

Annotating Generalising Statements in Literary Texts

Tillmann Dönicke, Luisa Gödeke, Hanna Varachkina

University of Göttingen

The Seventeenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation
16–17 June, 2021

Roadmap

Introduction

Challenges

Annotation

Corpus

Roadmap

Introduction

Challenges

Annotation

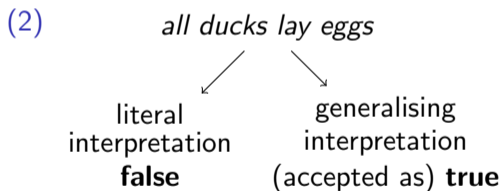
Corpus

Quantified Statements

- (1) *all students in the semantics class take notes*
true iff every single student takes notes
(in a non-collective reading)

Quantified Statements


- (1) *all students in the semantics class take notes*
true iff every single student takes notes
(in a non-collective reading)



Quantified Statements

(1) *all students in the semantics class take notes*
true iff every single student takes notes
(in a non-collective reading)

(2) *all ducks lay eggs*



literal interpretation
false

generalising interpretation
(accepted as) **true**

(3) *all triangles have three sides*
true by definition

Quantified Statements

(1) *all students in the semantics class take notes*
true iff every single student takes notes
(in a non-collective reading)

(2) *all ducks lay eggs*

```
graph TD; A["all ducks lay eggs"] --> B["literal interpretation"]; A --> C["generalising interpretation"]; B --- D["false"]; C --- E["(accepted as) true"]
```

literal interpretation **false**

generalising interpretation (accepted as) **true**

(3) *all triangles have three sides*
true by definition

generalising statements

1. involve assumed instances
2. are accepted as true
 - a) without evidence for trueness
 - b) despite evidence for falseness

Generalising Statements in Literary Texts

indicators for

- (self-)reflection of characters, narrator, or author:

“We are strange creatures,” said Edward, smiling. “If we can only put out of sight anything which troubles us, we fancy at once we have got rid of it.”

(Goethe, EA)

Generalising Statements in Literary Texts

indicators for

- (self-)reflection of characters, narrator, or author:
“We are strange creatures,” said Edward, smiling. “If we can only put out of sight anything which troubles us, we fancy at once we have got rid of it.”
(Goethe, EA)
- transfer from the fictional world to the actual world:
Heroism is the exception and mostly the product of a separate situation.
(Fontane, Stechlin)

Roadmap

Introduction

Challenges

Annotation

Corpus

Analysing Quantifier, Restrictor, and Scope

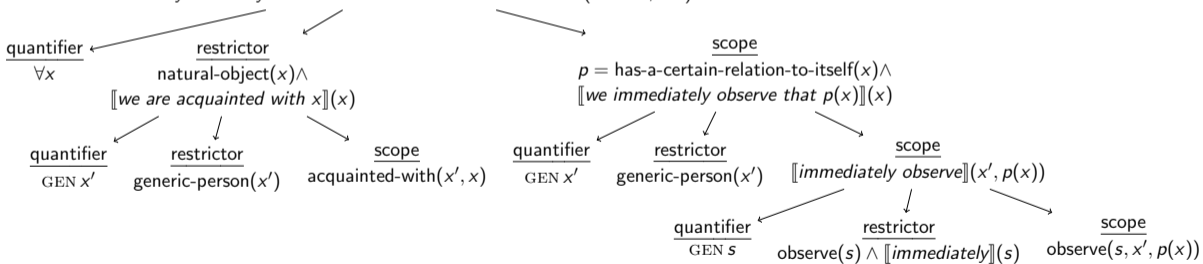
Short example:

In all natural objects with which we are acquainted, we observe immediately that they have a certain relation to themselves. (Goethe, EA)

Analysing Quantifier, Restrictor, and Scope

Short example:

In all natural objects with which we are acquainted, we observe immediately that they have a certain relation to themselves. (Goethe, EA)



Ambivalent Structures

Help upon the spot is the thing you often most want in the country.
(Goethe, EA)

Ambivalent Structures

Help upon the spot is the thing you often most want in the country.
(Goethe, EA)

- GEN[y : help-upon-the-spot(y)][OFTEN[s][GEN[x][most-want(s, x, y)]]]
- OFTEN[s][GEN[y : help-upon-the-spot(y)][GEN[x][most-want(s, x, y)]]]
- GEN[x][OFTEN[s][GEN[y : help-upon-the-spot(y)][most-want(s, x, y)]]]
- ...?

Absence of Overt Markers

Business requires earnestness and method; life must have a freer handling.
(Goethe, EA)

Absence of Overt Markers

Business requires earnestness and method; life must have a freer handling.
(Goethe, EA)

- how many covert quantifiers?
- $?_{\text{GEN}}[x : \text{business}][\text{GEN}[y_1 : \text{earnestness}(y_1)][\text{GEN}[s_1][\text{require}(s_1, x, y_1)]]] \wedge \text{GEN}[y_2 : \text{method}(y_2)][\text{GEN}[s_2][\text{require}(s_2, x, y_2)]]]$

Challenges and Objectives

Challenges:

1. complex sentence structures / older versions of German
↪ difficulties of analysing quantifier, restrictor, and scope
2. ambivalent syntactic structures
↪ scope ambiguities
3. absence of overt markers
↪ covert quantifiers are easily overlooked

Challenges and Objectives

Challenges:

1. complex sentence structures / older versions of German
↪ difficulties of analysing quantifier, restrictor, and scope
2. ambivalent syntactic structures
↪ scope ambiguities
3. absence of overt markers
↪ covert quantifiers are easily overlooked

Objectives:

1. complete statements
2. exhaustiveness of generalisation (i.e. quantification)
3. (fast) annotation based on surface text

Roadmap

Introduction

Challenges

Annotation

Corpus

Types of Quantification

- overt

- covert

Types of Quantification

- overt
 - precise
 - vague
- covert

Types of Quantification

- overt
 - precise
 - universal (ALL): *all, every, always, everywhere*
 - majority (MEIST): *most(ly), main(ly)*
 - existential (EXIST): *there is/are, exist*
 - numerical (ZAHL): *one, half, up to two thirds, at least five, dozen, percent, million*
 - vague
- covert

Types of Quantification

- overt
 - precise
 - universal (ALL): *all, every, always, everywhere*
 - majority (MEIST): *most(ly), main(ly)*
 - existential (EXIST): *there is/are, exist*
 - numerical (ZAHL): *one, half, up to two thirds, at least five, dozen, percent, million*
 - vague (DIV): *few, some, many, occasionally, commonly, often, as a rule, in general*
- covert

Types of Quantification

- overt
 - precise
 - universal (ALL): *all, every, always, everywhere*
 - majority (MEIST): *most(ly), main(ly)*
 - existential (EXIST): *there is/are, exist*
 - numerical (ZAHL): *one, half, up to two thirds, at least five, dozen, percent, million*
 - vague (DIV): *few, some, many, occasionally, commonly, often, as a rule, in general*
- covert
 - generic (BARE)

Annotation Span

- syntactic constraint: a statement is a sequence of (one or several) clauses
“[Our excellent superior commonly permits me to read the letters in which she communicates her observations upon her pupils to their parents and friends]_{DIV.} (Goethe, EA)

Annotation Span

- syntactic constraint: a statement is a sequence of (one or several) clauses
“[Our excellent superior commonly permits me to read the letters in which she communicates her observations upon her pupils to their parents and friends]_{DIV.} (Goethe, EA)
- semantic constraint: a statement contains quantifier, restrictor and scope
“It concerns our friend the Captain,” answered Edward; “you know the unfortunate position [in which he, like many others, is placed]_{DIV.} (Goethe, EA)

Annotation Span

- syntactic constraint: a statement is a sequence of (one or several) clauses
“[Our excellent superior commonly permits me to read the letters in which she communicates her observations upon her pupils to their parents and friends]_{DIV.} (Goethe, EA)
- semantic constraint: a statement contains quantifier, restrictor and scope
“It concerns our friend the Captain,” answered Edward; “you know the unfortunate position [in which he, like many others, is placed]_{DIV.} (Goethe, EA)
- same span: subsequent statements with 1) the same type of quantification and 2) the same restrictor or the same scope
[The students from the towns and universities are sufficiently clever and orderly, but they are deficient in personal experience]_{BARE.} (Goethe, EA)

Multiple Quantification

One statement can have multiple (generalising) quantifications.

- binary > counting: assign all possible tags but maximally once

Most Popes love all their subjects equally] $_{ALL+MEIST}$

Every Pope loves all his subjects equally] $_{ALL}$

Multiple Quantification

One statement can have multiple (generalising) quantifications.

- binary > counting: assign all possible tags but maximally once

Most Popes love all their subjects equally]ALL+MEIST

Every Pope loves all his subjects equally]ALL

- covert quantifiers are easily overlooked: only assign BARE if no other tag applies

The Pope loves all his subjects equally]ALL

The Pope loves his subjects equally]BARE

Quantification and Negation

- *no lion sleeps*
 - $\neg\exists[x : \text{lion}(x)][\text{sleep}(x)] \rightsquigarrow$ EXIST?
 - $\forall[x : \text{lion}(x)][\neg\text{sleep}(x)] \rightsquigarrow$ ALL?
 - $\neg\text{GEN}[x : \text{lion}(x)][\text{sleep}(x)] \rightsquigarrow$ BARE?
 - $\text{GEN}[x : \text{lion}(x)][\neg\text{sleep}(x)] \rightsquigarrow$ BARE?
- negation can apply to quantifier, restrictor or scope

Quantification and Negation

- *no lion sleeps*
 - $\neg\exists[x : \text{lion}(x)][\text{sleep}(x)] \rightsquigarrow$ EXIST?
 - $\forall[x : \text{lion}(x)][\neg\text{sleep}(x)] \rightsquigarrow$ ALL?
 - $\neg\text{GEN}[x : \text{lion}(x)][\text{sleep}(x)] \rightsquigarrow$ BARE?
 - $\text{GEN}[x : \text{lion}(x)][\neg\text{sleep}(x)] \rightsquigarrow$ BARE?
- negation can apply to quantifier, restrictor or scope
- tag NEG for all generalising statements with negation

“*[And there are many cases [...] in which we are obliged, and in which it is the real kindness, rather to write nothing than not to write]*_{NEG}.”
(Goethe, EA)

Tagset Overview

| Tag | Description |
|-------|---|
| ALL | overt universal quantification |
| MEIST | overt majority quantification |
| EXIST | overt existential quantification |
| ZAHL | overt numerical quantification |
| DIV | overt vague quantification |
| BARE | none of the above + covert quantification |
| NEG | any of the above + negation |

Table: Tagset

Tagset Overview

| Tag | Description |
|-------|---|
| ALL | overt universal quantification |
| MEIST | overt majority quantification |
| EXIST | overt existential quantification |
| (ZAHL | overt numerical quantification) |
| DIV | overt vague quantification |
| BARE | none of the above + covert quantification |
| NEG | any of the above + negation |

Table: Tagset

Roadmap

Introduction

Challenges

Annotation

Corpus

Modes of Narration and Attribution Corpus (MONACO)

- German fictional literature (1650–1950)
- 3 phenomena: generalising interpretation / comment / non-fictional speech
- first about 200 sentences per text
- initial annotation: 2 research assistants per phenomenon per text
- gold annotation: 2 researchers per phenomenon per text
- currently 10 texts with gold standards

Agreements and Statistics

| GI+Q | | GI | | Q | |
|-------------|----------|-----------|----------|----------|----------|
| κ | σ | κ | σ | κ | σ |
| .67 | .20 | .68 | .22 | .85 | .14 |

Table: Mean inter-annotator agreement (κ) over all texts and standard deviations (σ)

| Generalising statements | | | | | | |
|--------------------------------|-------|-------|-----|------|-----|-------|
| ALL | MEIST | EXIST | DIV | BARE | NEG | Total |
| 151 | 7 | 17 | 76 | 332 | 145 | 728 |

Table: Number of generalising statements in the gold standard

Conclusion

- complex sentence structures in literary texts
- shallow annotation scheme:
 - annotation on clause/statement-level
 - only annotate type of quantification
- annotation results:
 - substantial agreement for generalising interpretation
 - almost perfect agreement for quantification
- MONACO: <https://gitlab.gwdg.de/mona/korpus-public>
- future work: generalisation tagger

References

- Theodor Fontane. *The Stechlin*. Translated by William L. Zwiebel. Germ Series. Camden House, 1995. ISBN 9781571130242. URL: <https://books.google.de/books?id=7yh0ZJjNgxoC>.
- Johann Wolfgang von Goethe. *Elective affinities : a novel*. 19--? URL: <https://archive.org/details/electiveaffiniti00goetuoft/page/12/mode/2up>.
- Sarah-Jane Leslie, Sangeet Khemlani, and Sam Glucksberg. Do all ducks lay eggs? the generic overgeneralization effect. *Journal of Memory and Language*, 65(1):15–31, 2011. ISSN 0749-596X. doi: <https://doi.org/10.1016/j.jml.2010.12.005>. URL <https://www.sciencedirect.com/science/article/pii/S0749596X10001154>.
- Sebastian Löbner. *Quantoren im GWDS*, pages 171–192. Max Niemeyer Verlag, 2005. doi: [doi:10.1515/9783110925876.171](https://doi.org/10.1515/9783110925876.171). URL <https://doi.org/10.1515/9783110925876.171>.