

Appendix B

The Annotation Guidelines

Introduction

Whenever we say something, we carry out a social action—we *do* something with the words we speak. These social actions are called *speech acts*. The purpose of these guidelines was to aid me in annotating sentences with their speech acts. However, that was in the past. Now their purpose is to document and report how I carried out this annotation work, in order for potential future annotators to perhaps replicate this.

Annotating Speech Act Labels

There are six speech act labels: *Assertive*, *Question*, *Directive*, *Expressive*, *Unsure*, and *Other* (see below for further details on each). These are annotated with a web-browser-based tool⁵ and are done in sessions of 50 sentences at a time.

Label: Assertive

An Assertive is a sentence where the speaker holds the propositional content to be true to some degree. The expression of the content can be absolute (“Bilen är grå”), uncertain (“Bilen kanske är grå”), unknowing (“Jag vet inte om bilen är grå”), or anywhere in between.

Assertives often end with a period (“.”), but other punctuations can also be used.

Some examples:

- “Bilen var grå.”
- “Det var ju han som sa att han ville åka.”
- “Han sa att det var en fin bil.”
- “Instämmer om avataren.”
- “Men det ska nog lösa sig”
- “Det är nog så.”
- “Jag vet inte om han såg den.”

Label: Question

A Question is a sentence where the speaker requests information about whether the propositional content is true, or under what conditions it is true. There are different types of questions:

- *Yes/no question*: the speaker asks whether the propositional content is true (“Har du sett den nya filmen?”).
- *Searching question*: the speaker asks under what conditions the propositional content is true (“När går tåget?”).
- *Question mark*: the sentence is only a question mark (“?”) or several (“???”).

⁵ Available on GitHub: <https://github.com/Daniel-B-Tufvesson/speech-act-classifier>

- *Rhetorical question*: the speaker knows the answer and is not actually requesting information from the listener.
- *Echo question*: the speaker repeats what was said before (or parts of it) as a way to request the listener to elaborate.

Questions often end with a question mark (“?”), but other punctuations can also be used.

Some examples:

- “Vill du se en film?”
- “Varför gick han hem?”
- “Hur gick det?”
- “Då är frågan bara vad man ska göra.”
- “Va?”
- “Kalle?”
- “Jag undrar vilken fisk man ska köpa?”
- “Varför??”
- “Vem fan gör så?”

Label: Directive

A Directive is a sentence where the speaker tries to, through the speech act, get the listener to carry out the action that the sentence describes. This always concerns a future action of the listener (compare “Du ska göra detta” and “Du borde ha gjort detta”). Directives are often indirect, since it can sometimes be considered rude to give a direct request (compare “Jag undrar om man kan få lite salt.” and “Skicka hit saltet!”).

There are different types of directives:

- *Command*: the listener is not allowed to refuse.
- *Appeal*: the action is assumed to be in the listener's interest.
- *Offer*: the listener is allowed to refuse, but is not allowed to carry out the action without the directive.
- *Advice/recommendation*: the listener is allowed to refuse or act without the directive.

Directives often end with an exclamation mark (“!”), but other punctuations can also be used.

Some examples:

- “Hämta bilen imorgon!”
- “Kolla in denna sida.”
- “Jag kan rekommendera en större mobil.”
- “Om du gillar action-filmer så kan du se den nya Mission Impossible-filmen”
- “Har hört att det kan vara ganska riskfyllt med sådana investeringar, så du kanske bör läsa på om olika strategier innan du börjar.”
- “Du är inbjuden till mitt bröllop.”
- “Gå in i affären och fråga bara?”
- “Vill du skicka saltet?”
- “Men kan du inte ge dig?”
- “Här har du en intressant länk, som du kanske kan ha nytta av!”

Label: Expressive

An expressive is a sentence where the speaker expresses some feeling or emotional attitude about the propositional content. There are different types expressives:

- *Expressions of emotion*: the speaker expresses an emotional attitude towards the propositional content which the speaker also holds to be true. Some examples:
 - “Vad bra!”
 - “Vilken fin bil!”
 - “Det är en bra film.”
 - “Att du inte bara kan komma i tid!”
 - “Förlåt”
 - “Boken är svårbegriplig”
 - “Det är livsfarligt att låta politiker bestämma hur skattepengar ska spenderas!”
 - “Fy vad hemskt!”
 - “Det är helt ok.”
 - “Inget fel med det!”
- *Wishes*: the speaker holds the proposition not to be true, but wants it to be true. Some examples:
 - “Jag vill ha en bil!”
 - “Snus ska vara billigt!!”
 - “Vad fint det hade varit med en sådan!”
 - “Känner bara för att stanna hemma istället.”
- *Fears*: the speaker holds the proposition not to be true, and does not want it to be true. Some examples:
 - “Skulle ALDRIG välja att föda hemma.”
 - “Detta ska självklart inte ske.”
- *Greetings*: the speaker greets the listener. For example: “Hej!”, “Välkommen!”, “Adjö!”.
- *Congratulations*: the speaker expresses praise for an achievement or good wishes. For example: “Grattis!”, “Lycka till!”.
- *Thanks*: the speaker expresses gratitude.
- *Laughter*: the speaker expresses lively amusement or sometimes contempt or derision by laughing.
- *Emotional exclamations*: the speaker expresses an emotion through a short cry or remark. For example: “Va!”, “NEJ!”, “Usch”, “Fy”, “Aj”, “Jaa!”, “Fan”, “Herregud!”.

Expressives sometimes end with an exclamation mark (“!”), but other punctuations are also used.

Label: Unsure

The Unsure label should be used for ambiguous sentences. A sentence is ambiguous if it is not clear what speech act it has, or if it has several speech acts and it does not belong to any of them more than the others.

Some examples:

- “Apropå ingenting...”
- “Ärligt nu.”
- “O så ÄTER man kött o kykling o sådant !”
- “puff”
- “anime o manga kontakter och man kolla på de sj så”
- “Inte där kommer det hem fulla elektriker, så varför tar du upp så konstiga exempel som är dessutom kassa.” (Is this a question, expressive or assertive?)

Label: Other

The Other label is a garbage label and should be used on, for example, broken sentences, merged sentences, foreign sentences, emojis, text that are not sentences, computer-generated text, table data, etc. It is essentially used for texts that are not valid sentences. However, valid is here not meant as grammatically correct. Even ungrammatical sentences should be labeled with a proper speech act. The Other label is hence intended only for texts that are clearly not Swedish sentences, or where it is clear that the sentence segmentation is wrong (see below about broken and merged sentences).

Some examples:

- “FÖDDA..... Fyrabarnsmor, då 42 år (nära 43)... fick en son 051116 mobygirl, då 40 år ... fick en son 060122 Sannastina, då 41 år fick en son 061104 SusanneN, 42 år fick en son 071018 Tittija, 42 år fick en dotter 080102 Spigge, då 40 årfick en son 080209 xanni, 41 årfick en dotter 080215 Svarten 42 år och 7 månader fick en tjej 080710”
- “Midka skrev 2009-05-29 20:32:40 följande:Dorian Erty mexx skrev 2009-05-29 19:04:59 följande: Är det fysiskt möjligt för dig?”
- “BF 080808 ulle4, 41 år...”
- “21 september - Sagoskog”
- “Open up the control panel and search for plugins.”
- “(5 år)”
- “.:D”
- “.:-)”
- “😂”
- “!”
- “, eller vad vet jag”
- “(Självgående, svett-jocke osv) En gammal självgående”

Problematic and Special Cases

There are gray areas that can be problematic when annotating sentences. Here are some common problematic cases and how to handle these.

Complex Sentences with Multiple Speech Acts

Complex sentences, such as sentences with several main clauses or with subordinate clauses, should be labeled with the speaker’s intended speech act. As an example, the sentence “Jag såg den nya Spider-man trailern och undrar om du skulle vilja se den med mig?” is a complex

sentence consisting of both an Assertive and a Question. However, the Assertive only exists to give context to the Question, so the whole sentence should therefore be labeled as a Question. This involves some human intuition and interpretation. In uncertain cases, label it instead as Unsure.

Merged Sentences

Sometimes the automatic sentence segmenter fails to divide sentences from each other, and they end up as one. And sometimes people use commas (",") instead of periods (".") to separate sentences, which the sentence segmenter then fails to account for. Sometimes do not write any punctuation marks at the end of the sentence, which again confuses the automatic segmenter. Cases such as these should be labeled with Other.

Some examples:

- "Innan detta va jag väldigt hypokondrisk.Tänk om jag får en hjärtinfarkt."
- "Jag är ju ingen expert på hur mycket de olika företagen säljer, så om du menar vem som är världsledande i.o.m vems som säljer flest produkter så kan jag inte svara på det.Jag kan ju tala om vilka märken som jag och mina kollegor använder mest och vad som dominerar i press och annan media."
- "Jag gick i skogen, Bladen hade börjat komma, Sen såg man många som tränade,"

Broken Sentences

Sometimes the automatic sentence segmenter has divided sentences where they should not have been divided. These should be classified as Other.

Some examples:

- "Apropå ingenting..."
- "(Självgående, svett-jocke osv) En gammal självgående"
- "Ärligt nu."
- ", eller vad vet jag."
- ") men inte på länge."

Foreign Language and Mixed Language

As mentioned earlier, sentences written in a foreign language should be labeled with Other. However, sometimes sentences are mixed with both Swedish and foreign words. If more than 50% of the words in a sentence are foreign, then it should be labeled with Other. Otherwise, it should be labeled with its appropriate speech act.

Some examples:

- "Den nya filmen var awesome!" (Should be labeled with Expressive.)
- "This resulted in a sample representative of the population." (Should be labeled with Other.)

Links, Dates, Numbers, and Names

Some sentences consist of only a web link or a name. These should be labeled as an Assertive. This also applies to non-random numbers (“10” and “2”, but not “48749187419”) and dates (“2017-08-22”). There is of course the exception if there is a question mark at the end (“Henrik?”, “15?”), in which case the sentence should be labeled as a Question.

Some examples:

- “10”
- “https://www.google.com/”
- “2017-08-22”
- “Lotta”
- “New York”

The Difference Between Unsure and Other

It is worth noting that sentences labeled as Unsure and Other are not to be included in the final data set. As a consequence, there is not much reason to differentiate these two. The Unsure label exists only for diagnostic purposes since it shows what sentences are not covered by the guidelines. This was only relevant for the development of the guidelines, but not for the final data set.