

Introducing Frege to Fillmore

A FrameNet Dataset that Captures both Sense and Reference

Levi Remijnse, Piek Vossen, Antske Fokkens, Sam Titarsolej

LREC 2022

Introduction

Widely supported claim in semantics and philosophy:

Meaning arises from the combination of both sense and reference

(Kenny, 1995; Devitt & Sterelny, 1999; Frege, 1948)

We derive meaning from a linguistic expression by both disambiguating its sense and grounding its referent in the real world.

Introduction (2)

Within and across texts, we use a variety of conceptual representations for referencing the same entity through different lexemes.

Likewise, the same lexeme in the same sense can be used to refer to different entities.

- 1) A **member** of a Dutch wine tour [...] **tested** positive for **Covid-19** at the weekend. (DOC1)
- 2) He's probably the **Patient Zero** of the Winelands. (DOC1)
- 3) The virus could be **spread** to humans. (DOC2)
- 4) A Group of Ministers (GOM) on **COVID-19** is expected to convene on Monday. (DOC3)

Introduction (3)

Examples of subfields in NLP investing in reference part:

- (event) co-reference resolution (Filatova & Hatzivassiloglou, 2004; Choubey et al., 2018)
- Entity-linking (Hachey et al., 2013; Getman et al., 2018)

Examples of subfields in NLP investing in the conceptual part:

- Abstract Meaning Representation (AMR) (Banarescu et al., 2013)
- FrameNet (Ruppenhofer et al., 2010; Baker et al., 2003)

Yet, datasets that combine sense and reference are scarce.

Introduction (4)

Dutch FrameNet (DFN) annotation tool:

- Supports referential grounding of entity and event mentions across texts of the same event instances
- Provides the evoked frames and their frame elements.
- Variation of linguistic framing of real-world event instances within and across documents.
- Language independent.

This paper evaluates the first output of the tool:

- Inter-annotator-agreement (IAA)
- DFN lexicon

Referentially grounded corpora

- Most corpora contain a relatively small number of reference texts with low event co-reference. (Vossen et al. 2018)
- **Text-to-data**: labour intensive, time consuming, insufficient, no links to structured data
- Multilingual Wiki Extraction Platform (MWEP) (Vossen et al.,2020):
 - **Data-to-text**
 - Takes an identifier of an event type and queries Wikidata for event instances and corresponding reference texts
 - Returns both structured Wikidata and reference texts

FrameNet

- Lexicographic project
 - predicates -> semantic **frames**, e.g., schematic representations of meaning.
- Each frame contains a set of **lexical units (LUs)**
- Each LU is said to **evoke** that frame
 - e.g., Commerce_sell is evoked by *sell.v*, *auction.n*, *retail.v*
- Each frame contains **Frame Elements**, i.e., specified semantic roles

“Abby bought a car from Robin for \$5,000”

bought evokes the Commerce_buy frame

Commerce_buy consists of Frame Elements:

Abby = Buyer
a car = Goods
from Robin = Seller
for \$5,000 = Money

FrameNet (2)

Subsequent FrameNet initiatives:

- Salto (Burchardt et al., 2006)
- Webanno (Eckart de Castilho et al., 2016)
- Global FrameNet (Torrent et al., 2018)
 - Contributions from, e.g., Germany (Burchardt et al., 2009), Japan (Ohara et al., 2004) and France (Djemaa et al., 2016)

First attempt to initiate a **DFN lexicon** (Vossen et al., 2018):

- SoNaR corpus (Oostdijk et al., 2008)
- Dutch frame annotations on 116 documents
- with the aid of PropBank (De Clercq et al., 2012) relations
- Text-to-data fashion
- IAA: 47% for frames, 79% for frame elements

DFN annotation tool

- Enables researchers to analyze how in-text mentions and their evoked frames vary with respect to the entities in the world they reference, or how similar words reference different entities
- Loads a linguistically processed event corpus + structured data
- Displays one text at a time paired with the incident's structured data
- Two annotation types:
 - Instance-linking of in-text mentions to structured data
 - FrameNet annotation
- Annotation departures from structured data:
 - Discourse annotation: frame elements across sentences
 - Context: enhances frame identification
- Language independent with the use of markable correction:
 - Idioms, phrasal verbs, compounds

test_release mass shooting@en (Q21480) 2019 Utrecht shooting@en (en Deadly Utrecht attack leads

Select Annotation Type: Save Clear selection

Select Annotation Task:

Select Frame Type:

Select Relation type:

Select Typicality range:

Deadly Utrecht attack leads to heightened security for Dutch Jews (source)

AMSTERDAM — Security around Dutch Jewish community buildings was increased significantly as police searched for the gun man suspected of killing at least three people in what the prime minister said may have been a terrorist attack in Utrecht . The incident , in which several people were wounded , happened Monday morning inside a tram in the central city near Amsterdam . Police released a picture of a 37-year - old Turkey - born man named Gokman Tanis whom they suspect of shooting multiple people before fleeing . Using the handle " HateDemocrat , " Tanis has threatened violence in the past against PowNed , a right - wing news website . The Federative Jewish Netherlands group wrote on Twitter that elite police troops are patrolling Jewish community buildings with semi - automatic weapons . In Utrecht , the government for the first time ever raised the terrorism threat level to 5 , its highest state . Prime Minister Mark Rutte said it " may be a terrorist attack and an attack on our democracy . " cnxps.cmd.push(function () { cnxps ({ playerId : ' 36af7c51 - 0caf-4741 - 9824 - 2c941fc6c17b ' }) .render('4c4d856e0e6f4e3d808bbc1715e132f6 ') ; }) ; if(window.location.pathname.indexOf("656089 ") != -1){console.log("hedva connatix");document.getElementsByClassName("divConnatix")[0].style.display = " none " ; } Mosques in Utrecht were temporarily shuttered , De Telegraaf . Utrecht 's Chabad House also was not operating on Monday , but only because there were no activities scheduled for the day , a spokesman told the Jewish Telegraphic Agency . But Chabad Utrecht said Purim celebrations are to take place later this week as scheduled .

Structured Data
 incident type: mass shooting@en (Q21480300)
 incident ID: 2019 Utrecht shooting@en (Q62090804)
 hasPlace: Utrecht
 hasActor: Gökmen Tanis, victims

Chosen Frame
 Label:
 PreMON URI:
 FrameNet URL:

Selected predicate
 Label: Attack
 Term POS: NOUN
 Premon: Click here
 FrameNet: Click here
 Predicate ID: pr49
 Frame Relation: type

Frame Element	Role Type	Annotated	Expressed
Victim	Core	true	true
Assailant	Core	true	true

Dutch police arrest Turkish man suspected of killing three in tram shooting (source)

UTRECHT, Netherlands (Reuters) - Dutch police arrested a **Turkish man suspected of shooting dead three people and wounding five** on a tram in the Dutch city of **Utrecht** on Monday . Police said **the suspect** , **37-year - old Gokmen Tanis** , had been taken into custody after an hours - long manhunt and had earlier run - ins with authorities in the Netherlands . The city was put into lockdown after **the shooting** , shortly after the morning rush hour , which authorities initially said was an apparent terrorist attack . Police conducted raids in several locations after issuing an image of **Tanis** and warning the public not to approach **him** . But hours after **the shooting** , **the gunman** 's motive remained unclear . A prosecutor said it could be for " family reasons " and Turkey 's state - run Anadolu news agency , quoting relatives of **the gunman** , said **he** had fired at a **relative** on the tram and had then shot at **others who tried to help her** . The Turkish intelligence agency said it is investigating whether **the attack** was personally motivated or an act of terrorism , President Tayyip Erdogan said . Earlier on Wednesday , Dutch authorities had raised the terrorism threat in Utrecht province to its highest level , schools were told to shut their doors and paramilitary police increased security at airports , other vital infrastructure and at mosques . The threat level was reduced by one notch to where it had been after **the suspect** was arrested , the National Counter - Terrorism Agency said . Slideshow (23 images) Dutch Prime Minister Mark Rutte convened crisis talks immediately after **the incident** , which came three days after a lone gunman killed 50 people in mass shootings at two mosques in the city of Christchurch , New Zealand . " Our country has today been shocked by **an attack** in Utrecht ... A terrorist motive can not be excluded , " Rutte said . " The first reports have led to disbelief and disgust . **Innocent people** have been struck by violence . " The mayor of Utrecht , Jan van Zanen , said **three people** had been killed and **five** injured , **three** of them critically . CRIMINAL RECORD **The suspect** had previously been arrested , regional prosecutor Rutger Jeurden told reporters , without giving further details . **The shooting** took place in **Kanaleneiland** , **a quiet residential district on the outskirts of Utrecht with a large immigrant population** . Slideshow (23 images) " It 's frightening that something like **this** can happen so close to home , " said Omar Rahhou , who said his parents lived on a street cordoned off by police . " These things normally happen far away but this brings it very close , awful . " Witness Daan Molenaar , who said he had been sitting at the front of the tram when **the shooting** started , told national broadcaster NOS he did not believe **it** was a terrorist attack . " The first thing I thought was , **this** is some kind of revenge or something , or **somebody** who 's really mad and grabbed a pistol , " he said . Utrecht , the Netherlands ' fourth largest city with a population of around 340,000 , is known for its picturesque canals and large student population . Gun killings are rare in Utrecht , as elsewhere in the Netherlands .

Structured Data

incident type: [mass shooting@en \(Q21480300\)](#)

incident ID: [2019 Utrecht shooting@en \(Q62090804\)](#)

hasPlace: [Utrecht](#)

hasActor: [Gökmen Tanis, victims](#)

Notes

Save Notes

Figure 3: Snapshot of the DFN annotation tool's reference text box (left) with instance-linked mentions in bold. They are linked to instances in the structured data box (right). By tagging *the shooting*, it marks green and its annotated link to the incident ID marks purple.

Dutch police arrest Turkish man suspected of killing three in tram shooting (source)

UTRECHT , Netherlands (Reuters) - Dutch police **arrested** a Turkish **man** **suspected** of **shooting** dead three **people** and **wounding** five on a tram in the Dutch city of Utrecht on Monday . Police said the **suspect** , 37-year - old Gokmen Tanis , had been **taken into custody** after an hours - long **manhunt** and had earlier **run - ins** with authorities in the Netherlands . The city was put into **lockdown** after the **shooting** , shortly after the morning rush hour , which authorities initially said was an apparent **terrorist attack** . Police conducted **raids** in several locations after **issuing** an image of Tanis and **warning** the public not to **approach** him . But hours after the **shooting** , the **gunman** 's motive remained unclear . A prosecutor said it could be for " family reasons " and Turkey 's state - run Anadolu news agency , quoting relatives of the gunman , said he had **fired** at a relative on the tram and had then **shot** at others who tried to **help** her . The Turkish intelligence agency said it is **investigating** whether the **attack** was personally motivated or an **act of terrorism** , President Tayyip Erdogan said . Earlier on Wednesday , Dutch authorities had **raised** the **terrorism threat** in Utrecht province to its highest level , schools were **told to shut** their doors and paramilitary police **increased security** at airports , other vital infrastructure and at mosques . The **threat** level was **reduced** by one notch to where it had been after the **suspect** was **arrested** , the National Counter - Terrorism Agency said . Handout still image taken from CCTV footage shows a man who has been named as a **suspect** in Monday 's **shooting** in Utrecht , Netherlands , in a still image from CCTV footage released by the Utrecht Police on March 18 , 2019 REUTERS / Utrecht Police / Handout via Reuters Dutch Prime Minister Mark Rutte **convened** crisis **talks** immediately after the incident , which came three days after a lone gunman killed 50 people in mass shootings at two mosques in the city of Christchurch , New Zealand . " Our country has today been **shocked** by an **attack** in Utrecht ... A **terrorist** motive can not be excluded , " Rutte said . " The first reports have led to **disbelief** and **disgust** . Innocent people have been struck by **violence** . " The mayor of Utrecht , Jan van Zanen , said three people had been **killed** and five **injured** , three of them critically . CRIMINAL RECORD The **suspect** had previously been **arrested** , regional prosecutor Rutger Jeurden told reporters , without giving further details . The **shooting** took place in Kanaleneiland , a quiet residential district on the outskirts of Utrecht with a large immigrant population . " It 's **frightening** that something like this can **happen** so close to home , " said Omar Rahhou , who said his parents lived on a street **cordoned** off by police . " These things normally **happen** far away but this brings it very close , awful . " Slideshow (22 Images) Witness Daan Molenaar , who said he had been **sitting** at the front of the tram when the **shooting started** , told national broadcaster NOS he did not believe it was a **terrorist attack** . " The first thing I thought was , this is some kind of **revenge** or something , or somebody who 's really mad and **grabbed** a **pistol** , " he said . Utrecht , the Netherlands ' fourth largest city with a population of around 340,000 , is known for its picturesque canals and large student population . Gun **killings** are rare in Utrecht , as elsewhere in the Netherlands .

Figure 4: Snapshot of the DFN annotation tool's reference text box (same text as Figure 3) with frame annotations and examples of frame element annotations on discourse level. The blue markables are frame annotated predicates. *shooting* is tagged as a lexical unit evoking KILLING. *man* is annotated as expressing the frame element Killer, *people* is annotated with Victim and *pistol* is annotated with Instrument. All frame elements occur in different clauses than their frame-evoking predicate.

ID	Sentence	Evokes
1	Dit is geen gewone liquidatie <i>This is no ordinary liquidation</i>	KILLING
2	er is geschoten in een tram <i>there was a shooting in a tram</i>	HIT_TARGET
3	[...] de slachtoffers van de aanslag in Utrecht <i>[...] the victims of the attack in Utrecht.</i>	ATTACK
4	[...] de krankzinnige moord partij in Utrecht <i>[...] the insane murder party in Utrecht</i>	PARTICIPATION
5	Hij wordt verdacht van betrokkenheid bij het schiet incident <i>He is suspected of involvement in the shooting incident</i>	CATASTROPHE

Table 4: An example of variation in framing of an event instance. The Dutch example sentences are taken from reference texts referencing the Utrecht shooting (Wikidata identifier: Q62090804). The first column indicates the sentence identifier. The second column shows the example sentence with English translation in italics and the frame-evoking predicates in bold. The value of the third column is the evoked frame. The boldfaced predicates are all instance-linked to the main event in structured data and thus show co-reference.

Recourses

Corpus data and corresponding structured data obtained by the use of MWEP: reference texts for incidents grouped under 13 event types.

event type (QID)	#Li	#En.Ri	#Du.Ri	Avg. #Ri per Li
riot (Q124757)	73	494	65	7.7
mass shooting (Q21480300)	88	822	70	10.1
legal case (Q2334719)	39	455	4	11.8
auto race (Q24050099)	9	62	0	6.9
economic crisis (Q290178)	4	123	0	30.8
disease outbreak (Q3241045)	2	198	358	278
royal wedding (Q63442071)	17	350	0	20.6
aircraft shootdown (Q6539177)	1	183	135	318
natural disaster (Q8065)	1	64	19	83
storm (Q81054)	60	318	0	5.3
presidential election (Q858439)	111	420	0	3.8
music festival (Q868557)	14	650	49	49.9

Table 1: Descriptive statistics regarding the DFN corpus. The first column indicates the event types and corresponding Wikidata identifier. The second column, Li, indicates the number of event instances that belong to the event type. The third and fourth columns, Ri, present the total number of English and Dutch reference texts, each referring to one of the event instances. Finally, the average number of reference texts per event instance are shown.

Annotation process

- Month 1-3
 - Four annotators
 - Two event types: *mass shooting, aircraft shootdown*
- Month 4
 - Six event types: *mass shooting, aircraft shootdown, disease outbreak, riot, natural disaster, music festival*
- Month 5-6
 - Six annotators

Annotation process (2)

Per reference text:

- Instance-linking to structured data
- Frame annotation (frames and frame elements)
- Markable correction

- Absent frame elements from text are annotated as **unexpressed**
- Remaining **subevents** with no instance-link (following criteria from O’Gorman et al., 2016; Caselli & Vossen, 2017)

For the Dutch texts, the tool’s dynamic lexical lookup initiated a DFN lexicon in which every novel entry is saved and continuously proposed with every tag of the same markable.

Results and discussion

Descriptive stats:

- 326 annotated texts
 - 276 Dutch
 - 50 English
- 27.533 mentions:
 - 9.220 tokens
 - 2.729 different lexical units
 - 574 different frames (avg. 16.06 annotations per frame) ←
 - 7.457 (27%) instance-links
- 1.840 (19.9%) markable corrections (avg. 5.6 per text):
 - 699 multi-words
 - 1141 compounds ←

IAA

- 15 Dutch reference texts annotated by multiple annotators
- Instance-links and frames almost perfect
 - Structured data guides the annotators
- Frames with no instance-links almost perfect
 - Knowledge of event type still provides sufficient context for frame identification
- Frame elements moderate
- Strong differences from Vossen et al. (2018)

1. instance-links	
span matches	89.5%
agreement on span matches	89.4%
2. frames	
span matches	73.7%
agreement on span matches	91.9%
similarity in disagreement	0.59
similarity p-value	0.07
3. frames and instance-links	
joint agreement	97.58%
disjoint agreement	89.94%
joint:disjoint ratio	1:2.3
4. frame elements	
agreement incl. unexpressed FE's	69.5%
agreement excl. unexpressed FE's	94.0%

Table 2: Overview of IAA, providing the following information: span matches and agreement for those span matches on instance-links and frames; similarity score and similarity p-value for the confused frame annotations (range 0,1); joint agreement (mentions with both instance-link and frame) and disjoint agreement (mentions with only a frame), and joint:disjoint ratio; and the agreement for frame elements, both including and excluding *unexpressed* frame elements. To take the number of annotations per document into account, the averages over the documents are weighted by the number of annotations.

Disagreement on frames

- Similarity score: 0.6
(Sikos & Padó, 2018)
- When no absolute agreement, the frames still show strong conceptual feature overlap

frame pair	s
OPERATE_VEHICLE RIDE_VEHICLE	0.80
EXPERIENCE_BODILY_HARM CAUSE_HARM	0.75
SCRUTINY INSPECTING	0.72
ARREST BECOMING_AWARE	0.49
NOTIFICATION_OF_CHARGES JUDGMENT	0.47
OBJECTIVE_INFLUENCE IMPACT	0.45

Table 3: Examples of frame pairs the annotators confused, including their similarity score.

Discourse annotation

- Low IAA on frame elements due to discourse annotation
- 99.8% of all annotated frames contains at least one sentence-external frame element
- IAA sentence versus discourse
 - Highest agreement when sentence-internal
- Peak at 30-40 sentences: participants are introduced at the onset and implicated later on

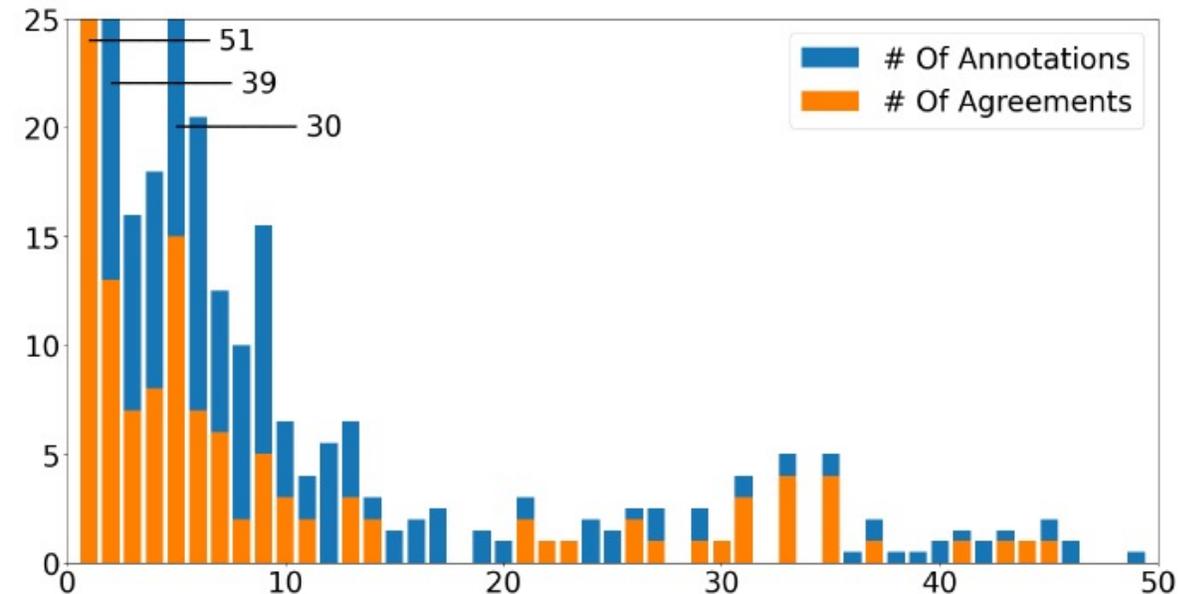


Figure 1: Number of sentence-external frame elements with the distance in sentences to the annotation of the frame. The figure includes the agreement score for each distance to the frame.

DFN lexicon

- Continuous steep increase of lexical entries despite working longer within the same event types
- Small bump when new event types are introduced
- Long tail of variation

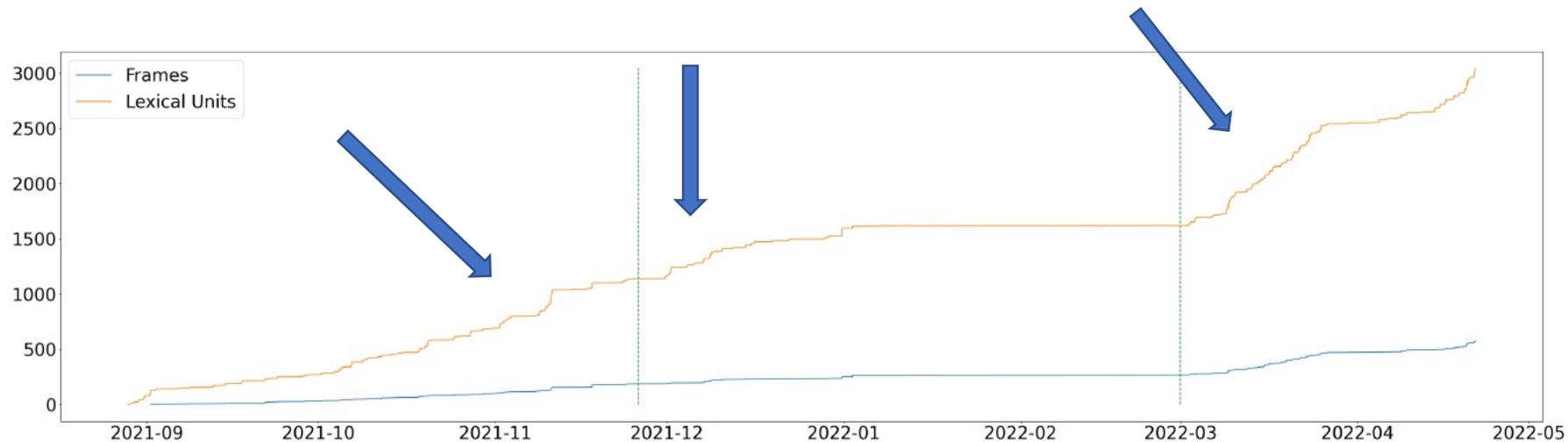


Figure 2: Distribution of new DFN lexicon entries over time, from the beginning to the end of the annotation appointment. The first green vertical line indicates the moment that the annotations of event types extended from four to six. The second green line indicates the moment two more annotators were added.

Conclusions

- Frame annotation in data-to-text fashion is by far more consistent compared to traditional text-to-data methods
- Frame element annotation suffers due to the ambiguity of candidates in discourse
- The annotation output (DFN lexicon) shows a head of domain-specific frames, but also a long tail of variation