# University of Groningen

## Towards Reference-Aware FrameNet Annotation

Remijnse, Levi; Minnema, Gosse

**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.**

*Document Version*
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

# Towards Reference-Aware FrameNet Annotation

**Levi Remijnse**[a] **and Gosse Minnema**[b]

[a]Vrije Universiteit Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
l.remijnse@vu.nl
[b]Rijksuniversiteit Groningen
Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, The Netherlands
g.f.minnema@rug.nl

## Abstract

In this paper, we introduce the task of using FrameNet to link structured information about real-world events to the conceptual frames used in texts describing these events. We show that frames made relevant by the knowledge of the real-world event can be captured by complementing standard lexicon-driven FrameNet annotations with frame annotations derived through pragmatic inference. We propose a two-layered annotation scheme with a 'strict' FrameNet-compatible lexical layer and a 'loose' layer capturing frames that are inferred from referential data.

## 1. Introduction

Written narratives can describe a single real-world event in different ways. In particular, an event of great cultural importance often generates a growing portion of written referential texts over time, all displaying various linguistic forms when referring to that same event or components of the event (Vossen et al., 2018a). These linguistic forms activate conceptual representations displaying perspectives, goals and motivations. In order to systematically investigate how the components of a single event are conceptually represented across texts, large-scale resources are needed that, on the one hand, link knowledge about real-world events to event mentions in text, and on the other hand link these mentions to conceptual information of that event. FrameNet can be a useful resource for linking event mentions to conceptual information, given that it provides a rich database of conceptual knowledge about event and situation types, which are linked both to each other and to lexical expressions evoking this conceptual knowledge.



Figure 1: Fictitious example of the same event described in different ways with different frames

In this paper, we will show how FrameNet annotations can be used as a resource for showing how structured knowledge about real-world events is conceptualized in text. Figure 1 shows an example of how FrameNet can be used to analyze how a single event can be described from different perspectives. While both texts mention the basic fact that a killing took place, the lower text stays close to the facts and provides details about the shooting event itself, whereas the upper text is less detailed and takes a more interpretative perspective by telling us that the event came to be seen as a crime (murder). This is reflected in the frame annotations: both texts evoke KILLING, but only the upper text evokes OFFENSES, whereas the lower one evokes USE_FIREARM.

In this fictitious example, the frames that are expressed by the lexical items in the texts fit well with the conceptual information needed to understand the perspective taken by these texts. However, this is not always true in natural texts, as in many cases, event descriptions are implicit. For example, "John was shot and died" does not contain any particular lexical item expressing a killing event, and would not be annotated with KILLING following FrameNet annotation standards. Yet, the sentence clearly refers to such an event.

In this paper, we will analyze such challenges, and propose a way to more comprehensively annotate the relationship between frames and referential data. In short, we will introduce an *inferred frame layer* of annotation on top of a 'regular' FrameNet annotation layer. In this way, we can annotate event mentions that standard frame annotation would not be able to capture, while preserving a standard FrameNet layer, thus contributing to the global FrameNet effort. We will illustrate the challenges and proposals we discuss with examples in English and Dutch, but we expect them to be relevant cross-linguistically.

**Contributions** The main contributions of our work are:

- We identify challenges for performing FrameNet annotation guided by referential data (Section 4.);

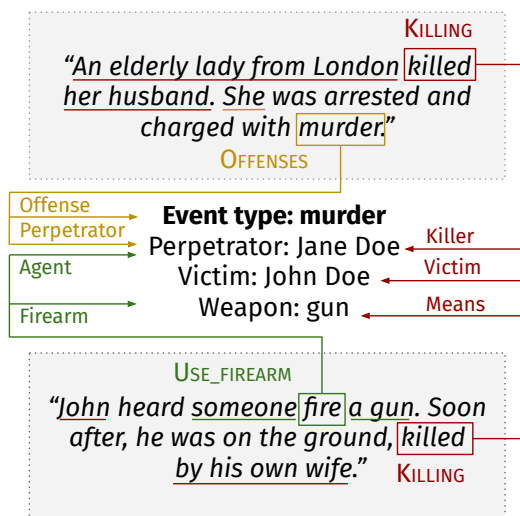- We propose a solution in the form of an extra anno-

tation layer for pragmatically inferred frames (Section 5.);

- We show the implications of our approach for pragmatics and frame semantics (Section 6.);

- We implement the inferred frame layer in an annotation tool as part of the Dutch FrameNet project[1] – for more details, see Postma et al. (this workshop).

## 2. Terminology

In order to avoid confusion between concepts from the 'reference world' and the 'frame world', some key terminology that we will rely on throughout this paper is given in Box 1. While these definitions might seem obvious, when linking frame annotations to information about real-world events, it is important to make an explicit distinction between events and frames on one hand and types, instances, and mentions on the other hand. Not doing so could easily cause confusion in an example like (1):

(1)  a. He killed the murderer of JFK, who was assassinated two days earlier.
b. He shot the murderer of JFK, who had died two days earlier.
c. He murdered someone yesterday, and did it again today.

In (1a), "killed", "murderer", and "assassinated", all describe the same event type (murder) but refer to two different instances of this event type (the "murderer" and "assassinated" refer to the murder of JFK, "killed" refers to the murder of JFK's killer). They also all evoke the same frame type (KILLING), while each of them is a separate mention of this frame. On the other hand, in (1b), "shot", "murderer" and "died", again refer to two event instances of the type murder,[2] but evoke three different frame types (HIT_TARGET, KILLING, DEATH), introducing a single mention of each of these. Finally, in (1c), "murdered" and "did it again" describe two instances of the murder event type, but only "murdered" is a mention of the KILLING frame type.

## 3. Background

### 3.1. FrameNet and conceptual information

FrameNet (Baker et al., 2003; Ruppenhofer et al., 2010a) provides a useful paradigm to analyze how conceptual information is encoded in language. Within this paradigm, *lexical units* (word forms with a specific sense) can *evoke* frame types, which are schematic representations of situations involving participants and other conceptual roles. These semantic roles (*frame elements*, or FEs) are expressed by constituents. Frame mentions are analyzed within clause boundaries. Two typical examples are given in (2):

---

> **Frames vs. events**
>
> **Event type**: category of real-world events
> *Example:* murder, election
>
> **Event instance**: individual event in the real world
> *Example:* the murder of JFK
>
> **Frame type:** frame entry in the FrameNet database, formally a tuple $\langle T, E, R \rangle$ ($T$: set of target LUs, $E$: set of frame elements, $R$: set of frame-frame relations). *Example:* KILLING $= \langle \{\text{kill.v}, \ldots\}, \{\text{Killer}, \ldots\}, \{\langle \text{Inherited\_by}, \text{EXECUTION} \rangle, \ldots\} \rangle$.
>
> **Frame mention:** expression of a frame type in text, formally a tuple $\langle f, t, e \rangle$ ($f$: frame type, $t$: target LU, $e$: set of (frame element name, frame element span) pairs. *Example:* given "He killed JFK", annotate: $\langle \text{KILLING}, \text{killed}, \{\langle \text{Killer}, \text{he} \rangle, \langle \text{Victim}, \text{JFK} \rangle\} \rangle$.

Box 1: Key terminology for our annotation task

(2)  a. COMMERCE_SELL
[Time Yesterday], [Seller John] ⊙sold [Buyer Mary] [Goods a book].
b. COMMERCE_BUY
[Buyer A woman] ⊙bought [Goods a novel] [Place in the shop].

In (2a), "sold" is a lexical unit that evokes COMMERCE_SELL. This frame comes with an inventory of frame elements, some of which are necessary for the reader to process the frame (*core elements*). For COMMERCE_SELL, these are the Buyer, "Mary", the Seller, "John", and the Goods, "a book". Similarly, in (2b), "bought" evokes COMMERCE_BUY, which has the same frame elements: a Buyer, expressed by 'a woman', Goods, expressed by "a novel", and a Seller, which is unexpressed in this sentence.

The overlap of semantic roles between these two frame types indicates that both COMMERCE_SELL and COMMERCE_BUY have a Perspective_on relation with the abstract ('non-lexical') frame type COMMERCE_GOODS-TRANSFER. This relation encodes the fact that both frame types describe the same abstract concept, but from different perspectives: COMMERCE_SELL takes the point of view of the Seller, whereas COMMERCE_BUY takes that of the buyer. In this way, FrameNet provides us with rich information about variation in framing on a conceptual level.

### 3.2. Reference-driven annotation

Besides representing conceptual knowledge, a point of interest is to capture variations in the way that texts frame components of the real-world event that they refer to. We want to know, for instance, whether the sentences in (2) describe the same event instance, and hence, whether "Mary" in (2a) and "a woman" in (2b) both refer to the same participant of this event instance. In order to annotate texts with this type of information, we need a resource providing structured data about events in the real world and texts describing these events.

---

[1] www.dutchframenet.nl

[2] Note that knowledge of the real-world event is necessary to recognize that "shot" and "died" both describe a murder event: the lexical content of these words does not imply murder (one can be shot without dying, and one can die without having been murdered), but in this context they do refer to (subevents of) murders.

We make use of the *data-to-text* method (Vossen et al., 2018b; Vossen et al., in press) in order to establish such a resource. This method inverts the usual process of annotating data: instead of starting from (unstructured) text and then annotating it with referential information, we start from structured information about real-world event instances and then match these to texts describing these instances. More concretely, we query Wikidata (Vrandečić and Krötzsch, 2014) for a set of event instances belonging to a particular event type. The Wikidata API then returns records of such instances, accompanied by structured data (minimally: the event type, data, location and participants). Wikidata also provides the Wikipedia text pages in various languages, which in turn provide hyperlinks that point to other texts referring to the same event. We aggregate the Wikipedia texts themselves with the texts they point to, to build a corpus of *reference texts* linked to event instances.

Next, we prepare the corpus for manual FrameNet annotation. FrameNet contains a large number of different frame types (1224 in Berkeley FrameNet for English).[3] In order to efficiently annotate large corpora, we restrict the scope of our FrameNet annotations to only include frame types that are known to be relevant for the event types in our dataset. To achieve this, we first automatically annotate the acquired corpus using Open-SESAME, a state-of-the-art frame semantic role labeler (Swayamdipta et al., 2017). Then, by analyzing the frequency distribution of the frame types found in the automatic annotations, we define a list of *typical frames* containing the frame types that are most dominant in texts referring to a particular type of event.[4] To summarize, utilizing the data-to-text method results in the following data:

- Records of a set of event instances belonging to one event type (e.g. 'murder');

- A corpus of reference texts for each event instance;

- Structured data for each event instance;

- A list of typical frames belonging to the event type.

The next subsection elaborates on the integration of frame annotations and referential annotations.

### 3.3. Integrating FrameNet in Referential Annotations

The product of the data-to-text method enables the annotator to annotate frame mentions representing the conceptual content of each text, and then link these mentions to structured data about the corresponding event instance. Returning to the examples in (2), we see that on the conceptual (frame) level, "Mary" is the Buyer of COMMERCE_SELL in (2a) and that "a woman" is the Buyer of COMMERCE_BUY in (2b). Next, let us assume that the structured data we

found, tells us that the two sentences refer to the same event instance in the real world. This allows us to make the link to the referential level by annotating "buy" and "sell" as referring to the event instance, and "Mary" and "a woman" as referring to the same participant in that event instance. In integrating these annotations, we find that the two sentences show conceptual variation in framing of the same event instance in the real world.

The typical frames generate expectations about the frame types to be found in the reference texts. Often, these frame types are also conceptually necessary for recognizing the for instance, a text can only be interpreted as describing a murder event if the conceptual content of KILLING is somehow expressed in the text. Hence, in addition to guiding expectations of the most probable frame types to be found in the texts, the typical frames function as a 'checklist' for the annotator to explore to what extent the typical frame types are encoded in the text. Annotating whether or not each typical frame is indeed expressed provides much information about the perspective of a text; for instance, in the example texts discussed in the introduction (Figure 1), OFFENSES is a typical frame for describing murder events, but the fact that only one of the texts expresses this frame type tells us something about the different perspectives of the two texts.

As we will show in Section 4., in some cases, typical frames are expressed in the text, but do not have a target word corresponding to a lexical unit in FrameNet, nor can they be derived through frame-to-frame relations. In such cases, we run into an inherent limitation of FrameNet: FrameNet is, at heart, a lexicographical project; conceptual information is always 'activated' through a direct correspondence between a lexical unit and a frame. This limitation has been noted even from within the field of frame semantics: Fillmore himself has allowed for the possibility that frame types, in some cases, are not evoked by lexical units, but by other linguistic features (Andor, 2010, p. 158). If we want to account for the way in which frame types related to the referential level are activated in corpora, we need to complement the lexical semantic approach of FrameNet with a broader view that takes into account compositional semantics and pragmatics. In Section 4., we motivate this view.

### 3.4. FrameNet and Inference

The notion of 'inference' is crucial for the annotation approach proposed in this paper: we aim to annotate frame mentions that are not directly evoked by a lexical unit but whose relevance can be inferred from the textual and referential context of an event. Inference in the context of frame semantics has been studied in the literature, but the notion we use in this paper is subtly different. Here, we provide a brief overview of notions of inference found in the FrameNet literature and how our notion differs from it.

**Frame-to-frame relations** In the FrameNet literature, inference is often connected to frame-to-frame relations. For example, Chang et al. (2002) propose a scheme for modeling *shared inferential structure* between frame types. An example of frame types with shared inferential structure

---

are COMMERCE_buy and COMMERCE_sell: both refer to the same type of event in the real world; hence, when one of these frame types is used, it can be inferred that the other frame type is also conceptually 'active'. Different frame-to-frame relations give rise to different kinds of inferences; for example, Sikos and Padó (2018), investigate the Using relation as a source for paraphrases. This allows, for example, for the inference of LABELING ("he ⊙called him a hero") from JUDGEMENT_COMMUNICATION ("he ⊙praised him for being a hero").

For the purposes of this paper, we focus on a different kind of inference: we are interested in frame types whose conceptual content is 'activated' by a text, but cannot be annotated as being evoked by a lexical unit. While, in a subset of such cases, there might be a frame-to-frame relationship between the frame type of interest and other frame types that are evoked in the text, this is not always the case. Moreover, even if such a frame-to-frame relationship is present, this might not be sufficient for licensing the inference. In the example "John was shot and died" (discussed in the introduction), "die" evokes DEATH, which has a Causative relation with KILLING, but this relation alone is not enough to make the inference: the fact that someone died does not imply that this person was also killed. Instead, we can infer that a killing did take place from the textual context ("John was shot").

**Cognitive frames**   The idea of frames that are present but not evoked by a lexical unit is also known from the literature about cognitive frames,[5] as is evident in the following famous example from Minksy:

(3)   Mary was invited to Jack's party. She wondered if he would like a kite.   (Minsky, 1974)

Here, the lexical unit "party" evokes SOCIAL_EVENT. The second sentence, "she wondered if he would like a kite" gives us reason to think that the party described is of a specific kind: most likely a birthday party. This would suggest the relevance of a frame type such as BIRTHDAY_PARTY (not currently existent): from our cultural knowledge, we know that parties at which gifts are given are typically birthdays or some other type of commemorative event.

However, this notion of inference goes beyond what we are aiming for in this paper. In the above example, it could be guessed what kind of party is at play, but the inference does not follow directly from the text: it could be some other party where, for whatever reason, gifts are given. This means that annotators would have to rely on their cultural knowledge. By contrast, within our framework, world knowledge can play a role in deriving inferred frame types, but their conceptual content should always be fully specified by the linguistic cues in the text. However, unlike in standard FrameNet annotation, these cues are not limited to single lexical items, but can comprise larger constructions.

---

[5]We assume the distinction between cognitive and linguistic frames proposed by Fillmore (2008).

## 4.    Challenges for Reference-Aware Annotation

In this section, we detail and motivate the main challenges that we see for structured-data-driven frame annotation that cannot be solved within the standard framework of FrameNet. We first motivate the general problem, and then discuss a number of concrete problems that we would like to address. An overview of these problems is shown in Box 2.

---

**Annotation Challenges**

**Problem: how to link *n* LUs to *m* frame types**

**Many-to-One**
*Compositionality:* $\geq 2$ LUs, $\geq 1$ frame type
*Complex Verbs:* verb components, $\geq 1$ frame type(s)

**One-to-Many**
*Frame Overlap:* 1 LU, $\geq 2$ frame types
*Lexical Gaps:* out-of-vocab LU, $\geq 1$ frame type(s)

---

Box 2: Overview of the annotation challenges

### 4.1.    The Coverage Problem

A general issue of FrameNet that has been noted in the literature is that it covers many frame types while only a limited number of number of annotations are available per frame type and per lexical unit (Palmer and Sporleder, 2010; Vossen et al., 2018b). As a logical consequence, when annotating texts with a limited set of frame types, as in our approach, the number of annotations per text would be expected to be small. Indeed, results from the CALOR project for French (Marzinotto et al., 2018), in which a small subset (53 frame types) of all possible FrameNet frame types was annotated, show that the number of sentences with at least one frame mention varied between 21%–34%, depending on the topic of the annotated texts.

One of the texts that we annotated in preliminary annotation experiments, describing the killing of visitors of a Christmas market in Berlin during a terror attack in 2016, is shown in Table 1. Our aim is to show whether each of the referential attributes of the event is expressed in the text, and if so, how it is conceptualized with frame mentions. For this particular text, reasoning from structured data, one would expect at least KILLING to be activated, and possibly also OFFENSES, USE_firearm, and/or WEAPONS (depending on whether the event is seen as an offense and whether the authors choose to mention the weapon). Surprisingly, it turns out that none of these frame types is evoked in the text in relation to the event mention of interest; even though "he was ⊙killed in a shootout . . . " contains a KILLING frame mention, this is in relation to a secondary event mentioned in the text (i.e. the killing of the perpetrator of the main murder event described in the text). Moreover, none of the frame types evoked by the lexical units in the text can be linked to the typical frames through a frame-to-frame relation; if this had been the case, we might have been able to indirectly annotate the frame types of interest,

| [Wikidata] Q28036573 | [Text] "2016 Berlin truck attack" | [Typical Frames] |
|---|---|---|
| Event type: *murder* <br> Time: *2019-12-19* <br> Location: *Berlin* <br> Participant: *Annis Amri* <br> Number_deaths: *12* <br> Weapon: *truck* | *On 19 December 2016, a truck was deliberately driven into the* <br> *Christmas market next to the Kaiser Wilhelm Memorial Church* <br> *at Breitscheidplatz in Berlin, leaving 12 people dead and 56 oth-* <br> *ers injured. [. . . ] The perpetrator was Anis Amri, a Tunisian* <br> *failed asylum seeker. Four days after the attack, he was killed in* <br> *a shootout with police near Milan in Italy. [. . . ]* | { KILLING, <br> USE_FIREARM, <br> OFFENSES, <br> WEAPON, <br> COMMIT_CRIME } |

Table 1: Example output of the data-to-text pipeline.

as discussed in Section 3.4..

However, from a close examination of the text, we find that each of the referential attributes from the structured data is in fact mentioned, but without using any lexical units belonging to one of the typical frames. We argue that the conceptual content of these frame types is still relevant for describing how the event instance is expressed in the text, and that this should be reflected in the annotations. For example, in FrameNet, the definition of KILLING is given as "A *Killer* or *Cause* causes the death of the *Victim*". A 'killing' event is very clearly expressed in the text by "a truck was deliberately driven into the Christmas market . . . leaving 12 people dead". However, it is difficult to specify which lexical unit(s), if any, evokes this particular frame mention in the standard FrameNet sense.

Work on what has become known as the *implicit semantic role labeling* task (Ruppenhofer et al., 2010b) addresses a related problem: semantic roles are sometimes 'missing' in the sentence of their associated predicate, but are conceptually 'activated' by this predicate and expressed elsewhere in the discourse. In example (4), the Charges role of "cleared" is not explicitly expressed, but can be inferred because "murder" is still active from the previous sentence:

(4) In a lengthy court case the defendant was tried [Charges for murder]. In the end, he was ⊙cleared. (Ruppenhofer et al., 2010b, p. 107)

The challenges we address in this paper are also related to implicit semantic roles, but in a more abstract way: in our case it is not the fillers of semantic roles, but the frame types defining these semantic roles that are unexpressed and have to be inferred. In the remainder of this section, we will discuss these challenges in more detail. In Section 5., we will propose a solution to these challenges.

### 4.2. 'Many-to-One' Problems

In the first class of challenges we encountered, at least one frame type is relevant for describing how an event instance is conceptualized, but there is no lexical unit in the text that, under standard FrameNet assumptions, would evoke this frame type. Instead, several items in the text together allow the reader to infer that the frame type is relevant, and give rise to annotating a mention of this frame.

**Compositionality** The Compositionality Problem occurs when multiple lexical items, through the composition of their meanings, 'activate' a single frame type. The sentence

in (5) (already briefly discussed above) is a clear example of this:

(5) KILLING <br> [Cause a truck] was deliberately ⊙?driven . . . <br> ⊙?leaving [Victim 12 people] ⊙?dead . . .

The sentence describes an action ("drive") with the consequence ("leaving") of people dying ("dead"); while none of these is a 'killing word' per se, the sum of these components imply (or even entail) that a killing event took place. We would like to capture in our annotations that (the conceptual content of) KILLING is relevant for this sentence, but standard FrameNet annotation does not allow us to annotate this, since there is no lexical target for KILLING, nor can KILLING be derived through other frame types that are evoked in the text.[6]

**Complex Verbs** A special case of the Compositionality Problem is the Complex Verbs Problem, in which the targets that jointly activate a frame type are all part of a complex (prepositional) verb:

(6) a. OPERATE_VEHICLE <br> . . . [Vehicle a truck] was deliberately ⊙driven [Goal into the Christmas market] . . . <br> b. IMPACT <br> . . . [Impactor a truck] was deliberately ⊙?driven ⊙?into [Impactee the Christmas market] . . .

Since FrameNet lists "drive", but not "drive into", as a lexical unit, the canonical analysis of (6) should be (6a). However, in this sentence, "into" does not simply add a destination to "drive", but modifies the meaning of "drive" so that it expresses not just a driving event, but also a hitting event. Hence, one would like to annotate a mention of IMPACT as well as of OPERATE_VEHICLE.

The Complex Verb Problem is particularly relevant in Dutch, which has many complex verbs that are often discontinuous:[7]

(7) Toen **reed** een vrachtwagen **op** het publiek **in** <br> then drove a truck on the crowd into

---

[6]For example, "dead" evokes DEAD_OR_ALIVE, which is (distantly) related to KILLING, but does not imply its relevance (the fact that someone dies does not imply that someone was killed).

[7]From the Dutch version of the Wikipedia article about the Berlin Christmas market attack (https://nl.wikipedia. org/wiki/Aanslag_op_kerstmarkt_in_Berlijn_ op_19_december_2016).

'Then, a truck (deliberately) drove into the crowd'

   a. OPERATE_VEHICLE
     [Time toen] ⊙?reed [Vehicle een vrachtwagen] [Goal op het publiek in]

   b. IMPACT
     [Time toen] ⊙?reed [Impactor een vrachtwagen] [Impactee ⊙?op het publiek] ⊙?in

Here, the verb *inrijden (op)* "(deliberately) drive into" expresses the same two meanings (i.e., driving and hitting) as "drive into" in (6). However, "in" in "inrijden" is arguably 'more part of the verb' than "into" in "drive into"; thus, it is likely that "inrijden" would be a separate lexical unit in the (still to be developed) Dutch FrameNet. Hence, in (6), the correct analysis under standard FrameNet annotation would be to use OPERATE_VEHICLE (because "drive", not "drive into" exists in FrameNet). By contrast, in Dutch FrameNet, "inrijden" would most likely be a lexical unit of IMPACT. Hence, (6) and (7) have an almost identical semantic content but would get very different analyses, where one of the relevant frame types is lost. Ideally, in our annotations we would like to capture both of the two relevant frame types.

### 4.3. 'One-to-Many'

The second class of challenges that we identify applies in the inverse situation of the 'many-to-one' challenge: these consist of cases with a certain number of relevant frame types, but not enough lexical units to evoke all of these frame types.

**Frame Overlap** Under the Frame Overlap Problem, a single lexical unit is relevant for more than one frame type. An example is given in (8):

(8) HOSTILE_ENCOUNTER
   [Side 1 he] was killed in a ⊙shootout [Side 2 with police]

In FrameNet, "shootout" is listed as a lexical unit of HOSTILE_ENCOUNTER. However, the lexical semantics of "shootout" clearly involves the use of a firearm, which makes USE_FIREARM conceptually relevant as well. Since USE_FIREARM is part of the typical frames for murder events, we would like our annotations to reflect the fact that the text indeed expresses a USE_FIREARM event. A naive solution would be to add a lexical unit "shootout" to USE_FIREARM so that we could annotate that frame type. This would not work well, since USE_FIREARM, though conceptually relevant, does not fit well with the structure of the sentence: a typical context of USE_FIREARM are sentences like "[Agent she] ⊙fired [Firearm her gun]", with the firearm and the shooter, rather than the participants in a conflict, as core roles.

An even more subtle version of the Frame Overlap Problem arises from the hypothetical example in (9):

(9) OFFENSES
   [Perpetrator He] was convicted for the [Offense ⊙murder] of [Victim JFK].

"Murder" is a lexical unit in both OFFENSES and KILLING, and has an almost identical meaning in both of them. Which of the two frame types should be annotated depends on the context: OFFENSES.murder is activated only when there is a governing verb such as 'convict' or 'accuse'; in other contexts KILLING.murder is activated. In (9), we clearly have an OFFENSE context rather than a KILLING context, but this does not mean that the meaning of KILLING is not also active: while the sentence, through a mention of OFFENSES, tells us that someone was convicted of a crime (further specified as the 'murder of JFK'), it also tells us that the murder happened in the first place, which we would like to capture using a mention of KILLING.

**Lexical Gaps** An extreme case of the Frame Overlap Problem occurs when a particular lexical unit does not exist in FrameNet, but would be a potential target for some frame type. We call this the Lexical Gaps Problem: a single lexical unit is associated with zero frame types in FrameNet, but at least one frame type is relevant for annotation. For example, in (10), "perpetrator" is not listed as a lexical unit for COMMIT_CRIME, but is a very likely target for it, especially because the verb "perpetrate" is listed under that frame type.[8]

(10) COMMIT_CRIME
   The ⊙?perpetrator was [Perpetrator Anis Amri] . . .

It is well-known that the FrameNet lexicon is incomplete, especially when annotating out-of-domain corpora (Hartmann et al., 2017). In this sense, the Lexical Gaps problem seems more superficial than the other problems discussed in this section. Yet, the lexical gaps detected by using our method of structured-data-driven annotation require some kind of inference on the part of the annotator. Namely, the list of typical frames guides the annotator in inferring frame types from potential lexical units currently missing in FrameNet.

## 5. Towards a Workable Solution for Annotating Inferred Frames

In this section, we aim to address the challenges previously explained by proposing an extra annotation layer (next to, not instead of, traditional FrameNet annotation) for capturing *inferred frames* whose conceptual content is expressed in the text without explicitly using one of the frame type's lexical units, but through inference. This layer would allow annotators to use any combination of words in the text as a 'trigger' for any number of frame mentions. While this idea is conceptually simple, some challenges need to be overcome for implementing it in practice: how do we make sure we get enough data? How do we apply the annotations in a consistent way?

### 5.1. Introducing Inferred Frame Annotation

The overall annotation pipeline that we propose is shown in Figure 2. The process starts with choosing event types

---

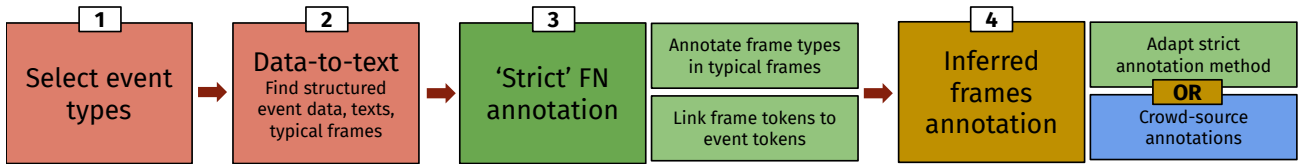[8]For comparison: in KILLING, both "murder" and "murderer" are listed as targets.

Figure 2: Overall annotation pipeline



(a) Screen 1: explanation of frame types



(b) Screen 2: selecting target words



(c) Screen 3: selecting frame elements

Figure 3: Mockup of a crowd-sourcing interface (possible user input marked in bold)

of interest and running the data-to-text pipeline (see Section 3.2.) to obtain linked event data and texts (Steps 1 and 2). Then, 'strict' FrameNet annotation is applied (Step 3): this annotation step will be done following standard FrameNet guidelines, except that (i) only frame types in the typical frames, selected by the data-to-text algorithm, will be taken into consideration and that (ii) frame mentions will be linked to event instances and their attributes in the structured data, much like in the initial example we gave in Figure 1. This step will be done by annotators, who need to be trained in applying FrameNet annotation guidelines. Finally (Step 4), we will annotate the inferred frame layer that we have motivated in this paper.

Annotation on this layer is much 'looser' than the annotation done in step 3. Annotators do not need to respect the FrameNet rule of 'one lexical unit, one frame mention', but are free to annotate any number of frame mentions based on any combination of lexical items in the text. An inherent risk of this type of 'free-style' annotation is that inter-annotator agreement is likely to be lower, simply because the number of possible annotation decisions is much larger and less constrained than under standard FrameNet annotation.

## 5.2. Annotation Strategies

Currently, we see two possible (not necessarily exclusive) paths to mitigating this risk. The first option involves a *qualitative* approach that aims to make the procedure that annotators follow as consistent as possible. Alternatively, a *quantitative* approach would use crowd-sourcing for gathering as much data for every text as possible, and then comparing and aggregating the annotations from different annotators.

Under the qualitative option, we would integrate annotation of Step 3 and Step 4: the annotators would annotate both layers in the same way, using the same tools. The advantage would be that the annotators are trained in doing FrameNet annotation, which improves the consistency of the annotations. However, due to the looseness of the task, we still expect considerable disagreements between different annotators. Moreover, training and deploying expert annotators is costly and time-consuming.

On the other hand, the quantitative option would 'embrace' the unconstrained nature of the inferred frame layer, and use crowd sourcing to gather as much data as possible. This would mean moving further away from standard FrameNet annotation, given that the annotators would be unfamiliar with FrameNet and its philosophy. Annotations are also

likely to be less consistent: different annotators might have different standards for what words are relevant for each frame mention.

However, previous studies have shown that annotation tasks similar to FrameNet annotation, such as PropBank-style semantic role labeling, can be successfully addressed using (partial) crowd sourcing (Wang et al., 2017). Moreover, the task of annotating the inferred frame layer is potentially more suitable for crowd sourcing than standard FrameNet annotation is: since there are no strict guidelines that the annotations need to adhere to, it is not clear how consistent the annotations need to be with one another in order to be acceptable. In fact, provided that enough data points are collected, it might be interesting to get a wide range of possible annotations from different annotators applying slightly different strategies, and then to look for patterns that apply across annotators. After the annotation process, a 'canonical' representation of the annotations could be obtained by filtering out infrequent annotations.[9]

A possible way to present the task to crowd annotators would be as shown in the example in Figure 3. In the first screen, the sentence to be annotated would be shown together with simple explanations of the frame types in the typical frames (which could be called 'mini-stories' for people unfamiliar with FrameNet). Next, for every frame type, the annotators would be asked to indicate if they think the text expresses it, and if so, which words in the sentence contribute to it. Finally, if the frame type is indeed expressed, the same question is asked for all of the (core) frame element.

For implementing the crowd-sourcing task, we propose making use of the Wordrobe gamification platform (Venhuizen et al., 2013). In Wordrobe, annotators get scores based on how consistent they are with other annotators, and are encouraged (e.g. through 'leader boards') to aim for higher scores. This encourages consistency and makes the annotation task more interesting for participants.

## 6. Discussion

The output of the inferred frame layer forms a scheme displaying a group of $n$ words for each frame mention that, according to the annotator, activates the corresponding frame type. In this section, we will argue that the inferences that led to each of these annotations can be categorized as either 'conventional' (i.e., always apply) or 'situational' (i.e., only apply in a specific context). We expect that most conventional inferences indicate coverage gaps in FrameNet. Once identified, these could be used to enrich the database. On the other hand, we expect the situational inferences to

be pragmatic instead of lexical in nature. In the following subsections, we will elaborate on the potential benefits of categorizing the output in this way.

### 6.1. Conventional Inferences and FrameNet Coverage

Certain annotations can be categorized as conventional. These annotations could not be performed in traditional FrameNet, but nevertheless seem to show a consistent mapping to the same targets across texts, and therefore might show a lexical coverage problem. These conventional inferences can provide useful insights for enriching or adapting the FrameNet database. The most typical examples of annotations that reveal coverage problems, are the ones related to the Lexical Gaps Problem (see Section 4.3.). When a word that is not yet listed in FrameNet is consistently annotated as activating a particular frame type, this word might be a lexical unit that is still missing in the frame type's inventory and could be added to it. However, because of the 'looseness' of the inferred frame layer, it is also possible that a word is very often annotated with a particular frame type, but does not qualify for being a lexical unit in the standard FrameNet sense.

For instance, "perpetrator" is currently not listed in FrameNet, but is conceptually relevant for OFFENSES, so it is conceivable that many annotators would annotate it as activating this frame type, even though it does not fit well with the structure of OFFENSES (which exclusively lists kinds of offenses such as "murder.n", "robbery.n"). However, the fact that the word is frequently annotated still suggests it should be added to FrameNet. A potential strategy to deal with this is to look for a better fit in frame types directly related to the one that is annotated. In this case, a good fit could be COMMITTING_CRIME (as we argued previously), which is connected to OFFENSES through the Is_used_by relation.

Another type of conventional annotation that provides cues for enriching FrameNet is related to the Frame Overlap Problem (see Section 4.3.): if annotators consistently annotate particular frame types on the inferred frame layer as activated by the same targets, this could be a strong indicator that there exists a relation between these frame types. For instance, if OFFENSES is often annotated for the same lexical items as KILLING, then these frame types are likely to be related.

Finally, the Frame Overlap problem can also provide cues that some lexical units are conceptually related to more than one frame type. Even when one of these frame types clearly fits best (e.g., HOSTILE_ENCOUNTER for "shootout", see example (8)), the conceptual content of another frame type may still be relevant to such a degree that it can be viewed as part of the lexical meaning of the target word. This could be suggested by a large number of annotations of this frame type on the inferred frame layer (e.g., USE_FIREARM for "shootout"). A possible way for encoding this in the lexicon would be to introduce frame-lexical unit relations in FrameNet. Currently, lexical units can only be related to frame types through the 'evoke' relationship, which means

---

[9]This should be done on different levels. For example, in Figure 3b, there should be a mention of KILLING in the final representation only if a majority of annotators answers "yes" (frame level); "shootout" should be kept as a target word for this mention only if a majority of annotators included it in their target span (target level); and "police" should be kept as a mention for the Killer frame element only if a majority of annotators included it (frame element level).

that every lexical unit can be related to only one frame type. However, as we have shown, lexical units can make the conceptual content of more than one frame type relevant without, strictly speaking, evoking all of these frame types. Allowing for secondary frame-lexical unit relations would allow us to model one-to-many mappings without weakening the existing 'evoke' relationship.

## 6.2. Situational Inferences and Pragmatics

The remainder of the annotations in the inferred frame layer can be categorized as situational. For instance, from the conceptually related linguistic components in sentence (5), KILLING is inferred with the aid of situational knowledge about the incident. This inference differs from the inference leading to BIRTHDAY_PARTY in (3), which is derived from both cultural knowledge and cues that are not conceptually related but frequently co-occur in the context of this frame type.

In the field of Gricean pragmatics, annotations like the one in (5) can be analyzed with respect to the means of inference (entailment, implicature, etc., see Levinson (1983) and Grice (1975)) by which frame mentions are pragmatically derived. Also, one could investigate external factors, such as historical distance and cultural background, underlying these inferences.

Another way in which this type of situational inference is relevant for pragmatics is by exposing discourse relations. This crucially depends on the observation that event instances, after being introduced in the beginning of a text, may be implicated in the remainder of the text. By annotating the referential relationship between the initial event mention and implicated event mentions, we implicitly capture this discourse relation and use it to combine the conceptual content from the frame types they evoke. For example, in the text in Table 1, once the murder event instance has been introduced (by "a truck was deliberately driven into the Christmas market … leaving 12 people dead", which under our approach could be annotated with KILLING), it will be implicitly active in the remainder of the text. This leads words like "perpetrator" (which evokes COMMITTING_CRIME) to be interpreted against the background of this event. Marking the two event mentions as referentially related then allows us to connect their associated frame mentions as well. Given that KILLING is still 'active' in the discourse, we can infer that "perpetrator" refers to a murder, and not to some other crime.

The incidental nature of these inferences makes it hard for researchers to model them in such a way that they can be added to FrameNet. One could wonder if researchers want incidental relations between frame types to be implemented in such a lexicographical project at all. Rather, situationally inferred frames show that even a fully developed version of FrameNet would not allow us to annotate all frame mentions referring to an event instance.

## 7. Summary

In this paper, we introduced a new use case of FrameNet: using frame annotations for showing how a single event instance in the real world can be conceptualized in text in different ways using frames. We showed that, in some cases (e.g. in the example in Figure 1), this can be done within the standard FrameNet annotation framework. However, in many cases the annotation scheme needs to be extended in order to allow for annotating frame mentions without an explicit lexical target. As a general solution, we proposed adding an *inferred frame* layer that allows arbitrary text spans to serve as a 'trigger' for any number of frame mentions, and suggested two possible ways to annotate the layer: either using a traditional FrameNet annotation process with annotators trained specifically for the task, or using crowd-sourcing. Finally, we show that the output of the inferred frame layer could be used as a basis for pragmatic analysis, and for extending the lexical coverage of FrameNet.

## 9. Bibliographical References

Andor, J. (2010). Discussing frame semantics: The state of the art: An interview with Charles J. Fillmore. *Review of Cognitive Linguistics*, 8(1):157–176.

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

Chang, N., Petruck, M. R. L., and Narayanan, S. (2002). From frames to inference. In *In Proceedings of the First International Workshop on Scalable Natural Language Understanding*.

Fillmore, C. J. (2008). The merging of "frames". In Favretti R. Rossini, editor, *Frames, corpora, and knowledge representation*, pages 2–12. Bologna: Bononia University Press.

Grice, H. P. (1975). Logic and conversation. In Peter Cole et al., editors, *Speech acts*, pages 41–58. Brill, Leiden.

Hartmann, S., Kuznetsov, I., Martin, T., and Gurevych, I. (2017). Out-of-domain FrameNet semantic role labeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 471–482.

Levinson, S. C. (1983). *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Marzinotto, G., Auguste, J., Bechet, F., Damnati, G., and Nasr, A. (2018). Semantic frame parsing for information extraction: the CALOR corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Minsky, M. (1974). A framework for representing knowledge. *MIT-AI Laboratory Memo*, 306.

Palmer, A. and Sporleder, C. (2010). Evaluating FrameNet-style semantic parsing: the role of coverage gaps in FrameNet. In *COLING 2010: Posters*, pages 928–936.

Postma, M., Remijnse, L., Ilievski, F., Fokkens, A., Titarsolej, S., and Vossen, P. (this workshop). Combining

conceptual and referential annotation to study variation in framing.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., and Schefczyk, J. (2010a). FrameNet II: Extended theory and practice.

Ruppenhofer, J., Sporleder, C., Morante, R., Baker, C., and Palmer, M. (2010b). SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. In *Proceedings of the 5th International Workshop on Semantic Evaluation*.

Sikos, J. and Padó, S. (2018). Framenet's using relation as a source of concept-based paraphrases. *Constructions and Frames*, 10(1):38–60.

Swayamdipta, S., Thomson, S., Dyer, C., and Smith, N. A. (2017). Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold. *arXiv preprint arXiv:1706.09528*.

Venhuizen, N. J., Basile, V., Evang, K., and Bos, J. (2013). Gamification for word sense labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers*, pages 397–403.

Vossen, P., Caselli, T., and Cybulska, A. (2018a). How concrete do we get telling stories? *Topics in cognitive science*, 10(3):621–640.

Vossen, P., Ilievski, F., Postma, M., and Segers, R. (2018b). 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)*.

Vossen, P., Ilievski, F., Postma, M., Fokkens, A., G., M., and L., R. (in press). Large-scale cross-lingual language resources for referencing and framing. Paper to be presented at LREC 2020.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: a free collaborative knowledge base.

Wang, C., Akbik, A., Chiticariu, L., Li, Y., Xia, F., and Xu, A. (2017). CROWD-IN-THE-LOOP: A hybrid approach for annotating semantic roles. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark.

## 10.   Language Resource References

Collin F. Baker. (2015). *FrameNet*. International Computer Science Institute, Berkeley, version 1.7.

Vrandečić, Denny and Krötzsch, Markus. (2014). *Wikidata: a free collaborative knowledge base*.