium intensity

high

(37) *Lage beurskoersen zijn heel goed voor beleggers die in willen stappen.*

⇨ the positive sentiment expressed by *zijn heel goed* towards *lage beurskoersen* has a high intensity

Note that when assessing the sentiment being expressed, annotators should **bring the modifiers affecting the polar expression into the equation**.

It is possible for a polar expression to have **multiple targets**. This is the case if the polar expression describes an event which can be seen from the perspective of different entities in the sentence, thus expressing different sentiment towards these entities (targets). We determine the sentiment expressed towards all of these targets. Examples of this can be found in sentences (48) and (49). Often these polar expressions are comparative constructions (e.g. sentence (49)).

(48) *Technologiebedrijf zet rijzende sterren als Monster Beverage in de schaduw.*

⇨ polar expression *zet ... in de schaduw* expresses positive sentiment (with medium intensity) towards the target *technologiebedrijf*

⇨ polar expression *zet ... in de schaduw* expresses negative sentiment (with medium intensity) towards the target *rijzende sterren als Monster Beverage*

(49) *Vegan options are more expensive than their nonvegetarian counterparts.*

⇨ polar expression *are more expensive than* expresses negative sentiment (with medium intensity) towards the target *vegan options*

⇨ polar expression *are more expensive than* expresses positive sentiment (with medium intensity) towards the target *their nonvegetarian counterparts*

BRAT HOWTO

To mark a target, highlight the corresponding text span and select '**Entity**'⁷. Subsequently, link the polar expression to all of its targets by means of drag and drop. You will then be asked to determine the polarity and intensity of the sentiment being expressed towards this target by

⁷If the target consists of non-consecutive tokens, mark all parts as separate entities, then link them to each other by dragging one of the entities on top of the other and select **in span with**.

selecting one of the following options: **has low/medium/high positive/negative/other target** or **has unknown polarity** in the case of a polar expression with unknown or ambiguous polarity (see Section 2.4).

3.2.5 Causativity

Private state or polar fact expressions sometimes express causativity in that they refer to a positive or negative effect being exerted on a target entity or entities by another entity. In sentence (10), for instance, the private state expression ‘could damage’ indicates that a negative effect is exerted on the target ‘the global economy’ by the entity ‘higher oil prices’. We refer to private state and polar fact expressions like this as polar resultative causatives. In contrast to simple causatives, which only refer to the causal link between two entities, resultative causatives are causative constructions which refer to the causal link plus a part of the resulting situation⁸. If a causative construction is resultative, and the resulting situation referred to by that resultative causative results in a positive or negative impression of the target entity or entities (i.e. conveys a private state or polar fact), it is a **polar resultative causative**. In sentence (10), ‘could damage’ is a polar resultative causative because the resulting situation (= ‘the global economy could be damaged’) results in a negative impression of the target ‘the global economy’. ‘Higher oil prices’ is identified as the **cause** entity of this polar resultative causative. Note that causes are not included in the polar expression text span. Furthermore, if a sentence (or text) contains multiple references to the cause, we always annotate the cause mention syntactically linked to or closest to the polar expression (cf. sources and targets).

(50) *Meanwhile, investor worries that higher oil prices could damage the global economy gathered pace.*

Further examples of polar resultative causatives are given in sentences (51) to (54).

(51) *Apple trekt Nasdaq over 3.000 punten*

⇨ *trekt ... over 3.000 punten* = polar resultative causative - indicating that cause *Apple* has a positive effect on target *Nasdaq* [because the resulting situation (= ‘Nasdaq rises above 3.000 points’) results in a positive evaluation of *Nasdaq*]

(52) *Syria terrorists attack Palestinian Refugee Camp.*

⇨ *attack* = polar resultative causative - indicating that cause *Syria terrorists* has a negative effect on target *Palestinian Refugee Camp* [because the resulting situation (= ‘the Palestinian Refugee Camp is destroyed’) results in a negative evaluation of *Palestinian Refugee Camp*]

(53) *Dit is de moordenaar van Kadhafi.*

⇨ *is de moordenaar van* = polar resultative causative - indicating that cause *dit* has a negative effect on target *Kadhafi* [because the resulting situation (= ‘Kadhafi is dead’) results in a negative evaluation of *Kadhafi*]

⁸In (Girju, 2003), the term resultative causative is used to refer to verbal constructions, but we also identify other causative constructions.

(54) *De door de gemeente verhoogde belastingen zijn niet enkel Groen in het verkeerde keelgat geschoten.*

⇨ *verhoogde* = polar resultative causative - indicating that cause *de gemeente* has a negative effect on target *belastingen* [because the resulting situation (= 'de belastingen zijn gestegen') results in a negative evaluation of *de belastingen*]

Our annotation scheme also covers simple causatives and other non-verbal causative constructions which do not convey positive or negative sentiment. Chapter 5 describes how these **non-polar causal links** are annotated and how they can be distinguished from polar resultative causatives.

BRAT HOWTO

If a private state or polar fact expression is a polar resultative causative, this is indicating by means of the polar expression attribute '**Is_causative**'. To mark the cause of the polar resultative causative, highlight the corresponding text span and select '**Entity**'⁹. Subsequently, link the polar expression to the cause by means of drag and drop and select '**has_cause**'.

3.3 Identification of Embedded Polar Expressions

Polar expressions can be identified on different sentence levels. This means that one polar expression can be embedded in another. Annotators should **start by identifying polar expressions on the lowest sentence level** using the procedure in Section 3.2. Subsequently, they can determine whether a polar expression is enclosed by another polar expression on a higher sentence level by means of the procedure described below. We analyze sentence (55) as an example.

(55) *HTC maakt een kleine winst.*

1. Detect the lowest-level polar expression.

⇨ *kleine* = polar expression - expressing negative sentiment towards *winst*

2. Look for potential polar expressions on a higher level, in which the low-level polar expression is embedded.

⇨ *maakt een kleine winst* = polar expression expressing positive sentiment towards target *HTC*?

3. Analyze the higher-level polar expression making abstraction of the low-level expression.

⇨ *maakt een kleine winst* = polar expression?

4. If the higher-level expression still expresses sentiment, annotate it.

⇨ does *maakt (een) winst* still express positive sentiment towards *HTC*?

YES ⇨ *maakt een kleine winst* = polar expression

⁹If the cause consists of non-consecutive tokens, mark all parts as separate entities, then link them to each other by dragging one of the entities on top of the other and select '**in_span_with**'.

We analyze some additional sentences to further illustrate the use of the procedure.

(48) *Technologiebedrijf zet rijzende sterren als Monster Beverage in de schaduw.*

∓ *rijzende* = polar expression - expressing positive sentiment towards *sterren*

∓ *rijzende sterren* = polar expression - expressing positive sentiment towards *Monster Beverage*
[because *rijzende sterren* also expresses positive sentiment towards *Monster Beverage*]

(56) *Will the high price of the Galaxy S III stop you from buying?*

∓ *high* = polar expression - expressing negative sentiment towards *price*

∓ *high price* = NOT a polar expression expressing negative sentiment towards *the Galaxy S III* [because *high price* does not express negative sentiment towards *Galaxy S III*]

(57) *Federer had een uitstekende start genomen.*

∓ *uitstekende* = polar expression - expressing positive sentiment towards *start*

∓ *had een uitstekende start genomen* = NOT a polar expression expressing positive sentiment towards *Federer* [because *had een uitstekende start genomen* does not express positive sentiment towards *Federer*]

Chapter 4

Annotation of Coreferential and Feature Relations

4.1 Coreferential Relations

A sentence (or text) can contain multiple references to certain elements of a polar expression. **For sources, targets and causes**, we always annotate the mention that is syntactically linked to or closest to the polar expression (see Sections 3.2.3, 3.2.4 and 3.2.5). In sentence (47) for instance, *wat* is marked as the target of the polar expression "onbegrijpelijk".

(47) *CD&V-minister Joke Schauvliege besliste een milieuvergunning toe te kennen, wat sp.a-collega Ingrid Lieten kort nadien "onbegrijpelijk" noemde.*

Consequently, text spans annotated as sources, targets or causes are not always very meaningful; often they are pronouns (e.g. *wat* in example (47)). For this reason, we perform coreference annotation and **link entities with little or no intrinsic meaning to the closest meaningful antecedent**. In sentence (47), the target *wat* is linked to its antecedent *CD&V-minister Joke Schauvliege besliste een milieuvergunning toe te kennen*. Annotators should be careful not to directly mark this antecedent as the target of the polar expression.

BRAT HOWTO

To link a source, target or cause to its meaningful antecedent, highlight the antecedent text span and select 'Entity'. Subsequently, link the source/target/cause to it and select 'refers.to'.

4.2 Feature Relations

Sometimes the sentiment being expressed towards a certain target **also affects another entity, because the target is a feature of this entity**, i.e. a part or property of it. In sentence (1), we identify three different polar expressions.

(1) *The restaurant was terrific: the best chocolate ice cream ever and platters upon platters of breakfast choices.*

- ⇨ *was terrific* = polar expression - expressing positive sentiment towards target *the restaurant*
- ⇨ *the best ... ever* = polar expression - expressing positive sentiment towards target *chocolate ice cream*
- ⇨ *platters upon platters* = polar expression - expressing positive sentiment towards target *breakfast choices*

Each of these expressions has a different target, but the sentiment being expressed towards *chocolate ice cream* and *breakfast choices* is also indirectly aimed at *the restaurant*: a positive impression of the former two will usually result in a positive impression of the latter. For this reason, we indicate *chocolate ice cream* and *breakfast choices* as being features of *the restaurant*.

Note that we only mark a feature relation between a target and another entity if the sentiment being expressed towards the target actually affects the other entity. In sentence (58), *was erg slecht* is a polar expression which expresses negative sentiment towards *ons hotel*. This target could be considered a feature of *Barcelona*, because the hotel is located there. However, the source's negative impression of *ons hotel* does not result in a negative evaluation of the city. Consequently, *ons hotel* is not annotated as a feature of *Barcelona*.

(58) *We vonden Barcelona super, maar ons hotel was erg slecht!*

As can be deduced from examples (1) and (58), feature relations are common in user reviews. However, they also occur in other text types (e.g. sentence (59)).

(59) *BATA has an increasing profit margin.*

- ⇨ *increasing* = polar expression - positive sentiment is expressed towards *profit margin* and indirectly towards *BATA*
- ⇨ *profit margin* = feature of *BATA*

BRAT HOWTO

In order to indicate that a polar expression target is a feature of another entity, mark this other entity as '**Entity**', link the target to it using drag and drop and select '**feature_of**'.

Chapter 5

Annotation of Non-Polar Causal Relations

Besides causative polar expressions i.e. polar resultative causatives, our annotation scheme also covers **simple causatives** i.e. **causal constructions which only refer to the causal link between two entities (events)**. Although (Girju, 2003) uses the term for verbal constructions, we also identify other causal constructions.

5.1 Elements

A causal relation consists of three elements:

- **Cause expression:** the linguistic expression of the causal link
- **Cause:** the entity referring to the causing event
- **Effect:** the entity referring to the resulting event

5.2 Annotation Procedure

To annotate a causal relation, annotators must first mark the cause expression. This is followed by the identification of cause and effect, both of which are then linked to the cause expression. Examples of causal relations can be found in sentences (60) to (64).

(60) *Global warming leads to melting of polar ice caps.*

⇨ *leads to* = cause expression - *global warming* = cause, *melting of polar ice caps* = effect

(61) *Hover over each satellite photo to view the devastation caused by the earthquake and tsunami.*

⇨ *caused by* = cause expression - *the earthquake and tsunami* = cause, *the devastation* = effect

(62) *Elastische kleding is de oorzaak van het obesitasprobleem.*

⇨ *is de oorzaak van* = cause expression - *elastische kleding* = cause, *het obesitasprobleem* = effect

(63) *Sneller lopen dankzij afgebonden benen.*

⇨ *dankzij* = cause expression - *afgebonden benen* = cause, *sneller lopen* = effect

(64) *Door de sterke groei van de Turkse economie vorig jaar is de inflatie ontspoord.*

⇨ *door* = cause expression - *de sterke groei van de Turkse economie vorig jaar* = cause, *is de inflatie ontspoord* = effect

Note that for causes and effects, the **full event** is to be marked (e.g. *is de inflatie ontspoord* in sentence (64)) and **no overlap with the causal expression is possible**.

Annotators should be careful **not to confuse non-polar causal relations with polar resultative causatives**. The former only refer to the causal link between two entities, whereas the latter also refer to a part of the resulting situation (see Section 3.2.5). Sometimes the resulting event of a causal relation contains a polar expression; in sentence (64), *is ... ontspoord* is a polar expression expressing negative sentiment towards the target *de inflatie*. However, we do not identify a polar resultative causal link between *de sterke groei van de Turkse economie* and *de inflatie*, because a clear distinction can be made between the expression of the causal link (*door*) and the expression of sentiment (*is ... ontspoord*). In sentence (65) on the other hand, annotators should mark *deed ... ontsporen* as a causative polar expression with *de sterke groei van de Turkse economie* being the cause and *de inflatie* being the target.

(65) *De sterke groei van de Turkse economie vorig jaar deed de inflatie ontsporen.*

BRAT HOWTO

To mark a cause expression, highlight the corresponding text span and select ‘CauseExpr’¹. Subsequently, identify the cause and effect and mark them as ‘Entity’. Finally, link the cause expression to the cause/effect and select ‘has_cause’/‘has_effect’.

¹If the cause expression consists of non-consecutive tokens, mark all parts as separate cause expressions, then link them to each other by dragging one of the expressions on top of the other and select ‘in_span_with’.

Chapter 6

Extended Annotation Examples

In this chapter, we present the full annotation for a couple of the example sentences discussed in the previous chapters and for some additional sentences.

(54) *De door de gemeente verhoogde belastingen zijn niet enkel Groen in het verkeerde keelgat geschoten.*

⇨ polar expression 1 = *verhoogde*

- **Type:** 0
- **Modifier of past:** *verhoogde*
- **Author is source**
- **Target:** *belastingen* [**sentiment:** **polarity** = negative, **intensity** = medium]
- **Is causative**
- **Cause:** *de gemeente*

⇨ polar expression 2 = *zijn ... in het verkeerde keelgat geschoten*

- **Type:** 3
- **Modifier of past:** *zijn ... geschoten*
- **Is also opinionated source expression**
- **Source:** *niet enkel Groen*
- **Target:** *de door de gemeente verhoogde belastingen* [**sentiment:** **polarity** = negative, **intensity** = high]

(45) *Mijn naam is onterecht door het slijk gehaald.*

⇨ polar expression = *is onterecht door het slijk gehaald*

- **Type:** 3
- **Modifier of past:** *is ... gehaald*
- **Impersonal source**
- **Is also opinionated source expression**
- **Target:** *mijn naam* [**sentiment: polarity** = negative, **intensity** = high]

(48) *Technologiebedrijf zet rijzende sterren als Monster Beverage in de schaduw.*

⇨ polar expression 1 = *rijzende*

- **Type:** 3
- **Author is source**
- **Target:** *sterren* [**sentiment: polarity** = positive, **intensity** = medium]

⇨ polar expression 2 = *rijzende sterren*

- **Type:** 3
- **Author is source**
- **Target:** *Monster Beverage* [**sentiment: polarity** = positive, **intensity** = medium]

⇨ polar expression 3 = *zet ... in de schaduw*

- **Type:** 3
- **Author is source**
- **Target 1:** *technologiebedrijf* [**sentiment: polarity** = positive, **intensity** = medium]
- **Target 2:** *rijzende sterren als Monster Beverage* [**sentiment: polarity** = negative, **intensity** = medium]

(66) *Vandaag zijn technologiebedrijven opnieuw gigantisch populair, maar ze leggen ook sterke prestaties voor.*

⇨ polar expression 1 = *zijn opnieuw gigantisch populair*

- **Type:** 3
- **Modifier of increase:** *gigantisch*
- **Author is source**
- **Target:** *technologiebedrijven* [**sentiment: polarity** = positive, **intensity** = high]

⇨ polar expression 2 = *sterke*

- **Type:** 3
- **Author is source**
- **Target:** *prestaties* [**sentiment:** **polarity** = positive, **intensity** = medium] - **feature** of *technologiebedrijven*

(67) *Kiev profits from Euro 2012.*

⇨ polar expression 1 = *profits from*

- **Type:** 3
- **Author is source**
- **Target:** *Kiev* [**sentiment:** **polarity** = positive, **intensity** = medium]
- **Is causative**
- **Cause:** *Euro 2012*

(68) *Ingrediënt in cola is kankerverwekkend.*

⇨ polar expression = *is kankerverwekkend*

- **Type:** 0
- **Author is source**
- **Target:** *Ingrediënt in cola* [**sentiment:** **polarity** = negative, **intensity** = medium]

(69) *Fortis and KBC are having difficulties.*

⇨ polar expression = *are having difficulties*

- **Type:** 3
- **Author is source**
- **Target:** *Fortis and KBC* [**sentiment:** **polarity** = negative, **intensity** = medium]

(70) *Ik vrees dat ik een rood sein heb gemist.*

⇨ polar expression 1 = *vrees*

- **Type:** 3
- **Is also source expression**
- **Source:** *Ik*
- **Target:** *dat ik een rood sein heb gemist* [**sentiment:** **polarity** = negative, **intensity** = medium]

⇨ polar expression 2 = *een rood sein heb gemist*

- **Type:** 0
- **Author is source**
- **Target:** *ik* [sentiment: **polarity** = negative, **intensity** = high]

References

- Ding, X., B. Liu, and P.S. Yu. 2008. A Holistic Lexicon-Based Approach to Opinion Mining. In Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM 2008), pages 231–240, Palo Alto, California, USA.
- Girju, R. 2003. Automatic Detection of Causal Relations for Question Answering. In Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering, Volume 12 (MultiSumQA 2003), pages 76–83, Sapporo, Japan.
- Hu, M. and B. Liu. 2004. Mining and Summarizing Customer Reviews. In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004), pages 168–177, Seattle, Washington, USA.
- Musat, C. and S. Trausan-Matu. 2010. The Impact of Valence Shifters on Mining Implicit Economic Opinions. In Proceedings of the 14th International Conference on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA 2010), volume 6304 of Lecture Notes in Computer Science, pages 131–140. Springer.
- Nigam, K. and M. Hurst. 2004. Towards a Robust Metric of Opinion. In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, pages 598–603, Palo Alto, California, USA.
- Ounis, I., M. de Rijke, C. Macdonald, G. Mishne, and I. Soboroff. 2006. Overview of the TREC-2006 Blog Track. In Proceedings of the 15th Text REtrieval Conference (TREC 2006), pages 17–31, Gaithersburg, Maryland, USA.
- Pang, B. and L. Lee. 2008. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2):1–135.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. A Comprehensive Grammar of the English Language. Longman, London.
- Seki, Y., D.K. Evans, L. Ku, H. Chen, N. Kando, and C. Lin. 2007. Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proceedings of NTCIR-6, pages 265–278, Tokyo, Japan.
- Stenetorp, P., S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii. 2012. brat: a Web-based Tool for NLP-Assisted Text Annotation. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012), pages 102–107, Avignon, France.
- Toprak, C., N. Jakob, and I. Gurevych. 2010. Sentence and Expression Level Annotation of Opinions in User-Generated Discourse. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 575–584, Uppsala, Sweden.
- Turney, P.D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002), pages 417–424, Philadelphia, Pennsylvania, USA.
- Wiebe, J., T. Wilson, and C. Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation, 39(2-3):165–210.
- Wilson, T. 2008. Annotating Subjective Content in Meetings. In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), pages 2738–2745, Marrakech, Morocco.