

# Introducing Frege to Fillmore: A FrameNet Dataset that Captures both Sense and Reference

Levi Remijnse<sup>♣</sup>, Piek Vossen<sup>♣</sup>, Antske Fokkens<sup>♣◇</sup>, Sam Titarsolej<sup>♣</sup>

<sup>♣</sup> Vrije Universiteit Amsterdam, Computational Linguistics and Text Mining Lab

<sup>◇</sup> Eindhoven University of Technology, Dep. of Mathematics & Computer Science

{l.remijnse, piek.vossen, antske.fokkens}@vu.nl, s.titarsolej@gmail.com

## Abstract

This article presents the first output of the Dutch FrameNet annotation tool, which facilitates both referential- and frame-annotations of language-independent corpora. On the referential level, the tool links in-text mentions to structured data, grounding the text in the real world. On the frame level, those same mentions are annotated with respect to their semantic sense. This way of annotating not only generates a rich linguistic dataset that is grounded in real-world event instances, but also guides the annotators in frame identification, resulting in high inter-annotator-agreement and consistent annotations across documents and at discourse level, exceeding traditional sentence level annotations of frame elements. Moreover, the annotation tool features a dynamic lexical lookup that increases the development of a cross-domain FrameNet lexicon.

**Keywords:** frame semantics, reference, events, annotation tool, lexicon

## 1. Introduction

A widely supported claim in the fields of semantics and philosophy is that meaning arises from the combination of sense and reference (Kenny, 1995; Devitt and Sterelny, 1999; Frege, 1948). We derive meaning from a linguistic expression by both disambiguating its sense and grounding its referent in the real world. This way, when we talk or write about real-world event instances, we use language to construct or interpret narratives around those event instances and their participants.<sup>1</sup> Within and across written texts, we use a variety of conceptual representations for referencing the same real-world entity through different lexemes and expressions. Likewise, the same lexeme in the same sense can be used to refer to different entities. See the examples in (1).

- (1) a. A **member** of a Dutch wine tour [...] **tested** positive for **Covid-19** at the weekend. (2020)
- b. He 's probably the **Patient Zero** of the Winelands. (2020)
- c. The virus could be **spread** to humans. (2020)
- d. A Group of Ministers (GOM) on **COVID-19** is expected to convene on Monday. (2020)

In (1a) and (1b), taken from the same document, *member* and *Patient Zero* co-refer to the same referent, one focusing on a membership aspect, the other focusing on a medical aspect. Similarly, *tested positive* and *spread* in 1a and 1c, taken from different documents, could co-refer to the same event instance. Yet, in (1a) and (1d),

the same mention *COVID-19* in the same sense (that of disease) refers to different entities: in (1a) the referent is the virus itself, but in (1d), the referent is the pandemic, the outbreak of the virus. These examples are a small share of the continuous variation in both sense and reference we produce in order to derive meaning.

Different subfields of NLP invest in the reference part of semantics, e.g., event co-reference resolution annotates data with information about which mentions co-refer to an event (Filatova and Hatzivassiloglou, 2004; Choubey et al., 2018); entity-linking is the task of linking in-text mentions to entries of entities in a knowledge base (Hachey et al., 2013; Getman et al., 2018). Other subfields are concerned with the conceptual part of annotated data, e.g., FrameNet, a paradigm initiated by Charles Fillmore, (Ruppenhofer et al., 2010; Baker et al., 2003) disambiguates words with semantic frames;<sup>2</sup> Abstract Meaning Representation (AMR) represents a sentence's lexical concepts and their typed relations as a graph (Banarescu et al., 2013). Yet, a dataset that does justice to the above explained claim about semantic meaning would exhibit annotations of both sense and reference, thus informing us about an expression's referent and on top of that, describing its conceptual representation through evoked frames. This dataset would then require both structured information where the expressions can be linked to annotate reference, and a lexicographic knowledge base for annotating conceptual representations.

We applied these requirements while producing data in the Dutch FrameNet (DFN) project,<sup>3</sup> for which we created a dedicated annotation tool that supports referential grounding of entity and event mentions across texts

<sup>1</sup>In this paper, we use the term *event instance* for event instances of a specific event type, e.g., an instance of shooting.

<sup>2</sup>Explore the database in <https://framenet.icsi.berkeley.edu/fndrupal/>

<sup>3</sup><http://dutchframenet.nl>

of the same event instances, while providing the evoked frames and their frame elements.

This paper evaluates the first output of using the DFN annotation tool (Postma et al., 2020) to combine FrameNet annotations with links of in-text mentions to structured data. This way, annotators are enabled to create a dataset that captures variation of linguistic framing of real-world event instances within and across documents. Furthermore, the tool is language independent, which means that texts and FrameNet lexicons of any language can be loaded and utilized (in this paper we focus on English and Dutch).

The data and corpora are compiled with the use of the Multilingual Wiki Extraction Platform (MWEP), a **data-to-text** method that takes identifiers of event types as input to collect structured Wikidata of corresponding event instances and crawl reference texts (Vossen et al., 2020). The structured information, visualized in the tool, guides the annotator in frame identification, resulting in high inter-annotator-agreement (IAA). Furthermore, the tool facilitates a dynamic lexical lookup: whenever a markable is annotated with a frame, the lexicon continues to propose its entry when a different token of the same type of markable is tagged. We show that, since the tool loads corpora on the basis of specified event types, the development of a DFN lexicon follows domain-specific annotations. Loading and annotating cross-event type corpora then results in a cross-domain database.

Our contributions are as follows:

- we created and pre-processed a referentially grounded corpus in both English and Dutch, with multiple texts per event instance and the event instances grouped under event types;
- we present the first output of the DFN annotation tool, which consists of both links of in-text mentions to structured data and frame annotations on our corpus;
- we show that annotation with the aid of structured data results in high agreement on both the linking of events and entities, and frame identification;
- we show that the frame annotations increase the volume of the Dutch FrameNet lexicon in cross-domain fashion.

This paper is further structured as follows. We first discuss related work and background in Section 2. We then introduce our methodology in Section 3. Section 4 provides the results of the annotation process, which we discuss in Section 5. We conclude in Section 6.

## 2. Related work

In this section, we discuss previous work regarding referentially grounded corpora (2.1), AMR with Wikification (2.2), FrameNet (2.3) and the DFN annotation tool (2.4).

### 2.1. Referentially grounded corpora

With respect to referential grounding, in recent years, researchers became aware of coverage issues with respect to corpora. Most of the corpora contain a relatively small number of reference texts. Vossen et al. (2018b) evaluate the nine most prominent text corpora (e.g., OntoNotes (Pradhan et al., 2007), ECB (Bejan and Harabagiu, 2010), ECB+ (Cybulska and Vossen, 2014), ACE2005 (Peng et al., 2016)) and show that their sum comprises of less than four thousand documents of which the largest part (over 1100 documents) is subsumed by OntoNotes. Event co-reference within and across reference texts remains low (10 mentions per document on average, and only a subset of texts contains cross-document reference), even after recent attempts of extensive manual annotation (Cybulska and Vossen, 2014; Song et al., 2015; O’Gorman et al., 2016). Event co-reference for these datasets is established in **text-to-data** fashion, i.e., starting from text to derive annotation sentence-by-sentence, which is evaluated by Vossen et al. (2018b) as labour intensive, time consuming and insufficient, with no link to structured event instance information and thus no way of grounding references in the real world.

Vossen et al. (2020) reversed the text-to-data method of building event corpora with the aim to make the process less labour intensive and more efficient, with attention to a high ratio of reference texts per event instance and the aim to incorporate structured information to link references. They implemented the Multilingual Wiki Extraction Platform (MWEP) as the first **data-to-text** method in NLP that results in referential grounding. This pipeline takes identifiers denoting event types and specified languages as input to query Wikidata (Vrandečić and Krötzsch, 2014) for corresponding event instances. Per event instance, MWEP not only returns structured Wikidata, but also primary reference texts crawled from the event instance’s language specific Wikipedia page, linguistically processed by SpaCy (Honnibal et al., 2020) and stored in the NLP Annotation Format (NAF) (Fokkens et al., 2014). The first run of MWEP returned tens of thousands of event instances for ten event types, with Wikipedia pages and their primary reference texts in English, Italian and Dutch.

Remijnse et al. (2021) created Historical Distance Data (HDD) as a second implementation of MWEP. They aggregated hundreds of reference texts for event instances of the contrasting event types *presidential election*, *storm*, *music festival* and *gun violence*. They adapted the pipeline to import the gun violence subcorpus from the Gun Violence Archive (Ko, 2018), and applied statistics of historical distance between the document creation time and the date of its event instance. For their research purposes, they only crawled English reference texts and did not utilize the structured data.

Since we want to create an annotated dataset in both Dutch and English, in which text mentions are

grounded in the real world, we used MWEP to create a diverse corpus with a large volume of reference texts in the respective languages accompanied with structured Wikidata.

## 2.2. Combining sense data with referential data

Earlier work on linking sense data with referential data can be found in the field of AMR (Banarescu et al., 2013). AMRs are English graph-based representations of sentence-level lexical concepts and their typed relations. They integrate several semantic features, e.g., coreference, modality and negation, in a single structure, abstracting over syntactic variation. AMR structures align with PropBank (De Clercq et al., 2012) predicate-argument semantics. Mostly in unsupervised fashion, AMR has been enriched with extensions for entity linking to Wikipedia. Specifically, proper names as arguments are linked to Wikipedia entries (for examples of this implementation, see (Pan et al., 2015; Van Noord and Bos, 2017; Damonte et al., 2017)). In this paper, for the first time, we use FrameNet as a lexical semantic resource in combination with reference data. FrameNet is designed to be cross-linguistically applicable, thus also to Dutch. Moreover, The entity links in our dataset are manually acquired, including both proper names and pronouns. This way, we get a complete overview of variation in framing of all referents across documents.

## 2.3. FrameNet

FrameNet is a lexicographic project built on the hypothesis that people interpret the conceptual meaning of words against semantic **frames**. In the English FrameNet database (Ruppenhofer et al., 2010), frames represent schematized situations involving highly specified semantic roles called **frame elements**. Frames are **evoked** by **lexical units**, i.e., words in one of their senses. FrameNet takes predicates as a point of departure when performing frame annotations, which results in datasets consisting of isolated sentences with the predicate being a lexical unit evoking the frame and syntactic arguments expressing the frame elements. (2) shows a FrameNet representation of (1a) and (1c).

- (2) a. EXPERIMENTATION  
 [*Subjects* A member of a Dutch wine tour]  $\odot$ tested [*Result* positive] [*Topic* for Covid-19] [*Time* at the weekend].
- b. MOTION  
 [*Theme* The virus] could be  $\odot$ spread [*Goal* to humans].

In (2), *tested* is annotated with the EXPERIMENTATION frame and *spread* is annotated with MOTION. The syntactic arguments are annotated with frame elements. *Subjects*, *Topic*, *Theme* and *Goal* are considered **core** frame elements, i.e., their overt expression is mandatory for the reader to cognitively process the frame.

Other frame elements, like *Result* and *Time* are peripheral, i.e., they modify the frame but are not mandatory. English FrameNet was the first implementation of a frame semantic annotated resource. All subsequent initiatives of FrameNet annotation tools are variations of the English FrameNet setup. Moreover, all FrameNet lexicons for other languages employ the English FrameNet’s database, keeping its frames and altering its lexical units. Salto (Burchardt et al., 2006) is a multi-level annotation tool that after a syntactic analysis of a sentence provides the option for dragging and dropping frames and frame elements to the annotated constituents. Webanno (Eckart de Castilho et al., 2016) is a web-based annotation tool that mainly focuses on the relation between syntactic and semantic structures, with the option to introduce constraint settings that increase annotation pace. Global FrameNet (Torrent et al., 2018) takes the annotation setup of English FrameNet to the multilingual level. FrameNets of many different languages have already contributed to this project, such as German (Burchardt et al., 2009), Japanese (Ohara et al., 2004) and French (Djemaa et al., 2016).

As a first attempt to initiate a DFN corpus and lexicon, Vossen et al. (2018a) used SoNaR (Oostdijk et al., 2008), a corpus exhibiting a large variety of Dutch documents, and performed Dutch frame annotations on 116 documents with the aid of previously annotated PropBank (De Clercq et al., 2012) relations. In text-to-data fashion, they annotated 4,755 different lexical units distributed over 671 frames and showed 47% IAA, which is generally considered weak. Yet, given agreement on the frame, the agreement on the frame elements was a moderate 79%. The authors conclude that the low agreement is an effect of the text-to-data method because of which the annotators were unaware of the context or text genre, and needed to continuously consider all FrameNet frames. The higher agreement on frame element annotation is a consequence of the agreement on the frame, which makes frame element identification easier. In the current study, we aim to solve this problem by the use of a data-to-text method, enhancing the annotators with structured data and hence, guiding the annotation process.

All FrameNet implementations discussed in this section operate on a conceptual level, which means deriving word meaning and semantic role distribution from text. FrameNet does not focus on the reference part of meaning. Thus, while we obtain the semantic frames for the mentions *tested* and *spread* in (2), we do not gain insight as to whether the mentions reference the same real-world event instance. Similarly, the mention *member* is a *Subject* of EXPERIMENTATION, but we do not know whether it co-refers with *Patient Zero* in (1b) to the same real-world entity. In the next section, we discuss the annotation tool introduced by Postma et al. (2020) that meets the requirements for such data creation.

## 2.4. Dutch FrameNet annotation tool

In order to enable 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, Postma et al. (2020) present the Dutch FrameNet (DFN) annotation tool. This tool loads a linguistically processed event corpus aggregated by MWEP and displays dropdowns in its interface leading the annotator to a subcorpus of reference texts in a specified language and belonging to a specific event instance of a specific event type. The corpus is accompanied with structured data per event instance. The tool displays one text at a time, paired with the event instance’s structured data. For the presented text, it then facilitates two annotation types: linking of in-text mentions to both structured events and entities (Wikidata entries), which we cover by the notion **instance linking**, and frame annotation with the use of the canonical version 1.7 of FrameNet. Table 4 in the Appendix shows an illustrative example of the combined annotation. The resulting annotation scheme of the text that is saved in NAF exhibits the instance-links as well as the frames and frame elements. Annotating the whole document collection of an event instance then results in a collection of annotation schemes in which all possible mentions of the structured data are both instance-linked and frame annotated.

Postma et al. (2020) import the English FrameNet frame database and therefore follow the tradition of regarding these frames as the universal standard when it comes to creating coverage. Yet, the DFN annotation tool shows a major deviance from English FrameNet in its setup: the annotation departs from an event instance’s structured data. This has the following implications.

On a technical level, the annotator has to look for instance-links across sentences and thus also for frames and frame elements across sentences. While frames are still evoked by predicates, their frame elements can be looked for throughout the discourse and across coreferential mentions. The aim of this tool is to capture how event instances and their participants are framed within and across reference texts, which entails annotation across sentences. Core frame elements that are absent from a text are registered as *unexpressed*.

On a cognitive level, the structured data provides the annotator with context. We believe this context enhances frame identification, resulting in high IAA. We will evaluate this agreement in Section 5.

The tool is designed to accommodate language independent corpora by facilitating manual markable correction of multiple tokens forming one semantic unit, e.g., idioms and phrasal verbs (Lexicon of Linguistics, 2020b), (Quirk, 2010). Likewise, the annotator can apply this feature to split single tokens that are composed of multiple semantic units, e.g., endocentric compounds (Lexicon of Linguistics, 2020a), making it possible to annotate those units with frames or

frame elements in line with the proposal and dataset by Ponkiya et al. (2018) and Ponkiya et al. (2021). This is of particular importance to Dutch compounds, since they orthographically form one unit.

This paper builds upon Postma et al. (2020) by discussing the first output of the DFN tool. In the following section, we discuss the procedure of corpus acquisition, annotation process and data analysis.

## 3. Methodology

In this section, we describe the methodology used in order to get the first DFN annotation tool output and data analysis. This includes resources (3.1), the annotation process (3.2) and evaluation (3.3).

### 3.1. Resources

Following (Fokkens et al., 2013), the model for our data relies on three main concepts: event type, event instance, and reference text with event mentions<sup>4</sup> Let  $E$  be a set of event types, let  $I$  be a set of real-world event instances, and let  $R$  denote a registry of reference texts. Each real-world event  $L_i \in I$  is an instance of one or more event types. Also, there can be reference texts that refer to a particular real-world event instance  $L_i$ .

The reference texts are located, retrieved, and processed by applying the following steps. First, we make use of the Internet archive Wayback Machine.<sup>5</sup> Second, we apply news-please (Hamborg et al., 2017) to crawl the reference text. Finally, we process the text using spaCy (Honnibal et al., 2020) for sentence splitting, tokenization, lemmatization and dependency parsing. For Dutch reference texts, spaCy trains the syntactic dependency parser Dutch LassySmall v2.5 (Bouma and van Noord, 2017; Van Noord et al., 2013) to unite the components of phrasal verbs.

Following our model, we obtained data for our DFN corpus by applying MWEP on thirteen Wikidata event types. We selected event types that differ in conceptual features in order to enrich the DFN lexicon with annotations covering different domains.

Table 1 shows descriptive statistics of the application of our software. We obtained strong variation in both the number of event instances per event type and the average number of reference texts per event instance. Similar to Vossen et al. (2020), we find that event types generating less event instances return a higher number of documents per event instance, e.g., compare *aircraft shutdown* to *presidential election*. We also find that Wikidata and Wikipedia facilitate mostly English texts. Our software returns the reference texts paired with structured data per event instance. Both data are loaded in the DFN annotation tool.

<sup>4</sup>In the following, we use event and event instance interchangeably.

<sup>5</sup><https://web.archive.org/>

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.

### 3.2. Annotation process

Four annotators performed annotations on English and Dutch reference texts in the DFN annotation tool for four months, eight hours per week. In the first months, they annotated texts grouped under the event types *mass shooting* and *aircraft shootdown*. From the onset of the annotation process, the tool’s dynamic lexical lookup initiated a DFN lexicon, in which every novel annotated entry is saved and continuously proposed with every tag of the same markable. During the fourth month, the event types were extended with *disease outbreak*, *riot*, *natural disaster* and *music festival*. After a short break, the annotators continued with the same event types for one more month while they were joined by two more annotators.

Per reference text, the annotators first performed instance-linking between in-text mentions and structured data (see Figure 3 in the Appendix). Then, on the frame level, the annotators performed frame annotations - including frames and core frame elements - on all event instance mentions that were previously instance-linked. Note that here, instead of traditionally annotating all predicates in a text as mentions, the annotator is guided by the instance-linked mentions, resulting in annotations restricted to text segments that are relevant to the event instance’s main narrative. Finally, the annotators frame annotated all mentions of subevents of the main event instance that were not previously linked to structured data. Even though they are not part of the original structured data, they still contribute to both the main frame representation of the event instance’s narrative and the DFN lexicon. Annotators were instructed to consider the temporal and causal containers of the main event as a criterion to decide on including events as relevant subevents, following (O’Gorman et al., 2016; Caselli and Vossen, 2017). If core frame elements were not found in the sentence of their frame, the annotators annotated the

first mention of the frame element in the text (see Figure 4 in the Appendix). If no mention of the frame element was found, it was annotated as *unexpressed*. When needed to complete frame annotation, the annotators performed markable correction. If many frame elements in a document were ascribed to mentions that did not exist in the structured data box, the annotators were able to update the box with the corresponding Wikidata entry.

### 3.3. Evaluation

Throughout the annotation process, the annotators regularly worked on the same documents as to compare their output and measure their IAA. This is then compared to the IAA measured in Vossen et al. (2018a). Agreement is calculated for complete and partial span overlap, taking into account that annotators vary in their inclusion of function words. As the FrameNet ontology includes 1075 lexical frames (categories to normalize for), the probability of agreement by chance is neglectable for the task of frame annotation. Also, we compare agreement between two parallel annotation tasks (frame annotation and instance-linking) which makes normalizing scores a complex assignment. Thus, we choose to compute the agreement in percentages, instead of applying metrics such as Cohen’s Kappa (Cohen, 1960), allowing interpretable comparisons between the two annotation tasks. We expect overall high agreement, due to guidance by the structured data. In particular, we expect a correlation between high agreement on both annotation levels, whereas frame annotation of subevents without an instance-link might show lower agreement due to lack of the instance-link’s assistance. In line with the findings of Vossen et al. (2018a), we expect high agreement on frame element annotations as an effect of the agreement on their frames.

On the frame level, for each (partially) overlapping pair

of annotations showing disagreement, we computed the cosine similarity score between the two frames, utilizing word2vec embeddings based on English FrameNet definitions and annotations, as introduced in Sikos and Padó (2018). The cosine similarity is also used to compute the p-value of all annotations that show disagreement, using the distribution of the cosine similarities between all embeddings of all frames in FrameNet. Using this p-value, the probability of annotating similar frames by random chance can be taken into account for our evaluation. Both the similarity scores and p-value give us insight in the conceptual similarity between the frames that annotators disagreed upon.

Finally, we investigate the effect of domain-specific frame annotation by analyzing the distribution of the lexical entries over the timespan of the annotation process. We expect a strong increase of lexical units in the first weeks, followed by stabilization. Since the annotators continuously work within the same event types, we assume those to generate a demarcated set of domain-specific lexemes. Then, after the event types are extended, we expect the growth of the lexicon to boost again, assuming that those event types generate different domain-specific lexemes.

## 4. Results

In total, the annotated output consists of 326 annotated reference texts, 276 Dutch and 50 English. 27533 mentions were annotated with 9220 tokens of 2729 different lexical units, covering 574 different frames (avg. 16.06 annotations per frame). In order to enable correct frame annotation, 1840 (19.9%) mentions received markable correction (avg. 5.6 per text): 699 multi-words and 1141 compounds. Also, 7457 (27,0%) of these mentions were annotated with instance-links. In the following subsections, we will present data analysis of IAA (4.1), discourse annotation (4.2), and the DFN lexicon (4.3).

### 4.1. Inter-annotator-agreement

Throughout the annotation process, 15 Dutch reference texts were annotated by multiple annotators. Table 2 displays IAA on different annotation levels.

With respect to the instance-links and the frames as separate annotation levels, the agreement percentages range from 73.7% to 91.9%, which can be considered strong to almost perfect. When considering the mentions that were jointly instance-linked and frame annotated, we observe that the agreement increases from 91.9% to 97.58%, while the agreement decreases to 89.9% when considering disjoint annotations. Although this increase in agreement is significant, the baseline of 89.9% is already considered strong. Furthermore, the ratio between the number of joint and disjoint annotations shows that this increase in agreement is not caused by a decreasing number of annotations.

On the frame annotation level, we computed a similarity score of 0.6 for the mentions on which the annotators disagreed about the frame candidates. Compared

<b>1. instance-links</b>	
span matches	89.5%
agreement on span matches	89.4%
<b>2. frames</b>	
span matches	73.7%
agreement on span matches	91.9%
similarity in disagreement	0.59
similarity p-value	0.07
<b>3. frames and instance-links</b>	
joint agreement	97.58%
disjoint agreement	89.94%
joint:disjoint ratio	1:2.3
<b>4. frame elements</b>	
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.

to the distribution of similarity scores, this results in a p-value of 0.07, showing that when no absolute agreement between two frame annotations can be found, the annotators still strongly agree on a conceptual level on the sense of a mention. Table 3 shows examples of confused frame pairs, along with their similarity score. We find that frame pairs with a higher similarity score show stronger conceptual feature overlap between the frames, e.g., compare the top and the bottom pairs.

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.

Contrasting annotations of expressed and unexpressed frame elements, we find that annotators disagree most on whether a frame element is expressed, lowering the agreement to 69.5%.

### 4.2. Discourse annotation

The main explanation for the lower agreement on the frame element annotation compared to traditional text-to-data annotations of Vossen et al. (2018a) is the com-

plexity of the discourse annotation that we applied. 27.1% of all frame elements (including *unexpressed*) were annotated as not occurring in the same sentence as their frame. However, 99.8% of all annotated frames contain at least one of such sentence-external frame elements (avg. 1.59 frame element per frame). Figure 1 shows both the distance in sentences of the annotated frame elements to the sentence of their frames, and the average level of agreement between annotators on that distance. Most sentence-external frame elements were annotated in surrounding sentences. A small peak shows around 30-40 sentences distance. We can see that there is a very high agreement for elements within the same sentence (distance 0), whereas agreements vary for more distant sentences containing frame elements.

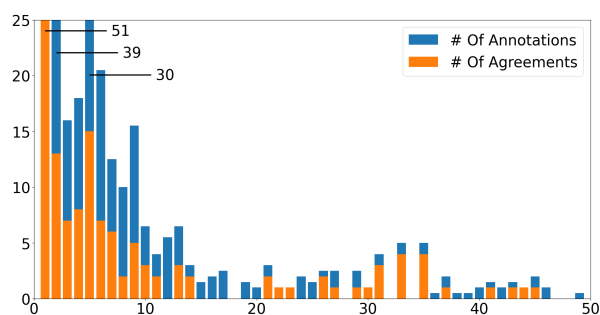


Figure 1: Number of sentence-external frame elements with the distance in sentences to the annotation of the frame. The figure includes the agreement score for each distance to the frame.

### 4.3. DFN lexicon

Figure 2 shows the distribution of added DFN lexicon entries over the appointment period. In the first two months, the graph shows a gradual increase of new entries. Then, from the second month onward, we observe a more steep increase of lexical entries while the annotators are still annotating texts within the same event types. After the point of extension to more different event types, we see another increase, particularly around 2021-12. The flat line from 2022-02 to 2022-03 reflects the annotators’ break after which they continued for two more months with two additional annotators. Here, the growth of the lexicon receives another boost. At 2022-04, the annotators had another short break. Furthermore, we see a more steep growth of the lexical entries as compared to the annotated frames, which makes sense given the sizes of lexemes in Dutch and frames in FrameNet.

## 5. Discussion

With respect to DFN coverage, the overall descriptive statistics show on average 17.6 annotations for 599 different frames (55.7% of all available frames). Thus, while our output does not reach a significant level of

FrameNet coverage, each annotated frame does entail a considerable amount of annotations. Since the annotations are grouped under specific event types, the resulting data is suitable for training domain-specific machine learning models.

19.9% of the frame annotations were performed with the aid of markable correction, showing the need of language-independent FrameNet annotation tools for this feature. Also, the high amount of Dutch compound splitting reveals that this morphological feature plays an important role in framing of events and entities in Dutch.

Recall that Vossen et al. (2018a) conclude that the low agreement in their study is a result of the text-to-data method, in which annotators start from text, without the aid of context and thus continuously considering all FrameNet frames. In the study of this paper, we utilize a data-to-text method. From the findings in Table 2, it becomes clear that the data-to-text method results in considerable higher agreement on frame annotation than was observed in the aforementioned study (47% versus 91.9%, taking into account that mentions were given in the text-to-data method). We conclude that starting the annotation process from structured data guides the annotator in frame identification, even when they are free to choose mentions, correct markables and apply out-of-sentence relations. Therefore, we observe a strong correlation between the agreement on the instance-linking and frame layer: where annotators agree on the referent of the mention, they also agree on the frame that the mention evokes. Moreover, the similarity score shows that the frames that the annotators disagreed upon still show strong overlap in conceptual features.

For mentions with frame annotations but no instance-links, we expected low agreement, since the annotators are not guided by an instance-link. These are the subevents of the main event that the annotators were instructed to annotate after frame annotation of all instance-linked mentions. Yet, we still observe an agreement of 89.7% for these frame annotated subevents. We assume that the structured data, in particular the event type, still provides sufficient context for the annotator to identify the frame.

The moderate agreement score of 69.5% for frame elements can be ascribed to the facilitation of discourse annotation, as it increases ambiguity on whether frame elements are expressed throughout the entire document, as opposed to the traditional FrameNet annotation process in which only one sentence has to be considered. Note that the agreement on frame elements in the study of Vossen et al. (2018a) is higher (79%). This means that frame element annotation still profits from the text-to-data method, since the annotator only has to consider the semantic roles within the predicate’s sentence. The benefit of discourse annotation is then the notion of *unexpressed* for those core frame elements that are completely absent from text. Those frame el-

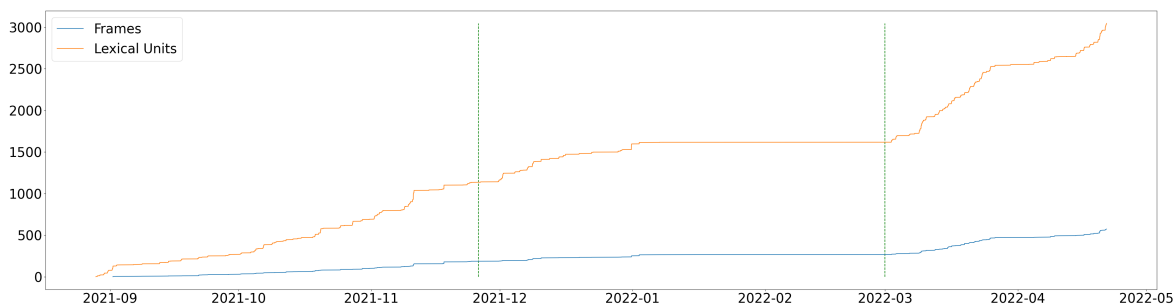


Figure 2: Distribution of new DFN lexicon entries over time, from the beginning to the end of the annotation appointment. The first green vertical line indicates the moment that the annotations of event types extended from four to six. The second green line indicates the moment two more annotators were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

ements are assumed to be implicatures. Thus, the tool contributes to the field of Natural Language Inference by enriching the output with this pragmatic data.

Figure 1 shows that most sentence-external frame elements occur in surrounding sentences. Since the annotators were instructed to look for the first mention of sentence-external frame elements in the reference text, the small peak of 30-40 sentences distance seems to point to those frame elements that are introduced at the onset of the text to establish the main topic and participants. We also learn that the agreement is strongest for sentence-internal frame elements and varies for sentence-external frame elements.

With respect to the DFN lexicon, Figure 2 displays a strong increase of lexical entries over time, with a peak around the moment that reference texts of different event types are introduced. This is an indication that annotation of a corpus that follows an event type-based model, generates event type-specific FrameNet entries. Annotation of texts across multiple event types would then lead to a cross-domain lexicon.

Even though the annotators worked most of the time within the same event types, before the switch to different event types, the figure shows no stabilization of new lexical entries in the lexicon, while one would expect that at some point, the annotators would annotate more and more tokens of the same lexical units. This stands out the most at around 2022-03, where two more annotators are joined, but the event types remain the same. This suggests a high amount of variation in framing of the event instances within and between texts as is displayed by the examples in Table 4 in the Appendix.

## 6. Conclusions

This paper reports on the first results of annotating FrameNet frames with frame elements and references following a data-to-text approach in which the texts are referentially grounded. Our data is freely available under the license CC-BY-SA4.0 as release-1.1 on our website: <http://dutchframenet.nl/data-releases/>. The annotation tool is freely available on our GitHub: <https://github.com/>

`cltl/frame-annotation-tool`.

We provided evidence that the frame annotation is by far more consistent compared to traditional text-to-data approaches despite the fact that we followed a discourse approach for frame and frame elements that exceeds the sentence boundary. We expect that the referential grounding across different sources in relation to the same event instance and across different event instances of the same event type provides new insights in the variation of framing and henceforth into the structural and pragmatic factors that dictate framing choices. We also described the growth of the Dutch FrameNet lexicon in relation to the annotation in terms of size and richness as a function of the volume of annotated text and the diversity of the event instances that are annotated.

In future work, we will extend the annotation to cover more event types and more languages and we will experiment with new ways of pre-annotating texts on the basis of the data and lexicon that has been created so far. A down-side of our approach is that the coverage of the annotations and lexicon is driven by the event types and the data that is available through Wikidata and Wikipedia. We will therefore explore additional approaches to increase the coverage to underrepresented event instances and situations.

## 7. Acknowledgements

The research reported in this article was funded by the Dutch National Science organisation (NWO) through the project *Framing situations in the Dutch language*, VC.GW17.083/6215. We would like to express our gratitude to our student annotators Sanne Hoeken, Iris Lau, Adrielli Lopes Rego, Olga Pela, Dorien Renting and Sharona Badloe.

## 8. Bibliographical References

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013). Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop*



- and Interoperability with Discourse, pages 178–186, Sofia, Bulgaria, August. Association for Computational Linguistics.
- Bouma, G. and van Noord, G. (2017). Increasing return on annotation investment: The automatic construction of a Universal Dependency treebank for Dutch. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 19–26, Gothenburg, Sweden, May. Association for Computational Linguistics.
- Burchardt, A., Erk, K., Frank, A., Kowalski, A., Pado, S., and Pinkal, M. (2006). SALTO-A Versatile Multi-Level Annotation Tool. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*.
- Choubey, P. K., Raju, K., and Huang, R. (2018). Identifying the most dominant event in a news article by mining event coreference relations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 340–345, New Orleans, Louisiana, June. Association for Computational Linguistics.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.
- Damonte, M., Cohen, S. B., and Satta, G. (2017). An incremental parser for Abstract Meaning Representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 536–546, Valencia, Spain, April. Association for Computational Linguistics.
- Devitt, M. and Sterelny, K. (1999). *Language and reality: An introduction to the philosophy of language*. mit Press.
- Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., and Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 76–84, Osaka, Japan, December. The COLING 2016 Organizing Committee.
- Filatova, E. and Hatzivassiloglou, V. (2004). Event-based extractive summarization. In *Text Summarization Branches Out*, pages 104–111, Barcelona, Spain, July. Association for Computational Linguistics.
- Fokkens, A., van Erp, M., Vossen, P., Tonelli, S., van Hage, W. R., Serafini, L., Sprugnoli, R., and Hoeksema, J. (2013). GAF: A grounded annotation framework for events. In *Workshop on Events: Definition, Detection, Coreference, and Representation*, pages 11–20, Atlanta, Georgia, June. Association for Computational Linguistics.
- Fokkens, A., Soroa, A., Beloki, Z., Ockeloen, N., Rigau, G., Van Hage, W. R., and Vossen, P. (2014). NAF and GAF: Linking Linguistic Annotations. In *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*.
- Frege, G. (1948). Sense and reference. *The philosophical review*, 57(3):209–230.
- Getman, J., Ellis, J., Strassel, S., Song, Z., and Tracey, J. (2018). Laying the groundwork for knowledge base population: Nine years of linguistic resources for tac kbp. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Hachey, B., Radford, W., Nothman, J., Honnibal, M., and Curran, J. R. (2013). Evaluating entity linking with wikipedia. *Artificial Intelligence*, 194:130–150. Artificial Intelligence, Wikipedia and Semi-Structured Resources.
- Hamborg, F., Meuschke, N., Breiting, C., and Gipp, B. (2017). news-please: A generic news crawler and extractor. In *15th International Symposium of Information Science (ISI 2017)*, pages 218–223.
- Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.
- Kenny, A. (1995). *Frege: An introduction to the founder of modern analytic philosophy*. Wiley-Blackwell.
- Pan, X., Cassidy, T., Hermjakob, U., Ji, H., and Knight, K. (2015). Unsupervised entity linking with abstract meaning representation. In *Proceedings of the 2015 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pages 1130–1139.
- Ponkiya, G., Patel, K., Bhattacharyya, P., and Palshikar, G. K. (2018). Towards a standardized dataset for noun compound interpretation. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).
- Ponkiya, G., Kanojia, D., Bhattacharyya, P., and Palshikar, G. (2021). FrameNet-assisted noun compound interpretation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2901–2911, Online, August. Association for Computational Linguistics.
- Postma, M., Remijnse, L., Ilievski, F., Fokkens, A., Titarsolej, S., and Vossen, P. (2020). Combining conceptual and referential annotation to study variation in framing. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 31–40.
- Quirk, R. (2010). *A Comprehensive Grammar of the English Language*. Pearson Education India.
- Remijnse, L., Postma, M., and Vossen, P. (2021). Variation in framing as a function of temporal reporting distance. In *Proceedings of the 14th International*

- Conference on Computational Semantics (IWCS)*, pages 228–238.
- Sikos, J. and Padó, S. (2018). Using embeddings to compare FrameNet frames across languages. In *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing*, pages 91–101, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.
- Van Noord, R. and Bos, J. (2017). Neural semantic parsing by character-based translation: Experiments with abstract meaning representations. *arXiv preprint arXiv:1705.09980*.
- Van Noord, G., Bouma, G., Van Eynde, F., De Kok, D., Van der Linde, J., Schuurman, I., Sang, E. T. K., and Vandeghinste, V. (2013). Large scale syntactic annotation of written dutch: Lassy. In *Essential speech and language technology for Dutch*, pages 147–164. Springer, Berlin, Heidelberg.
- Vossen, P., Fokkens, A., Maks, I., and van Son, C. (2018a). Towards an open dutch framenet lexicon and corpus. In *Proceedings of the LREC 2018 Workshop International FrameNet Workshop*, pages 75–80.
- Vossen, P., Ilievski, F., Postma, M., and Segers, R. (2018b). Don’t 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)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).
- Vossen, P., Ilievski, F., Postma, M., Fokkens, A., Minnema, G., and Remijnse, L. (2020). Large-scale cross-lingual language resources for referencing and framing. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3162–3171, Marseille, France, May. European Language Resources Association.
- Vrandečić, D. and Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.
- web.archive.org/web/20200317180116. (2020). <https://www.thehindu.com/news/national/coronavirus-number-of-confirmed-cases-in-india-crosses-110/article31073984.ece>.
- web.archive.org/web/20210124162405. (2020). <https://www.dispatchlive.co.za/news/2020-03-17-patient-zero-of-the-winelands-cape-estates-close-over-coronavirus>.
- web.archive.org/web/20210308133541. (2020). <https://www.dutchnews.nl/news/2020/04/coronavirus-identified-on-two-mink-farms-in-the-netherlands>.
- Bejan, C. and Harabagiu, S. (2010). Unsupervised event coreference resolution with rich linguistic features. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1412–1422, Uppsala, Sweden, July. Association for Computational Linguistics.
- Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2009). 8. using framenet for the semantic analysis of german: Annotation, representation, and automation. In *Multilingual FrameNets in computational lexicography*, pages 209–244. De Gruyter Mouton.
- Caselli, T. and Vossen, P. (2017). The event storyline corpus: A new benchmark for causal and temporal relation extraction. In *Computing news storylines and events workshop, ACL-2017*.
- Cybulska, A. and Vossen, P. (2014). Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, pages 4545–4552, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- De Clercq, O., Hoste, V., and Monachesi, P. (2012). Evaluating automatic cross-domain Dutch semantic role annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 88–93, Istanbul, Turkey, May. European Language Resources Association (ELRA).
- Djemaa, M., Candito, M., Muller, P., and Vieu, L. (2016). Corpus annotation within the French FrameNet: a domain-by-domain methodology. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 3794–3801, Portorož, Slovenia, May. European Language Resources Association (ELRA).
- Ko, J. (2018). Gun Violence Data. <https://www.kaggle.com/jameslko/gun-violence-data>.
- Lexicon of Linguistics. (2020a). Endocentric compound. <https://lexicon.hum.uu.nl/?lemma=Endocentric+compound&lemmacode=778&lemma=Endocentric+compound&lemmacode=778>.
- Lexicon of Linguistics. (2020b). Idiom. <https://lexicon.hum.uu.nl/?lemma=Endocentric+compound&lemmacode=778&lemma=Endocentric+compound&lemmacode=778>.
- Ohara, K. H., Fujii, S., Ohori, T., Suzuki, R., Saito, H., and Ishizaki, S. (2004). The japanese framenet project: An introduction. In *Proceedings of LREC-04 Satellite Workshop “Building Lexical Resources from Semantically Annotated Corpora” (LREC 2004)*, pages 9–11.
- Oostdijk, N., Reynaert, M., Monachesi, P., Van Noord, G., Ordelman, R., Schuurman, I., and Vandeghinste,

## 9. Language Resource References

- Baker, C. F., Fillmore, C. J., and Cronin, B. (2003). The Structure of the FrameNet Database. *International Journal of Lexicography*, 16(3):281–296.

- V. (2008). From D-coi to SoNaR: a reference corpus for Dutch. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).
- O’Gorman, T., Wright-Bettner, K., and Palmer, M. (2016). Richer event description: Integrating event coreference with temporal, causal and bridging annotation. In *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*, pages 47–56.
- Peng, H., Song, Y., and Roth, D. (2016). Event detection and co-reference with minimal supervision. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 392–402, Austin, Texas, November. Association for Computational Linguistics.
- Pradhan, S., Loper, E., Dligach, D., and Palmer, M. (2007). SemEval-2007 task-17: English lexical sample, SRL and all words. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 87–92, Prague, Czech Republic, June. Association for Computational Linguistics.
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., and Schefczyk, J. (2010). FrameNet II: Extended theory and practice.
- Song, Z., Bies, A., Strassel, S., Riese, T., Mott, J., Ellis, J., Wright, J., Kulick, S., Ryant, N., and Ma, X. (2015). From light to rich ERE: Annotation of entities, relations, and events. In *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 89–98, Denver, Colorado, June. Association for Computational Linguistics.
- Torrent, T. T., Ellsworth, M., Baker, C., and Matos, E. (2018). The Multilingual FrameNet Shared Annotation Task: a preliminary report. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*.

## 10. Appendix

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 **Gökmen 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 Jeuken told reporters , without giving further details . The **shooting** took place in **Kanalenland**, 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 Ralhhou , 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 Gökmen 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 Jeuken told reporters , without giving further details . The **shooting** took place in **Kanalenland**, 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 Ralhhou , 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 <b>liquidatie</b> <i>This is no ordinary <b>liquidation</b></i>	KILLING
2	er is <b>geschoten</b> in een tram <i>there was a <b>shooting</b> in a tram</i>	HIT_TARGET
3	[...] de slachtoffers van de <b>aanslag</b> in Utrecht <i>[...] the victims of the <b>attack</b> in Utrecht.</i>	ATTACK
4	[...] de krankzinnige <b>moordpartij</b> in Utrecht <i>[...] the insane murder <b>party</b> in Utrecht</i>	PARTICIPATION
5	Hij wordt verdacht van betrokkenheid bij het schiet <b>incident</b> <i>He is suspected of involvement in the shooting <b>incident</b></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.