SentiML corpus annotation guidelines

Marilena Di Bari
Contents

1 Introduction 5

2 What you need to know about annotation 7
  2.1 The concept of TARGET ................................. 7
  2.2 The concept of MODIFIER ............................. 10
  2.3 The concept of APPRAISAL GROUP .................. 10
  2.4 Using the annotation tool ............................ 10
  2.5 Annotating a TARGET ............................... 13
  2.6 Annotating a MODIFIER ............................. 13
  2.7 Annotating an APPRAISAL GROUP .................. 14
  2.8 Saving your work .................................... 14

3 Whether to use the TARGET tag 17

4 Attributes for TARGET 21
  4.1 Type attribute ........................................ 21
  4.2 Orientation attribute ................................ 21

5 Whether to use the MODIFIER tag 23

6 Attributes for MODIFIER 27
  6.1 Attitude attribute ................................... 27
  6.2 Orientation attribute ................................ 28
  6.3 Force attribute ...................................... 28
  6.4 Polarity attribute ................................... 29

7 Whether to use the APPRAISAL GROUP tag 31

8 Attributes for APPRAISAL GROUP 35
  8.1 Orientation attribute ................................. 35
CONTENTS

9 Special cases 39

Appendices

Appendix A Appendix A 45

Bibliography 47
Chapter 1

Introduction

This document describes the annotation guidelines for the SentiML corpus. The annotated texts are in English and belong to three different categories:

- **Political speeches.** Mainly American presidents addresses available on the web.

- **Talks.** Collected from the website TED (Technology, Entertainment, Design). Talks are on a wide range of topics under the slogan “ideas worth spreading”.

- **News.** Available in the MPQA opinion corpus and manually annotated for opinions and other private states (e.g., beliefs, emotions, sentiments, speculations). We decided to focus on human rights topic first in order to have a link with political themes.

The goal of this annotation schema is to identify and classify sentiment groups (positive and negative) at the sentence level. In order to do this, we focus on three categories: target (expression the sentiment refers to), modifier (expression conveying sentiment) and appraisal group (couple of modifier and target).

For example, the sentence “I will eliminate capital gain taxes for the small businesses” is annotated as

“I will [(eliminate)\textsubscript{\textsc{M}} capital gains [(taxes)\textsubscript{\textsc{T}}]_{AG} from the small businesses]"
where $T$, $M$ and $AG$ respectively refer to target, modifier and appraisal group.

Because we take into account important linguistic issues such as the polarity reversal “eliminate” in this case, we can potentially classify the sentence as positive, whereas current approaches would be mislead by the presence of two apparent negative words (“eliminate” and “taxes”). We will see later in the document how to handle other linguistic issues (e.g., negation, false sentiment words, opinions only implied).

Targets and modifiers have a number of attributes (e.g., type for targets and orientation for targets/modifiers) that need to be annotated as well. The present annotation schema follows Bloom et al. [1] in the way such attributes are extracted from the Appraisal Framework [4] in the context of Systemic Functional Linguistics [3] and the difference in prior and contextual polarity by Wilson [9], which we call prior and contextual orientation (see Chapters 4, 6 and 8).

There is no preference for the order in which different layers of information for each unit are annotated, but it is necessary to annotate targets and modifiers first in order to create an appraisal group. Please see Chapter 7 on further suggestions on annotation strategy.

These guidelines should be used as follows. Chapter 2 provides a Quick Start guide to start annotating. The other chapters contain the full annotation guidelines, whereas Appendix A contains an example of annotation output in XML form.

The same annotation guidelines will be used to annotate the SentiML corpus versions in Italian and Russian. When possible, translations of the already annotated English texts will be used, i.e. political speeches and talks, whereas news produced in each of the mentioned source languages will be taken into consideration.
Chapter 2

What you need to know about annotation

The goal of this project is to create a manually annotated sentiment corpus in which the targets carrying sentiment and the related modifiers, along with the relations among them (called appraisal groups) are explicitly marked.

The annotation task involves three steps:

1. Identifying the targets
2. Identifying the modifiers
3. Determining the appraisal groups

For this reason, the guidelines considers three tags. Each tag has a set of attributes that need to be annotated as well. The attributes specify different properties of each tag (e.g., type of target, force of the modifier, etc.). For a quicker annotation, the most common attribute value is selected (see Table 2.1 for a complete list).

In this section, each of these entities will be introduced, and a quick tutorial on their annotation using the tool MAE [6] will follow.

2.1 The concept of TARGET

A target is any object, person or concept that is implicitly or explicitly regarded as positive or negative by the author of the text. There is often more than one target in a sentence. Consider three types of situation you will find yourself during the annotation:
CHAPTER 2. WHAT YOU NEED TO KNOW ABOUT ANNOTATION

<table>
<thead>
<tr>
<th>Tag</th>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>Type</td>
<td>Thing</td>
</tr>
<tr>
<td></td>
<td>Orientation</td>
<td>Neutral</td>
</tr>
<tr>
<td>Modifier</td>
<td>Orientation</td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td>Attitude</td>
<td>Affect</td>
</tr>
<tr>
<td></td>
<td>Orientation</td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td>Force</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>Polarity</td>
<td>Unmarked</td>
</tr>
<tr>
<td>Appraisal group</td>
<td>Orientation</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

Table 2.1: Attributes and default values. When tags are created in the annotation tool, some attributes are assigned default values. These values need to be manually changed or verified for each tag.

1. **One target with one feature.** It is the most common case, for example in “This article is useless”, “useless” is feature of the target “article”.

2. **One target with several features.** For example, in the sentence “The camera has a good quality, but a considerable weight”, the two features referred to the same object (camera) would be “quality” and “weight”.

3. **Different targets.** For example, in the sentence “I bought a good camera from that website, but a very bad phone”, there are potentially three targets (I, camera, phone), of which however “I” is not involved in a sentiment expression (I bought).

The targets of interest are usually in *sentiment* sentences. In order to explain this concept, we propose a list of definitions by Pang and Lee [5] with related examples (of which the first by the authors). However it is important to bear in mind that this is just for reference, as we do not use such a detailed classification.

- **Opinion** implies a conclusion thought out yet open to dispute (e.g., “Each expert seemed to be competent”).

- **View** suggests a subjective opinion (e.g., “Very assertive in stating his views”, “Very good performance”).
2.1. THE CONCEPT OF TARGET

- **Belief** implies often deliberate acceptance and intellectual assent (e.g., “A firm belief in her party’s platform”, “She thinks that souls survive after death”).

- **Conviction** applies to a firmly and seriously held belief (e.g., “The conviction that animal life is as sacred as human”, “A life devoted to the environment”).

- **Persuasion** suggests a belief grounded on assurance (as by evidence) of its truth (e.g., “Was of the persuasion that everything changes”, “Parents say that children make life better”).

- **Sentiment** suggests a settled opinion reflective of one’s feelings (e.g., “Her feminist sentiments are well-known”, “He hates his father with all his heart”).

To avoid any possible confusion, when talking about the category of **sentiment sentences**, it is also important to bear in mind that our annotation does not take into account whether they are **subjective** or **objective**.

**Subjective** sentences express subjective views and opinions [8], like in the examples above.

**Objective** sentences express factual information, which however hide an opinion. For example “The amount of oil import has doubled in the last ten years”.

After giving a broad overview of where to look for targets, we will now talk specifically about what targets are.

For what concerns the logical function, they can be either a **subject** or an **object**. It depends on which carries the sentiment together with the modifier (e.g., in “we share beliefs”, only the couple “share beliefs”).

For what concerns the grammatical function, they can be:

- **Common nouns**. By far, the most common category (e.g., children, cat).

- **Proper nouns**. Sub-categories include person (e.g., Olivia), place (e.g., China), animal (e.g., Sparky).

- **Verbs**. Usually when the modifier is an adverb (e.g., to talk too fast). However we will explain in Section 4 that this represents an exemption.

Targets have two attributes, **type** and **orientation**, which we will see in Section 4.
2.2 The concept of MODIFIER

The modifier is what modifies the target. Depending on the grammatical category of the target it refers to, it may belong to different categories:

- **Adjective.** When the target is a noun, for example “[beautiful]M [car]T”.

- **Verb.** When the target is a noun, for example “[obtain]M [victory]T” or “[victory]T [obtained]M”.

- **Adverb.** When the target is a verb and the adverb is too important to be implicit, for example “foolishlyM soughtT”. We will see in detail in Section 5 what this means.

- **Noun.** When the target is another noun and they are linked by a preposition, for example “[alliance]T for [progress]M”, which is equivalent to say “progress alliance”.

2.3 The concept of APPRAISAL GROUP

The appraisal group represents the opinion on a specific target. For this reason it contains both the target and the modifier. As mentioned previously (Subsection 2.2, in our annotation appraisal groups always match one of the following combinations:

- **A noun with an adjective.** For example “[wonderful]M [example]T”.

- **A pronoun with a noun.** For example “[they]T [cliches]M”.

- **A noun with a noun when linked by prepositions.** For example “[stigmatization]T of [people]M”.

- **A verb with an adverb.** For example “[strongly]M [support]T”.

- **A noun with a verb.** For example “[children]T [love]M”.

2.4 Using the annotation tool

In this annotation project we will use a toolkit called *Multi-purpose Annotation Environment Tool* (MAE) [6], freely available to download\(^1\). Before

\(^1\)https://code.google.com/p/mae-annotation/
2.4. USING THE ANNOTATION TOOL

you can run MAE, you need to make sure that Java\(^2\) is installed on your system. Once you have Java installed, on Windows, you simply double-click on the file with extension .jar. On Mac or Unix open a terminal window and, under the directory where you unzipped it, run the command “java –jar Mae_v0.9.1.jar”.

The window in Figure 2.1 will appear upon the launch of the program.

![MAE graphical interface](image)

Figure 2.1: MAE graphical interface

To start the annotation, you need to first load the annotation schema into MAE. This is a DTD file called “AppraisalAnnotation.dtd”, available for download\(^3\).

If you want to speed up a bit the annotation, you can also avoid annotating the prior orientation of targets and modifiers, by using another version of the dtd\(^4\).

Click on the “File” menu and select “Load DTD”. In the pop-up window, navigate to the “sample” folder and select the “AppraisalAnnotation.dtd”, and click the “Open” button to load the DTD. Notice that three tabs appear in the bottom screen, “APPRAISALGROUP”, “MODIFIER” and “TARGET”, which means that the schema has been successfully loaded. You will need to load this DTD file every time you run MAE.

\(^{2}\text{http://java.com/en/download/index.jsp}\)
\(^{3}\text{http://smlc09.leeds.ac.uk/marilena/public/AppraisalAnnotation.dtd}\)
\(^{4}\text{http://smlc09.leeds.ac.uk/marilena/public/AppraisalAnnotation_new.dtd}\)
CHAPTER 2. WHAT YOU NEED TO KNOW ABOUT ANNOTATION

Once the DTD is loaded, we can open a text file and start our annotation. It is of extreme importance to remember that, in case you want to annotate documents not already in the corpus, you have to pre-process them. Here is a list of suggested pre-processing tasks:

1. Split in sentences. Consider doing it on full stops, interrogation and exclamation marks.

2. Eliminate tags. In TED talks, it is likely you will have to eliminate brackets containing the tags *applause* or *laughs*.

3. Eliminate XML mark-up.

4. Eliminate unnecessary spaces.

In case you are annotating a text already in the corpus, you can just click on the “File” menu and select it. For this tutorial the text “Kennedy Inaugural.txt” has been chosen. The text will appear in the text area of the screen, like in Figure 2.2.

We can now begin the annotation. This should be done keeping in mind the author of the text’s perspective in order to reduce as much as possible the number of borderline cases. Reading the full text in advance might help.

![Figure 2.2: MAE: DTD loaded and document successfully loaded.](image-url)
2.5 Annotating a TARGET

To annotate a target, select the span of text (paying attention not to include white spaces before and after the word), right click on it, and select “TARGET” from the drop-down menu. See Figure 2.3 for an example.

![Figure 2.3: MAE Target annotation](image)

You will notice that the text “clergy” will be highlighted in red. If you click on the “TARGET” tag in the bottom screen, you will notice a new line inserted into the table. The columns of the table list several attributes of the TARGET (see Section 4). You always have the possibility to delete it, by selecting the span in the text or the relative row, right click and select “Remove”. Do the same for all of the other targets, even if they repeat themselves. To change the type or orientation attribute of your TARGET, you would click on the default attribute and select the desired value from the drop-down list.

2.6 Annotating a MODIFIER

Similarly, to annotate a MODIFIER in MAE, select the span of the text that you want to mark (paying attention not to include white spaces before and after the selected words), and then right click on it. Choose “MODIFIER” from the drop-down menu and it should be highlighted in blue. If you click on the “MODIFIER” tag in the bottom screen, you will notice a new line
inserted into the table. The columns of the table list several attributes of the MODIFIER (see Section 6). You always have the possibility to delete it, by selecting the span in the text or the relative row, right click and select “Remove”. Do the same for all of the other modifiers, even if they repeat themselves. To change the attitude, orientation, force or polarity attribute of your MODIFIER, you would click on the default attribute and select the desired value from the drop-down list.

2.7 Annotating an APPRAISAL GROUP

To annotate an APPRAISAL GROUP, you would hold down the CRTL button (or Command on Mac) and click on the two entities that you would like to link. A pop-up window will appear. Select the first entity (e.g., TARGET) in the “Link from” field, the second entity (e.g., MODIFIER) in the “Link to” field, and select APPRAISAL GROUP in the “Link Type” field. Then click “Create Link” (see Figure 2.4). An important function in MAE is the possibility of annotating the same word both as modifier and as a target, and decide which function to take when the appraisal group is created.

The orientation attribute of APPRAISAL GROUPs can be edited under the APPRAISAL GROUP tab in the bottom screen. As for targets and modifiers, you always have the possibility to delete it, this time only by selecting the row, right click and select “Remove”. This action will not imply the removal of individual targets and modifiers, whereas the removal of targets/modifiers will imply the cancellation of the related appraisal group. Do the same for all of the other appraisal groups, even if they repeat themselves. After all the APPRAISAL GROUPs in this text are annotated, your screen should look like that in Figure 2.5.

2.8 Saving your work

To save your annotation, click on “File” menu and select “Save file as XML”. We will save the file as “Kennedy Inaugural.xml” under the sample folder. Next time, when you are ready to continue your annotation, use the “Load File” function under “File” menu to load the XML file, and continue your work.
2.8. SAVING YOUR WORK

![Figure 2.4: Create Appraisal Groups in MAE](image1)

![Figure 2.5: Annotated text in MAE](image2)
Chapter 3

Whether to use the TARGET tag

As seen in Subsection 2.1, accepted grammatical categories are:

- **Common nouns.** For example “journals”, “meals”, “peoples”.

- **Proper nouns.** Sub-categories include person (e.g., Olivia), place (e.g., China), animal (e.g., Sparky). In multi-word expressions such as “the United States of America”, annotating “America” is enough, whereas “U.S.” and “United States” are annotated as they are.

- **Pronouns.** Usually personal (I, you, s/he, we, they).


An extremely important thing to bear in mind is that the target tag should be assigned to one word per time.

When two targets are linked by a preposition such as *and*, like in the case of “Luke and Isabelle”, please annotate them separately. This case can happen with common nouns as well (e.g., girls and boys) or a mix of common and proper nouns (e.g., cooperation with China). In all cases, remember that they must be linked to the same modifier(s). For example, in the expression “Luke and Isabelle are nice”, the groups are “[Luke]T [nice]M” and “[Isabelle]T [nice]M”.

The cases in which the tag must not be used are:

- **Non-sentiment words**, i.e. words that are not carrying actual sentiment. For example, “divine religions”, “double standards”, “worldly
possessions” or “scale of our ambitions”. When there is ambiguity, what should predominate is the author of the text’s perspective. For example:

- “Samuel Pisar, an Auschwitz survivor said”. In this case the expression “Auschwitz survivor” only indicates who this person is, so we cannot really assign a positive or negative connotation to “survivor” or “Aushwitz”.

We can see the difference with targets carrying sentiment (sometimes implied) in following examples:

- “People with disease are stigmatized”. In this case, the negative opinion is on the fact that this group of people is badly treated, so we need to annotate “people” as part of the group “[people]T [stigmatized]M”, but not the group “disease people”.

- “The Holocaust teaches us that nature, even in its cruelest moments, is benign in comparison with man, when he loses his moral compass and his reason”. In this case, “Holocaust” is implicitly negative because of the description of the reproachable human behaviour.

- “In the 1960s and 1970s, the Vietnam War was basically shown in America’s living rooms day in, day out”. In this case the fact that a negative event was constantly shown is considered negative itself. So we annotate the target “war” and include it into two groups “[Vietnam]M [war]T” and [war]T [shown]M”.

- “This is practically the model of TED”. Because this statement was made during a TED presentation, we understand that “[TED]M [model]T” has positive connotation.

- “The demand for pharmaceuticals is going to increase tremendously”. As intuition (confirmed by the broader context of this sentence) we know that we need to link the target “demand” to the two negative groups “[pharmaceuticals]M [demand]T” and “[demand]T [increase]M”.

- “Artists and innovators, many of the people you’ve seen on this stage”. Because of the implicit admiration towards “artists” and “innovators”, we annotate these two targets.

- “We will see an Einstein in Africa in this century”. Because of the implicit admiration towards “Einstein”, we annotate it.
– “Is China drinking our milkshake?” In this case, the target “milkshake” has to be annotated as part of the negative group “[drinking]_M [milkshake]_T”.

• **Actions that are simply statement of facts.** For example:
  
  – “The leader rejected the report”.
  – “Where the answer is yes, we intend to move forward. Where the answer is no, programs will end”.
  – “Those of us who manage the public’s dollars will be held to account”.
Chapter 4

Attributes for TARGET

4.1 Type attribute

The possible type values are: person, thing, place, action, other.

Animals are included in the category “thing”. Countries are usually annotated as place (e.g., Italy), whereas “world” can be either thing or place, depending on whether it carries the action or not.

The value “other” is only used when an adjective is marked as target, for example “[easily]M [imaginable]T”.

4.2 Orientation attribute

The orientation attribute is prior if a word typically evokes something positive (e.g., “peace”) or something negative (e.g., “pessimism”) when taken out of context. It can be positive, negative, neutral or ambiguous.

Ambiguous orientation is given if it depends on the context (e.g., for “example” in the cases “bad example” and “good example”). The actual orientation is annotated in the appraisal groups.

Some words are already considered ambiguous by sentiment dictionaries (e.g., “influence”, “intervention”, “retirement” and “revolution”). However, more common is the case of words that do not seem ambiguous at first (e.g., “growth”, which seems always positive, but is actually negative when modified by “slow”).

An important thing to underline is that, when prior orientation is marked, the sense of the word (grammatical function) should be taken into account (e.g., “light” as verb in “light the world” is positive a priori, “light” as noun in “give light” might be neutral).
Chapter 5

Whether to use the MODIFIER tag

The modifier tag usually corresponds to one word. The category of single words include the hyphenated ones such as low-budget).

As mentioned in Chapter 2.2, the accepted grammatical categories are:

- **Adjective.** When the target is a noun, for example “[beautiful]_M [car]_T ”. Apart from qualifying adjective, we can also have possessive adjectives (e.g., “[our]_M [star]_T”.

- **Verb.** When the target is a noun, for example “[obtain]_M [victory]_T ” or “[victory]_T [obtained]_M ”.

- **Adverb.** When the target is a verb and the adverb is too important to be implicit, for example “foolishly_M sought_T”. We will see in detail in Section 5 what this means.

- **Noun.** When the target is another noun and they are linked by a preposition, for example “[alliance]_T for [progress]_M”, which is equivalent to say “progress alliance”.

However, there are special cases in which one word does not carry the entire sentiment:

- **Phrasal/multi-word verbs.** These are annotated as single tokens (e.g. “cast off”) and put in a group (e.g., cast off worries). Other examples are “get rich”, “put forward”, “worried about”.

Multi-word verbs such as “symbolizing end”, “signifying renewal”, “proclaim an end” are included.
Because MAE allows annotating only words close to each other, in case the phrasal verb is split (e.g., “carried us up”, “tore the world apart”) but it still conveys the meaning, either the verb (e.g., “carried”) or the preposition (e.g., “apart”) can be annotated.

Another example is “pull ourselves out of this abyss” in which the group “[out]M [abyss]T” has been annotated.

- **Multi-word expressions.** These are annotated as single tokens (e.g. “at issue”, “at odds”, “in practice”, “out of control”, “under pressure”, “in the light”, “go off a cliff”).

- **Modified modifiers.** Modifiers modified by adverbs of degree (e.g., significantly, slightly) or affirmation (e.g., absolutely, seemingly). For example:
  - “Not very happy”, in which we keep the information related to “not” and “happy” as marked polarity and high force respectively (see Chapter 6).
  - “Far more iron”, in which “far more” is considered high force of “iron”.
  - “Far worse destructive power”, in which we annotate two groups “[far worse]M [power]T” and “[destructive]M [power]T” because “worse” is a high force modifier by itself.
  - “Strongly supporting”, in which “supporting” has got high force because of “strongly”.
  - “Too firmly focused”, in which the modifier “focused” is double modified.

The cases in which the modifier tag must not be used are the following:

- **Auxiliary verbs**
  - “Be” when followed by a non-finite verb (e.g., “it is working”, “it was chosen”).
  - “Have” when followed by a past participle (e.g., “crisis has re-minded”, “work has to be done”).
  - “Do” when followed by an infinitive (e.g., “I don’t like”, “you did know”).
  - “Used”, “dare”, “need” (e.g., “we dare not [meet]M a powerful [challenge]T”).
This category also includes **modal** verbs:

- “Can/could” (e.g., “he can really sing”).
- “May/might” (e.g., “that may be a problem”).
- “Must” (e.g., “we must change”).
- “Will/would” (e.g., “we will do”).
- “Shall/should” (e.g., “you should stop that”).
- “Ought” (e.g., “they ought to respect the law”).

**Catenative verbs**

- “Get” (e.g., “she got chosen for the job”).
- “Keep” (e.g., “she kept disturbing, to keep our [education]_T system [globalized]_M”).
- “Start/begin” (e.g., “she started to blame”).
- “Help” (e.g., “she helped improving”).

In the case in which two type of verbs follow each other other, preference is given to the one carrying the sentiment. For example:

* “Trade [walls]_T began to come [tumbling down]_M”
* “Have [caused]_M [change]_T to happen”

**Copulas/Operators**

- “be” followed by a noun (e.g., “heritage is a strength”) or an adjective (e.g., “she is ready”)
- “to have” in the sense of “possess” (e.g., “she has money”)

**Words not placed next to each other.** For example *never* and *before* in the expression “never seen our planet from this perspective before”.

No annotation should be done also in the case of phrasal verbs whose meaning is ruined when the second part is split (e.g., “pick ourselves up”, “dust ourselves off”).

**Modifier not linked to anything in the sentence.** For example in the short paragraph

“The common disease or the rare disease? Common. The common”

the modifier “common” makes a sentence by itself in two cases.
Chapter 6

Attributes for MODIFIER

6.1 Attitude attribute

These three values are taken from the Appraisal Framework [4]:

- **Affect**, which deals with personal emotions and opinions (e.g., “happy”, “sad”) and is further qualified by polarity;

- **Judgement**, which concerns author’s attitude towards people’s behaviour (e.g., “heroic”, “craven”);

- **Appreciation**, which considers the evaluation of things (e.g., “ugly”, “useful”).

We basically follow Taboada and Grieve [7] in order for an opinion to be easily classified under one of these three labels: comment on a thing (appreciation), a person (judgement), or one’s self (affect).

The value affect is definitely the most difficult to apply. We use it not only for personal states (e.g., I’m optimistic), but also for opinions (e.g., “if we are divided, we won’t achieve much”).

Because any grammatical category (noun, verb, preposition, adverb) can fall under the modifier label, we choose appreciation, judgement and affect according to what such word refers to (e.g., “out of control” would be classified as appreciation if linked to an action or to judgement if linked to a person).
6.2 Orientation attribute

As in the case of target, the orientation attribute is prior if a word typically evokes something positive (e.g. “beautiful”) or something negative (e.g. “horrid”) when taken out of context. It can be positive, negative, neutral or ambiguous.

Ambiguous orientation is given if it depends on the context (e.g., for the verb wishes in the cases “wishes well” and “wishes ill”). The actual orientation is annotated in the appraisal groups.

6.3 Force attribute

Force is:

- **Normal** as standard input. For example “good” because it is not modified itself. This applies to short questions as well, e.g., “But Jefferson wasn’t quite right, was he?”

- **High** when there is an adverb such as “very”, “extremely”, “not only..., but also...”.

- **Low** when there is an adverb such as “little”, “poorly”.

- **Reverse** when there are words called reversals such as “abolish” followed by targets usually associated to negative feelings (e.g., taxes, slavery), thus making the result positive.

Reverse category is used for nouns (e.g., “termination”), and verbs (e.g., “to decrease”, “to limit”, “to diminish”, “to abolish”, “to remove”), as well as for prepositions (e.g., “without”, “despite”) and condition operators that neutralize the attitude of related words (e.g., “if”, “even though”).

For example “If the world were as wealthy as the United States now”, the group “world wealthy” has reverse force, whereas “United States wealthy” has normal force. Same for “If you’d been a little more optimistic”, “optimistic” takes reverse force in the group “you optimistic” marked as negative.

The expression “definitely sparked outrage” is analysed in table 6.1 for reference.
Table 6.1: Appraisal group derived from the sentence “They definitely sparked outrage”

<table>
<thead>
<tr>
<th>Category</th>
<th>Expression</th>
<th>Orientation</th>
<th>Type</th>
<th>Attitude</th>
<th>Force</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modifier</td>
<td>sparked</td>
<td>Ambiguous</td>
<td>Action</td>
<td>Judgement</td>
<td>High</td>
<td>Unmarked</td>
</tr>
<tr>
<td>Target</td>
<td>outrage</td>
<td>Negative</td>
<td>Thing</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Appraisal group</td>
<td>sparked outrage</td>
<td>Negative</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

An important thing to bear in mind is that force signals are not annotated. For example in “She, better than most, knew the power of an image”, “knew” has got high force, because of “better”.

However, in those cases in which the adverb radically modifies the force of the group like in the expression “those who foolishly sought power”, a new group is created (“foolishly sought”), apart from the main one (“sought power”), provided that is falls into one of the combinations mentioned in Chapter 7. In “foolishly sought”, “foolishly” has high force, whereas “sought” has normal force.

We will now list some adverbs that do not need to be included in an appraisal group, i.e not to be marked as modifiers. They usually indicate intensity.

- Those to which corresponds low force: scarcely, merely, somewhat, badly, unsurprisingly, so.

- Those to which corresponds high force: extremely, a lot, well, again, more than ever before/ even more than ever before.

We will also list some adverbs that are advisable to put in groups. They usually indicate mode: safely, tirelessly, grudgingly, gladly, easily, at least, comprehensively, peacefully, rightly, responsibly.

6.4 Polarity attribute

The polarity of the modifier is marked when there is a negation, for example in the phrase “observe today not a victory”, “not” has not been included in the annotation, but to keep that information the polarity of “observe” is marked.

This is the case for every kind of negation: local (e.g., “not good”), involving longer-distance dependencies (e.g., “does not look very good”) or the subject (e.g., “no one thinks that it is good”, “nothing so satisfying”).
Auxiliary verbs give the polarity (e.g., “system cannot tolerate”), but they are not annotated either (see Chapter 5).
Chapter 7

Whether to use the APPRAISAL GROUP tag

The APPRAISAL GROUP tag has to be used with:

- **A noun with an adjective.** For example “[wonderful]M [example]T”. If a target has more than one *sentiment* modifier, like in the expression “cultural and spiritual origins”, one group for each modifier is created (“[cultural]M [origins]T”, “[spiritual]M originsT”). This applies also for modifiers not necessarily close to the target, for example in the sentence below:

  
  Let the word go forth from this time and place, to friend and foe alike, that the torch has been passed to a new generation of Americans born in this century, tempered by war, disciplined by a hard and bitter peace, proud of our ancient heritage and unwilling to witness or permit the slow undoing of those human rights to which this Nation has always been committed, and to which we are committed today at home and around the world.

  
  Figure 7.1: Sentence 1: several modifiers referring to one target

  

- **A noun with a verb.** For example “[children]T [love]M”.

- **A verb with an adverb.** For example “[rightly]M [marked]T” (see Subsection 6.3).

- **Nouns linked by a preposition:** For example the expression “victims of war” is annotated as “[war]M [victims]T. The usual prepositions are *of, for, in against, with, towards, between.* This case includes
32 CHAPTER 7. WHETHER TO USE THE APPRAISAL GROUP TAG

<table>
<thead>
<tr>
<th>id</th>
<th>fromID</th>
<th>fromText</th>
<th>toID</th>
<th>toText</th>
<th>orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A14</td>
<td>T13</td>
<td>Americans</td>
<td>M15</td>
<td>tempered</td>
<td>positive</td>
</tr>
<tr>
<td>A15</td>
<td>T13</td>
<td>Americans</td>
<td>M14</td>
<td>disciplined</td>
<td>positive</td>
</tr>
<tr>
<td>A16</td>
<td>T13</td>
<td>Americans</td>
<td>M16</td>
<td>proud</td>
<td>positive</td>
</tr>
<tr>
<td>A17</td>
<td>M17</td>
<td>hard</td>
<td>T14</td>
<td>peace</td>
<td>negative</td>
</tr>
<tr>
<td>A18</td>
<td>M18</td>
<td>long</td>
<td>T14</td>
<td>peace</td>
<td>negative</td>
</tr>
<tr>
<td>A19</td>
<td>M19</td>
<td>ancient</td>
<td>T15</td>
<td>heritage</td>
<td>positive</td>
</tr>
<tr>
<td>A20</td>
<td>M19</td>
<td>human</td>
<td>T16</td>
<td>rights</td>
<td>positive</td>
</tr>
<tr>
<td>A21</td>
<td>M21</td>
<td>slow</td>
<td>T17</td>
<td>undoing</td>
<td>negative</td>
</tr>
<tr>
<td>A22</td>
<td>M21</td>
<td>permit</td>
<td>T17</td>
<td>undoing</td>
<td>negative</td>
</tr>
<tr>
<td>A23</td>
<td>T13</td>
<td>Americans</td>
<td>M23</td>
<td>unwilling</td>
<td>positive</td>
</tr>
<tr>
<td>A24</td>
<td>T18</td>
<td>male</td>
<td>M24</td>
<td>committed</td>
<td>positive</td>
</tr>
<tr>
<td>A25</td>
<td>T19</td>
<td>we</td>
<td>M16</td>
<td>committed</td>
<td>positive</td>
</tr>
<tr>
<td>A183</td>
<td>T159</td>
<td>word</td>
<td>M195</td>
<td>go forth</td>
<td>positive</td>
</tr>
<tr>
<td>A194</td>
<td>T160</td>
<td>each</td>
<td>M196</td>
<td>passed</td>
<td>positive</td>
</tr>
</tbody>
</table>

Figure 7.2: Appraisal groups in Sentence 1 with several modifiers referring to one target

also the possessive case such as “man’s power”, “mankind’s war”, “expediency’s sake”, “women’s rights”.

• **Superfluous modifier.** For example:

  – “To set aside childish things”, in which “set aside things” does not have a completely different orientation with respect to “set aside childish things”.

  – “To choose our better history”, in which “choose history” has the same orientation as “choose better history”.

• **Numbers, percentages and words related.** The are annotated only if they carry sentiment for the speaker. For example:

  – “There is twice as much light available for everyone”, with the group “[twice][light]”.

  – “Several hundred documents”, with the groups “[several][documents]” and “[hundred][documents]”.

  – “Income less than 1,000 dollars per year”, with the group “[incomes][less]”.

  – “Less than 1/10th of one percent of the world’s population are scientists and engineers”, with the group “[less][percent]”. The other ways of annotating this would have been “less population” (wrong subject) or “less scientists/less engineers” (grammatically incorrect).
“1996, less than one million new university students in China, per year. 2006, over five million”, with the groups “[one]_M [million]_T” (negative) and “[five]_M [million]_T” (positive).

“Drug treats 1,000 people, 100,000 people, or a million people”, with the groups “[drug]_T [treats]_M” and “[treat]_M [people]_T”. In this case, because it does not make much sense to annotate the numbers (1,000, 100,000 and 100,000), we do not annotate the other occurrences of “people” other than the first either.

“Billions of our processors are off-line”, with the group “[processors]_T [off-line]_M”

The cases in which the appraisal group tag must not be used are:

- **Non-sentiment modifier.** For example, “growth of six percent”. Only “growth” is annotated as target.

- **Outside fixed combinations.** For example:
  - “Merely as a struggle” or “beyond doubt” because of the combination “adverb + noun”.
  - “Expect to see” because of the combination “verb + verb”.

- **Group assuming a wrong attitude.** For example in the sentence “to proclaim an end to false promises” the group “[end]_M [promises]_T” would be negative, so not equivalent to “end false promises”.

- **Group with no sense.** Unlike in “greatest economic disaster”, which can be easily split into “[greatest]_M [disaster]_T” and “[economic]_M [disaster]_T”, in
  - “Full measure of happiness”, the groups “measure of happiness” and “full measure” have not been annotated because they do not make sense without each other.
  - “First World War”, only the group “World War” has been annotated, not “First War”.

- **Same group in the same sentence.** For example in the sentence “They threw up walls, political walls, trade walls, transportation walls, communication walls, iron curtains”
only one group “[walls]T [threw up]M” has been created because all
the other occurrences (i.e. political walls, trade walls, transportation
walls, communication walls) do not convey different meanings when
put in a group without their modifiers.

• **Numbers and related words with no sentiment.** As always, we
take into account the author’s perspective, like in the following case:
“Even at that rate by 2100, average GDP per capita in the world will
be 200,000 dollars”.

Chapter 8

Attributes for APPRAISAL GROUP

8.1 Orientation attribute

The orientation of the appraisal group is *contextual*, i.e. it considers the context in which the word appears. The values it might assume are: positive, negative, neutral and ambiguous.

There are cases in which the orientation completely depends on the context. For example:

- “[Hungry]_M [minds]_T”, a positive appraisal group.
- “Provoke us to step up” with the appraisal group “[provoke]_M [us]_M” as ambiguous. The group “[us]_T [step up]_M” would instead be positive.

Other cases in which the prior orientation of a word is changed when put in a group. For example:

1. “All of this has been tremendous for the world” with the appraisal group “[all]_T [tremendous]_M” as positive (see Figure 8.1 for the context).

2. “Lack of freedom”, which is negative although “freedom” is positive a priori.

3. “Prohibition of terrorism”, which is positive although “terrorism” is negative a priori.
CHAPTER 8. ATTRIBUTES FOR APPRAISAL GROUP

We globalized the world. And what does that mean? It means that we extended cooperation across national boundaries. We made the world more cooperative. Transportation walls came tumbling down. You knew in 1950 the typical ship carried 5,000 to 10,000 tons worth of goods. Today a container ship can carry 150,000 tons. It can be manned with a smaller crew, and unloaded faster than ever before. Communication walls, I don’t have to tell you, the internet have come tumbling down. And of course the iron curtains, political walls wall have come tumbling down. Now all of this has been tremendous for the world. Trade has increased.

Figure 8.1: Context for the sentence “All of this has been tremendous for the world”

The neutral value is actually never used during the annotation, as we are interested only in targets which carry sentiment (see Chapter 1 for further details).

The ambiguous value can be assigned for two reasons:

- **Modifier cannot be included.** For example:
  
  
  
  
  - “[Increasing]M [demand]T” in the expression “increasing demand for ideas”.
  
  - “[They]T [met]M” in the expression “they will not be met easily”. The group “[met]T [easily]M” would instead have positive orientation.

- **Unsure connotation given by the speaker.** For example in:
  
  - “America bigger than the sum of individual ambitions, greater than all the difference of birth and wealth”, it is not clear whether “ambitions” and “differences” are negative or neutral.
  
  - “How many Ramanujans are there in India today toiling in the fields, barely able to feed themselves?”, it is not clear whether “feed themselves” is positive or negative.
8.1. ORIENTATION ATTRIBUTE

In case the orientation given by the author is clear, but it does not sound right outside the context, it must not be changed. For example in:

- “Misery does love company” the group “[love]_M [company]_T” is marked as negative.
- “US losing leadership”, the group “[US]_T [losing]_M” is positive and “[losing]_M [leadership]_T” is negative.

Groups that are ambiguous will assume the orientation of the context. For example:

- “[Showered upon]_M [favours]_T”, negative according to the context.
- “[Idea]_T [passed on]_M”, positive.
- “[War]_T against [network]_M” in the expression “Our nation is at war, against a far-reaching network of violence and hatred”. The groups “war against violence” and “war against hatred” have not been annotated because only logical groups, not grammatical.
- “[Band]_T of [patriots]_M in the expression “a small band of patriots huddled”.

Chapter 9

Special cases

We will now list a number of cases not very easy to annotate.

- **External references.** Especially in TED talks, it might happen to find external references to the screen (e.g., this photo, first few bars here). In this case, the annotation is as usual if the target carries sentiment.

- **Difficulty in extracting sentiment through couples.** For example:
  
  - “I think it’s very unlikely they were far from the minds of Americans”, which can be annotated as “[they]_T [far]_M” with marked force and negative orientation, as it was equivalent to “they were not far”.
  
  - “I wish you were here”, which can be annotated as “[wish]_M [you]_T ”.
  
  - “The horrifying images from Abu Ghraib as well as the images from Guantanamo have a profound impact”, with the groups “[horrifying]_M [images]_T, [Abu Grahib]_M [images]_T, [Guantanamo]_M [images]_T, [profound]_M [impact]_T.
  
  - “Thank you to all photographers”, annotated as “[thank]_M [photographers]_T”.

- **Too long modifiers.** For example, “in a short span of time” and “from generation to generation” are too long and cannot be annotated. In addition, in the first case, it is impossible to select “in time” because of software limitations.
CHAPTER 9. SPECIAL CASES

- **Adjective followed by a noun with words in the middle.** For example “resistant to growth”, annotated as “[growth]_T [resistant]_M”.

- **Useful negation markers.** For example in the sentence
  “heritage is a strength, not a weakness”
  one group contains “not” (“[not]_M [weakness]_T”), because otherwise the only verb “is” would have carried both positive and negative polarity.

- **Useful force markers**, especially because of the lack of subject. For example in:
  - “And over the next 18 years have almost tripled”, the group “[almost]_M [tripled]_T” should be annotated because the subject is missing.
  - “Absolutely incredible”, the group has necessarily to be that one because the sentence consists of these two words.
  - “Sometimes even poorer than their grandparents had been”, the group “[sometimes]_M [poorer]_T” should be annotated because the subject is missing.

- **Punctuation.** Because the sentence is usually split on (1) full stops, (2) exclamation mark and (3) interrogation marks, annotating words separated by colon and semi-colon is allowed. For example in:
  1. “Some data on tariffs: coming down to” the group “[tariffs]_T [coming down]_M” has been annotated because it is not just a fact for the speaker.
  2. “America bigger than the sum of our individual ambitions; greater than all the differences”, the group “[America]_T [greater]_M” has been created although the modifier appears after the semi-colon. See Figure 9.1 for detailed analysis.
  3. “HIV/AIDS” in the sentence below are separated by double dash.
     “Images have [[power]_T to [shade light]_M]_AG on [understanding]_M on [suspicion]_T, [ignorance]_T, and in particular – I’ve given [[a lot of]_M [talks]_T]_AG on this but I’ll just show one image – the issue of [HIV]_T/[AIDS]_T”.

- **Appraisal groups with marked polarity or reverse force** appear to be of opposite orientation. For example:
Figure 9.1: Appraisal groups annotated in sentences separated by semi-colon

- “Not a [bad]_M [thing]_T”. Although positive, marked as negative.
- “No more [kept]_T [in dark]_M”. Although positive, marked as negative.
- “Not [compromised]_M [principles]_T”. Although positive, but marked as negative.
- “Not [shrink from]_M [responsibilities]_T”. Although negative, marked as positive.
- “Not [one]_T of [shortcuts]_M”. Although negative, marked as positive.
- “Not [one]_T of [settling for less]_M”. Although negative, marked as positive.
- “Not a [path]_T for [faint-hearted]_M”. Although negative, marked as positive.

• **Co-reference**: When an element is referring to something else mentioned in the sentence before or after, as in the short paragraph
  “Let’s begin with some images. They’re iconic, perhaps cliches” the pronoun (“they”), rather than the actual subject (“images”), is annotated in two groups (“[they]_T [iconic]_M” and “[they]_T [cliches]_M”).
  This strategy is used only when two or more sentences are involved. If it happens in the same sentence, e.g., “a nation, whether it is good or bad”, annotating the logic subject (nation) instead of the actual one (it) is allowed.

• **Logical link**. For example in the following sentence:
  “People were actually getting poorer than their parents. And sometimes even poorer than their grandparents had been.”
  the groups “parents poorer” and “grandparents poorer” should not be annotated because the link is not direct. However, here are the cases in which a direct link works:
— “It is fortunate that we are becoming less of an idea leader”, in which the groups “[we] \(T\) [fortunate] \(M\)” and “[less] \(M\) [leader] \(T\)” are annotated, in the first case in order not to lose the information carried by “fortunate”.

— “Best efforts to help, because it is right”, in which the annotated group is “[help] \(T\) is [right] \(M\)”.

— “The many who are poor, the many who are rich”, in which the annotated groups are “[many] \(T\) [poor] \(M\)”, “[many] \(T\) [rich] \(M\)”.

• **Idioms and fixed expressions.** It is still advisable to find appraisal groups like in the following cases:

  — “The best is yet to come” as “[best] \(T\) to [come] \(M\)”.
  — “On the cutting edge” as “[cutting] \(M\) [edge] \(T\)”.
  — “We are not there” as “[we] \(T\) [there] \(M\)”.
  — “Settling for less” as “[settling] \(T\) [for less] \(M\)”.

In alternative, if it is important to link the target, they can be annotated as single unit (e.g., “[parents] \(T\) [had the last say] \(M\)”).

• **Sarcasm.** For example in “Pull yourself up by your own bootstraps, even if you don’t have boots”, the groups would be “[have] \(M\) [boots] \(T\)” as negative.
Appendices
Appendix A

Figure A.1 shows an example of the annotation output. It consists in a file in XML format respondent to the *Linguistic Annotation Framework* (LAF) developed by the *International Organization for Standardization* (ISO), which allows multilayer annotations and to have the text separated by the annotations.

![Figure A.1: Example of XML annotation output](image-url)
Bibliography


