Levi Remijnse
Piek Vossen
Antske Fokkens

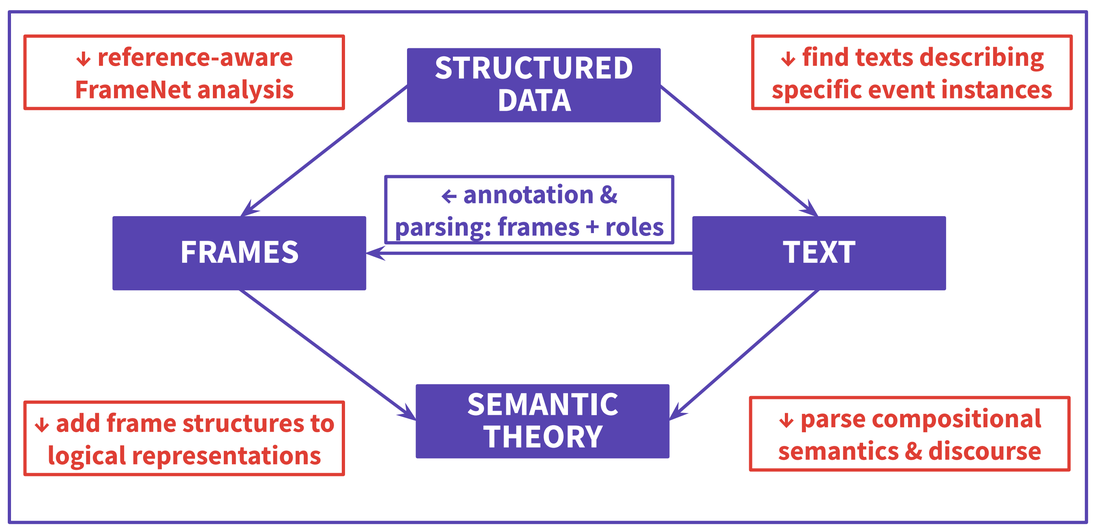
Gosse Minnema
Malvina Nissim
Johan Bos

l.remijnse@vu.nl / **www.dutchframenet.nl** / g.f.minnema@rug.nl

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



Berkeley FN: original project for English

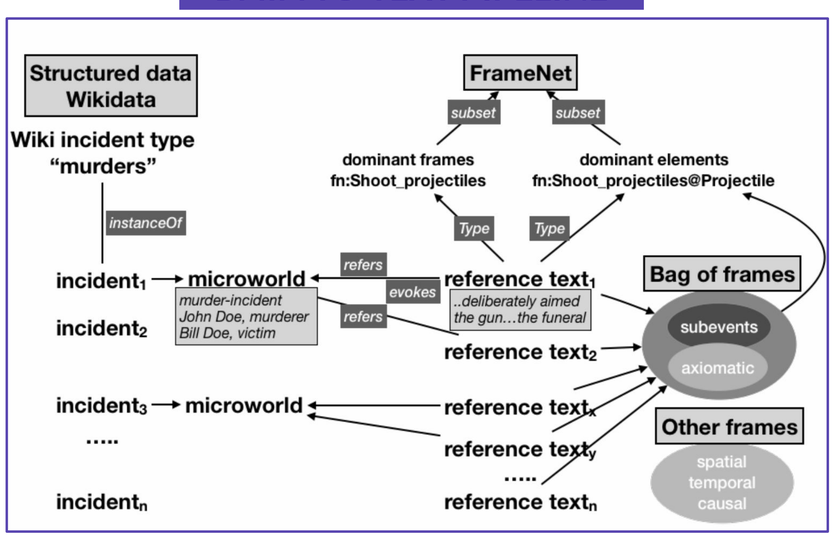
FN = *lexicon of frames*
Frame = *lexical units + semantic roles for situation/event/concept*

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

```
IncidentId:  Q28036573
EventType:   murder
Actor:       Anis Amri
Place:       Germany
TimeStamp:   2016-12-19
```

Scrape Wikidata for specific event types

Use Wikipedia to find reference texts

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

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.

"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

Attack — Commit_crime

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.

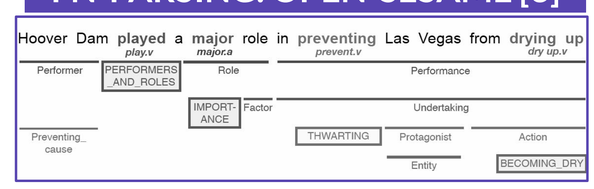
deliberately + Killing = Offense?

leave + dead = Killing?

Automatic parsing

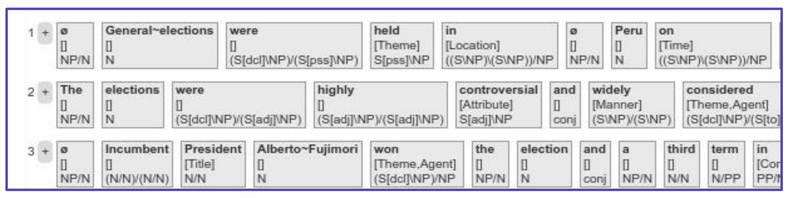
Integrate FrameNet and formal semantics (DRT)
Capture event and participant (co-)reference and pragmatic inference

FN-PARSING: OPEN-SESAME [3]



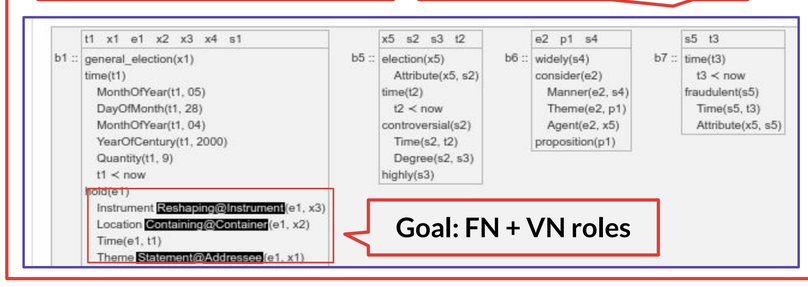
Current SOTA

DRT: PARALLEL MEANING BANK [4]



Tokens: CCG + VN roles

Logical representation



Goal: FN + VN roles

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)

References: [1] Vossen, P. Illievski, F. Postma, M., and Segers, R. (2018). Do not Annotate, but validate: a data-to-text method for capturing event data. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. [2] Morris, M. W. and Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & cognition*, 18(4);407-418. [3] Swayamdipta S., Thomson S., Dyer C., Smith N.A. (2017). Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold. https://arxiv.org/abs/1706.09528. [4] Abzianidze L., Bjerva L., Evang K., Haagsma H., Van Noord R., Ludmann P., Nguyen D., Bos J. (2017). The Parallel Meaning Bank: Towards a Multilingual Corpus of Translations Annotated with Compositional Meaning Representations. EACL. [5] Hartmann S., Kuznetsov I., Martin T., Gurevych I. (2017). Out-of-domain FrameNet Semantic Role Labeling.