SEVENTH FRAMEWORK PROGRAMME

OPEN POLARITY ENHANCED NAMED ENTITY RECOGNITION

OpeNER

Instrument: Collaborative Project
Thematic Priority: FP7- ICT -2011.4.1
Grant Agreement: 296451

WP5 - SENTIMENT- Development of Sentiment Lexicons and Opinion Mining Techniques

Deliverable Number	WP5 Extra Deliverable (Cycle 2)
Title	Guidelines for Opinion Annotations of Political News
Version	0.1
Date	4th June 2014
Status	Draft

EC Distribution

Project Partners: VICOMTECH-IK4, UPV/EHU, VUA, SYNTHEMA, OLERY, CNR

Contributors: VUA (Chantal van Son, Isa Maks)

Every effort has been made to ensure that all statements and information contained herein are accurate; however the Partners accept no liability for any error or omission in the same.

Document Control

Version	Date	Author	Modifications
0.1	4 June 2014	Chantal van Son / VUA	First edition.
0.1	4 June 2014	Isa Maks / VUA	First edition.

Project	Phase
Open Polarity Enhanced Named Entity Recognition	WP5 - SENTIMENT- Development of Sentiment Lexicons and Opinion Mining Techniques

Version	Date	Author
0.1	15 April 2014	VUA

Table of contents

1 Introduction

The task that is described in these guidelines is the identification of opinions expressed in political news texts. The basics of this task are similar to the identification of opinions in reviews (as described in the Quick Start Guide and the Annotation Guidelines for the annotation of reviews). However, political news texts are quite different from reviews in terms of complexity. Therefore, there are also some differences in the annotation framework.

The annotation framework we explain in these Guidelines is inspired by the work of Wiebe et al. (2005) and Wilson (2008), but has been simplified and extended in some respects. The annotation task that you will be performing can be divided into three subtasks:

    I.   Selection of good examples of opinionated sentences
    II. Selection of good examples of non-opinionated sentences
    III. Annotation of opinionated sentences (opinion triples)

The first step is to identify sentences that express a positive or negative attitude of someone towards something or someone. We will not annotate complete news articles with all the subjective evaluations that are present, as was done with the annotation of reviews. Instead, you are asked to find and select what will be regarded as 'good examples' of opinionated sentences in the news texts. What we mean by 'good examples' will be explained in Section 2.

Second, for each opinionated sentence you have selected, you will also select a non-opinionated sentence in the same article or –if needed- in another news item. This is needed in order to distinguish between opinionated and non-opinionated sentences.

The third step is to annotate *opinion triples* in the opinionated sentences that you have selected. A complete opinion triple consists of the following three components:

    I.   Opinion holder
    II. Opinion target
    III. Opinion expression

These concepts are also explained in the Annotation Guidelines for the annotation of reviews, so you are already familiar with them. However, there are some differences with respect to *what kind of words* can be annotated as a holder, target or expression. Furthermore, because opinions can be expressed in many different ways in news texts, we distinguish between two types of opinion expressions: *direct expression of attitude* and *indirect expression of attitude*. The difference between these two types will be explained later.

The opinion holder refers to the entity (a person or organization) to whom the opinion can be attributed. The opinion target refers to the entity or issue where the opinion is about. Finally the opinion expression refers to the span of words that indicate a positive or negative attitude of the holder towards the target. Sentence (1) shows an example of an opinion triple; we will use the colors green, purple and blue to mark the holder, target and expression respectively for all the examples throughout these Annotation Guidelines. Direct neutral attitude expressions (see Section 3.3.2) are marked grey.

(1) Germans are overwhelmingly opposed to military action in Syria.
(2) Merkel calls EU stance on Syria "invaluable".

2 Selection of sentences (opinionated/non-opinionated)

News articles often contain a lot of opinionated sentences. However, in this annotation project, we will not annotate complete news articles with all the subjective evaluations that are present. Instead, our aim is to create a collection of sentences that represents a balanced sample with good, clear and varied examples of opinionated text. It is up to you to select from different news articles the best examples. Furthermore, you are asked to select non-opinionated sentences as well. Non-opinionated sentences express only factual information without any attitude towards this information. Although these non-opinionated sentences are not the main focus in this project, we do need to identify them in order to distinguish between opinionated and non-opinionated sentences.

We are a bit limited in the amount of sentences we can select for each article due to copyright restrictions. For articles with less than 10 sentences, you may select only 1 sentence (opinionated or non-opinionated). For articles with more than 10 sentences, you may select around 10-15%. In some cases (if one article contains a lot of good sentences!) you are allowed to select some more, but be aware of these restrictions. Furthermore, because some articles contain a lot non-opinionated sentences, while other contain a lot opinionated sentences, you are allowed to select only non-opinionated sentences in one article and only opinionated sentences in another. Of course you may also select both non-opinionated and opinionated sentences in one article.

2.1 Selection of news articles
Before you select the sentences, you have to select news articles from which you are going to collect sentences. News articles have been collected from all kinds of newspapers: internet newspaper, news sites, and online versions printed (offline) newspapers. The folder names refer to the source where they come from. You are free to choose sentences from all sources although the articles from online newspapers and news sites seem to be more appropriate than the news articles of online versions of the printed newspapers. Usually, sentences meant for online readings are shorter and less complex.
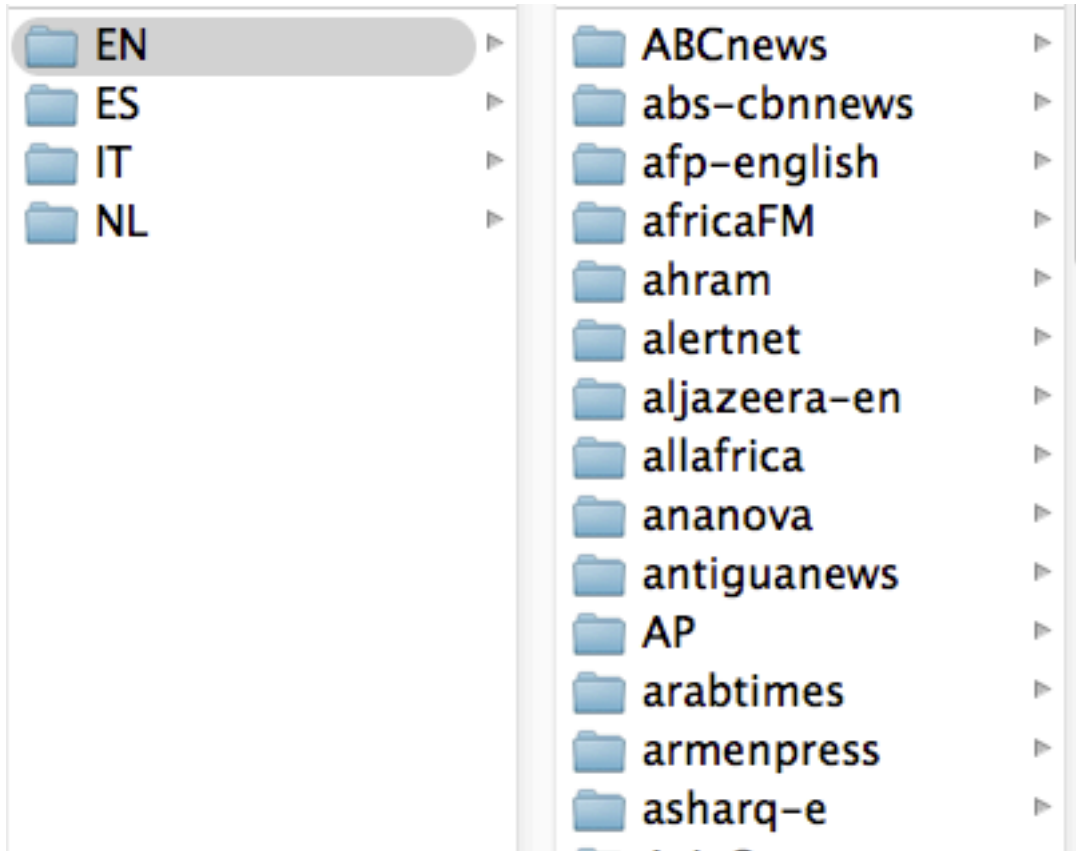
Figure 1: Folder structure of English news articles

2.2 Non-opinionated sentences

Non-opinionated sentences express only factual information without any attitude towards this information. Examples of non-opinionated sentences are listed below. "Non-opinionated" does not mean that there may not be some positive or negative fact mentioned in the sentence (e.g. *more than 270 rockets had struck Israel, injured, took heavy losses*), but rather that there is no sentimental evaluation towards this (inherently) positive or negative fact present in the sentence.

(3) The Netherlands is Russia 's biggest bilateral trade partner after China.

(4) It said more than 270 rockets had struck Israel since the start of the operation, with its Iron Dome interceptor system shooting down more than 130 rockets bound for residential areas.

(5) Angela Merkel injured in skiing accident.

(6) The anti-immigrant Geert Wilders' Eurosceptic party took heavy losses.

(7) Syrian and U.N. officials are working to finalize the timing of the visit.

However, something may be presented as a fact, but still represents an attitude of someone. For example, in (8) the newspaper reports about the fact that Guido Westerwelle spoke with Angela Kane, but the content of the conversation contains a positive attitude of the U.N. and Syria towards experts visiting the site. Therefore, this sentence is annotated as an opinionated sentence.

(8) German Foreign Minister Guido Westerwelle spoke with U.N. disarmament chief Angela Kane about the agreement between the U.N. and Syria to allow experts to visit the site.

Tag level 5 in the KAF Annotation Tool is used to mark non-opinionated sentences. We only annotate complete sentences as non-opinionated.
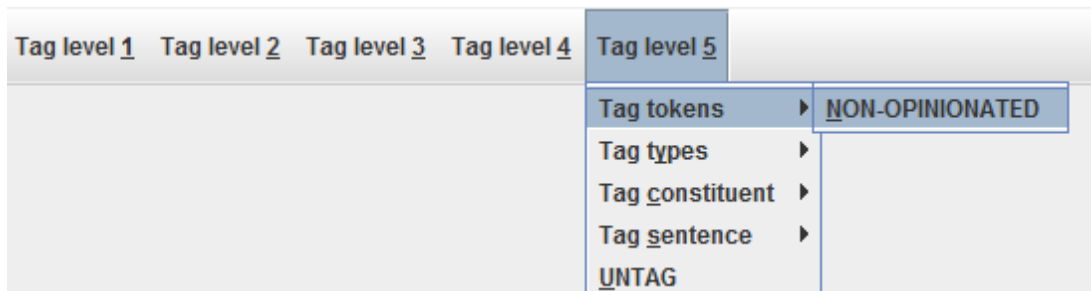
Figure 2: Tag level 5 for non-opinionated sentences

2.3 Clearly opinionated sentences

Clearly opinionated sentences are sentences that express a positive or negative attitude from an opinion holder towards an opinion target. The attitude may refer to an evaluation of the current (cf ex. 9, 13), past (ex. 12) or future (ex. 10, 11) situation. It may be expressed by an emotion (ex. 13), an evaluation (ex. 9, 10, 11) or by stance-taking expressions (ex. 12).

(9) Merkel calls EU stance on Syria "invaluable".
(10) However, Conservative Justice Minister Kenneth Clarke said Britain's exit from the EU would be "disastrous".
(11) He said in a Cuban newspaper: "If a war breaks out there, there would be a terrible slaughter of people with no benefit for either of them.
(12) Angela Merkel didn't want Joachim Gauck to be president.
(13) More than 70 % of the Greeks are angry and resentful at Germany for continuous austerity measures.

In these sentences you will annotate the opinion triples. This is explained in Section 3 and 4.

2.4 Sentences to avoid

2.4.1 Sentences that are not tokenized correctly

Unfortunately, not all sentences in the articles are tokenized correctly. This is caused by missing punctuation marks. This can be seen by the colours in the KAF Annotation Tool, by the words that have

been merged together, or words that do not belong in the sentence (e.g. data of publication), as in Figure 3. Do <u>not</u> annotate these sentences.

Id	Word token	Pos	Tag1	Tag1 id	T
w9	Russia	R		0	
w10	Urges	R		0	
w11	US	R		0	
w12	to	P		0	
w13	Await	V		0	
w14	Syria	R		0	
w15	Inspection	R		0	
w16	FindingsMO...	R		0	
w17	August	R		0	
w18	25	O		0	
w19	,	O		0	
w20	2013	O		0	
w21	(	O		0	
w22	AP	R		0	
w23	)	O		0	
w24	By	P		0	
w25	LAURA	R		0	
w26	MILLS	R		0	
w27	Associated	R		0	
w28	PressRussia	R		0	
w29	urged	V		0	
w30	the	D		0	
w31	U.S.	R		0	
w32	and	C		0	
w33	its	Q		0	
w34	allies	N		0	
w35	Sunday	R		0	
w36	to	P		0	
w37	await	V		0	
w38	the	D		0	
w39	findings	N		0	

Figure 3: Sentence that is not tokenized correctly: do not annotate

2.4.2 Too complex sentences

Some sentences are just too complex to annotate or contain opinions that are difficult to interpret. Do not annotate these kind of sentences; rather choose sentences that you are sure of that you can annotate them and are as unambiguous as possible.

(14) One old boy, whom I had always regarded as a friend, wrote that I was only supporting the bill because I wanted my eight-year-old son to be sodomised.

(15) If only those Conservative backbenchers who are now so exercised about the evils of same-sex marriage would swivel their eyes back to the speech David Cameron made a few years ago, they might just understand what the majority of the public really think.

Project	Phase
Open Polarity Enhanced Named Entity Recognition	WP5 - SENTIMENT- Development of Sentiment Lexicons and Opinion Mining Techniques

Version	Date	Author
0.1	15 April 2014	VUA

(16) I might add –further setting the scene of Brooks' professional and social circle– that in my conversation with Brooks that evening (she was not yet married to Charlie Brooks, so was still officially Rebekah Wade), she was discussing what made Rupert Murdoch Rupert Murdoch.

(17) He said: I'll do everything for Greece to stay in the euro and have the resources it needs by the end of the year, without it having to be necessary to inflict new conditions other than these already admitted by the Samaras government.

2.4.3 Expressing neutral opinions

You may come across sentences that do contain some opinion, but are not positive or negative. For example, phrases like (17) and (18) can be used when someone argues about something (convincing the other that this is true), but it does not carry any sentimental value. Another example is (19), in which you cannot interpret "national attention" as being either positive or negative.

(18) Saying there is strong evidence…

(19) We believe that…

(20) While Kentucky's implementation of the Affordable Care Act has received national attention.

3   Annotation of opinion triples

A complete opinion triple consists of an opinion holder, an opinion target and an opinion expression. In the following paragraphs we will explain exactly what kind of words and phrases can be tagged with

TIPS

- Only annotate complete sentences as non-opinionated (no opinion present at all); opinionated sentences may contain non-opinionated parts (do not annotate these parts).
- Choose short sentences with not too complex syntactic constructions.
- Avoid sentences with opinions that are difficult to interpret.
- Only annotate sentences that you can annotate without too much difficulties.
- Headlines can be chosen. They are often highly opinionated. As they are not always grammatical, we prefer a mix of headlines and other, grammatically well formed, sentences.
- Try to choose sentences with holder and target, or at least one of them.

these labels and what kind of difficulties you may come across.

3.1 Opinion Holder

The opinion holder is the entity to whom the opinion can be attributed. In news texts, opinions can be attributed to a variety of entities. Sometimes, the holder can be lexicalized with a lot of adjectival phrases; the rule is to try to annotate as precise as possible and not to annotate very 'large' holders. In most cases, the head of the NP will be sufficient.

3.1.1 The author

First, the opinion holder can be the writer of the news article himself. In this case, the holder is either lexicalized with a first person pronoun (rarely) or a phrase like "in my opinion", or implicitly present as the 'default' opinion holder, as in (21).

(21) So what do we have? A prime minister who has had the courage and leadership to do "the right thing". (implicit – do not annotate any holder)

3.1.2 Specific persons, organizations, etc.

In most cases however, the writer of a news article will present the opinions of others. Examples of opinion holders that you may come across are Named Entities, that is, specific (groups of) persons, organizations, or locations (countries or cities that actually refer to organizations).

(22) Military strikes against Syria would threaten prospects of any international peace talks, Russian Foreign Minister Sergei Lavrov said.

(23) New German minister lashes Russia for exploiting Ukraine.

(24) Germans are overwhelmingly opposed to military action in Syria.

(25) Moscow will hope it keeps Kiev in its political and economic orbit.

(26) As anti-government protests continue in Kiev, Russia agrees to buy $ 15 billion of Ukrainian debt and to reduce the price of natural gas.

3.1.3 Pronouns

In some cases however, there may be a pronoun functioning as an opinion holder that refers to an entity that is mentioned earlier or later in the sentence. For example, in (27) *he* refers back to *Finnish PM Jyrki Katainen*. This happens a lot with quotations, as in (28). In this sentence, *we* refers forward to Van Rompuy and, presumably, some of his colleagues. In these cases, you have to annotate the pronoun as an opinion holder.

(27) Finnish PM Jyrki Katainen also said he wants to keep the ban.

(28) "We simply could not ignore the extremely difficult economic realities across Europe," Van Rompuy told reporters.

31.4 Other opinion holders

Finally, you may come across opinion holders that are not a Named Entity, but refer to some (more or less 'vague') group of people. Examples of these kind of opinion holders are shown in the following sentences.

(29) But it seemed a loss for many of the newer -and generally poorer- members , who see Europe as a club that is only as strong as its weakest member. (note that we only annotated the head of the NP *members,* because the whole phrase *many of the newer –and generally poorer-* members is too large).

(30) That group, led by Poland and France, argued that Europe meant nothing if the budget were not used to bridge the wealth gap between rich and poor members and help restart growth.

(31) The world's top political and religious leaders have expressed shock and regret at Pope Benedict XVI 's resignation, but declared respect for his tenure and leadership of 1.1 billion Catholics.

3.2 Opinion Target

The opinion target refers to 'whom or what' the opinion is about. As with opinion holders, a target may be a Named Entity referring to a specific person, organization, or location. The same rule of 'annotating as precise as possible' (not much more than the head of the NP) also applies for the target.

(32) Ms. Karr, who works at the House Education and the Workforce Committee, described Carlyle as the "best of the best." (++)

(33) The Turks and "some Arab countries" have been especially supportive, he said, without providing details.

(34) "Germany is an engine of growth, an anchor for stability," said Dr Merkel.

(35) Germany, along with the European Union, blames the Syrian government for the attack.

However, targets may also be issues or events. Note that, again, in some cases pronouns can function as the target when they refer to something mentioned earlier or later in the text, as in (39).

(36) The joint banking supervisory body is also a breakthrough.

(37) "The effort was worth it," said German Chancellor Angela Merkel.

(38) The proposed budget has also been criticised for cutting too deeply into aid for poor countries or other areas critical for Europe.

(39) But it seemed a loss for many of the newer -and generally poorer- members , who see Europe as a club that is only as strong as its weakest member.

(40) Merkel calls EU stance on Syria "invaluable".

(41) The world 's top political and religious leaders have expressed shock and regret at Pope Benedict XVI 's resignation, but declared respect for his tenure and leadership of 1.1 billion Catholics.

3.3 Opinion Expressions

The opinion expression refers to the span of words that indicate a positive or negative attitude of the holder towards the target. The opinion expressions are subcategorized according to their type of expression (direct vs. indirect), their orientation (positive vs. negative) and their strength (default vs. strong). The following paragraphs will explain what direct and indirect attitude expressions are, how to decide whether the opinion has a positive or a negative polarity and whether the opinion is strong or 'normal'. Furthermore, both polarity and strength of the expression can be modified by *intensifiers*, *weakeners* and *polarity shifters*, which will also be explained. Again: try to be as specific as possible when annotating the opinion expression (choose only the words that express the negative or positive attitude).

3.3.1 Direct and indirect attitudes

Opinions can be expressed in several ways. In this annotation framework, we distinguish between two types of opinion expressions:

I.   Direct attitude expression
II.   Indirect attitude expression

The difference between these two types of expressions is illustrated in the following examples:

(42)   Amnesty International has criticized the sentencing of 13 Catholic activists. (direct)
(43)   Amnesty International called the punishment "unacceptable". (indirect)

In both sentences there is a negative attitude expressed of the opinion holder *Amnesty International.* In (42) the negative attitude towards the target is expressed by the verb *criticized*, which directly expresses the type of attitude held by the opinion holder. Also, these kind of attitude expressions directly imply the opinion holder: there needs to be someone that "is criticizing". We will call these kind of words *direct attitude expressions*. Other examples of verbs that belong to this category are *to disapprove, to fear* and *to support*, but it may also be nouns or adjectives, like *disagreement, (to have) hope, support (for), (to be) afraid, (to be) satisfied,* etc. These are words or phrases that will probably reoccur quite often. TIP: Direct expressions are often emotions (*he is angry about*), speech acts (*he refuses to; his refusal to*) or position-taking statements (*he wants; he is against; he has strong views against*).

In (43) on the other hand, the negative attitude is expressed by the words between quotation marks: "*unacceptable".* In this case, the negative attitude is not as directly expressed as in (42), but rather can be inferred from the content of the words. The inferred attitude would be something like 'disapproval'. We will call these kind of opinion expressions *indirect attitudes.* Indirect attitude expressions do not imply an opinion holder: there does not have to be someone that thinks that this is the case, it might just simply be so. However, in many cases these indirect attitude expressions will be combined with what we consider to be *neutral direct attitude expressions,* like *called* in (43). These words or phrases introduce an opinion of someone without having a negative or positive meaning themselves (see also

	Project	Phase
	Project	**Phase**
	Open Polarity Enhanced Named Entity Recognition	WP5 - SENTIMENT- Development of Sentiment Lexicons and Opinion Mining Techniques

Version	**Date**	**Author**
0.1	15 April 2014	VUA

Section 3.3.2). Finally, in contrast to direct attitude expressions, indirect attitude expressions are often phrases that you will probably rarely find in another text again.

If you are not sure whether you are dealing with a direct or an indirect attitude, it may help to paraphrase the sentences as in Figure 3. In these sentences, you fill in the attitude expression that you have identified. For example, you can paraphrase (42) as "There is someone (*AI*) that *criticizes* something (*the sentencing*)", but not as "There is someone that finds that something is *criticizes(?)*." Therefore, *criticizes* is a direct attitude. On the other hand, (43) can be paraphrased as "There is someone (*AI*) that finds that something (*the punishment*) is *unacceptable* ", but not as "There is someone that *unacceptable(?)* something." Therefore, the expression in (43) is an indirect attitude. Note that you can also paraphrase nouns as verbs for the direct attitudes. For example, *disagreement* could be paraphrased as "There is someone that *disagrees* (about) something". Adjectives can be paraphrased in combination with *to be* or *to have* ("There is someone that *is happy/has hope for* something").

Direct attitude	There is someone that	attitude	something/someone
Indirect attitude	There is someone that finds that something/someone is		attitude

Figure 3: Paraphrasing the sentence to decide whether the attitude is direct or indirect

Other examples of direct and indirect attitude expressions are shown in the following sentences.

Examples of direct attitudes

(44) Germany does not want the fund to spend billions in exchange for collateral from ruined banks.

(45) European leaders have hailed an agreement to use the eurozone 's collective bailout funds to recapitalise struggling banks directly , as stock markets reacted positively and the euro surged against the dollar over the summit deal.

(46) During their phone call, the two leaders "expressed their willingness .. . to continue their close cooperation to meet the challenges of the European project", the French presidency said in a statement.

(47) While British Prime Minister David Cameron is seeking to keep payments into EU coffers down as low as possible, French President Francois Hollande called for sustained subsidies for farming and development programs for poorer nations.

(48) Mr Cameron said he was "disappointed" not to have been told about the Algerian military operation before it began.

Examples of indirect attitudes

(49) The joint banking supervisory body is also a breakthrough.

(50) The Turks and "some Arab countries" have been especially supportive, he said, without providing details.

(51) "Germany is an engine of growth, an anchor for stability," said Dr Merkel.

(52) A bailout would not be the apocalypse," said Jose Maria Beneyto.

(53) A flustered Mr Cameron yesterday said it was "completely unacceptable" and ordered an internal party inquiry.

3.3.2 Polarity: positive vs. negative (and neutral)

An attitude is in principal either positive or negative, as is illustrated in the following examples:

(54) In Fitch's opinion, the debt ceiling is an ineffective and potentially dangerous mechanism for enforcing fiscal discipline. (-)

(55) According to the report, the Japanese government backed the whale hunt with ¥3.06 billion, or $33 million, in 2011, a ¥2.28 billion jump from previous years. (+)

When deciding whether an attitude is positive or negative, you should look at the expression as a whole. It is possible that there are words in the expression that carry a positive or negative meaning, but occur in combination with a so-called 'polarity shifter', that is, a word (or span of words) that shifts the polarity of an opinion word from positive to negative or vice versa, like negations or *too (little/much/etc.)*. For example, the positive phrases *to be pleased with* and *to pay attention to* are shifted to negative expressions by the polarity shifters *not* and *too little*.

(56) While Kentucky's implementation of the Affordable Care Act has received national attention, state lawmakers from both parties are not pleased with parts of the legislation. (-)

(57) With all eyes on Iran, *too little attention* is being paid to the fact that Ukraine is being forced back under the control of the Kremlin. (-)

In some cases however, an indirect attitude expression is combined with a direct attitude expression that cannot be classified as positive or negative. Rather, they introduce an opinion. Examples are *saw* and *considers* in (58) and (59). These kind of expressions are annotated as *direct neutral attitude expressions*.

(58) Merkel saw (D-neutral) Great Britain as an ally (I-positive).

(59) Merkel considers(D-neutral) EU stance on Syria invaluable(I-negative).

3.3.3 Strength: default vs. strong

Some opinion words, phrases or clauses express stronger polarity than others, like *hate* (vs. *dislike*) and *excellent* (vs. *good*). Furthermore, opinion expressions may contain so-called 'intensifiers', that is, words or spans of words that intensify the strength of an opinion word (e.g. *very, really, severely*). So opinion expressions can either express a 'normal' (default) positive or negative attitude, or a 'strong'

positive or negative attitude, depending on the strength of the positive or negative words themselves and the presence or absence of intensifiers. Examples of strong positive and negative attitudes are presented in (60-65). Intensifiers are underlined.

(60) Ms. Karr, who works at the House Education and the Workforce Committee, described Carlyle as the "best of the best." (++)

(61) Germans are overwhelmingly opposed (--) to military action in Syria.

(62) A flustered Mr Cameron yesterday said it was "completely unacceptable" (--) and ordered an internal party inquiry.

(63) But Israel said the agreement was an "historic mistake" (--).

(64) "I deeply regret (--) any impression of impropriety arising from my bluster in that conversation."

(65) "People are sick and tired of paying over the odds."

3.4 Tag levels in KAF Annotation Tool

3.4.1 Tag levels 1 and 2

As with the review annotations, Tag level 1 and 2 are used for annotating and linking the three components of an opinion triple. Tag level 1 is used for the tagging of the three components of an opinion triple. For the opinion expressions, you first have to choose whether it is a direct or an indirect attitude expression. Then you choose between the labels Negative, Positive, StrongNegative and StrongPositive. For the direct attitude expressions you can also choose the Neutral label (as explained in 3.3.2). Tag level 2 contains two labels. The 'Opinion' is the one you will use most: this label is used for linking the separate components of an opinion together. The 'OH1-OP' label is used for sequences of opinions with one opinion holder; how you can use this label is explained in Section 4.3.
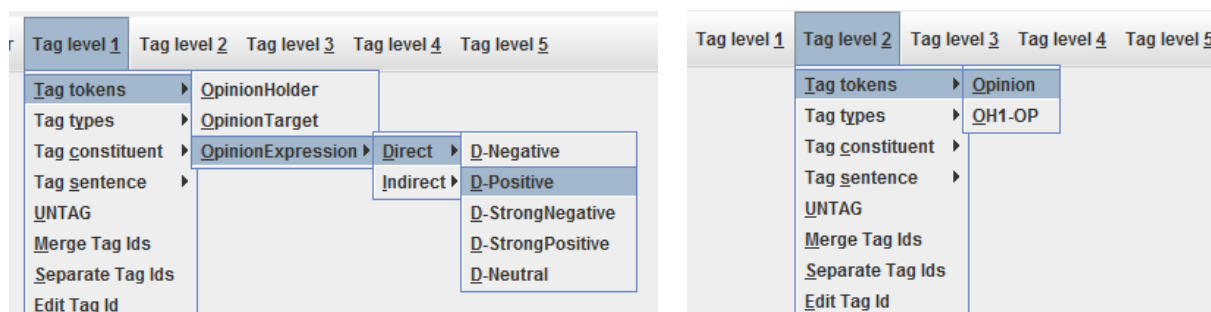
Figure 4: Tag level 1 and 2

For tagging an

opinion you will thus use both Tag level 1 and Tag level 2. First annotate the separate components of the opinion in Tag level 1. Then select the span of words that form the opinion as a whole and use the tag 'Opinion' or 'OH1-opinion' from Tag level 2 to link them together. Figure 5 shows an example of an annotated sentence.

Id	Word token	Pos	Tag1	Tag1 id	Tag2	Tag2 id
w234	"	N		0	Opinion	5
w235	I	Q	OpinionHolder	1	Opinion	5
w236	am	V	D-StrongPos...	2	Opinion	5
w237	very	A	D-StrongPos...	2	Opinion	5
w238	excited	G	D-StrongPos...	2	Opinion	5
w239	about	P	D-StrongPos...	2	Opinion	5
w240	the	D	OpinionTarget	4	Opinion	5
w241	prospects	N	OpinionTarget	4	Opinion	5
w242	for	P	OpinionTarget	4	Opinion	5
w243	these	D	OpinionTarget	4	Opinion	5
w244	negotiations	N	OpinionTarget	4	Opinion	5
w245	,	O		0	Opinion	5
w246	"	N		0	Opinion	5

Figure 5: Example of annotated sentence

3.4.2 Tag levels 3 and 4

Tag level 3 and 4 are a repetition of Tag level 1 and 2 and contain the same labels. These levels are only used for the annotation of more complicated opinions, in which Tag level 1 and 2 are not sufficient enough. This may be the case for the annotation of a sentence with multiple holders, target and/or expressions (not always) and for the annotation of nested opinions. How to annotate these more complex opinions and how to use Tag level 3 and 4 will be explained in the next Section.

4 Annotation of complicated opinions

4.1 Multiple expressions, holders or targets: Tag level 1 and 2

An opinion may consist of multiple expressions, holders or targets. For example, in (66) there is one opinion holder (*several other countries*), one opinion target (*an increase in the bloc's spending plan*), and the negative opinion of the holder towards this target is expressed by two separate negative expressions (*the concerns of* and *do not want*).

> (66) Cameron voiced the concerns of several other countries that do not want to see an increase in the bloc's spending plan at a time when many member states are cutting budgets at home.

In cases like this, all of the components are all linked together with the "Opinion" tag from Tag level 2. However, make sure that in Tag level 1 the components get their own tag id. So *the concerns of* gets another tag id than *do not want*. However, they are part of the same opinion, so they have the same tag id at Tag level 2. See Figure 6 for the correct annotation of this sentence.

Id	Word token	Pos	Tag1	Tag1 id	Tag2	Tag2 id
w196	Cameron	R		0	Opinion	6
w197	voiced	V		0	Opinion	6
w198	the	D	D-Negative	2	Opinion	6
w199	concerns	N	D-Negative	2	Opinion	6
w200	of	P	D-Negative	2	Opinion	6
w201	several	G	OpinionHolder	3	Opinion	6
w202	other	G	OpinionHolder	3	Opinion	6
w203	countries	N	OpinionHolder	3	Opinion	6
w204	that	O		0	Opinion	6
w205	do	V	D-Negative	4	Opinion	6
w206	not	A	D-Negative	4	Opinion	6
w207	want	V	D-Negative	4	Opinion	6
w208	to	P		0	Opinion	6
w209	see	V		0	Opinion	6
w210	an	D	OpinionTarget	5	Opinion	6
w211	increase	N	OpinionTarget	5	Opinion	6
w212	in	P	OpinionTarget	5	Opinion	6
w213	the	D	OpinionTarget	5	Opinion	6
w214	bloc	N	OpinionTarget	5	Opinion	6
w215	's	O	OpinionTarget	5	Opinion	6
w216	spending	N	OpinionTarget	5	Opinion	6
w217	plan	N	OpinionTarget	5	Opinion	6
w218	at	P		0	Opinion	6
w219	a	D		0	Opinion	6
w220	time	N		0	Opinion	6

Figure 6: Annotation of a sentence with multiple opinion expressions

4.2 Multiple expressions, holders or targets: Tag level 1, 2, 3 and 4

However, there may be sentences in which you need an additional level, because some elements are shared in two separate opinions. This is illustrated in (67) and (68). In (67), there is one target (*the healthcare law*), but two opinion holders that have their own attitude towards this target. For the *54 percent* this is a negative attitude (*disapprove*), and for the *38 percent* this is a positive attitude (*approve*). Another example of such a sentence is (68). In this case, there are two opinion holders (*Germany* and *the European Union*) that express a negative opinion about one target (*blames the Syrian government for the attack*), and a positive opinion about another target (*urged waiting for a report from UN weapons inspectors)*. Especially this situation, with a sequence of opinions of one opinion holder, occurs very often.

(67)　The survey found that 54 percent disapprove of the healthcare law, against only 38 percent who approve.

(68)　Germany, along with the European Union, blames the Syrian government for the attack but urged waiting for a report from UN weapons inspectors before any US-led military response .

This example is different from (66) (which also contained two opinion expressions) in that the two opinion expressions both have their own target, and each attitude is only towards its 'own' target and not towards the other. More specifically: the Syrian government is blamed, but the waiting for a report from the UN weapons inspectors is not. At the same time, the opinion holders urge waiting for a report from UN weapons inspectors, but not the Syrian government. Therefore they cannot be annotated as one opinion, as in Figure 7.

Id	Word token	Pos	Tag1	Tag1 id	Tag2	Tag2 id
w335	Germany	R	OpinionHolder	1	Opinion	9
w336	,	O		0	Opinion	9
w337	along	P		0	Opinion	9
w338	with	P		0	Opinion	9
w339	the	D	OpinionHolder	2	Opinion	9
w340	European	R	OpinionHolder	2	Opinion	9
w341	Union	R	OpinionHolder	2	Opinion	9
w342	,	O		0	Opinion	9
w343	blames	V	D-Negative	3	Opinion	9
w344	the	D	OpinionTarget	4	Opinion	9
w345	Syrian	G	OpinionTarget	4	Opinion	9
w346	government	N	OpinionTarget	4	Opinion	9
w347	for	P		0	Opinion	9
w348	the	D		0	Opinion	9
w349	attack	N		0	Opinion	9
w350	but	C		0	Opinion	9
w351	urged	V	D-StrongPos...	5	Opinion	9
w352	waiting	V	OpinionTarget	6	Opinion	9
w353	for	P	OpinionTarget	6	Opinion	9
w354	a	D	OpinionTarget	6	Opinion	9
w355	report	N	OpinionTarget	6	Opinion	9
w356	from	P	OpinionTarget	6	Opinion	9
w357	UN	R	OpinionTarget	6	Opinion	9
w358	weapons	N	OpinionTarget	6	Opinion	9
w359	inspectors	N	OpinionTarget	6	Opinion	9
w360	before	P		0		0

Figure 7: How not to annotate this sentence: one opinion at Tag level 1

However, annotating two different opinions at Tag level 2 in the 'normal' way is not satisfying either, since (both) the opinion holders are part of both opinions but this is not reflected in the annotations this way.

Id	Word token	Pos	Tag1	Tag1 id	Tag2	Tag2 id
w334	.	O		0		0
w335	Germany	R	OpinionHolder	1	Opinion	9
w336	,	O		0	Opinion	9
w337	along	P		0	Opinion	9
w338	with	P		0	Opinion	9
w339	the	D	OpinionHolder	2	Opinion	9
w340	European	R	OpinionHolder	2	Opinion	9
w341	Union	R	OpinionHolder	2	Opinion	9
w342	,	O		0	Opinion	9
w343	blames	V	D-Negative	3	Opinion	9
w344	the	D	OpinionTarget	4	Opinion	9
w345	Syrian	G	OpinionTarget	4	Opinion	9
w346	government	N	OpinionTarget	4	Opinion	9
w347	for	P		0	Opinion	9
w348	the	D		0	Opinion	9
w349	attack	N		0	Opinion	9
w350	but	C		0		0
w351	urged	V	D-StrongPos...	5	Opinion	10
w352	waiting	V	OpinionTarget	6	Opinion	10
w353	for	P	OpinionTarget	6	Opinion	10
w354	a	D	OpinionTarget	6	Opinion	10
w355	report	N	OpinionTarget	6	Opinion	10
w356	from	P	OpinionTarget	6	Opinion	10
w357	UN	R	OpinionTarget	6	Opinion	10
w358	weapons	N	OpinionTarget	6	Opinion	10
w359	inspectors	N	OpinionTarget	6	Opinion	10
w360	before	P		0		0

Figure 8: How not to annotate this sentence: two opinions at Tag level 2

Therefore, one opinion should be annotated at Tag level 1 and 2, and the other one at Tag level 3 and 4. This way, the element (in this case, the opinion holder) that is a functional part of two different opinions is also annotated twice in two separate opinions. See Figure 9 for the illustration of the annotation of this sentence. Sentence (67) could be done in a similar way, with the target instead of the holder as the functional part of two different opinions.

Id	Word token	Pos	Tag1	T...	Tag2	Ta...	Tag3	Tag3 id	Tag4	Tag4 id	T.
w_321	.Germany	R	OpinionHolder	6	Opinion	10	OpinionHolder	11	Opinion	19	
w_322	,	O		0		0		0		0	
w_323	along	A		0		0		0		0	
w_324	with	P		0		0		0		0	
w_325	the	D		0		0	OpinionHolder	12		0	
w_326	European	R	OpinionHolder	7	Opinion	10	OpinionHolder	12	Opinion	19	
w_327	Union	R	OpinionHolder	7	Opinion	10	OpinionHolder	12	Opinion	19	
w_328	,	O		0		0		0		0	
w_329	blames	V	D–Negative	8	Opinion	10		0		0	
w_330	the	D	OpinionTarget	9	Opinion	10		0		0	
w_331	Syrian	G	OpinionTarget	9	Opinion	10		0		0	
w_332	government	N	OpinionTarget	9	Opinion	10		0		0	
w_333	for	P		0		0		0		0	
w_334	the	D		0		0		0		0	
w_335	attack	N	D–Negative	21	Opinion	20		0		0	
w_336	but	C		0		0		0		0	
w_337	urged	V		0		0	D–Positive	16	Opinion	19	
w_338	waiting	V		0		0	OpinionTarget	18	Opinion	19	
w_339	for	P		0		0	OpinionTarget	18	Opinion	19	
w_340	a	D		0		0	OpinionTarget	18	Opinion	19	
w_341	report	N		0		0	OpinionTarget	18	Opinion	19	
w_342	from	P		0		0		0		0	

Figure 9: How to annotate this sentence: two opinions at Tag level 1 ,2, 3 and 4

4.3 Dealing with ellipsis of the opinion holder in complex clauses: another solution

As mentioned in the previous section, many sentences are sequences of opinions of one opinion holder. However, this opinion holder is often only explicitly mentioned once. The following sentence is an even more extreme example:

(69) EU foreign ministers meeting in the Lithuanian capital Vilnius on Saturday urged "a clear and strong response " to the attack , saying there was "strong evidence that the Syrian regime is responsible " but stopping short of endorsing a strike on the regime.

In this example, *the EU foreign ministers* is the opinion holder of 4 different opinions:

a. Urge a strong and clear response
b. The Syrian regime is responsible
c. Stopping short of endorsing
d. endorsing a strike

A sentence like this cannot be solved as explained in Section 4.2, because you would need even more Tag levels. Therefore, to be able to annotate this in an easy way, the possible tags of Tag level 2 (and Tag level 4, if needed) include a label 'OH1-op'. The opinion holder of the opinion labelled with 'OH1-op', is identical to the first opinion holder of the sentence. Figure 10 (next page) shows the annotation of this sentence using the 'OH1-op' label. The reason that in this sentence you still need to use Tag level 3 and 4 is because there is a nested opinion present in the phrase *stopping short of endorsing a strike*; this is explained in the next Section.

This solution is especially useful for sentences in which there are more than two opinions linked to one opinion holder, but of course you may also annotate sentences like (68) this way.

Id	Word token	Pos	Tag1	Tag1 id	Tag2	Tag2 id	Tag3	Tag3 id	Tag4	Tag4 id
w_48	EU	R	OpinionHolder	17	Opinion	20		0		0
w_49	foreign	G	OpinionHolder	17	Opinion	20		0		0
w_50	ministers	N	OpinionHolder	17	Opinion	20		0		0
w_51	meeting	N		0	Opinion	20		0		0
w_52	in	P		0	Opinion	20		0		0
w_53	the	D		0	Opinion	20		0		0
w_54	Lithuanian	G		0	Opinion	20		0		0
w_55	capital	N		0	Opinion	20		0		0
w_56	Vilnius	R		0	Opinion	20		0		0
w_57	on	P		0	Opinion	20		0		0
w_58	Saturday	R		0	Opinion	20		0		0
w_59	urged	V	D–Positive	18	OH1–OP	39		0		0
w_60	"a	D	OpinionTarget	19	OH1–OP	39		0		0
w_61	clear	G	OpinionTarget	19	OH1–OP	39		0		0
w_62	and	C	OpinionTarget	19	OH1–OP	39		0		0
w_63	strong	G	OpinionTarget	19	OH1–OP	39		0		0
w_64	response	N	OpinionTarget	19	OH1–OP	39		0		0
w_65	"	O		0		0		0		0
w_66	to	O		0		0		0		0
w_67	the	D		0		0		0		0
w_68	attack	N	D–Negative	21	Opinion	22		0		0
w_69	,	O		0		0		0		0
w_70	saying	V		0		0		0		0
w_71	there	O		0		0		0		0
w_72	was	V		0		0		0		0
w_73	"strong	G		0		0		0		0
w_74	evidence	N		0		0		0		0
w_75	that	P		0		0		0		0
w_76	the	D	OpinionTarget	23	OH1–OP	41		0		0
w_77	Syrian	G	OpinionTarget	23	OH1–OP	41		0		0
w_78	regime	N	OpinionTarget	23	OH1–OP	41		0		0
w_79	is	V		0	OH1–OP	41		0		0
w_80	responsible	G	I–Negative	24	OH1–OP	41		0		0
w_81	"	O		0		0		0		0
w_82	but	C		0		0		0		0
w_83	stopping	V	D–Negative	26	OH1–OP	40		0		0
w_84	short	G	D–Negative	26	OH1–OP	40		0		0
w_85	of	P	D–Negative	26	OH1–OP	40		0		0
w_86	endorsing	V	OpinionTarget	27	OH1–OP	40	D–Positive	28	Opinion	30
w_87	a	D		0		0	OpinionTarget	29	Opinion	30
w_88	strike	N	D–Negative	33	Opinion	32	OpinionTarget	29	Opinion	30
w_89	on	P	OpinionTarget	34	Opinion	32		0		0
w_90	the	D	OpinionTarget	34	Opinion	32		0		0
w_91	regime	N	OpinionTarget	34	Opinion	32		0		0
w_92	.	O		0		0		0		0

Figure 10: The correct annotation of this sentence

4.4 Nested opinions

Opinions can be nested, that is, one opinion may be expressed *within* another opinion. Because news articles report about many different perspectives, these nested opinions occur quite often. More specifically, these nested opinions can be found in the *target* of the higher-level opinion. For example, in (70) there is one opinion with *agreement* as the target, "*historic mistake"* as strong negative expression, and *the agreement* as the target. The target itself, however, is some positive expression as well.

   (70) But Israel said the agreement was an "historic mistake".

In cases like this, some words serve several functions. *Agreement* is both part of the target of the higher-level opinion, but a positive expression in the lower-level opinion. Since the Annotation Tool does not allow for multiple annotations within one tag level, we need another level to annotate the other functions of the words; Tag level 3 is used for this annotation of the target, holder and expression in nested opinions. Tag level 4 is used to link the components of the nested opinions together with the tag 'Opinion'. See Figure 11 for the correct annotation of (70).

Id	Word token	Pos	Tag1	Tag1 id	Tag2	Tag2 id	Tag3	Tag3 id	Tag4	Tag4 id
w122	.	O		0		0		0		0
w123	But	C		0	Opinion	5		0		0
w124	Israel	R	OpinionHolder	2	Opinion	5		0		0
w125	said	V		0	Opinion	5		0		0
w126	the	D	OpinionTarget	3	Opinion	5		0		0
w127	agreement	N	OpinionTarget	3	Opinion	5	D-Positive	8	Opinion	9
w128	was	V		0	Opinion	5		0		0
w129	an	D		0	Opinion	5		0		0
w130	"	G	I-StrongNegative	4	Opinion	5		0		0
w131	historic	G	I-StrongNegative	4	Opinion	5		0		0
w132	mistake	N	I-StrongNegative	4	Opinion	5		0		0
w133	"	N	I-StrongNegative	4	Opinion	5		0		0
w134	.	O		0	Opinion	5		0		0

Figure 11: How to annotate a nested opinion

4.5 Be specific!

As mentioned in Section 3, it is important to annotate the holder, target and expressions as specific as possible. That is: do only annotate the words that 'really matter' and leave the words that are less important unannotated. For the holders and targets, this will often mean that you only annotate the head of the NP instead of the whole phrase; for expressions, this means that you annotate only the words that carry the positive or negative attitude or are enough to understand that it is positive or negative. So for example, in (71) *the plan* is annotated as the target, instead of *the plan to double the sales tax to 10 percent over three years*. For the (indirect) positive expression, only *a first step* is annotated, instead of *a first step towards curbing Japan's snowballing public debt.*

(71) The plan to double the sales tax to 10 percent over three years is seen as a first step towards curbing Japan's snowballing public debt.

Not only is this way of annotating better for the machine learning, it can also be very useful in nested sentences like (72). In this sentence, there is a positive attitude expressed towards the target *block the absurd use of force*. However, in this opinion there is a nested opinion, and even within the nested opinion there is another nested opinion. That is: within the target of the highest-level opinion (*block the absurd use of force*), *block* expresses a negative attitude towards the target *absurd use of force*. But within this target, *absurd* expresses a negative attitude towards *use of force*.

(72) Assad's government denies carrying out a chemical assault, and has urged the United Nations Security Council to [block the "[absurd [use of force]],"] according to state-run news agency ANA.

To solve this, we annotate *to block* as the target of the opinion expression *has urged to* (instead of the whole VP *to block the absurd use of force*) in Tag level 1 and 2, which makes it possible to annotate *the absurd use of force* as a separate opinion in Tag level 3 and 4. The resulting annotation of this sentence is shown in Figure 12.

Id	Word to...	Pos	Tag1	Tag...	Tag2	Tag...	Tag3	...	Tag4	T...	...
w_94	Assad	R	OpinionHolder	1	Opinion	5	OpinionHolder	6	Opinion	9	
w_95	's	O	OpinionHolder	1	Opinion	5	OpinionHolder	6	Opinion	9	
w_96	govern...	N	OpinionHolder	1	Opinion	5	OpinionHolder	6	Opinion	9	
w_97	denies	V	D–Negative	2	Opinion	5		0		0	
w_98	carrying	V	OpinionTarget	3	Opinion	5		0		0	
w_99	out	O	OpinionTarget	3	Opinion	5		0		0	
w_100	a	D		0		0		0		0	
w_101	chemical	N	I–StrongNegati...	11	Opinion	12		0		0	
w_102	assault	N	I–StrongNegati...	11	Opinion	12		0		0	
w_103	,	O		0		0		0		0	
w_104	and	C		0		0		0		0	
w_105	has	V		0		0		0		0	
w_106	urged	V		0		0	D–Positive	7	Opinion	9	
w_107	the	D		0		0		0		0	
w_108	United	R	OpinionHolder	16	Opinion	17		0		0	
w_109	Nations	R	OpinionHolder	16	Opinion	17		0		0	
w_110	Security	R	OpinionHolder	16	Opinion	17		0		0	
w_111	Council	R	OpinionHolder	16	Opinion	17		0		0	
w_112	to	O		0	Opinion	17	OpinionTarget	8	Opinion	9	
w_113	[block	V	D–Negative	13	Opinion	17	OpinionTarget	8	Opinion	9	
w_114	the	D	OpinionTarget	14	Opinion	17		0		0	
w_115	"	O	OpinionTarget	14	Opinion	17		0		0	
w_116	[absurd	G	OpinionTarget	14	Opinion	17	I–StrongNegative	...	Opinion	20	
w_117	use	N	OpinionTarget	14	Opinion	17	OpinionTarget	...	Opinion	20	
w_118	of	P	OpinionTarget	14	Opinion	17	OpinionTarget	...	Opinion	20	
w_119	force]	O	OpinionTarget	14	Opinion	17	OpinionTarget	...	Opinion	20	
w_120	,	O		0		0		0		0	
w_121	"	O		0		0		0		0	
w_122	]	O		0		0		0		0	

Figure 12: Example of the annotation of a nested opinion: annotating only the heads of phrases

REFERENCES

- Janyce Wiebe, Theresa Wilson, and Claire Cardie (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation, volume 39, issue 2-3,* pp. 165-210.
- Theresa Wilson (2008). Fine-Grained Subjectivity Analysis. PhD Dissertation, Intelligent Systems Program, University of Pittsburgh.

Confidentiality: EC Distribution