

# Guidelines for sentence and comment level annotations for health care feedback

LTG, FHI

## 1 Introduction

This document represents the guidelines for the annotation project undertaken as part of the FHI-SA/SANT-PREMS project. As the first of two annotation efforts we will annotate comments and sentences with polarity, relevance and comparatives. This document only describes the first, sentence- and comment-level annotations.

## 2 Background and Annotation Categories

In this section we will describe the four parts of this annotation procedure with examples. The guidelines for annotating sentence and comment level polarity and intensity are the same, but comments are not annotated for relevancy and comparatives.

### 2.1 Opinions and Polar Expressions

Before focusing on the four main parts of the annotation, we will have a quick but a bit more detailed look at how opinions (sentiment) behaves. Although the annotations described in this document are at the sentence and comment level, and do not require us to identify the following terms, they are useful when looking at relevant categories such as Relevance. An **opinion** is a reflection of some inner state, i.e. the unobtainable thoughts and feelings of a person. This reflection will in our case be represented by written words. People write feedback to indicate what they are thinking, but much can be lost when it is written, and we must remember that the written word is not always a true indicator of a person's opinion. The most common indicator of an opinion is what we call a **polar expression**. This expression is usually a common indicator of polarity such as *hate*, *love*, *like*, *good*, *bad*, etc. Words like these, which are often inherently understood as positive or negative, are called *sentiment lexemes*, or simply, *sentiment words*. It is however not always the case that an opinion

has explicit sentiment lexemes. **Experiences** and even **qualitative observations** can be understood as indicating opinions. When annotating, the annotator should consider the sentence as a whole, rather than looking for specific expressions, and annotate according to what impression the sentence gives. An opinion also always has a **target**, which can be seen as the goal of the polar expression, or what the opinion is about. Some times the target is not mentioned explicitly, it can also be implicit. In the context of patient reviews, the target is usually the GP, center of treatment or other people or organizations related to the patient’s healthcare. An important note about the target, is that if it is *not* the expected one, then this has to do with the Relevancy of the document. See details below. Finally, a polar expression has a **holder**, or **source**, which is the person who has the inner state we are trying to describe. The holder is often expressed by pronouns such as “Jeg” or “Vi”, but is also very often not mentioned explicitly. In the context discussed here the holder is the patient, or in some cases people speaking on behalf of the patient. Note that while perhaps not common in this domain, it is possible for someone to refer to other people’s opinions, in which case the holder is not the same as the writer, as in the example in 1. This is not marked in our case. To sum up, we can say that when looking at sentiment analysis, we try to look for opinions, which consists of some words that give us the feeling that they are positive or negative, written by someone who holds these opinions, towards some person, object or quality which causes these opinions in the person writing them.

- (1) *Og det er sikkert en del mennesker som foretrekker muntlig*  
 And there are surely a part people who prefer oral  
*kommunikasjon fremfor skriftlig, [...]*  
 communication rather than written, [...]  
 ‘And I am sure there are several people who prefer oral communication to  
 written [communication], [...]

Below we will look more closely at some examples of different types of polar expressions, but note that the lines between some of these types can be blurry, and since we do not indicate polar expression type, they are just meant as guides for what to look for when annotating.

### 2.1.1 Typical Sentiment Lexemes

Words that we normally perceive as positive or negative are usually clear indicators of opinions. For example, in sentence 2, we see the word *fantastisk* ‘fantastic’ is used towards the target *lege* ‘doctor’. *Fantastisk* is a word with typical positive associations, and it is a quite clear case, we believe.

- (2) *Fantastisk lege som jeg har hatt hele livet.*  
 fantastic medical.doctor whom I have had whole life.the.  
 ‘(My GP is a) Fantastic doctor whom I have had my entire life.’

### 2.1.2 Experiences

An experienced opinion is a description of something the holder has experienced, for example over time. In the sentence 3, we see that the source describes their experiences in *aldri hadde tid*, ‘never had the time’. None of the words here *aldri*, *hadde*, *tid* are sentiment lexemes, but together they indicate the source’s experience, and based on what we know of the situation, we understand that this reflects negatively on the specialist, indicating an experience-based polar expression.

- (3) *Jeg fikk en henvisning til spesialist som aldri hadde tid .*  
 I got a referral to specialist who never had time .  
 ‘I received a referral to a specialist who never had the time (for me).’

Typical types of experience-related polar expressions are indicated by temporal adverbs such as *ofte* ‘often’, *aldri* ‘never’, *så langt* ‘so far’, etc.

### 2.1.3 Objective Descriptions

While the comments are likely to contain more subjective opinions, it is not uncommon to see that objective descriptions are used to indicate opinions. If this is the case, it is usually understandable from context whether this objective description should be interpreted as polar or not. By definition, objective descriptions do not contain sentiment lexemes, i.e. words that usually clearly indicate polarity. In sentence 4, the sentence states a fact about the doctor’s habit of calling.

- (4) *Legen ringer meg når prøvesvar foreligger .*  
 Doctor.the calls me when test-results are.ready .  
 ‘The doctor calls me when (my) test results are ready.’

From context we can chose to interpret this action as favourable, as we like to be informed when test-results are ready. Correct understanding of this type of polar expression is dependent on domain knowledge, and can be subject to discussion. We urge annotators to have the patient in mind, and think about the following: What is the patient to gain from writing this? What is *not* said? We usually assume that there is intent behind sentences written, and it can be helpful to assume that if a seemingly objective fact reflects positively on the target, then it is likely a reflection of the holder’s opinion, and should be annotated as such.

## 2.2 Polarity

Polarity refers to something being **positive** or **negative**. A sentence can be void of polarity, in which case it is **neutral**. Note that positive and negative opinions cannot cancel each other out; A sentence containing both positive and negative opinions is classed as both positive and negative, in the way described below. Some times, this category is referred to as **mixed**. A sentence is annotated with positive polarity if it contains one or more positive opinions, and it is annotated as negative if it contains one or more negative opinions. The annotator does not need to be able to identify a specific polar expression or polar lexeme in the sentence to classify it as positive or negative. If the annotator feels that the patient (or other person, person speaking on behalf of the patient, etc.) writes something that feels positive or negative, it should be annotated as such. Care should be made to see each sentence and the words in it in context. Even if a sentence feels positive in one case, it can express something negative in a different context. A single opinion cannot have more than one polarity, but the same words can be included in several different opinions (see for example comparative expressions further down).

## 2.3 Intensity

We employ three levels of intensity in this annotation project: **slight**, **standard** and **strong**. Most polarities should be standard. A slight polarity could be an objective fact that is only slightly positive, or it might be a polar expression preceded by some sort of **hedge** expression, meaning expressions that make it less strong. Examples of this are *litt*, *noe,ikke så*, etc. Strong intensity is the opposite of slight intensity. It is often preceded by intensifiers such as *veldig,utrolig, helt*, etc. However, note that some expressions can be seen as inherently slight or strong. Typically slight polar expressions might include words such as *ok,grei, passe*, while inherently strong ones might include *fantastisk,hate,elske*, but note that for either end of the polarity scale, these words should be seen and treated in the context they appear in. As some multi-opinion sentences are unavoidable, the intensity should cover all opinions of the polarity it indicates in those cases. For example, given a sentence with two positive polar expressions and one negative, the annotator will have to decide on the intensity best fitting of the two positive expressions for positive polarity, but only on the single negative one for the negative polarity. For example in example 5 below, we see one standard positive and a slight negative expression.

- (5) *Fastlegen min er snill , men kommer noen ganger for seint.*  
GP.my is nice , but comes some times too late.  
'My GP is nice, but some times (s)he comes too late.'

Below are some examples of different intensities for different polarities.

Sentence 6 shows a strong positive evaluation. We see that words with very positive associations such as *grundig* ‘thorough’ are used, and we also see another positive word *seriøs* ‘serious’ with an intensifier *veldig* ‘very’ modifying it.

- (6) *Fastlegen min er grundig , forklarer ulike*  
 GP.my is thorough , explains different challenges  
*utfordringer på en god måte og er veldig seriøs i*  
 on a good way and is very serious in work.the  
*arbeidet sitt*  
 his  
 ‘My GP is thorough, explains different challenges in a good way and is very serious in his work.’

In example 7 we see a softer evaluation. *fornøyd* is a quite positive word, but the adverb *generelt* ‘generally’ restricts its meaning somewhat. This gives us a standard polarity.

- (7) *Generelt er jeg fornøyd*  
 generally am I pleased  
 ‘Generally I am pleased.’

In this last positive example 8 we see the patient describe a bare minimum of what is expected. We also see the expression  *greit nok* ‘ok enough/fair enough, just so so’, indicating a positive evaluation, but a slight one.

- (8) *Det er " greit nok " i og med at jeg får de tjenestene*  
 it is " ok enough " in and with that I get those services  
*jeg trenger , hva angår sykemeldinger , resepter og annen*  
 I need , what concerns sick.leave , prescriptions and other  
*medisinsk behandling*  
 medical treatment.  
 ‘It’s “ok enough”, given that I get the services I need, concerning sick leave, prescriptions and other medical treatment.’

Sentence

- (9) *Når du kommer er han alltid forsinket , flere ganger opp*  
 When you come is he always delayed , several times up  
*til 45 minutter med venting*  
 until 45 minutes with waiting  
 ‘When you come, he is always late , several times for as much as 45 minutes of waiting.’

- (10) *Får følelsen av at de ikke ønsker at du ringer (*  
 Get feeling.the of that they not want that you call ( *sekretærene ) .*  
 secretaries.the)  
 ‘I get the feeling that they do not want you to call (the secretaries)’
- (11) *Jeg har flere ganger fått vondt av henne .*  
 I have several times got bad of her.  
 ‘I’ve felt bad for her several times.’

## 2.4 Relevancy and Questionnaire-related Information

A common phenomena in patient feedback comments is the tendency to include comparisons and feedback to related health services, earlier general practitioners, or other types of information that the patient feels is relevant. If all polar expressions of one polarity (positive or negative) are about something other than the main topic of a comment, then they are marked as irrelevant. That is to say, that if the *target* is something other than the main topic, then it is irrelevant. The topic of the comment or sentence should be understood as the topic being evaluated by the questionnaire where it comes from. Irrelevant in this sense only means irrelevant to the domain, for example *general practitioner* or *psychiatric care*, it is still relevant to the annotation project. If a sentence contains one relevant and one irrelevant opinion of the same polarity, then it is **not** marked as irrelevant, regardless of whether it is in relation to a polar expression or not. Irrelevant should be understood as irrelevant to the topic of the comment. We distinguish between two different types of irrelevant information: information related to other or earlier health services, for which we use the label ‘i’, and polar expressions or information related to the questionnaire or any of the data collection processes themselves, for which we use ‘q’. For example, GP comments should be about the patient’s current GP office, the GP or current substitute GPs, and the services provided at the office. The labels are usually used in combination with a polar annotation, giving one of the following: si,ssi,sssi. It is possible to add them to the ‘other’ column, but this can only be done if the sentence contains descriptions of these irrelevant cases, without any polarity. An example of the use of ‘q’ can be found in example 12. Here we see that the patient refers to a specific question in the questionnaire, and then they give some reason for their insecurities. Since this is about the questionnaire, and since the sentence was otherwise not found to contain any polarity, a ‘q’ is added to the “Other” column. If irrelevant or questionnaire-related themes are tied to a specific polar expression, it is also possible to add “i” or “q” together with “s,ss or sss” in the polarity columns.

- (12) *Usikker på svaret i spm 48 , jeg har nettopp fått ny fastlege (dato) og har ikke vært i kontakt med min nye fastlege*  
 un-sure on reply.the in question 48 , I have just received new GP (date) and have not been in contact with my new GP  
 ‘Not sure about my reply in question 48, (as) I have just received a new GP and have not been in contact with my new GP.’

## 2.5 Comparatives

Comparative expressions can be difficult in a sentiment analysis setting, as the interpretation of the expression is always dependent on some (usually) unknown polarity of that with which the target is being compared. If any opinion in a sentence contains a comparative expression, then the sentences (not the individual opinions) are marked as containing a comparative expression. A comparative always consists of a comparison between two elements A and B, but it is possible for one of the elements to be implicit (known from context). Very often the comparisons are of opposing polarities, but it is also possible to have comparisons between elements of the same polarity, but different intensities. Comparative expressions always contain a comparative element, but there are many different possible elements. The most common ones are comparative subjunctions such as *enn* ‘than’, comparative adjective or adverbial forms, certain comparing lexical items such as *foretrekke* ‘prefer’ or adverbs such as *heller* ‘rather’, *enda* ‘even’ or *mer* ‘more’, as in examples 13 and 14.

- (13) *Andre Turnus leger er mye bedre enn fastleger på alle måter*  
 other on.duty doctors are much better than GPs in all ways  
 ‘Other doctors on duty are much better than GPs in all respects.’
- (14) *Jeg foretrekker å gå til min egen fastlege*  
 I prefer to go to my own GP  
 ‘I prefer going to my own GP’

When it comes to analysing comparative expressions for polarity and intensity, context is important. As with objective statements, we assume that the holder has done the comparison for some reason. It can be difficult to interpret the absolute polarity that was intended, but usually we are able to identify to some degree that what is compared is more or less positive or negative than what it is compared to. Note that although usually there will be a positive and a negative part, this is not always the case. For example, expressions with *enda* ‘even’ might just indicate different degrees of the same polarity, as in the following artificial example 15.

- (15) *Denne fastlegen er enda verre enn den forrige fastlegen min.*  
 This GP.the is even worse than the previous GP.the mine.  
 ‘This GP is even worse than my last GP’

In this example, both polar expressions must be negative, as *enda* indicates that the polarity is the same, but the degree is stronger.

## 2.6 Summary of Placement of k, q and i

In the case of “q” and “i”, the annotator must decide whether these themes are tied to a specific polar expression. If a sentence contains no sentiment, but is irrelevant or contains comments on the questionnaire, it receives i or q, respectively. However, if there is one or more polar expression, then the annotator must decide whether the irrelevant or questionnaire-related themes are tied to any of the polar expressions or not. It is possible to have a sentence that contains comments on irrelevant themes or the questionnaire, but where these comments are not linked to the polar expression(s). In this case, the relevant letter is only placed in the “Other” column.

# 3 Annotation Procedure and Format

## 3.1 Extra information

The original comments are anonymized by NIPH before they are sentence tokenized using Stanza. The tokenized sentences are each given one line in a document in the .tsv format, where the first column holds the sentences, and the four next columns are used for annotations and comments. The first column is used to annotate positive polarity, the second to annotate negative polarity, the third (called “Other”) is used to indicate the three labels: k for comparatives, i for irrelevant information and q for questionnaire-related information. The final column (called “Comments”) is used for comments related to annotation difficulties or problems with sentence tokenization. Note that the columns containing the text must not be modified in any way, as this might lead to issues with postprocessing.

The degrees of polarity are indicated as follows: A polarity of standard degree is indicated with a double ‘s’ (ss), a polarity of strong degree is indicated with three ‘s’ (sss), while a polarity of slight degree is indicated with a single ‘s’ (s). Irrelevant polarities are indicated with the letter ‘i’ or ‘q’, depending on the type, always directly following the degree indicators in the appropriate column. The presence of a comparative expression is indicated in the third column by the letter ‘k’. If a sentence does not contain any polarity, or if it is lacking polarity of one type but not



the other, the column not containing polarity is marked with the letter ‘b’, for *blank*. If a sentence contains no polarity, the labels i and q can only be placed in the third column. In addition to these, annotators can write questions or comments related to annotation, or problems with preprocessing, in a final column. This must be added by the annotators themselves, after the third column.

## 3.2 Context

When giving evaluations to a sentence, it is recommended to first look at the sentence by itself. If the polarity is understandable, then it can be annotated without taking the context into account. However, if there are any words, phrases, facts or other elements that cannot be evaluated as positive, negative or neutral in the context of the sentence itself, the annotator is advised to look at the whole comment (or the sentences before or after), to understand what the patient meant. The context should not influence a sentence if does not need it. Note that some expressions can be interpreted as both positive and negative in the same context and in the same sentence. In these cases, both positive and negative polarity should be added. If you are unsure, it is better to add polarity than to not add it.

## 3.3 Referring expressions and context dependence

In some cases, a sentence uses a referring expression that refers back to a polar expression of known polarity. Words such as *det*, *sånn* and *slik* are examples of this. In example 16, we see that the polarity of the second sentence is reliant on the interpretation of the anaphoric “slik”. In these cases, we should annotate the polarity as it is, even though the actual polar expression is unavailable in the sentence in question. This is also the case if a word can only be interpreted in the context it is in.

- (16) *Hos den forrige fastlegen min var det mye uro. Slik*  
 At previous GP mine was it much unrest. Like.this is  
*er det også her.*  
 it also here.  
 ‘There was a lot of unrest at my previous GP. It’s like that here as well.’

This is the case even if we can see that the expression in question could be interpreted differently in a different context. The assumed inner state of the patient takes precedence over other possible contextual interpretations.

### 3.4 Errors with sentence splitting

While initial investigations show that there are few errors with sentence splitting, there is bound to be some errors in the text. While it is difficult to fix these errors in the dataset, we would like the annotators to give feedback

However, be aware of cases where the sentence split looks strange, but is actually correct. This might happen in cases where the patient overuses subject-less sentences, for example in cases like example 17

- (17) *Var hos legen forrige uke. Grei behandling. Hyggelig oppførsel.*  
Was at GP last week. OK treatment. Nice demeanour.  
*Snill.*  
Kind.  
'(I) visited my GP last week. (I had) OK treatment. (The GP had a) nice demeanour. (He was) kind.'

In example 17 we see 4 sentences, but they are all subject-less, and in some cases the verb is also left out. While these are not complete sentences in the strict sense, they are all individual utterances, and we keep them separated.

It is when the authors of the texts have errors with punctuation that the issues become more visible. For example, a wrongly placed period or comma might lead to sentences being split before they should, or not split when they should have been. If the annotators note errors related to this, they can mark this in the Comments column.

### 3.5 Comment level annotations

When all sentences of a comment have been annotated, the annotator will have a final line to write the polarity of the whole comment. This annotation does not make use of the indicators for comparatives or irrelevancy, so the annotator only needs to add the appropriate intensity (s,ss,sss) or no polarity (b) in the positive and negative columns. When annotating at the comment level, one should consider the totality of the comment and what the annotator feels the intention of the comment as a whole is. The annotator should try to avoid relying on earlier sentences mathematically, i.e. the comment level should not necessarily be positive just because there are more positive sentences, etc. In the summary, it is the relevant information that should be summarized. Annotators do not need to take evaluations marked as 'i' or 'q' into account.

### 3.6 Final Examples

In addition to the examples presented throughout the guidelines, the following artificial examples in Table 1 can be a useful guide to how the annotations might look like. In the first comment we have 6 sentences. Three of the sentences have irrelevant polarities (two ‘i’ and one ‘q’), one of which is in a comparative sentence, and two which are exclusively irrelevant. Since none of the negative expressions are relevant, the overall summary for the comment is positive. The second comment has more negative sentences that are relevant. There are two positive expressions, one standard one (P:ss) and a strong one (P:sssq), but the strong one is about the questionnaire, and is therefore not relevant to the comment level summary, which is more negative, but slightly positive because there are aspects of their GP that the patient indicate that they like.

ID	Comment	P	N	Other
0 start		–	–	–
0-1	Jeg liker fastlegen min	ss	b	–
0-2	Den forrige legen min tok meg for mye på skulderen.	b	sssi	–
0-3	Fastlegen er ok	s	b	–
0-4	Jeg synes det er helt greit å gå til fastlegen min	s	b	–
0-5	Fastlegen min er bedre enn den forrige fastlegen min.	s	si	k
0-6	Jeg stod opp kl. 8 for å skrive denne tilbakemeldingen	b	sq	–
0 end	– <b>Summary</b> –	<b>ss</b>	<b>b</b>	–
1 start		–	–	–
0-5	Jeg liker ikke fastlegen min	b	ss	–
0-5	Legen kommer alltid for seint, men smiler og er hyggelig.	ss	sss	–
0-5	Da jeg ble lagt inn fikk jeg dårlig behandling.	b	ssi	–
0-5	Jeg er veldig glad for å kunne skrive fritekst!	sssq	b	–
0 end	– <b>Summary</b> --	<b>s</b>	<b>ss</b>	–

Table 1: Table showing artificial examples of annotation

## 4 Other Notes

### 4.1 Spelling Mistakes

When annotating, we do not distinguish between correctly written text and text containing spelling mistakes. As long as the annotators understand what it is meant, it should be annotated accordingly.

## 4.2 Languages

There are at least three languages in the dataset: Norwegian Bokmål, Norwegian Nynorsk and English. Annotators should annotate all languages the same way.

## 5 Norsk sammendrag

Når vi annoterer sentiment er det forfatterens meninger vi har i bakhodet. Vi prøver å tenke hva den som skrev kommentaren mente med setningen. Det vil aldri være mulig å vite dette med absolutt sikkerhet. Når et uttrykk blir tatt med som polart, vil det være et tegn på at pasienten mener noe med den setningen. I mange tilfeller vil det være vanskelig å skille mellom objektive påstander, og det vi kaller objektbasert sentiment. Pasientens observasjoner kan være et tegn på polaritet, selv om det ikke er noen tydelige “polare” uttrykk, altså ord som vi oftere forbinder med sentiment enn ikke.

Her er ei lita liste med tips som kan hjelpe:

- Det kan hjelpe å tenke over hva som er polart i setningen? Hva uttrykker polaritet, hvem/hva er det mot, og hvem sier det?
- Man kan ikke alltid være enig i styrken i et uttrykk, men et tips er å se om man lett kan gjøre uttrykket svakere eller sterkere. Hvis man kan det, så kanskje det ikke er så sterkt/svakt allikevel. Merk allikevel at det ikke finnes noen absolutte regler her. Disse avgjørelsene er domenespesifikke, og annotørene må bli enige sammen.
- Objektiv polaritet, altså tilfeller hvor objektive opplysninger indikerer polaritet, kan være vanskelig. Tenk over at det ofte er en grunn til at en pasient nevner noe, og se på konteksten rundt. Uttrykkes det faktaopplysninger som klart kan tolkes som positivt eller negativt i en gitt kontekst?
- Sammenligninger har alltid to elementer, men det er ikke alltid begge elementene er til stede.
- Det er ofte bedre å ta med noe som ofte kan være polart, selv om det er litt usikkert om det er det i et gitt tilfelle.
- Ikke bruk for mye tid på å diskutere intensitet. Det er veldig vanskelig å bli helt enige, men tenk over alle uttrykkene du har sett, og sammenlign litt.