

DiscoGeM: A Crowdsourced Corpus of Genre-Mixed Implicit Discourse Relations

Merel Scholman, Tianai Dong, Frances Yung, & Vera Demberg
Saarland University, Germany

LREC 2022



Introduction

- ▶ Discourse relations (DRs): logical links between segments of text
- ▶ Can be *explicit* (connectives) or *implicit*
- ▶ Implicit relations difficult to classify (automatic & manually)

Example:

- (1) I'm a feminist **because** I believe in gender equality.
- (2) I'm a feminist; **in other words**, I believe in gender equality.
- (3) I'm a feminist. I believe in gender equality.

- ▶ Parsers perform poorly on implicit relations (and out-of-domain text)
→ need for implicit relation annotations in different genres
- ▶ Obtaining manually annotated data is costly and time-consuming
- ▶ Crowdsourcing can provide solution
- ▶ Additional benefit: multiple observations per relation
→ Derive a distribution of relation senses per relation that might better represent the ambiguity of the relation

Goals of current contribution:

- ▶ Collect large, multi-genre, reliable discourse-annotated resource
- ▶ Provide distribution of relation senses; examine optimal aggregation method
- ▶ Compare distributions of implicit relations between genres

Introduction

Goals of current contribution:

- ▶ Collect large, multi-genre, reliable discourse-annotated resource
- ▶ Provide distribution of relation senses; examine optimal aggregation method
- ▶ Compare distributions of implicit relations between genres

DiscoGeM: PDTB3-style crowdsourced corpus of 6,505 implicit discourse relations

Table of Contents

① Method

- Data
- Task design
- Crowd annotators

② Results

Table of Contents

① Method

- Data
- Task design
- Crowd annotators

② Results

Method: Data

	Europarl	Literature	Wikipedia	Total
No. DRs	2,800	3,060	645	6,505

Table: Corpus size in number of discourse relations per genre and in total.

Method: Data

	Europarl	Literature	Wikipedia	Total
No. DRs	2,800	3,060	645	6,505

Table: Corpus size in number of discourse relations per genre and in total.

Europarl

- ▶ Political speech, oftentimes prepared; tends to be argumentative
- ▶ Data from Europarl (Koehn 2005) and Europarl Direct (Cartoni & Meyer, 2012)

Method: Data

	Europarl	Literature	Wikipedia	Total
No. DRs	2,800	3,060	645	6,505

Table: Corpus size in number of discourse relations per genre and in total.

Europarl

- ▶ Political speech, oftentimes prepared; tends to be argumentative
- ▶ Data from Europarl (Koehn 2005) and Europarl Direct (Cartoni & Meyer, 2012)

Literature

- ▶ Narrative writing; tends to be sequence of events
- ▶ Data from 20 novels

Method: Data

	Europarl	Literature	Wikipedia	Total
No. DRs	2,800	3,060	645	6,505

Table: Corpus size in number of discourse relations per genre and in total.

Europarl

- ▶ Political speech, oftentimes prepared; tends to be argumentative
- ▶ Data from Europarl (Koehn 2005) and Europarl Direct (Cartoni & Meyer, 2012)

Literature

- ▶ Narrative writing; tends to be sequence of events
- ▶ Data from 20 novels

Wikipedia

- ▶ Informative text; explains known facts about common topics
- ▶ Texts taken from first section of 69 Wikipedia entries
- ▶ Reference annotations available for this genre (3 expert annotators)

Table of Contents

① Method

- Data
- Task design
- Crowd annotators

② Results

Method: Task design

Crowdsourced annotations using **Two-step Discourse Connective (DC)**
Method

Method: Task design

Crowdsourced annotations using **Two-step Discourse Connective (DC) Method**

① Freely insert connective to express relation

I merely repeat, remember always your duty of enmity towards Man and all his ways.

type here

Whatever goes upon two legs is an enemy. Whatever goes upon four legs, or has wings, is a friend.

Method: Task design

Crowdsourced annotations using **Two-step Discourse Connective (DC) Method**

- 1 Freely insert connective to express relation

I merely repeat, remember always your duty of enmity towards Man and all his ways.

Whatever goes upon two legs is an enemy. Whatever goes upon four legs, or has wings, is a friend.

- 2 Choose from automatically provided list to disambiguate

I merely repeat, remember always your duty of enmity towards Man and all his ways.

Whatever goes upon two legs is an enemy. Whatever goes upon four legs, or has wings, is a friend.

Yung, Scholman & Demberg (2019), *LAW*.

Method: Connective bank

Created a connective bank for DC method to map connectives and labels

Contains >2,000 entries, including:

- ▶ typical connectives (e.g., *because*)
- ▶ frequent typos (*becuase*)
- ▶ variations (*largely because*)
- ▶ “alternative lexicalizations” (*the reason is that*)
- ▶ combinations (*and because*)

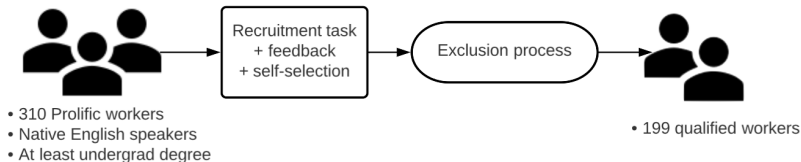
Table of Contents

① Method

- Data
- Task design
- Crowd annotators

② Results

Method: Selecting crowdsourced annotators



Participant qualification through recruitment task:

- ▶ Obtain annotations that allowed us to evaluate annotator potential
- ▶ Significantly improves quality of annotations (Scholman et al., 2022 *LREC*)

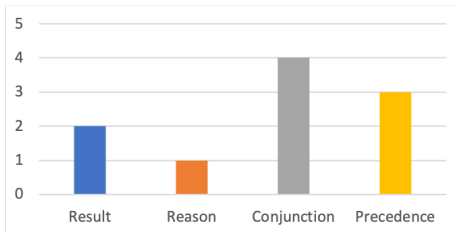
Table of Contents

① Method

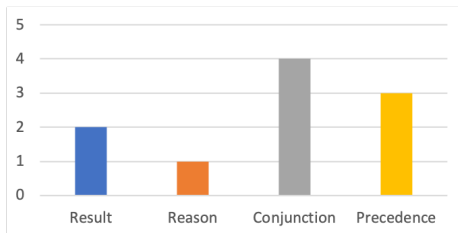
- Data
- Task design
- Crowd annotators

② Results

Results: Label aggregation

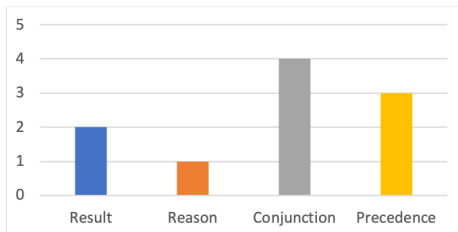


Results: Label aggregation



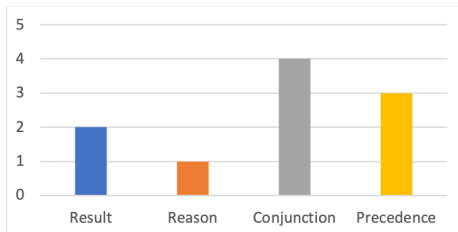
- ▶ **Majority-single**: sense with majority agreement

Results: Label aggregation



- ▶ **Majority-single:** sense with majority agreement
- ▶ **IRT-single:** highest probability sense, based on Dawid-Skene model (Passonneau & Carpenter, 2014)
 - ▶ Uses unsupervised learning to estimate the most likely sense for every item
 - ▶ Assumes single ground truth

Results: Label aggregation



- ▶ **Majority-single:** sense with majority agreement
- ▶ **IRT-single:** highest probability sense, based on Dawid-Skene model (Passonneau & Carpenter, 2014)
 - ▶ Uses unsupervised learning to estimate the most likely sense for every item
 - ▶ Assumes single ground truth
- ▶ **CrowdTruth-distribution:** all senses that reached threshold of 20% probability based on CrowdTruth 2.0 (Dumitrache et al., 2018)
 - ▶ Doesn't enforce agreement between annotators
 - ▶ Assumes there is no single ground truth

Results: Comparison with reference labels

Results: Comparison with reference labels

Aggregated measure	κ	%	P	R	F1
Majority - single	.55	67	.67	.49	.55
IRT - single	.53	64	.64	.46	.52
CrowdTruth - distribution	.75	82	.69	.66	.59

► Agreement good for implicit DR annotation

- Crowd & reference on 6 DR classes: F1=.51 (Kishimoto et al., 2019)
- PDTB & RST-DT on implicits: 37% (Demberg et al., 2019)
- 2 experts on implicits spoken text: 66%, κ =.58 (Hoek et al., 2021)

Results: Comparison with reference labels

Aggregated measure	κ	%	P	R	F1
Majority - single	.55	67	.67	.49	.55
IRT - single	.53	64	.64	.46	.52
CrowdTruth - distribution	.75	82	.69	.66	.59

- ▶ Agreement good for implicit DR annotation
 - ▶ Crowd & reference on 6 DR classes: F1=.51 (Kishimoto et al., 2019)
 - ▶ PDTB & RST-DT on implicits: 37% (Demberg et al., 2019)
 - ▶ 2 experts on implicits spoken text: 66%, κ =.58 (Hoek et al., 2021)
- ▶ Distribution measure better captures the reference label senses

Results: Distribution measures

Frequently co-occurring senses	% CrowdTruth
Conjunction & Result	13
ARG2-AS-DET. & CONJUNCTION	12
PRECEDENCE & RESULT	5

Table: Most frequent relation senses found to co-occur.

- ▶ The prevalence of some combinations unexpected, but data shows it is valid

Results: Distribution measures

Frequently co-occurring senses	% CrowdTruth
Conjunction & Result	13
ARG2-AS-DET. & CONJUNCTION	12
PRECEDENCE & RESULT	5

Table: Most frequent relation senses found to co-occur.

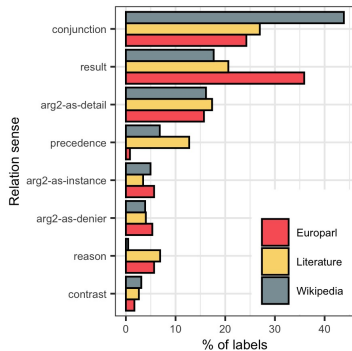
- ▶ The prevalence of some combinations unexpected, but data shows it is valid

Example:

- 1 *Cities are home to 80% of the EU inhabitants. // It is in cities that the great majority of jobs and companies are located.*
- ▶ CONJUNCTION – 4 annotators ; RESULT – 5 annotators

Results: Genre comparison

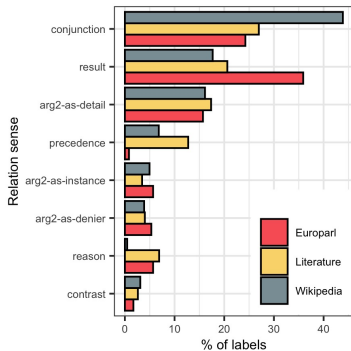
Clear differences in relational distribution between genres:



Highlights the importance of taking genre effects into consideration

Results: Genre comparison

Clear differences in relational distribution between genres:



- ▶ CONJUNCTION most prevalent in Wikipedia
- ▶ RESULT relations occur more in Europarl
- ▶ Most PRECEDENCE relations in literature

Highlights the importance of taking genre effects into consideration

Conclusion

- ▶ We created an awesome corpus! Go ahead and use it!

Conclusion

- ▶ We created an awesome corpus! Go ahead and use it!
- ▶ Many implicit DRs can express multiple relation senses. This is the first large resource that provides sense distributions → valuable for downstream tasks
- ▶ Genre effects in distributions of implicit relation types → classifiers need to take this into account.

Conclusion

- ▶ We created an awesome corpus! Go ahead and use it!
- ▶ Many implicit DRs can express multiple relation senses. This is the first large resource that provides sense distributions → valuable for downstream tasks
- ▶ Genre effects in distributions of implicit relation types → classifiers need to take this into account.

Thank you for your attention!

This research was funded by the German Research Foundation (DFG) as part of SFB 1102 at Saarland University.



UNIVERSITÄT
DES
SAARLANDES