



### University of Groningen

### Towards a Dutch FrameNet lexicon and parser using the data-to-text method

Minnema, Gosse; Remijnse, Levi

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Publication date: 2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Minnema, G., & Remijnse, L. (2020). Towards a Dutch FrameNet lexicon and parser using the data-to-text method. Poster session presented at Computational Linguistics in the Netherlands, Utrecht, Netherlands.

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverneamendment.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 19-11-2022



# Levi Remijnse

Piek Vossen Antske Fokkens

# Gosse Minnema

Malvina Nissim Johan Bos



<u>l.remijnse@vu.nl</u> / <u>www.dutchframenet.nl</u> / <u>g.f.minnema@rug.nl</u>

# Towards a Dutch FrameNet lexicon and parser using the data-to-text method



# Study variation in framing of situations in texts # Reference-aware FrameNet annotation # DRT-based cross-lingual FrameNet parsing



SCAN ME

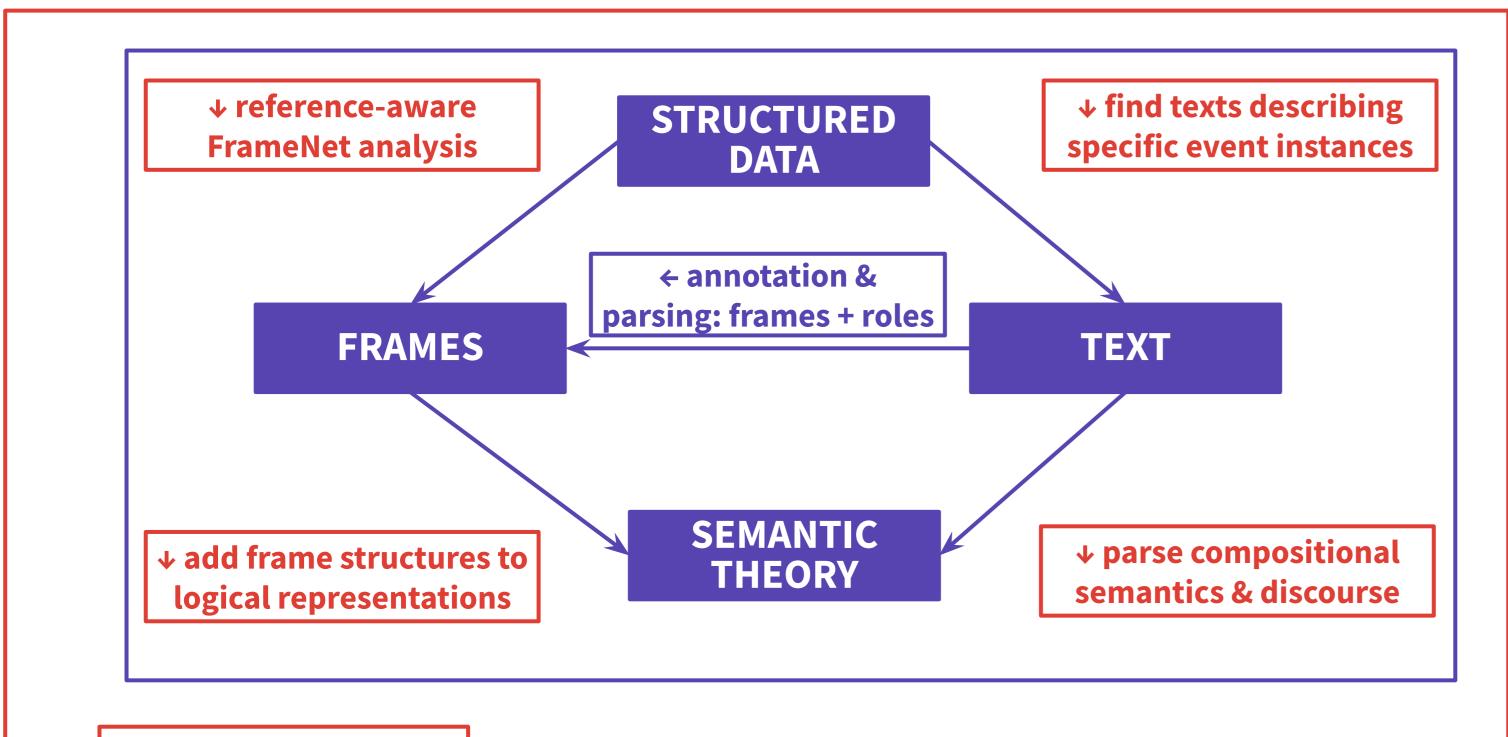
murder

**Germany** 

reference texts

Anis Amri

2016-12-19



### Berkeley FN: original FN = lexicon of framesproject for English **Frame** = lexical units + semantic roles for situation/event/concept FRAMENET BRASIL Frame Non-English projects

# Goals

- # How are similar situations framed by different texts?
- # Develop a FrameNet for Dutch
- # Semantic-pragmatic analysis of frame variation
- # Automatic frame parsing across languages

### **DATA-TO-TEXT PIPELINE** MICROWORLD Structured data **FrameNet** Wikidata Wiki incident type IncidentId: Q28036573 dominant frames dominant elements "murders" EventType: fn:Shoot\_projectiles@Projectile fn:Shoot projectiles Actor: Place: TimeStamp: Bag of frames incident<sub>1</sub> → microworld ← reference text<sub>1</sub> John Doe, murderer the gun...the funeral subevents **Scrape Wikidata for** incident<sub>2</sub> reference text<sub>2</sub> specific event types incident<sub>3</sub> → microworld <del><</del> reference textx Other frames reference texty temporal incident<sub>n</sub> reference text<sub>n</sub> Use Wikipedia to find

## **REFERENCE TEXT**

[EN] 2016 Berlin truck attack

On 19 December 2016, a truck was deliberately driven into the Christmas market next to the Kaiser Wilhelm Memorial Church at Breitscheidplatz in Berlin, leaving 12 people dead and 56 others injured. One of the victims was the truck's original driver, Łukasz Urban, who was found shot dead in the passenger seat. The truck was eventually stopped by its automatic brakes. The perpetrator was Anis Amri, a Tunisian failed asylum seeker. Four days after the attack, he was killed in a shootout with police near Milan in Italy. An initial suspect was arrested and later released due to lack of evidence. The event was designated as a terrorist attack. The Islamic State of Iraq and the Levant gave Anis Amri instructions. ISIL released a video of Amri pledging allegiance to the terror group's leader, Abu Bakr al-Baghdadi.

# Data-to-text [1]

- # Categorize basic level events [2]
  - # Query Wikidata for the registered incidents
- # Derive a microworld from the structured data
- # Obtain reference texts provided by Wikidata

# "BAG-OF-FRAMES"

Killing Offenses Weapon Use\_firearm Commit\_crime Attack

For every event type, automatically determine set of frames to be annotated

**Example:** murder event

# Annotation Problems

- # Show how frames provided by structured data are triggered in the text
- # Standard FrameNet annotation limited to one-to-one mapping targets ⇔ frames

## ONE TARGET, MANY FRAMES

HOSTILE\_ENCOUNTER [Side 1 he] was killed in a shootout [Side 2 with police] Weapon? Use\_Firearm?

### **DUTCH MORPHOLOGY**

Commit\_crime Attack

dat hij door de aanslag pleger was vermoord that he by the attack\_perpetrator was murdered "... that he was murdered by the perpetrator of the attack'

## ONE FRAME, MANY TARGETS

On 19 December 2016, a truck was deliberately driven into the Christmas Market [...] in Berlin, leaving 12 people dead and 56 others injured. ••••

deliberately + Killing = Offense? leave + dead = Killing?

# Automatic parsing

- # Integrate FrameNet and formal semantics (DRT)
- # Capture event and participant (co-)reference and pragmatic

NP/N (N/N)/(N/N) N/N

### FN-PARSING: OPEN-SESAME [3]

Hoover Dam played a major role in preventing Las Vegas from drying up

**Current SOTA** 

| and | a | third | term | in | [Cor | conj | NP/N | N/N | N/PP | PP/I

### **BOOTSTRAPPING DEEP FN PARSERS**

**BOOTSTRAPPING** STRATEGY #1 Use shallow parser, map to DRS structures

> **BOOTSTRAPPING** STRATEGY #2 Use SemLink [ref] to map **VerbNet to FrameNet**

**CHALLENGE #1:** Frame identification bottleneck [5]

**CHALLENGE #2:** Frame projection (English → Dutch, Italian)

# inference DRT: PARALLEL MEANING BANK [4] General~elections were + The elections were [Theme, Agent]

[Theme,Agent] [] [] (S[dci]\NP)/NP NP/N N

Tokens: CCG + VN roles Logical representation t1 x1 e1 x2 x3 x4 s1 x5 s2 s3 t2 e2 p1 s4 b7 :: time(t3) general\_election(x1) b5 :: election(x5) widely(s4) Attribute(x5, s2) t3 ≺ now MonthOfYear(t1, 05) Manner(e2, s4) Theme(e2, p1) Time(s5, t3) MonthOfYear(t1, 04) controversial(s2) Agent(e2, x5) Attribute(x5, s5) YearOfCentury(t1, 2000) Time(s2, t2) proposition(p1) Degree(s2, s3) Quantity(t1, 9) Goal: FN + VN roles