

AnnCor, 24 april 2019

## WP Discourse-annotatie

Jacqueline Evers-Vermeul  
in samenwerking met  
Ted Sanders  
Suzanne Bogaerds-Hazenberg  
Laura Brands  
Carla van Rooijen  
Nina Sangers



Universiteit Utrecht

Utrecht Institute of Linguistics OTS

## Wat hebben we gedaan?

1. Geannoteerde data geoptimaliseerd
2. Annotatie-experiment
3. Literatuur-review
4. CCR-theorie uitgebreid
5. CCR-theorie beter toegankelijk gemaakt
6. "Vertaal-tool" ontwikkeld voor annotatie-frameworks
7. Nederlands connectieflexicon ontwikkeld
8. In wording: website over CCR ontwikkeld



Universiteit Utrecht

## 1. Geannoteerde data geoptimaliseerd

- Geannoteerde data uit MODERN-onderzoek van Jet Hoek (zie haar PhD thesis *Making sense of discourse*, 2018)
- Selectie uit Europarl-corpus (Koehn 2005)
- Netjes afbakenen van alle vertalingen: Nina Sangers
- Beschikbaar te stellen via nieuwe website CCR en AnnCor-website



Universiteit Utrecht

## 1. Geannoteerde data geoptimaliseerd

- 1906 geannoteerde Engelse coherenterelaties met één van de volgende connectieven:
  - Because
  - So
  - Also
  - In addition
  - Although
  - But
  - If
  - Unless
- met hun vertalingen in het
  - Nederlands
  - Duits
  - Frans
  - Spaans



Universiteit Utrecht

## 1. Geannoteerde data geoptimaliseerd

You will forgive me, Prime Minister, if I say that the launching of the single currency on 2 and 3 May was not quite such a seamless operation, **although** all of us appreciate that you were seeking to reconcile seemingly irreconcilable positions, something we contend with in this House very often.

U zult me wel vergeven, mijnheer de eerste minister, als ik zeg dat de start van de eenheidsmunt op 2 en 3 mei niet geheel vlekkeloos is verlopen. **Maar** wij zien allemaal in dat dit kwam doordat u strijdige standpunten met elkaar heeft willen verzoenen, iets waar we in dit Parlement wel vaker mee te maken krijgen.

Verzeihen Sie mir die Bemerkung, Herr Premierminister, aber die Einführung der einheitlichen Währung am 2. und 3. Mai ist nicht gerade reibungslos verlaufen, **obwohl** wir alle Ihre Bemühung um die Annäherung scheinbar unvereinbarer Positionen zu schätzen wissen, womit wir es hier im Haus sehr oft zu tun haben.



Universiteit Utrecht

## 2. Annotatie-experiment


- MA-internship en -thesis van Laura Brands (2017), gebruikmakend van eerdere DiscAn-corpora:
  - Annotating coherence relations:*  
*Developing a method for naive annotators*
- Test van inzetbaarheid van "naieve annotatoren"



Universiteit Utrecht

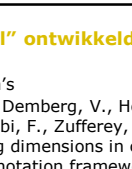
### 3. Literatuur-review


- Review van artikelen over de acquisitie van connectieven
- In samenwerking met Suzanne Bogaerds-Hazenberg
- Ontwikkelingsperspectief op
  - Temporele connectieven
  - Causale connectieven
- Rapportage: presentatie tijdens SSSR in Porto




Temporal connectives (no world knowledge)				
	Before Y, X	X, before Y	After X, Y	Y after X
1	picture selection (37-62%, C18) imitation (43%, K) enactment (20%, C)	picture selection (37-71%, C18) enactment (75-95%, F,C) imitation (49%, K)	picture selection (60-61%, B, C) enactment (85%, C) imitation (49%, K)	picture selection (52-61%, B, C) enactment (19-58%, C,F) imitation (43%, K)
4	enactment (74%, C) picture selection (61-68, B, C) imitation (55-57%, T4) video selection (58%, M <sup>9</sup> )	enactment (54-92-97%, W,R,C) video selection (82%, M <sup>9</sup> ) picture selection (68-70%, T4) imitation (49%, K)	enactment (60-61%, B, C) picture selection (60-61%, B, C) imitation (49%, K)	picture selection (60-89%, B,C) imitation (59-67%, T,K) video selection (50%, M <sup>9</sup> ) enactment (44%, C)
5	imitation (87%, T) picture selection (87%, T) video selection (87%, T) enactment (50%, J,A,R,N)	imitation (87%, T) picture selection (87%, T) video selection (87%, T) enactment (50%, J,A,R,N)	imitation (87%, T) picture selection (87%, T) video selection (87%, T) enactment (50%, J,A,R,N)	imitation (87%, T) picture selection (87%, T) video selection (87%, T) enactment (50%, J,A,R,N)
6	picture selection (63%, T) imitation (74%, T) enactment (72-80%, T)	picture selection (63%, T) imitation (74%, T) enactment (72-80%, T)	picture selection (63%, T) imitation (74%, T) enactment (72-80%, T)	picture selection (63%, T) imitation (74%, T) enactment (72-80%, T)
7	multiple choice (84%, P)	multiple choice (84%, P)	multiple choice (84%, P)	multiple choice (84%, P)
8	enactment (81%, A <sup>9</sup> )	enactment (81%, A <sup>9</sup> )	enactment (81%, A <sup>9</sup> )	enactment (81%, A <sup>9</sup> )
10	multiple choice (93%, P)	multiple choice (93%, P)	multiple choice (93%, P)	multiple choice (93%, P)
11	multiple choice (97%, P)	multiple choice (97%, P)	multiple choice (97%, P)	multiple choice (97%, P)
12	multiple choice (97%, P)	multiple choice (97%, P)	multiple choice (97%, P)	multiple choice (97%, P)

What can we conclude?






### Establishing coherence using connectives: A developmental overview




Jacqueline Evers-Vermeul,  
Suzanne Bogaerds-Hazenberg & Ted Sanders



### 4. CCR-theorie uitgebreid


- Samen met Jet Hoek en Merel Scholman
- CCR = Cognitive approach to Coherence Relations
- Oorspronkelijk ontwikkeld door Sanders, Spooren & Noordman 1992, 1993; in de loop der jaren uitgebreid
- Evers-Vermeul, J., Hoek, J. & Scholman, M.C.J. (2017). On temporality in discourse annotation: Theoretical and practical considerations. *Dialogue and Discourse*, 8(2), 1-20.
- Nieuwe dimensie toegevoegd: temporality

	+ Ordered in time	- Ordered in time
2	+ Sequential	- Sequential (= Synchronous)
3	+ Chronological - Chronological	




### 5. CCR-theorie beter toegankelijk gemaakt

- Met Jet Hoek en Ted Sanders
- Hoek, J., Evers-Vermeul, J., & Sanders, T.J.M. (submitted). Using the Cognitive Approach to Coherence Relations for Discourse Annotation. To appear in *Discourse & Dialogue*.
- Zie ook Chapter 3 in Hoek (2018), *Making sense of discourse*

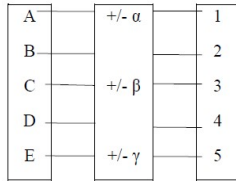


### 6. "Vertaal-tool" ontwikkeld voor annotatie-frameworks

- Met veel collega's
- Sanders, T.J.M., Demberg, V., Hoek, J., Scholman, M.C.J., Asr Torabi, F., Zufferey, S., & Evers-Vermeul, J. (2018). Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory*. First view online, doi: 10.1515/clit-2016-0078
- Met CCR als instrument, als een interlingua die onderzoekers in staat stelt onderzoekers van verschillende annotatie-frameworks met elkaar te laten communiceren/ van elkaars corpora gebruik te laten maken
- CCR-dimensies worden gebruikt om relaties van andere frameworks te typeren



## 6. "Vertaal-tool" ontwikkeld voor annotatie-frameworks



## Voorbeeld 1

- Startpunt: PDTB
- Relatie: *Chosen alternative*

*[Please hand in your expense sheet to Nicky now]  
or [mail it to her before April 10.]*

- Classificatie in termen van CCR-dimensies:
  - o negatief
  - o additief
  - o objectief/subjectief
  - o alternatief
- Equivalent in RST-DT met dezelfde waarden?  
=> *Disjunction*

## Voorbeeld 2

- Startpunt: RST-DT
- Relatie: *Cause* (met Satellite-Nucleus order)

*[Liesbeth rescheduled the MC meeting,]  
because [many people told her they had to leave early.]*

- Classificatie in termen van CCR-dimensies:
  - o positief
  - o causaal
  - o non-basic
  - o objectief
  - o anti-chronologisch
- Equivalent in PDTB met dezelfde waarden?  
=> *Reason* (met Arg1-Arg2 order)

## 7. Nederlands connectief-lexicon ontwikkeld

- Met Peter Bourgonje, Jet Hoek, Gisela Redeker, Ted Sanders en Manfred Stede
- Startpunt: Duits connectief-lexicon en een parallel corpus
- Daarna validatie met andere lijsten en handmatige controle door experts
- Met syntactische en semantische annotatie
- Lexicon: <https://github.com/discourse-lab/DisCoDict>
- Verantwoording van de totstandkoming:  
Bourgonje, P., Hoek, J., Evers-Vermeul, J., Redeker, G., Sanders, T., & Stede, M. (2018). Constructing a lexicon of Dutch discourse connectives. *Computational Linguistics in the Netherlands Journal*, 8, 163-175.

## 8. In wording: website

- Website over CCR als annotatie-framework
- Overzicht van artikelen over CCR
- Overzicht van geannoteerde corpora
- Overzicht van auteurs die met CCR gewerkt hebben