

# Annotation Guidelines for Dutch Quote Attribution

Frank van den Berg

July 9, 2022

## 1 How to annotate?

- Load the text into the annotation tool<sup>1</sup> designed by Muzny et al. (2017).
- Read the text from start to finish, mark and correct quote- and mention spans as you go.
- Each time you mark a span of text as a quote, you should also select the character that is the quote's speaker. Mentions should also be annotated with the character that they refer to. New characters can be added to the character list.
- Link each quote to the mention that is the most obvious indicator of that quote's speaker.
- Before saving the file to your machine, make sure that each quote is connected to a mention.

In the examples that will follow, correctly annotated spans are indicated with square brackets [ and ]; spans that should not be annotated are indicated with [ red brackets ].

## 2 Quotes

We are only focusing on the annotation of **direct speech**, as opposed to **indirect speech** and **free indirect discourse**. Direct quotes are used to indicate that the speaker said precisely what is written. These quotes appear entirely between quotation marks.

- A correct quotation span can look as follows:

[‘Waar wacht je op?']<sub>quote</sub> vraagt Lea.

- Quotes can also consist of multiple sentences:

[‘Höss,']<sub>quote</sub> had hij gezegd, [‘de commandant van Auschwitz. Boeiend. Had hij niet een verhouding met een gevangene in het kamp? Hij is opgehangen in Polen, niet?']<sub>quote</sub>

---

<sup>1</sup><https://github.com/muzny/quoteannotator>

- However, not all text between quotation marks is a direct quote:

Ze moesten zijn intelligentie en werkhouding peilen, uitvinden hoe hij hiervoor les had gehad, en of hij mogelijk [‘studiebeur-swaardig?’]<sub>quote</sub> zou blijken.

- Instead of quotation marks, quotes can also be introduced with a dash (-) sign:

De drie jongens rolden op een matras over elkaar. [- Was het mooi, mama?]<sub>quote</sub> vroeg Lili.

- If there is dialogue without quotation marks or a dash sign, we do not annotate it, as this is indirect speech (Van Boven and Dorleijn, 2003):

Ook daar werd hij door niemand herkend.  
Iemand zei [dat hij een bril droeg.]<sub>quote</sub>  
Een andere loketbeambte ontkende dat ten stelligste.

- Citations at the beginning of a book also do not count as dialogue, as they are paratext:

[“In dit geval haten prinsen de verrader, al houden ze van het verraad.”]<sub>quote</sub> - Samuel Daniel  
1  
Om zeven uur op een Caribische ochtend (...)

### 3 Mentions

The only mentions that should be annotated are the spans of text that refer to the speaker of a quote. Often, these mentions can be identified when the subject of a reported speech verb (says, replies, etc.) is found next to a quotation.

- If we apply this to our earlier example, we see that ‘Lea’ should be annotated as the mention:

[‘Waar wacht je op?’]<sub>quote</sub> vraagt [Lea]<sub>mention</sub>.

Not every quote is accompanied by a speech verb. In some cases, there can even be multiple mention candidates of a quote’s speaker. In these cases, we will consistently choose the mention that is the closest to the quote.

- Here both ‘Ouwe Joe Hunt’ and ‘een leraar’ refer to the same speaker, but we choose the mention ‘een leraar’, as it is the closest to the quote:

Op de derde ochtend van dat najaarstrimester hadden we geschiedenis van [Ouwe Joe Hunt]<sub>mention</sub>, droogkomisch innemend in zijn drielig pak, [een leraar]<sub>mention</sub> wiens systeem van orde berustte op het in stand houden van voldoende maar niet buitensporige landerigheid. [‘Jullie weten nog dat ik jullie gevraagd heb alvast iets te lezen over het koningschap van Hendrik viii.’]<sub>quote</sub>

Sometimes, the mention of a quote's speaker can even be found inside an earlier quote. We only annotate these mentions when they are **explicitly addressed**, e.g. by their name.

- The mention of the second quote can be found as the addressee in the first quote:

Ik keek vermoedelijk geïmponeerder dan [Dixon]<sub>mention1</sub> gezond achtte. ['[Webster]<sub>mention2</sub>, ga daar eens op door. ']<sub>quote1</sub>  
['Ik dacht dat het gewoon een gedicht was over een kerkuil, meneer.']<sub>quote2</sub>

However, we do not annotate mentions of speakers when they are addressed in an earlier quote only by their personal pronoun ('je', 'u' etc.).

- We annotate 'ze' as the mention of both the first and the third quote, as 'je' is not an explicit way of addressing the speaker of the third quote:

['Het is een oude man, denk ik,']<sub>quote1</sub> zei [ze]<sub>mention1,3</sub>.  
['Heb [je]<sub>mention3</sub> hem dan gezien?']<sub>quote2</sub> vroeg [ik]<sub>mention2</sub>.  
['Nee, dat kan ik horen. ']<sub>quote3</sub>

Although very rare, there are cases where the speaker of a quote is not mentioned.

- Here we only annotate the quote, but no mention is annotated:

Voetstappen kwamen op hem af.  
['Alstublieft, gaat u mee. U verstoort dit samenzijn.']<sub>quote</sub>  
Hardhandig werd hij aan zijn schouders weggevoerd.

As a final note: all the mentions we annotate should always be outside the quotes they are connected to.

## 4 Quote-mention links

Each quote should be linked to the mention of that quote's speaker. While each quote can only be linked to one mention, it can be the case that one mention is linked to multiple quotes.

- In the following sentence the mention 'ze' is linked to both quotation spans:
  - ['Die trappen,']<sub>quote1</sub> zegt [ze]<sub>mention1,2</sub>.  
['Die trappen van jou worden mijn dood nog eens.']<sub>quote2</sub>
- A similar situation happens when the same pair of speakers keeps taking turns:
  - [Ik]<sub>mention1,4</sub> zei: ['Ik wil het wel voor je doen.']<sub>quote1</sub>.  
['Nee,']<sub>quote2</sub> zegt [ze]<sub>mention2,3</sub>. ['Veel te duur.']<sub>quote3</sub>  
['O, je doet het liever zelf?']<sub>quote4</sub>

## References

- Muzny, G., Fang, M., Chang, A., and Jurafsky, D. (2017). A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 460–470, Valencia, Spain. Association for Computational Linguistics.
- Van Boven, E. and Dorleijn, G. J. (2003). *Literair mechaniek: Inleiding tot de analyse van verhalen en gedichten*. tweede, herziene druk (1e druk: 1999).