LREC-COLING 2024

The First Workshop on Reference, Framing, and Perspective

Workshop Proceedings

Editors
Pia Sommerauer, Tommaso Caselli, Malvina Nissim,
Levi Remijnse, Piek Vossen

25 May, 2024
Torino, Italia

Proceedings for the First Workshop on Reference, Framing, and Perspective

ISBN 978-2-493814-27-2
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Preface

When something happens in the world, we have access to an unlimited range of ways (from lexical choices to specific syntactic structures) to refer to the same real-world event. We can choose to express information explicitly or imply it. Variations in reference may convey radically different perspectives. This process of making reference to something by adopting a specific perspective is also known as framing. Although previous work in this area is present (see Ali and Hassan (2022)'s survey for an overview), there is a lack of a unitary framework and only few targeted datasets (Chen et al., 2019) and tools based on Large Language Models exist (Minnema et al., 2022). In this workshop, we propose to adopt Frame Semantics (Fillmore, 2006) as a unifying theoretical framework and analysis method to understand the choices made in linguistic references to events. The semantic frames (expressed by predicates and roles) we choose give rise to our understanding, or framing, of an event. We aim to bring together different research communities interested in lexical and syntactic variation, referential grounding, frame semantics, and perspectives. We believe that there is significant overlap within the goals and interests of these communities, but not necessarily the common ground to enable collaborative work.

The Workshop Organizers

Pia Sommerauer, Vrije Universiteit Amsterdam
Tommaso Caselli, Rijksuniversiteit Groningen
Malvina Nissim, Rijksuniversiteit Groningen
Levi Remijnse, Vrije Universiteit Amsterdam
Piek Vossen, Vrije Universiteit Amsterdam

Organizing Committee

Pia Sommerauer, Vrije Universiteit Amsterdam
Tommaso Caselli, Rijksuniversiteit Groningen
Malvina Nissim, Rijksuniversiteit Groningen
Levi Remijnse, Vrije Universiteit Amsterdam
Piek Vossen, Vrije Universiteit Amsterdam

Program Committee

Özge Alacam
Jelke Bloem
Oliver Czulo
Loic De Langhe
Mark Finlayson
Antske Fokkens
Fritz Günther
Sanne Hoecken
Eva Huber
Allison Lahnala
Lucy Li
Roser Morante
Seyed Mahed Mousavi
Sebastian Padó
Massimo Poesio
Marco Rovera Rovera
Thomnas Schmidt
Sara Tonelli
Tiago Torrent
Leonie Weissweiler
Philipp Wicke
Stella Verkijk
Alexander Willich
Alexander Ziem

Table of Contents

Tracking Perspectives on Event Participants: a Structural Analysis of the Framing of Real-World Events in Co-Referential Corpora

Levi Remijnse, Pia Sommerauer, Antske Fokkens, Piek Vossen
Vrije Universiteit Amsterdam, Computational Linguistics and Text Mining Lab
De Boelelaan 1105, 1081 HV Amsterdam
{l.remijnse, pia.sommerauer, antske.fokkens, p.t.j.m.vossen}@vu.nl

Abstract

In this paper, we present the outcome of a structural linguistic analysis performed on a referentially grounded FrameNet dataset. In this dataset, multiple Dutch events are referenced by multiple co-referential Dutch news texts. Mentions in those documents are annotated with respect to their referential grounding (i.e., links to structured Wikidata), and their conceptual representation (i.e., frames). Provided with each document's temporal reporting distance, we selected documents for two events - the Utrecht shooting and MH17 - and performed an analysis in which we tracked the events' participants over time in both their focalization (number of mentions) and their framing (distribution of frame element labels). This way, we use the carefully collected and annotated data to schematize shifts in focalization and perspectivization of the participants as a result of the constantly developing narrative surrounding the events. This novel type of linguistic research involves reference to the real-world referents and takes into account storytelling in news streams.

Keywords: FrameNet, narratology, referential grounding

1. Introduction

From the moment an event occurs in the world, it generates streams of co-referential news articles.[1] In particular, events that are of interest to society (e.g., mass shootings, music festivals, royal weddings) are reported on by large volumes of documents. Over time, they keep being reported on with regard to their aftermath (e.g., a shooting causes funerals, police investigations, arrests, trials). A narrative develops in which, with every new related topic, different people involved in the event might become the focus. In support of the changing narrative, writers might also change their perspective on those participants. See the examples below, taken from our data, with per sentence the historical distance to the main event, respectively the same day, one day later and the fourth day after:

(1) Vermoedelijk is er daarna **iemand** uit de tram gesprongen.
presumably is there that.after someone out the tram jumped
'Presumably someone jumped out of the tram afterwards.' (day 0) (2019)

(2) De **hoofdverdachte** van de aanslag [...] is de enige verdachte die nog vastzit.
The main.suspect of the attack [...] is the only suspect still that still stuck.sit.
'The main suspect in the attack is the only suspect still in custody.' (day 1) (2019)

(3) **Gökmen Tanis** bekent schietpartij in tram in Utrecht.
Gökmen Tanis confesses shooting.party in tram in Utrecht
'Gökmen Tanis confesses to shooting in tram in Utrecht.' (day 4) (2019)

All three example sentences stem from different documents with different temporal reporting distances, i.e., the temporal distance between the event date and the publication date. The boldfaced mentions co-refer the same entity participating in different events that are part of the same storyline. This entity is perspectivized accordingly. In (1), focus is on his involvement in a mass shooting in a tram; in (2), he is a suspect in custody; and in (3), he is the agent of a confession. Both events in (2) and (3) are related to the event in (1): the mass shooting.

The examples show that, over time, news streams reporting on a single significant event display a continuously developing narrative in which different aspects of the event and its aftermath are topicalized, and the same participant is perspectivized differently.

Suppose we want to perform a structural linguistic analysis in which we get a grip on the way news documents perform storytelling of real-world events. This requires a referentially grounded cor-

[1] We define an event as a specific event instance of a particular event type, e.g., a killing incident happening at a specific location, time, and involving certain participants.

First Workshop on Reference, Framing, and Perspective, pages 1–12

pus with multiple documents reporting on the same event. The documents themselves need to be annotated with information regarding both the referential grounding (which mentions co-refer to which event participant) and conceptual representation (what perspectives do the mention take on in reference to that event participant). Yet, NLP tasks and language resources only cover conceptual representation and reference separately. For instance, the tasks of co-reference resolution (Filatova and Hatzivassiloglou, 2004; Choubey et al., 2018) and entity-linking (Hachey et al., 2013; Getman et al., 2018) contribute to the study of reference. Both Abstract Meaning Representation (Banarescu et al., 2013) as a formal framework and FrameNet (Ruppenhofer et al., 2010; Baker et al., 2003) as a lexicographic paradigm focus on concept description. Yet, for the purpose of a linguistic analysis of referentially grounded data, a dataset should provide information regarding all three components: form, referent and concept.

Therefore, in this paper, we make use of the referentially grounded corpus collected by Remijnse et al. (2022), with documents reporting on real-world events. Focusing on perspectivization as operationalized within frame semantics, this corpus is annotated with FrameNet frames, modeling mentions of participants with semantic types (Remijnse et al., 2022; Postma et al., 2020). This combination of referential grounding and frame semantic information enables us to study the ways in which news streams frame their events over time, as the narrative surrounding those events develops. We take two events commonly known as the Utrecht shooting[2] and MH17[3], focus on the participants of those events, and analyze the ways in which they are framed in our corpus, as a reflection of the developing narrative over time.

We make the following contributions:

- We release a dataset with Dutch reference texts reporting on the Utrecht shooting and MH17. The documents are annotated with links to structured data and FrameNet frames.[4]
- We formulate a model of variation in framing of events that takes into account storytelling and temporal reporting distance.
- Given the dataset and our model, we provide a structural analysis of the linguistic framing of events' participants over time.
- We show patterns in our data of both focalization and perspectivization of participants across events.

This paper is structured as follows. We first discuss related work and background in Section 2. Building on that, we explain our model of computational storytelling in Section 3. We then discuss our selection of data and analysis method in 4. Section 5 provides the results and discussion of the data analysis. We conclude in section 6 and point out limitations in section 7.

2. Background

In this section, we cover related work that the research in this paper is built on, namely referentially grounded corpora (2.1), work in perspectivization (2.2) and narratology (2.3).

2.1. Referentially Grounded Corpora

If the aim is to perform a linguistic analysis that incorporates referential grounding of events, then this grounding affects the very first step of data collection: a corpus needs to exhibit a large variety of documents referencing the same events. Much work in corpus building follows a **text-to-data** method, i.e., starting from text to derive annotation sentence-by-sentence (e.g., OntoNotes (Pradhan et al., 2007), ECB (Bejan and Harabagiu, 2010) ECB+ (Cybulska and Vossen, 2014) and AEC2005 (Peng et al., 2016)). This work is evaluated as labour intensive and time-consuming, thus resulting in small numbers of annotated texts with low intra-document co-reference (10 mentions on average) and low cross-document co-reference (Vossen et al., 2018). Moreover, concrete links between mentions and structured data are absent.

In order to efficiently aggregate multiple co-referential texts per event, Vossen et al. (2020) reversed the process by developing **data-to-text** based software called MWEP, which takes a Wikidata identifier denoting an event type as input, and returns structured Wikidata concerning all events that are grouped under this event type in the Wikidata knowledge base (Vrandečić and Krötzsch, 2014). Per event, the structured output is accompanied with referential news texts crawled from corresponding Wikipedia pages (Simpson et al., 2010). By starting from structured data in aggregating documents, those documents are by default grouped under their event and annotation merely serves as validation.

For the purpose of our research, since we need both structured data and unstructured data grouped under events that have entries in Wikidata, we consider MWEP to be the most useful corpus compilation software.

[2] https://www.wikidata.org/wiki/Q62090804
[3] https://www.wikidata.org/wiki/Q17374096
[4] Our data is freely available at http://dutchframenet.nl/data-releases/

2.2. Perspectivization

The phenomenon of perspectivization (or "framing") has been analyzed in many different fields, e.g, cognitive linguistics (Horst, 2020; Ziem et al., 2018), political studies (Druckman, 2001; Iyengar, 1994; Entman, 1993), and media studies (Bryant and Finklea, 2022; Cacciatore et al., 2016; Van der Pas, 2014). Framing analysis applied to written text has been the focus of frameworks, such as Critical Discourse Analysis (CDA) (Van Dijk, 2015; Fairclough, 2013), a paradigm that views language as social practises and critically reads discourse as expressions of social power. Also, the above approaches to framing all have to some extent been gaining attention in computational research. For example, Mendelsohn et al. (2021) model political framing in immigration discourse on social media using multiple framing typologies from political communication theory. Walter and Ophir (2021) predict media framing of election candidate campaigns using variables at the level of candidate, state, and electoral race.

FrameNet has a more lexicographic focus, interpreting words in terms of **semantic** frames, i.e., schematized events with participants modeled as highly specified semantic roles, i.e., **frame elements** (Ruppenhofer et al., 2010).

The abovementioned fields of research use different definitions of framing, ranging from fine-grained semantic framing to coarse-grained political framing. Although these are distinct definitions, the types of framing are interrelated (Sullivan, 2023). For example, the fine-grained semantic framing of events and their participants as evoked by constructions is foundational to language, but can be used in combination with communicative means to shape political frames. In this paper, we focus on the more fine-grained semantic framing, and implement certain notions of narratology to see how this framing reflects some of the higher order developments in the narrative surrounding an event over time.

In recent decades, besides the creation of cross-linguistic FrameNets (Torrent et al., 2018; Djemaa et al., 2016; Burchardt et al., 2009; Ohara et al., 2004), the database has been used in many different NLP annotation platforms, like Webanno (Eckart de Castilho et al., 2016) and Salto (Burchardt et al., 2006). More recently, Xia et al. (2021) created LOME, a multilingual end-to-end frame parsing system. With this Large Language Model, texts from any target language can be parsed with both frames and frame elements. Minnema et al. (2022b) implemented LOME in a multilingual tool called SocioFillmore, which performs a large-scale analysis of perspectivization strategies across texts. All those different FrameNet databases and parsing implementations have contributed substantial insight in perspectivization of events. Minnema et al. (2022a) investigate how responsibility is framed by linguistic expressions in news texts reporting on events of gender-based violence. As far as we know, they are the first to use FrameNet to analyze perspectives on referents in corpora. Yet, they did not involve the aspect of narratology and how this affects the framing of a participant over time.

In order to track an event's participant with respect to its conceptual role in the narrative of a corpus, FrameNet can serve as a suitable proxy for modeling semantic framing evoked by the mentions in that corpus. Its definition of linguistic framing depicts frames as events, its database is extensive and cross-domain, and the labels and definitions of each frame's frame elements are highly specific. Given a frame, its frame elements are proxys for the perspectives taken on the event's participants within that frame. For example, in a news text reporting on a shooting, when a predicate evokes the **Killing** frame, the frame elements realize the perspectives: the **Killing**@KILLER perspective or the **Killing**@VICTIM perspective. Yet, on top of information about perspectives, we need referential grounding: information about the referential links between mentions and structured participants given a real-world event. This way, we can get insights about which perspectives are projected on which participants.

In evaluating the aforementioned FrameNet contributions, Remijnse et al. (2022) concluded that referential grounding is still absent from the annotated data. With only the frames, we lack information about who is mentioned and who is framed. The authors built the DFN annotation tool, a resource that combines frame annotations and co-referential annotations in a dual annotation layer. After loading documents grouped under their real-world event and paired with structured data in the interface, co-referential annotation is achieved by linking in-text mentions to that structured data. On top, the same mentions are annotated with semantic frames. In the resulting annotation scheme, per participant of an event, all mentions linked to that participant are schematized with their frame-annotations.

For the purpose of this paper, since we need our corpus to be annotated with information regarding both framing and referential grounding, we make use of the DFN annotation tool to annotate our corpus data.

2.3. Narratology

When analyzing the linguistic framing of participants of a real-world event in a referentially grounded corpus, we need to take into account that on a higher order, their fine-grained conceptual representations (i.e., the frames and frame elements) reflects their role in a continuously changing narrative. Describing events by means of creating narratives is an ability inherent to human nature (Boyd, 2009; Gottschall, 2012). In analyzing narratology as a discourse phenomenon, a text displays a sequence of causally related events involving participants, which constitutes a storyline (Mani, 2014; Bal and Van Boheemen, 2009; Forster, 1956). Vossen et al. (2021); Bal and Van Boheemen (2009) point out the following requirements that a storyline needs to meet in order to qualify as a narrative:

- The ordered events lead to a **climax**, which serves the document's topic.
- It follows a **focalizer**, i.e., one of the storyline's participants.
- The focalizer takes on a certain **perspective**, a certain role in the story.

Vossen et al. (2021) further break down the storyline's event sequence in a formal model that classifies pre-climax events and post-climax events, and derive a novel annotation scheme applied to news texts.

NLP tasks that model and extract narratological information from corpus data have been scarce. First attempts resulted in entailment recognition tasks (Dzikovska et al., 2013; Bowman et al., 2015), end-of-story prediction tasks (Mostafazadeh et al., 2016, 2017) and narrative chains (Chambers and Jurafsky, 2008, 2009). Although these NLP systems pose a relevant first step in getting structural insight into storytelling, they still have been evaluated as "limited and in their infancy" (Caselli et al., 2021, 2). Moreover, they still do not combine conceptual representation and referential grounding.

In the next section, we introduce a theoretical model of storytelling that involves referential grounding and framing.

3. A Model for Variation in Framing and Storytelling

In this section, we describe our theoretical model for variation in framing of real-world events with the incorporation of narratology. We start with a description of real-world events, referential grounding, and framing. Given an event instance in the world, this event instance involves structured data involving participants, location and time. Generally, we assume that people describe and report on an event instance at a granularity that fits their daily interest, and consisting of sequences of more fine-grained events. Besides structured data, this event instance generates a stream of co-referential texts with varying temporal reporting distances. The mentions in those texts can be linked to the structured data. On top of this referential relation, the mention also evokes a semantic frame or expresses a frame element. The set of mentions across documents that co-refer the same entity, can exhibit various frame elements. This way, we can model variation in framing.

We can apply the notions of narratology introduced in Section 2.3 as follows. The storyline is conveyed by the set of causally ordered mentions in a reference text. One of the entities in the structured data is selected by the writer as the document's focalizer. When mentioning this focalizer, it is perspectivized by the frame element expressed by the mention.

Instead of topicalizing the event instance, we expect documents to sometimes topicalize distinct yet related events. The motivation for reporting on those distinct events is their relevance to the more salient event instance. In other words, the salient event instance is always referenced, but not always as the document's climax. The climax can also be one of the events that affect the salient event instance or is caused by it in the aftermath. We label the salient event instance as the **anchor incident**. See an example of an anchor incident and both its reference texts and related events on a timeline in Figure 1.

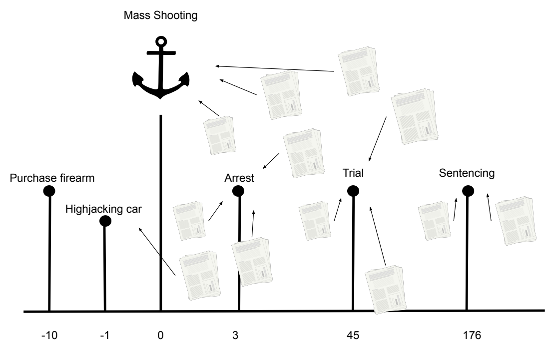


Figure 1: A timeline with a mass shooting as the anchor incident (indicated by an anchor) and different related events occurring before and after (indicated by black dots), together with reference texts (indicated by the newspapers). X-axis = number of days from the onset of the Anchor incident. The arrows from the reference texts to the events indicate the topic of writing, and thus a shift in narrative in the overall news stream over time.

Figure 1 displays a mass shooting as the anchor incident. On the timeline (x-axis), different yet related events occur before and after the shooting. Any of those events can form the climax of a reference text. Yet, in order for all those reference texts to show relevance of reporting their climax, the writers have to ground it in the related Anchor incident. We call this process **anchoring**: using at least minimal reference in your document to ground your climax incident in the anchor incident.[5]

With respect to the anchor incident's participants, we can make the following distinction:

- **DpA**: Directly related participant of the Anchor incident. This participant is present at the scene of the incident itself (e.g., shooters and victims of a mass shooting).
- **IpA**: Indirectly related participant of the Anchor incident. This participant is only indirectly involved in the Anchor incident (e.g., relatives and criminal investigators of a mass shooting).

The structured data exhibits both DpAs and IpAs. We expect IpA's to be referenced occasionally. Assuming that every reference text needs anchoring, we expect that the DpAs will always be referenced, if only briefly. Yet, when the narrative evolves around them, they might also undergo a change in their perspectivization, i.e., how they are framed.

To conclude, based on our theoretical model, we formulate the following hypotheses about the focalization and perspectivization of the participants of the anchor incidents in our data:

- On a referential level, we expect shifts in focalization between participants as a result of different topics that push the narrative surrounding the Anchor incident forward. We expect that the documents keep referencing DpAs over time, while IpAs are introduced occasionally.
- On a conceptual level, we expect that the focalized participants also show variation in framing: they show different perspectives as a result of their role in a new related topic. However, DpAs can be frequently referenced while not showing variation in framing. This is then a result of anchoring: they have to be mentioned to anchor the document's climax, but if they do not play an active role in the storyline, there is no need to change their framing.

In the next section, we describe our methodology with respect to the analysis of two anchor incidents.

4. Methodology

In this section, we describe our method. This includes corpus compilation (4.1), the annotation process (4.2), document clustering (4.3), participant selection (4.4) and data analysis (4.5).

[5]One could argue that anchoring is a product of following the Gricean maxim of Relevance, i.e., make your contribution, this news report, relevant to the reader. (see Grice (1991) and Grice (1975)).

4.1. Corpus Data

We make use of the corpus data collected by Remijnse et al. (2022). They used MWEP to query Wikidata with preset event type identifiers. For each event type, MWEP returned both structured and unstructured data for two anchor incidents. We selected two anchor incidents. The first incident is the 2019 Utrecht shooting (Q62090804), which is an instance of mass shooting (Q21480300). For this incident, MWEP returned 42 Dutch reference texts. The second incident is Malaysia Airlines Flight 17 (a.k.a MH17, Q17374096), which is an instance of aircraft shootdown (Q6539177). For this incident, MWEP returned 117 Dutch reference texts.

4.2. Annotation Process

Provided with both a reference text and the anchor incident's structured data in the annotation tool's interface, four annotators were trained to annotate different texts. Per text, they first linked in-text mentions to the structured data (entity-links). Then, they performed frame-annotations on the same text. Whenever the annotators could not find an antecedent for a frame element within its predicate's sentence boundaries, they were instructed to look for an antecedent across sentences. The main motivation for this instruction is that we assume that participants are sometimes implicitly involved in descriptions elsewhere in discourse, contributing to the storyline. As a result, mentions of participants get *n* annotations of frame elements belonging to frames evoked in different sentences.

For the Utrecht Shooting subcorpus, this process resulted in 1,459 links to 13 different entities, 1,830 frame-annotations and 5,807 assignments of frame elements. For the MH17 subcorpus, the annotation process resulted in 3,390 links to 37 different entities, 3.436 frame-annotations and 10,978 frame element assignments.

4.3. Temporal Distance Clustering

As discussed in Section 3, we expect that given an anchor incident, whenever a distinct yet related event occurs, this triggers a stream of reference texts reporting on this event as their climax. Those reports push the narrative of the anchor incident forward, possibly changing the framing of its participants. In order to model and vizualize this shift in narrative and analyze participant's framing, we first visualized the distribution of the reference texts per anchor incident, see Figure 2. We observed that in certain time periods, the documents appear to cluster. We take these clusters as a proxy for finding news streams reporting on a novel event that is topicalized as a new climax.

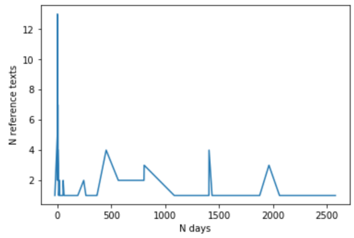


Figure 2: The distribution of reference texts reporting on MH17 over time, from the onset of the anchor incident.

For both anchor incidents, we selected the four periods on the timeline in which most reference texts were published.[6] Right after the incidents, those periods are days with few days of silence (i.e., no publications). Here, we started to widen the scope of a high frequent publication period, but already set borders when encountering a day of silence. The peaks of reference texts later on in the timeline show a larger spread of documents. Thus, we started to widen the scope of the cluster, but still included documents that were published after few days of silence. Here, the silence period had to be longer. We attempted to put the borders of the temporal classes where the reference texts would be equally balanced, as a means of normalization. Documents that fell between the borders of the classes, were removed from the final dataset. This results in the temporal distance classes (TDC) shown in Table 1.

Utrecht shooting	
Temporal Distance Class	N docs
1. Day 0	13
2. Day 1	11
3. Day 4-12	9
4. Day 37-703	9
MH17	
Temporal Distance Class	N docs
1. Day 0-1	18
2. Day 6-22	20
3. Day 333-1212	18
4. Day 1407-2581	18

Table 1: Per Anchor incident, the temporal distance classes with N of documents.

We checked the titles of the documents within each class to see if they would be reporting on the same events. Overall, this turned out to be the case, e.g., most reference texts on Day 0 of the Utrecht shooting cover a manhunt, while reference texts in Day 37-703 largely cover a trial.

[6]Considerable related work contributes to the process of clustering documents of various publication dates with respect to their topic (Wang et al., 2014; Wang and McCallum, 2006; Blei and Lafferty, 2006). Yet, the proposed models involve linguistic information from those documents in their techniques. Since the linguistic information is the object of our analysis, we had to find a different way of setting cluster boundaries, in order to avoid circularity.

4.4. Participant Selection

Both anchor incidents contain a large set of structured entities. For clear vizualization purposes, we selected all DpAs and the top three IpAs showing the highest number of entity-links. Table 2 shows the statistics in terms of number of entity-links and frame elements per participant (after temporal distance clustering). We will analyse these across TDCs in Section 4.5.

Utrecht shooting			
Participant	status	N entity-links	N FEs
Gökmen Tanis	DpA	329	983
victims	DpA	197	525
police officers	IpA	74	137
other suspects	IpA	60	89
Utrecht citizens	IpA	46	62
MH17			
Participant	status	N entity-links	N FEs
victims	DpA	239	462
suspects	DpA	126	356
relatives	IpA	123	169
Russia	IpA	166	283
Dutch Safety Board	IpA	69	96

Table 2: Selected participants per anchor incident with involvement status, number of entity-links and number of frame element annotations.

4.5. Data Analysis

We analyze the participants' referential grounding as well as their perspectivization separetely per anchor incident. With respect to referential grounding, we take frequency distribution of the entity-links as a proxy for focalization: we expect participants that are referenced most are focalized in storytelling. We distributed the number of entity-links of the participants per TDC to observe any shift in this focalization.

Regarding the perspectivization of the participants, we take frequency distribution of frame elements as a proxy for variation in framing: within a participant, a strong shift in frame element frequency shows a change in perspectivization. This new perspective that the participant is assigned with, is part of a change in narrative. Given that the frequency distribution of frame elements given a participant shows a long tail, we limited the vizualization of the data to the proportionally most salient frame elements. Per participant and per TDC, we sliced the top frequent 35% of frame elements. After experimenting with different percentages, this proportional number gave for every participant a sufficient number of different frame element types to interpret. Then, per participant, we took the union of those slices across TDCs. The resulting set contains the frame elements we assume convey most information about how the participant in general has been framed over time. Next, per participant, we plotted per frame element type its proportional frequency distribution across time buckets.

In the next section, we present and discuss the results of our data analysis.

5. Results and Discussion

In this section, we present and discuss the results of both the referential part and the framing part of our analysis. We first turn to the data of the Utrecht shooting. See Figure 3a for the referential grounding of the participants of the Utrecht shooting across TDCs. Overall, we find that the narrative evolves around the DpAs as they show the highest number of references. The focus shifts over time, from Gökmen Tanis (the perpetrator of the shooting) on Day 1 to the victims on Day 4-12, then back to Tanis on Day 37-703.

See Figure 3b-d for the proportional distribution of frame elements that were used to frame the participants across TDCs. We plotted this distribution for a selection of participants. A first observation is that all three participants show variation in framing over time. Their perspectives are not fixed, but subject to change. This is only possible if the narrative changes and writers choose to give the participants new perspectives.

Second, we notice a correlation between each participant's framing and its referential grounding. When a participant shows a significant change in number of references, its framing changes accordingly. For example, whereas Gökmen Tanis is focalized in Day 37-703 in 3a, in that same TDC in Figure 3b, frame elements such as **Hit_target**@AGENT decrease, while many frame elements, such as **Trial**@DEFENDANT, start to increase. Similarly, the victims are focalized in Day 4-12 in 3a, whereas their framing changes significantly in that same TDC in 3c. Finally, we see that the police officers show a decrease of references on Day 1 in Figure 3a, and similarly in Figure 3d, most frame elements drop to zero in that same TDC, while **Intentionally_act**@AGENT is introduced to henceforth frame this group.

Turning to the data of MH17, see Figure 4a for the referential grounding of the participants of MH17 across TDCs. This figure shows a shift in top frequent references between a variety of participants. Here, the IpAs play a larger role in the narrative as compared to the Utrecht shooting. Whereas the Dutch Safety Board (an investigation board) peaks in number of references at Day 333-1212, that number drops to zero in subsequent TDC, when the suspects are introduced. Together with the Russian government, they take over the narrative, completely backgrounding the Dutch Safety Board.

See Figure 4b-d for the proportional distribution of frame elements that were used to frame the participants across TDCs. We plotted this distribution for a selection of the participants. In Figure 4c-d, again we observe strong variation in framing over time. In fact, each TDC in Figure 4d shows different top frequent frame elements. Figure 4c and 4a show both a clear change in mention frequency and a change in frame element types in Day 333-1212.

Figure 4b shows no variation in framing over time. In fact, **Catastrophe**@PATIENT increases in the final TDC, whereas the participant simultaneously decreases in number of references in Figure 4a. This means that the victims do not play an active part in the narrative anymore and their perspective freezes. As a DpA, this group is still mentioned across documents in order to anchor the document's climax to MH17, but with a fixed set of frames.[7]

To conclude, we interpret the shift in frequency distribution of references in both Figures 3a and 4a as a shift in focalization between participants that play an active role in the development of the narrative. Furthermore, we notice that DpAs and IpAs behave differently between anchor incidents in their mention frequencies. In the Utrecht shooting, overall focus is on the DpAs, whereas in MH17, the IpAs are focalized to a stronger degree over time, and the DpA suspects is only introduced in the final TDC. It seems that the narrative surrounding each anchor incident is unique and affects different patterns of referential grounding of participants over time. The current analysis captures this development.

With respect to framing, we observe an overall correlation between a participant's focalization (sudden steep change in number of references) and variation in framing (shift in dominant frame type). The victims of MH17 are consistently framed with the same frame elements. Even when their number of references decreases over time, they are still necessarily mentioned to anchor the climax, but their part in the narrative has not changed, i.e., they do not require new perspectives.

[7] Remijnse et al. (2021) describe a similar process. They derive a fixed set of what they call **Anchor** frames that writers consistently evoke over time when anchoring a documents climax. This set of anchor frames is dependent of the anchor incident's event type.

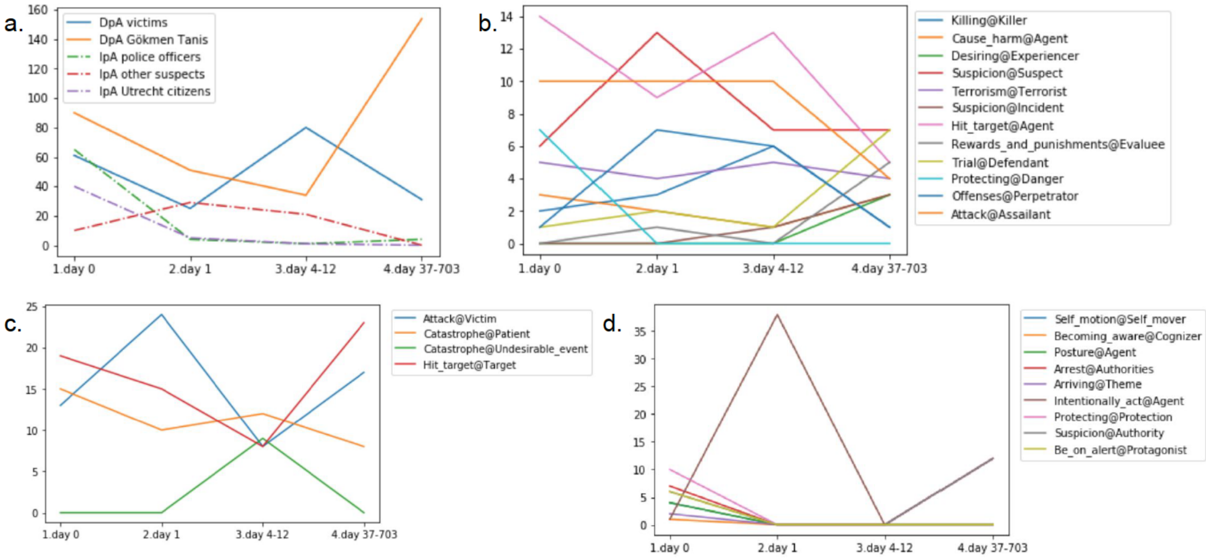


Figure 3: a. The frequency distribution of mentions of participants of the Utrecht shooting across TDCs; b-d. The proportional distribution of the top frequent frame elements across TDCs in reference to participants of the Utrecht shooting. b. Gökmen Tanis; c. victims; d. police officers. The frame element notations in the index can be read as frame@frame_element.

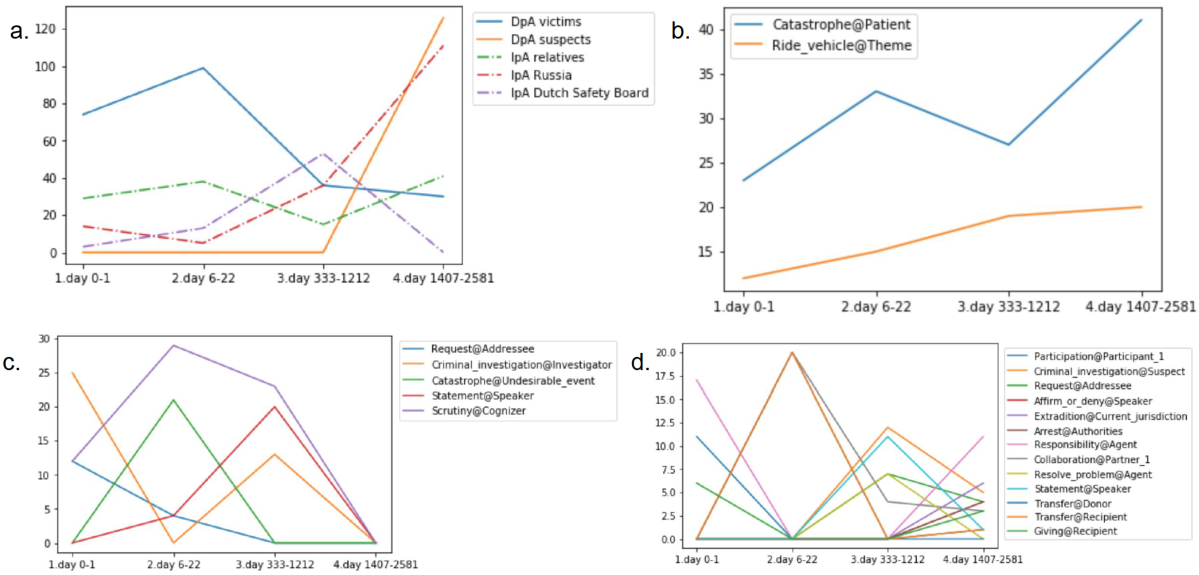


Figure 4: a. The frequency distribution of mentions of participants of MH17 across TDCs; b-d. The proportional distribution of the top frequent frame elements across TDCs in reference to participants of the Utrecht shooting. b. victims; c. Dutch Safety Board; d. Russian government. The frame element notations in the index can be read as frame@frame_element.

6. Conclusion

In this paper, we performed a structural linguistic analysis of variation in framing of participants of real-world events over time. In order to perform such an analysis, we met multiple requirements: collect a referentially grounded corpus accompanied with structured data; annotate the data with both entity-links and frames; and describe a theoretical model of variation in framing and narratology.

For two anchor incidents, we first analyzed the frequency distribution of the entity-links for the participants on a timeline, to observe a strong shift in focalization between participants, an indication of a change in narrative.

We then analyzed the frequency distribution of the participants' frame elements over time in order to measure the extent of variation in framing. Overall, we observe a correlation between shift in focalization between participants, and variation in framing within a participant: when a different participant is getting the focus of the narrative, this participant also gets a new perspective.

7. Limitations

The first limitation of our research is that it is limited to two incidents and one incident type. In order to find stronger patterns of focalization and perspectivization, we need to scale this to many more event types, incidents per event type and sources of reference text. We hope to do this in future work using automated techniques for frame annotation, coreference resolution and entity-linking. The MWEP tool can be used to collect large corpora of referentially grounded texts for this.

Another limitation is that we have now manually annotated the texts but need to rely on automatic techniques to scale our research. Automatic techniques will be less accurate and biased to assign more dominant frames and frame elements which may cause a bias in the analysis. Furthermore, current tools are not trained to annotate frame and frame element relation at the discourse level.

Finally, we hand-picked the TDCs for our analysis on the collected reference texts. This should be automated as well to apply this on a larger scale. However, what is a pause or not in the publication also depends on the ability to find all sources at any point of time that report on each incident. Furthermore, there could be multiple new events in the same TDC and related to the same anchor incident that are being reported.

8. Bibliographical References

Mieke Bal and Christine Van Boheemen. 2009. *Narratology: Introduction to the theory of narrative*. University of Toronto Press.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

David M Blei and John D Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120.

Ilona Bos, Amber Dujardin, Hans Marijnissen, Gidi Pols, and Petra Vissers. 2019. https://www.trouw.nl/home/om-kwart-voor-elf-staat-de-tram-stil-daarna-heel-utrecht a1d5dacf/.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. pages 632–642.

Brian Boyd. 2009. *On the origin of stories: Evolution, cognition, and fiction*. Harvard University Press.

Jennings Bryant and Bruce W Finklea. 2022. *Fundamentals of media effects*. Waveland Press.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, and Sebastian Pado. 2006. SALTO - a versatile multi-level annotation tool. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Michael A Cacciatore, Dietram A Scheufele, and Shanto Iyengar. 2016. The end of framing as we know it… and the future of media effects. *Mass communication and society*, 19(1):7–23.

Tommaso Caselli, Eduard Hovy, Martha Palmer, and Piek Vossen. 2021. *Computational Analysis of Storylines: Making Sense of Events*. Cambridge University Press.

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*, pages 789–797, Columbus, Ohio. Association for Computational Linguistics.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, page 602–610, USA. Association for Computational Linguistics.

Prafulla Kumar Choubey, Kaushik Raju, and Ruihong Huang. 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. Association for Computational Linguistics.

James N Druckman. 2001. The implications of framing effects for citizen competence. *Political behavior*, 23:225–256.

Myroslava Dzikovska, Rodney Nielsen, Chris Brew, Claudia Leacock, Danilo Giampiccolo, Luisa Bentivogli, Peter Clark, Ido Dagan, and Hoa Trang Dang. 2013. SemEval-2013 task 7: The joint student response analysis and 8th recognizing textual entailment challenge. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 263–274, Atlanta, Georgia, USA. Association for Computational Linguistics.

Richard Eckart de Castilho, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann. 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. The COLING 2016 Organizing Committee.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.

Norman Fairclough. 2013. *Critical discourse analysis: The critical study of language*. Routledge.

Elena Filatova and Vasileios Hatzivassiloglou. 2004. Event-based extractive summarization. In *Text Summarization Branches Out*, pages 104–111, Barcelona, Spain. Association for Computational Linguistics.

Edward Morgan Forster. 1956. *Aspects of the Novel*. Harcourt, Brace.

Roeland Franck. 2019. https://www.ad.nl/binnenland/gokmen-tanis-bekent-schietpartij-in-tram-in-utrecht a239a24a/.

Jeremy Getman, Joe Ellis, Stephanie Strassel, Zhiyi Song, and Jennifer Tracey. 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)*.

Jonathan Gottschall. 2012. *The storytelling animal: How stories make us human*. Houghton Mifflin Harcourt.

H Paul Grice. 1975. Logic and conversation. *Cole, P., and Morgan, J.(Eds.)*, 3.

Paul Grice. 1991. *Studies in the Way of Words*. Harvard University Press.

Ben Hachey, Will Radford, Joel Nothman, Matthew Honnibal, and James R. Curran. 2013. Evaluating entity linking with wikipedia. *Artificial Intelligence*, 194:130–150. Artificial Intelligence, Wikipedia and Semi-Structured Resources.

Kasper Hermans. 2019. https://www.metronieuws.nl/in-het-nieuws/2019/03/van-zanen-twee-verdachten-vrij-politie-ontkent/.

Dorothea Horst. 2020. Patterns 'we' think by? critical cognitive linguistics between language system and language use. *Yearbook of the German Cognitive Linguistics Association*, 8(1):67–82.

Shanto Iyengar. 1994. *Is anyone responsible?: How television frames political issues*. University of Chicago Press.

Inderjeet Mani. 2014. Computational narratology. *Handbook of narratology*, pages 84–92.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.

Gosse Minnema, Sara Gemelli, Chiara Zanchi, Tommaso Caselli, and Malvina Nissim. 2022a. Dead or murdered? predicting responsibility perception in femicide news reports. pages 1078–1090.

Gosse Minnema, Sara Gemelli, Chiara Zanchi, Tommaso Caselli, and Malvina Nissim. 2022b. SocioFillmore: A tool for discovering perspectives. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 240–250, Dublin, Ireland. Association for Computational Linguistics.

Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. 2016. A corpus and cloze evaluation for deeper understanding of commonsense stories. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 839–849, San Diego, California. Association for Computational Linguistics.

Nasrin Mostafazadeh, Michael Roth, Annie Louis, Nathanael Chambers, and James Allen. 2017. LSDSem 2017 shared task: The story cloze test. In *Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, pages 46–51, Valencia, Spain. Association for Computational Linguistics.

Levi Remijnse, Marten Postma, and Piek Vossen. 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, Groningen, The Netherlands (online). Association for Computational Linguistics.

Levi Remijnse, Piek Vossen, Antske Fokkens, and Sam Titarsolej. 2022. Introducing frege to Fillmore: A FrameNet dataset that captures both sense and reference. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 39–50, Marseille, France. European Language Resources Association.

Heather Simpson, Stephanie Strassel, Robert Parker, and Paul McNamee. 2010. Wikipedia and the web of confusable entities: Experience from entity linking query creation for TAC 2009 knowledge base population. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Karen Sullivan. 2023. Three levels of framing. *Wiley Interdisciplinary Reviews: Cognitive Science*, 14(5):e1651.

Daphne Van der Pas. 2014. Making hay while the sun shines: Do parties only respond to media attention when the framing is right? *The International Journal of Press/Politics*, 19(1):42–65.

Teun A Van Dijk. 2015. Critical discourse analysis. *The handbook of discourse analysis*, pages 466–485.

Piek Vossen, Tommaso Caselli, and Roxane Segers. 2021. A narratology-based framework for storyline extraction. *Computational Analysis of Storylines: Making Sense of Events*, 125.

Piek Vossen, Filip Ilievski, Marten Postma, Antske Fokkens, Gosse Minnema, and Levi Remijnse. 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. European Language Resources Association.

Piek Vossen, Filip Ilievski, Marten Postma, and Roxane Segers. 2018. 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. European Language Resources Association (ELRA).

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Dror Walter and Yotam Ophir. 2021. Strategy framing in news coverage and electoral success: An analysis of topic model networks approach. *Political Communication*, 38(6):707–730.

Jian Wang, Xianhui Liu, Junli Wang, and Weidong Zhao. 2014. News topic evolution tracking by incorporating temporal information. In *Natural Language Processing and Chinese Computing: Third CCF Conference, NLPCC 2014, Shenzhen, China, December 5-9, 2014. Proceedings 3*, pages 465–472. Springer.

Xuerui Wang and Andrew McCallum. 2006. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 424–433.

Patrick Xia, Guanghui Qin, Siddharth Vashishtha, Yunmo Chen, Tongfei Chen, Chandler May, Craig Harman, Kyle Rawlins, Aaron Steven White, and Benjamin Van Durme. 2021. LOME: Large ontology multilingual extraction. pages 149–159.

Alexander Ziem, Christian Pentzold, and Claudia Fraas. 2018. Medien-frames als semantische frames: Aspekte ihrer methodischen und analytischen verschränkung am beispiel der ‚snowden-affäre ‘. *Frames interdisziplinär: Modelle, Anwendungsfelder, Methoden*, pages 155–182.

9. Language Resource References

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

Cosmin Bejan and Sanda Harabagiu. 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. Association for Computational Linguistics.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 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.

Agata Cybulska and Piek Vossen. 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. European Language Resources Association (ELRA).

Marianne Djemaa, Marie Candito, Philippe Muller, and Laure Vieu. 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. European Language Resources Association (ELRA).

Kyoko Hirose Ohara, Seiko Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito, and Shun Ishizaki. 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.

Haoruo Peng, Yangqiu Song, and Dan Roth. 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. Association for Computational Linguistics.

Marten Postma, Levi Remijnse, Filip Ilievski, Antske Fokkens, Sam Titarsolej, and Piek Vossen. 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, Marseille, France. European Language Resources Association.

Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 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. Association for Computational Linguistics.

Levi Remijnse, Piek Vossen, Antske Fokkens, and Sam Titarsolej. 2022. Introducing frege to Fillmore: A FrameNet dataset that captures both sense and reference. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 39–50, Marseille, France. European Language Resources Association.

Josef Ruppenhofer, Michael Ellsworth, Miriam R L Petruck, Christopher R Johnson, and Jan Schefczyk. 2010. Framenet ii: Extended theory and practice.

Tiago Timponi Torrent, E Matos, Ludmila Lage, Adrieli Laviola, Tatiane Tavares, VG Almeida, and Natália Sigiliano. 2018. Towards continuity between the lexicon and the constructicon in framenet brasil. *Constructicography: Constructicon development across languages*, 22:107.

TimeFrame: Querying and Visualizing Event Semantic Frames in Time

Davide Lamorte[1], **Marco Rovera**[2], **Alfio Ferrara**[1], **Sara Tonelli**[2]
[1] Università Statale di Milano, Italy
[2] Fondazione Bruno Kessler, Trento, Italy
davide.lamorte@unimi.it, m.rovera@fbk.eu, alfio.ferrara@unimi.it, satonelli@fbk.eu

Abstract

In this work we introduce *TimeFrame*, an online platform to easily query and visualize events and participants extracted from document collections in Italian following a frame-based approach. The system allows users to select one or more events (frames) or event categories and to display their occurrences on a timeline. Different query types, from coarse to fine-grained, are available through the interface, enabling a time-bound analysis of large historical corpora. We present three use cases based on the full archive of news published in 1948 by the newspaper "Corriere della Sera". We show that different crucial events can be explored, providing interesting insights into the narratives around such events, the main participants and their points of view.

Keywords: Text Visualization, Event Extraction, Frame Parsing

1. Introduction

Event-based analysis of corpora has proven to be an effective way to distill information from large amounts of text, supporting tasks such as storytelling (Liu et al., 2020), text simplification (Barlacchi and Tonelli, 2013) and knowledge modelling (Rospocher et al., 2016; Vossen et al., 2016). However, making event information available to users, especially non-expert ones, represents a challenge due to the complex structure of event annotation paradigms, which typically include both event mentions and participants, often structured in some sort of taxonomy. Nevertheless, visualization is crucial to enable users to perform rapid, targeted and customized analysis of large event collections.

We address this limitation by presenting *TimeFrame*, an online tool that allows event-based browsing of large corpora in Italian based on the output of the EventNet-ITA's frame parser (Rovera, 2024). While existing text visualization tools often focus on term-based (Handler et al., 2022) or entity-based (Düring et al., 2021) queries, *TimeFrame* explores the opportunities events and their argument structure offer for event-based text mining and visualization, with particular reference to newspaper textual data. The tool takes in input documents analysed according to the FrameNet paradigm (Fillmore and Baker, 2001), i.e. annotated with semantic frames, each consisting of a trigger (so-called *lexical units – LU*) and a set of semantic roles (*frame elements – FE*). This annotation is further enriched with the temporal information related to the publication of each document, a crucial dimension to navigate historical archives and reconstruct stories.

Through *TimeFrame* it is therefore possible to perform event-based search of a document collection over time. Users can select the granularity of events they are looking for by choosing between the original (i.e. fine-grained) frames or opt for event categories, which have been manually created to group semantically similar events and provide different perspectives on the same event type.

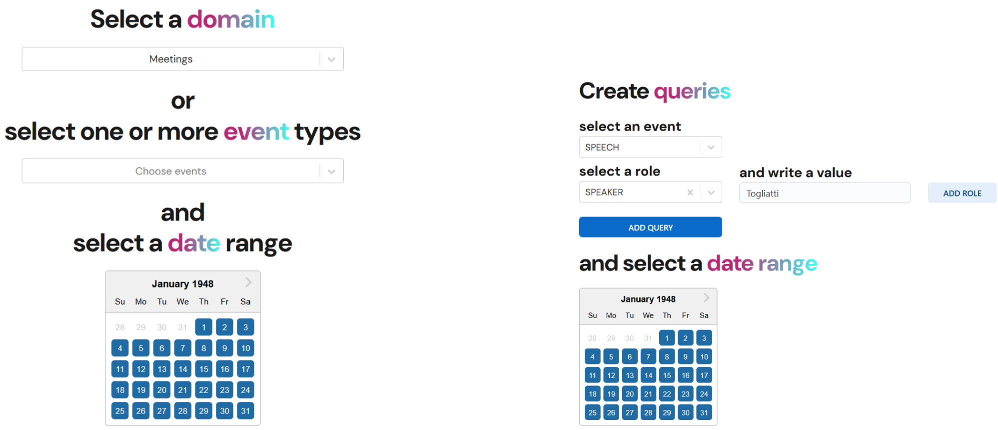
TimeFrame is structured on a back-end component, the frame parser, and on a front-end web application, providing the user interface and interaction. While the frame parser can be applied to any Italian text and the front-end can be used virtually for any frame-based dataset, in this paper we adopt the 1948 edition of the Italian newspaper *Corriere della Sera* as a case study.

2. Related Work

Searching and visualising textual corpora enriched with frame information is a challenging task, since this annotation framework foresees different information layers and a high number of frames and FEs. A notable example is the *Sociofillmore* tool (Minnema et al., 2022), which has been designed to analyse texts in different languages after performing frame-semantic annotation. Its main goal, however, is to study perspectives in written texts and give users the possibility to perform different linguistic analyses. *TimeFrame*, instead, is designed for a more serendipitous exploration of corpora and targets users interested also in the temporal dimension of events such as archivists and historians. Another available tool is *Smell Explorer* (Menini, 2024),[1] which annotates olfactory events inspired by frame semantics, but which is however only limited to one event type. Other search and

[1] https://smell-extractor.tools.eurecom.fr/

First Workshop on Reference, Framing, and Perspective, pages 13–17



(a) Interface for ‘easy’ search

(b) Interface for ‘advanced’ search

Figure 1: Different query options offered by *TimeFrame*.

visualization tools, albeit powerful, focus on specific aspects of investigation. *Clioquery* (Handler et al., 2022), for instance, is focused on term-based queries for historical investigation, while *Impresso* (Düring et al., 2021) is tailored to an entity-based search of archives from the past.

3. TimeFrame Platform

The *TimeFrame* platform is available at this link: `https://eventnetdemo.islab.di.unimi.it/`. Users can access the underlying database, where the frame annotated documents are saved, through different query types. We detail below the pre-processing step to create the corpus database, the query types to access it and the technical details of the platform implementation.

3.1. Data Pre-Processing

In order to use *TimeFrame*, it is necessary to pre-process the document collection of interest using a frame parsing tool. To analyse Italian data, we rely on EventNet-ITA (Rovera, 2024), a frame parser for Italian able to recognize and classify over 200 different types of event-denoting frames, along with their specific FEs. The model has been trained on a large, manually annotated corpus, sampled from the Italian Wikipedia edition. This annotated corpus is designed to cover fine-grained event frames over a bunch of different domains: movements, communications, war and conflicts, economics, geopolitics, arts, biographies, among others. The model takes in input one sentence at a time and parses it in a full-text fashion, extracting as many frames and FEs as there are in the sentence. On the Wikipedia corpus, the model achieves macro F1 = 0.90 for event frame classification and macro F1 = 0.73 on FE classification (at span level). After this pre-processing step, the analysed corpus is stored in a MongoDB database (see Section 3.3) and can be queried through the *TimeFrame* interface.

3.2. Query Types

After accessing the system, a user is displayed the possibility to navigate the processed corpus starting from event categories and frame(s). The ‘easy’ interface to perform different query types is displayed in Fig. 1 (a). An ‘advanced’ search is also available, which can be activated through the interface, that allows users to select also specific FEs and the corresponding fillers (i.e. strings) to be retrieved and displayed (see Fig. 1, b). Each search can be bound to a time period specified by the user. It is also possible to select whether a search should be performed only on document titles, to capture only the major events, or also on the text body.

Four types of queries with different granularities are made available in *TimeFrame*:

1) Query *by domain*: the user can select a domain, corresponding to a set of pre-defined event types in a thematic area, such as Economics, Geopolitics, Crime, Movements, Arts, Communication, Conflicts, among others. This query level provides an entry point for the user approaching the dataset, returning a general, domain-driven view on the data; these 13 categories are currently hardcoded and have been created by adapting the macro-categories that characterise the EventNet-ITA corpus. This query option is the first one displayed in Fig. 1 (a).

2) Query *by single or multiple event types* (or frames): in this case the user can pick one or more event types, in order to explore a more specific phenomenon; the set of event types chosen will define the particular view the user wants to elicit from the dataset. For example, a user could elicit information about declarations related to an election (or an electoral campaign) by using a combination like Election, Electoral_Campaign, Announcement, Statement, Commitment, Speech. This query option is the second one displayed in Fig. 1 (a), available in the 'easy' search.

3) Query by single event, constrained by FE: support for query composition is provided at this stage, where the user can constrain the query, based on a specific event, with one or more frame elements. The system will output all documents containing the target event where the selected frame element is realized. For example, we could ask the system to provide all documents where a Speaker is reported holding a Speech, or where a Designer is reported having designed some Object. This query option is the first one displayed in Fig. 1 (b) and is available in the 'advanced' search, like the following one.

4) Query by single event, constrained by FE and term search: this query model works as a further refinement of the previous one. In addition to role-based constraints, the user can provide a specific term that has to appear as filler of the given role. Continuing with the previous example, we could ask the system to return all Objects designed by 'Pininfarina' (a well-known Italian designer) or all speeches held by 'Togliatti' (leader of the Communist Party in Italy in 1948).

After a query is performed, results are displayed on a timeline, where events are chronologically ordered. Each retrieved document which contains an event matching the query is displayed via its title, and a button allows to access the full text of the article. This visualisation is introduced by a temporal heatmap, providing a synthetic overview of the distribution of the target phenomena over the chosen timespan (see the example in Fig. 2).

3.3. Implementation Details

TimeFrame is developed as a web application that exploits MongoDB as a DBMS, Node JS and Express as the server back-end and React for the front-end. In order to efficiently execute the types of queries illustrated in Section 3.2, the events extracted from EventNet-ITA are stored in the MongoDB database by associating them with the metadata and text available for each analyzed document.

Document metadata and text are stored in the `date`, `id`, `title` and `body` fields, respectively, while the list of events extracted from the document is stored in the `events` array. For each event in particular, in addition to information on its location in the document (`location` field), the record provides the `label` associated with the event and the list of `roles` associated with it, each characterized by a corresponding `label` and `text`.

This data structure makes it possible to formulate the queries supported by *TimeFrame* as simple searches in MongoDB. In particular, the *TimeFrame* interface supports finding documents whose `events` array contains at least one event with a label corresponding to the type of event specified by the user (e.g., `'events': {'$elemMatch': {'label ': 'DEATH'}}`). Starting from this query it is then possible to dynamically add one or more clauses concerning the semantic roles associated with the event, by selecting the documents whose events contain `roles` with labels and text corresponding to the user input.

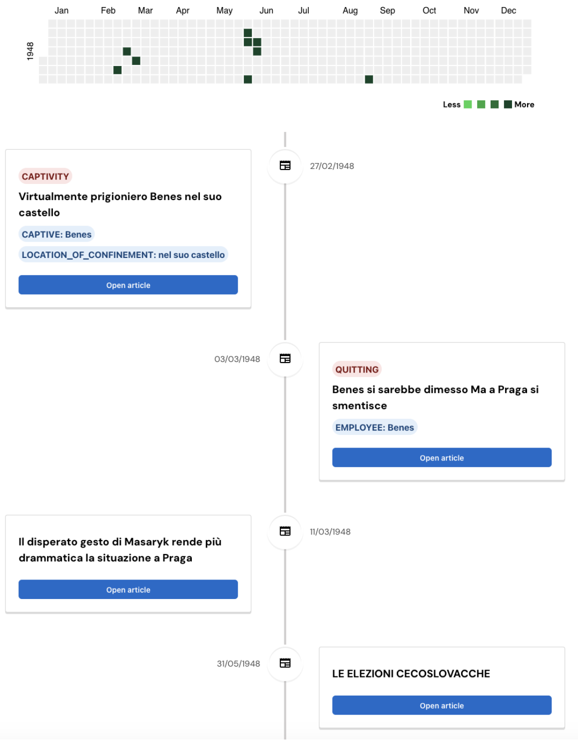


Figure 2: Example of query output after searching for the Quitting of Edvard Beneš, the former President of Czechoslovakia.

4. 1948 Archive of Corriere della Sera

1948 was a crucial year in world history and dense with events that would have far-reaching implications for future history. In Italy, the first republican constitution took effect in January, while in April the first democratic elections, after a fierce electoral campaign, resulted in the victory of the Christian Democratic party. In Europe, on the other hand, postwar political readjustments were in full swing. In March, the Brussels Treaty was signed, setting the stage for the future creation of international organizations like the European Union and NATO. In Eastern Europe, meanwhile, Soviet influence was growing and the region was shaken by coups and regime changes. In the United States, Harry Truman became president after defeating the Republican candidate Thomas Dewey. In India nonviolent leader Gandhi was assassinated in January, while early signs of decolonization were showing up in Burma and Indonesia.

With *TimeFrame* it is possible to perform a time-based exploration of different ways in which events are represented in public discourse. We do so by taking into exam the collection of all news articles published in 1948 by the Italian newspaper *Corriere della Sera*. This collection is currently not publicly available and has been granted for demonstration purposes by the newspaper's archive. The corpus contains 10,418 documents (5,111,000 tokens) which have been processed with EventNet-ITA, producing a database of 146,787 event occurrences (i.e. frame mentions) and 198,370 related FEs.

In public discourse, and particularly in journalistic narrative, different linguistic devices can be used to refer to a particular event, for example by mentioning only a specific aspect, or the outcome of such event. For example, there may be differences between the case where the event is the focus of the discourse (e.g., in the title of an article) or when the event is only mentioned, as part of a larger discourse focusing on a different topic. In some cases this is more or less intentional, while in other cases it is simply the result of the linguistic variety with which language refers to circumstances in the world.

We show below how *TimeFrame* can enable users to explore such referential variety, by retrieving different mentions of the same event. In all three use cases we employ the query type 4 described in Section 3.2. In brackets we provide the number of matches for each query.

4.1. UC 1: Assassination of Gandhi

The first use case concerns the assassination of Mohāndās Karamchand Gāndhī (Gandhi), the 30th January in New Delhi. Events like this (assassinations), can alternatively be presented as agentive (KILLING) or non-agentive (DEATH). We therefore perform two queries:

1. Frame: KILLING | FE: VICTIM, term: *Gandhi* (3 matches)

2. Frame: DEATH | FE: PROTAGONIST, term: *Gandhi* (4 matches)

By comparing the results, we observe that the DEATH frame is preferably used to factually describe the killing and the following sub-events (for instance how the British government reacted), while KILLING is mainly used for comments and more emotive reactions (for instance, he is described as 'glorious and heroic victim of world peace').

4.2. UC 2: U.S. Presidential Elections

Elections are often referred to from multiple perspectives, notably from the point of view of the winner, or of the loser, or, still, of the candidate appointed to the position. US presidential elections in 1948, covered by our newspaper corpus, provide a suitable example of how *TimeFrame* is able to capture such subtle variations in perspective:

1. Frame: ELECTION | [FE: ROLE, term: *presidenziali*] [FE: PLACE, term: *americane*] (4 matches)

2. Frame: APPOINTING_ELECTION | FE: APPOINTEE, term: *Truman* (6 matches)

3. Frame: WIN_ELECTION | FE: WINNER, term: *Truman* (5 matches)

4. Frame: LOSE_ELECTION | FE: LOSER, term: *Dewey* (2 matches)

As expected, much more space is given to discussing the elections from the point of view of the winner, while only two articles mention Dewey (one of which was written months before the election and presented the event as hypothetical).

4.3. UC 3: Czechoslovak Coup d'état

In many cases, events narrated in the press, or even in historical discourse, do not refer to a single occurrence, but to a chain of facts. In such cases, while events are still unfolding and therefore are not clear, journalistic practice resorts to general utterances, like "the recent events in Paris" or "the current happenings in Asia". This is the case, for example, with the "Czechoslovak coup d'état", a sequence of events that led to a regime change in the country on February 25.

1. Frame: COUP | FE: PLACE, term: *Cecoslovacchia*, *Praga* (5 matches)

2. Frame: CHANGE_OF_LEADERSHIP | FE: PLACE, term: *Cecoslovacchia*, *Praga* (2 matches)

3. Frame: QUITTING | FE: EMPLOYEE, term: *Benes* (9 matches)

4. Frame: EVENT | FE: PLACE, term: *Cecoslovacchia* (9 matches)

The output of the third Query is reported in Figure 2, and visually displays how the situation evolves over time. If we check the timeline, we observe that the four frames appear in sequence in the archive.

5. Conclusions

In this work we introduced the *TimeFrame* platform, which allows users to query corpora that have been analysed with a frame semantic parser. The tool makes it possible to perform both coarse-grained queries, at the level of event categories, and fine-grained ones, up to role fillers. Based on three use cases from the 1948 archive of Corriere della Sera newspaper, we show that the platform can be easily used to track specific events over time and capture different sub-events and points of view. The fact that each query type is bound to the temporal dimension makes this platform particularly valuable to users interested in historical analysis such as archivists and history scholars.

6. Acknowledgements

We thank the Archive of Corriere della Sera for making their documents available for our analysis.

7. Bibliographical References

Gianni Barlacchi and Sara Tonelli. 2013. Ernesta: A sentence simplification tool for children's stories in italian. In *Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part II 14*, pages 476–487. Springer.

Marten Düring, Roman Kalyakin, Estelle Bunout, and Daniele Guido. 2021. Impresso inspect and compare. visual comparison of semantically enriched historical newspaper articles. *Information*, 12(9):348.

Charles J. Fillmore and Collin F. Baker. 2001. Frame semantics for text understanding. In *In Proceedings of WordNet and Other Lexical Resources Workshop*.

Abram Handler, Narges Mahyar, and Brendan O'Connor. 2022. Clioquery: Interactive query-oriented text analytics for comprehensive investigation of historical news archives. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 12(3):1–49.

Bang Liu, Fred X Han, Di Niu, Linglong Kong, Kunfeng Lai, and Yu Xu. 2020. Story forest: Extracting events and telling stories from breaking news. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(3):1–28.

Stefano Menini. 2024. Semantic frame extraction in multilingual olfactory events. In *The 2024 joint international conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*.

Gosse Minnema, Sara Gemelli, Chiara Zanchi, Tommaso Caselli, and Malvina Nissim. 2022. Sociofillmore: A tool for discovering perspectives. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 240–250.

Teresa Paccosi, Stefano Menini, Elisa Leonardelli, Ilaria Barzon, and Sara Tonelli. 2023. Scent and sensibility: Perception shifts in the olfactory domain. In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, pages 143–152.

Marco Rospocher, Marieke Van Erp, Piek Vossen, Antske Fokkens, Itziar Aldabe, German Rigau, Aitor Soroa, Thomas Ploeger, and Tessel Bogaard. 2016. Building event-centric knowledge graphs from news. *Journal of Web Semantics*, 37:132–151.

Marco Rovera. 2024. EventNet-ITA: Italian frame parsing for events. In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, pages 77–90, St. Julians, Malta. Association for Computational Linguistics.

Nicole Sultanum, Anastasia Bezerianos, and Fanny Chevalier. 2021. Text visualization and close reading for journalism with storifier. In *2021 IEEE Visualization Conference (VIS)*, pages 186–190. IEEE.

Piek Vossen, Rodrigo Agerri, Itziar Aldabe, Agata Cybulska, Marieke van Erp, Antske Fokkens, Egoitz Laparra, Anne-Lyse Minard, Alessio Palmero Aprosio, German Rigau, et al. 2016. Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Knowledge-Based Systems*, 110:60–85.

Yunchao Wang, Guodao Sun, Zihao Zhu, Tong Li, Ling Chen, and Ronghua Liang. 2024. E2 storyline: Visualizing the relationship with triplet entities and event discovery. *ACM Transactions on Intelligent Systems and Technology*, 15(1):1–26.

Comparing News Framing of Migration Crises using Zero-Shot Classification

Nikola Ivačič[1], **Matthew Purver**[1,2], **Fabienne Lind**[3],
Hajo Boomgaarden[3], **Veronika Bajt**[4], **Senja Pollak**[1]
[1]Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
[2]School of Electronic Engineering and Computer Science
Queen Mary University of London,
London E1 4NS, UK
[3]University of Vienna, Department of Communication,
Computational Communication Science Lab
Kolingasse 14-16, 1090 Vienna, Austria
[4]Peace Institute
Metelkova 6, 1000 Ljubljana, Slovenia.
{nikola.ivacic, matthew.purver, senja.pollak}@ijs.si
{fabienne.lind, hajo.boomgaarden}@univie.ac.at
veronika.bajt@mirovni-institut.si

Abstract

We present an experiment on classifying news frames in a language unseen by the learner, using zero-shot cross-lingual transfer learning. We used two pre-trained multilingual Transformer Encoder neural network models and tested with four specific news frames, investigating two approaches to the resulting multi-label task: Binary Relevance (treating each frame independently) and Label Power-set (predicting each possible combination of frames). We train our classifiers on an available annotated multilingual migration news dataset and test on an unseen Slovene language migration news corpus, first evaluating performance and then using the classifiers to analyse how media framed the news during the periods of Syria and Ukraine conflict-related migrations.

Keywords: Transfer learning, Zero-shot classification, News Framing, Migrations

1. Introduction

News articles can portray topics by means of different styles of presentations and by emphasising different facets of the topic. News framing describes the selection of particular aspects of topics, people or events and rendering them salient to promote a particular interpretation, evaluation, and/or solution (Entman, 1993, 2003; de Vreese, 2005). Social scientists have long sought to computationally measure these frames, with researchers comparing various machine learning methods (Burscher et al., 2014; Eisele et al., 2023; Lind et al., 2021), despite the varying and informal definitions of framing. Although computational methods for detecting framing have been extensively explored (Ali and Hassan, 2022), they have recently gained significant attention in NLP research (Piskorski et al., 2023; Eisele et al., 2023), indicating a notable advancement in zero-shot computational framing research (Wu et al., 2023; Reiter-Haas et al., 2023).

We aimed to analyse and compare the framing of news in Slovenia during two distinct European migration waves, one triggered by the war in Syria and the other by the war in Ukraine. Both events saw a considerable rise of migrants entering the European Union (Kogovšek Šalamon and Bajt, 2016; Niemann and Zaun, 2023); yet the migrant groups differed markedly in terms of cultural and ethnic background. Both the context causing migration as well as cultural factors may play into how news media frame the migration issue during these two episodes (for an overview of the literature on media and migration, see Eberl et al., 2018). In the Slovenian context, Bučar Ručman, 2022 discusses how migrants from different migration waves were treated differently by authorities, the local population and the media. Therefore, we generally hypothesise that the way news was framed in Slovenia varied between these periods, which will also be reflected in a quantitative computational study. A related question was addressed by Caporusso et al., 2024, who investigated how the dehumanisation aspects of migrant dehumanisation changed in Slovenian newspapers during the Ukrainian and Syrian periods.

We used a manually annotated news corpus (Lind et al., 2020) developed for the REMINDER project[1] to train the multilingual frame classifiers. This corpus consisted of migration news articles in seven languages - yet not our target language, Slovene - and was manually annotated with four issue-specific frames. These frames have been

[1]https://www.reminder-project.eu/

First Workshop on Reference, Framing, and Perspective, pages 18–27

frequently studied in European news coverage of migration (Eberl et al., 2018; Chouliaraki and Zaborowski, 2017).

For the classification, we chose two multilingual pre-trained Transformer Encoder (Devlin et al., 2018; Conneau et al., 2019) models and fine-tuned them on the migration corpus for multi-label classification. While recent methods employing a contrastive learning approach exist (Reiter-Haas et al., 2023; Liao et al., 2023), we utilised classical techniques for our study. We used two transformation methods to tackle multi-label classification problems: Binary Relevance and Label Power-set (Ganda and Buch, 2018; see Section 3).

The Zero-shot technique, first used in classification tasks with a target to predict new unseen classes (Chang et al., 2008; Larochelle et al., 2008; Palatucci et al., 2009), has been applied in many NLP tasks and settings, including cross-lingual model transfer in which task-specific annotations in one language are used to fine-tune the model for evaluation in another language (Pires et al., 2019). Zero-shot cross-lingual model transfer has been demonstrated from Slovene to Croatian language on other tasks, e.g. offensive language detection (Pelicon et al., 2020) and for genre identification in Slovene texts (Kuzman et al., 2023).

We aimed to analyse Slovene news, but as no annotated training set exists, we tested whether zero-shot transfer would work. Consequently, we created a Slovene news corpus on migration for both periods, applied fine-tuned models to predict news framing, and analysed the results.

In summary, the contributions of this paper are two-fold: a) the development and testing of a multilingual news frames classifier for migration texts and b) the comparative analysis of Slovene news from two different migration-related periods.

2. Data Description

The following section presents the two corpora used in our experiments: The manually annotated REMINDER migration corpus and our Slovenia migration news corpus.

2.1. The REMINDER migration corpus

The manually annotated REMINDER corpus is a randomly selected sample of migration-related news articles published between January 2000 and December 2017. It contains 6,475 news articles from seven countries: Germany, Hungary, Poland, Romania, Spain, Sweden and the UK, with 925 samples per country. Each news article is marked with four labels showing whether an article contains aspects related to a specific migration-related frame. The labels were created by seven native speakers who coded the articles in their original language. These coders underwent joint training to ensure a shared comprehension of the four frame concepts. Intercoder reliability was evaluated. Those labels are coded as one if an article references the respective frame or zero if it does not. The labels in question are as follows:

- Economy: Does the article refer to economy/budget-related aspects of migration?
- Labour market: Does the article refer to labour market-related aspects of migration?
- Welfare: Does the article refer to welfare-related aspects of migration?
- Security: Does the article refer to security-related aspects of migrants/migration?

We filtered the corpus, removing double-occurring symbol characters that could negatively impact subword tokenisation and have no information value.

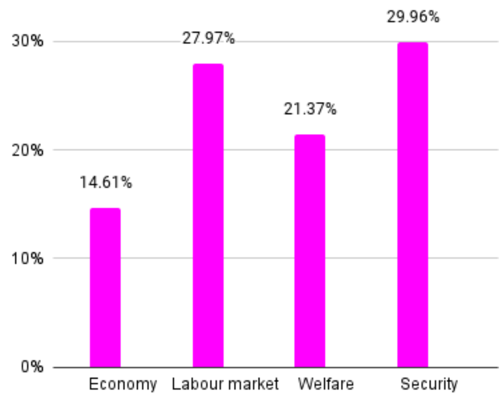


Figure 1: A bar for each label illustrates the Percentage of samples labelled with each frame.

In our examination of a corpus, we discovered two significant imbalances. Firstly, the distribution of individual labels within the corpus is uneven; for example, the label Economy is noticeably less frequent than other labels (see Figure 1). Secondly, there is a significant variation in the number of labels assigned to each sample, with multiple frame combinations less frequent than 0 or single frame labels, indicating an imbalance in label distribution per sample (see Figure 2).

Adding to these challenges is the issue that the most common label set is the empty set; this introduces a considerable bias towards unlabelled samples. Upon examining the label distribution, considering only single-label sets (which represent the largest group by label count), it is clear that the Security label appears far more frequently than any

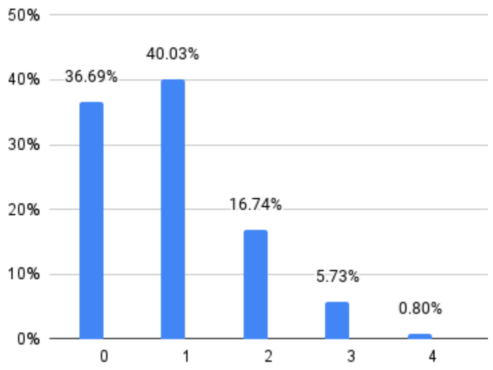


Figure 2: Distribution of the number of labels per sample, across all samples

other in this context (see Figure 3). In contrast, the Economy label set is notably smaller, highlighting a significant disparity in label representation.

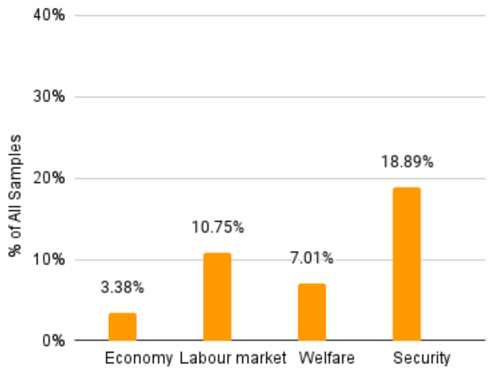


Figure 3: The distribution of single label sets across all samples

2.2. The Slovene migration corpus

For the transfer learning task, we collected the Slovene news corpus from 29 online news media outlets (see Appendix D). In this study, the selected media outlets encompassed major players and representative local media, ensuring a comprehensive analysis of the media landscape. We used a set of Slovene word prefixes frequently used in migration-related articles (shown in Table 1) and two distinct periods for the Syria and Ukraine migration crisis: August 2015 until April 2016 and February 2022 until March 2023.

These periods were selected to reflect the timeframe of increased migration from conflict areas to Europe and Slovenia. On the one hand, in the summer of 2015, the Balkans migration route from

Search Prefixes	English Translation
begunec, begunc, begunk, beguns	refugee
migracij, migrant, imigra	migration, migrant, emigrant
prebežni, pribežni	migration, migrant
azil	asylum

Table 1: Search prefixes used for Slovene corpus construction.

the Middle East to Turkey, Greece, Macedonia, Serbia, and Hungary also turned through Croatia and Slovenia (after Hungary closed its borders). According to the official Slovenian Police statistics, almost 400,000 migrants entered Slovenia between September 2015 and January 2016, most just passing through. On the other hand, following Russia's invasion of Ukraine on 24 February 2022, the European Union activated a Temporary Protection Directive that has been in effect in Slovenia since March 2022. More than 8,000 Ukrainian citizens have since received temporary protection in Slovenia. Both events resulted in pronounced media reporting about migration.

These search criteria yielded almost equal dataset sizes for the 2015/16 and the 2022/23 periods: 8617 and 8586, respectively.

Next, we manually annotated a small sample of 100 articles to act as an evaluation set for classification accuracy. We used our classifier to predict label values on the Slovene corpus and randomly selected equal numbers of positive and negative values for each of the four labels. We took 50 articles from both periods, resulting in 100 articles; these were then manually annotated to obtain a Slovene classification test set. Manual labelling was carried out by a single annotator following the coding instructions for the REMINDER corpus project.

3. Methodology

We employed BERT Multilingual Cased (Devlin et al., 2018) (BERTmc) and XLM-Roberta-base (Conneau et al., 2019) (XLMRb) pre-retrained Transformer models from HuggingFace (Wolf et al., 2020). We fine-tuned the models on the REMINDER corpus using two distinct combinations of news article fields: including the body with the title (T+B) and excluding the title (B).

Migration-related media frames were modelled as a multi-label classification problem, as multiple or zero frames may occur in the same news article. The small label count enabled the use of HuggingFace's built-in classification capabilities without the need for a custom neural network classification head to address the multi-label problem. This was achieved through the implementation of two problem transformation methods:

- Binary Relevance (BinRel), where we independently fine-tuned one transformer model per label.
- Label Power-set (LPSet), where we fine-tuned each label combination as a separate class.

We conducted a 10-fold cross-validation on the REMINDER corpus for six combinations involving two pre-trained models, two fields, and two transformation methods. Then, we compared the outcomes with those of the majority label-set and random classifiers.

The concluding phase involved classifying the Slovene migration corpus, choosing and manually annotating a small set of 100 articles, and then examining the outcomes.

In this work, we did not perform hyper-parameter optimisation; all the models are fine-tuned using the default set of hyper-parameters in the Transformers library, optimised for a large selection of common NLP tasks. More precisely, we used:

- AdamW optimiser with a learning rate of $2e-5$.
- Weight decay set to 0.01 for regularisation.
- Training for a maximum of 20 epochs.
- Batch size of 24.
- Maximum length of 512 sub-word tokens.
- Best model selection based on the validation set micro F1-score.

4. Evaluation

Here, we explain the measures used to assess our models, followed by an analysis of the fine-tuning results on the REMINDER corpus. Finally, we evaluate the Slovene language zero-shot classification, examining the effectiveness of our approach across different scenarios.

4.1. Evaluation Metrics

Following the work of Tsoumakas et al., 2010; Madjarov et al., 2012, we employed two categories of metrics:

- Example-based metrics, namely Hamming Loss and Accuracy, to assess the differences between the actual and predicted label sets across all samples.
- Label-based metrics, including Precision, Recall, and Macro-F1, to examine performance averaged across all labels.

We have selected the macro averaged Label-based metrics treating all labels of equal importance to have a better understanding of the model's performance on each label individually (see Appendices for micro-averaged results and formulas).

For a baseline comparison, we selected three straightforward classifiers: the $Majority\ \emptyset$ classifier, which assigns no labels to all samples; the $Majority\ L_1$, which labels all samples with the most common single label (Security); and the $Random$ classifier, which assigns labels based on the overall distribution of label-sets.

4.2. Fine-Tuning Results

This section will present the classification results of fine-tuning Transformer Encoder models across multiple languages on the REMINDER migration corpus.

The results in Table 2 and Table 3 show the manually annotated corpus 10-fold cross-validated classifier performance. Examining the results, we can see that the XLM-Roberta pre-trained models and the Binary Relevance problem transformation method outperform BERT and Label Power-set in both metrics categories. For the best model, we can see that almost 60% of example label sets are exactly matched while 13% of example-label pair is misclassified (see Appendix A for details).

Model	Method	Field	Accuracy	Hamming Loss
XLMRb	BinRel	B	$\mathbf{0.587} \pm 0.026$	$\mathbf{0.131} \pm 0.011$
XLMRb	BinRel	T+B	0.575 ± 0.028	0.133 ± 0.011
XLMRb	LPSet	B	0.571 ± 0.023	0.137 ± 0.009
XLMRb	LPSet	T+B	0.572 ± 0.016	0.138 ± 0.009
BERTmc	BinRel	B	0.557 ± 0.021	0.143 ± 0.008
BERTmc	BinRel	T+B	0.548 ± 0.025	0.144 ± 0.009
BERTmc	LPSet	B	0.531 ± 0.020	0.155 ± 0.009
BERTmc	LPSet	T+B	0.524 ± 0.021	0.159 ± 0.008
		Baseline models		
Majority $\emptyset$			0.367	0.235
Majority L_1			0.189	0.335
Random			0.192	0.355

Table 2: Example-based classifier performance - shows the pre-trained model used for fine-tuning, followed by a problem transformation method, selected article fields for training, classification accuracy (higher is better), and Hamming loss (lower is better). The results are compared to baseline models (the no-label, the most common label, and the random classifier).

It is evident that the choice of training fields from news articles has a minimal effect on classifier performance across both metric categories. Adding a title to the body negatively affects performance, suggesting that titles alone offer limited informational value.

Model	Method	Fields	Macro F1	Precision	Recall
XLMRb	BinRel	B	$\mathbf{0.709} \pm 0.019$	$\mathbf{0.714} \pm 0.027$	0.707 ± 0.026
XLMRb	BinRel	T+B	0.705 ± 0.016	0.705 ± 0.030	$\mathbf{0.709} \pm 0.026$
XLMRb	LPSet	B	0.686 ± 0.023	0.706 ± 0.030	0.670 ± 0.030
XLMRb	LPSet	T+B	0.683 ± 0.017	0.698 ± 0.029	0.672 ± 0.025
BERTmc	BinRel	B	0.674 ± 0.017	0.687 ± 0.020	0.663 ± 0.025
BERTmc	BinRel	T+B	0.675 ± 0.016	0.686 ± 0.021	0.669 ± 0.022
BERTmc	LPSet	B	0.649 ± 0.017	0.657 ± 0.028	0.648 ± 0.032
BERTmc	LPSet	T+B	0.642 ± 0.015	0.648 ± 0.028	0.643 ± 0.029
			Baseline models		
Majority L_1			0.115	0.075	0.250
Random			0.231	0.230	0.232

Table 3: Label-based classifier performance - shows the pre-trained model used for fine-tuning, followed by a problem transformation method, selected article fields for training, macro F1 score, macro precision and macro recall (higher is better for all three values). The results are compared to the baseline models (the most common label and the random classifier).

All models that underwent fine-tuning significantly surpassed the performance of the baseline models.[2] Their performance is also consistent regarding micro-averaged scores (see Appendix B).

4.3. Zero-Shot Results

Next, we needed to assess how well the classifier performed on the unseen Slovene language. We tested classifier performance on the 100 manually annotated Slovenian corpus articles. Although the model's performance fell short of expectations, it still surpassed the baseline on both categories of evaluation, indicating some level of effectiveness.

Model	Method	Fields	Accuracy	Hamming Loss
XLMRb	BinRel	B	0.340	0.255
XLMRb	BinRel	T+B	$\mathbf{0.370}$	0.253
XLMRb	LPSet	B	0.340	$\mathbf{0.250}$
XLMRb	LPSet	T+B	0.330	0.255
		Baseline models		
Majority $\emptyset$			0.210	0.308
Majority L_1			0.270	0.273
Random			0.160	0.353

Table 4: Example-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, classification accuracy (higher is better), and Hamming loss (lower is better). The results are compared to baseline models (the no-label, the most common label, and the random classifier).

It is noticeable that the poor recall values of the classifier impact the overall performance of the label-based metrics.

Model	Method	Fields	Macro F1	Macro P	Macro R
XLMRb	BinRel	B	0.422	0.605	0.341
XLMRb	BinRel	T+B	0.445	0.616	0.367
XLMRb	LPSet	B	$\mathbf{0.466}$	$\mathbf{0.640}$	$\mathbf{0.412}$
XLMRb	LPSet	T+B	0.461	0.595	0.403
			Baseline models		
Majority L_1			0.182	0.143	0.250
Random			0.267	0.275	0.265

Table 5: Label-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, macro F1 score, macro precision and macro recall (higher is better for all three values). The results are compared to the baseline models (the most common label and the random classifier).

4.4. Zero-Shot Predictions

Lastly, we proceeded to run predictions on the entire Slovene corpus with the best models from fine-tuning and zero-shot evaluation for both periods and examined distributions of the individual labels.

All predictions show consistent results regarding prevailing frame labels for each period regardless of model selection (see Table 6). We wanted to assess if the difference in predictions of our models for the two periods is significant. The difference between the predicted distributions for the two periods is statistically significant, with a χ^2 test showing a p-value nearly zero, well below the standard threshold 0.05 for all three selected models.

Model	Method	Fields	Period	Economy	Labour M.	Welfare	Security
			Best Fine-Tuning model predictions				
XLMRb	BinRel	B	Syria	796	590	973	2023
XLMRb	BinRel	B	Ukraine	517	895	1218	1661
			Best Zero-Shot model predictions				
XLMRb	BinRel	T+B	Syria	890	753	1230	2022
XLMRb	BinRel	T+B	Ukraine	600	1182	1434	1697
XLMRb	LPSet	B	Syria	1161	870	950	1908
XLMRb	LPSet	B	Ukraine	718	1272	1183	1797

Table 6: Best model predictions - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, number of predicted labels for Economy, Labour Market, Welfare and Security.

Figure 4 shows that the security frame is by far the most prevalent in both corpora, corroborating the existing research about migration, in general, becoming regarded as a security risk (Palidda, 2011; Bajt, 2019; Pajnik and Ribać, 2021). When comparing the distributions across the two subcorpora, it also shows that security has been a more

[2]The baseline classifier's performance excludes the $Majority\ \emptyset$ classifier for label-based metrics, as it does not generate any true positives.

exposed frame in media coverage of Middle Eastern migration to Europe than that of Ukrainians (see also sociological research by Bučar Ručman, 2022). Moreover, Muslims are stereotypically portrayed as dangerous as the idea of violent Muslims corresponds to racialised views of young Middle Eastern men as terrorists (Kundnani, 2015). Syrian refugees have thus been treated as a threat also in Slovenian media (Thiele et al., 2023).

On the other hand, the labour market and welfare aspects were more prevalent during the Ukraine migration wave. This pattern may be in line with what could be expected, given that the cultural and ethnic background of migrants from the Middle East tends to reproduce discourses that refer to alleged security issues posed by such migration (Sambaraju and Shrikant, 2023). By contrast, it is possible that the vast and very quick increase of migrants from Ukraine has posed challenges in terms of welfare provision for these migrants in particular. The prevalence of the welfare frame and the absence of the security frame also mirrors the findings for the UK press in their reporting on Ukrainian refugees (Nataliya Roman, 2020).

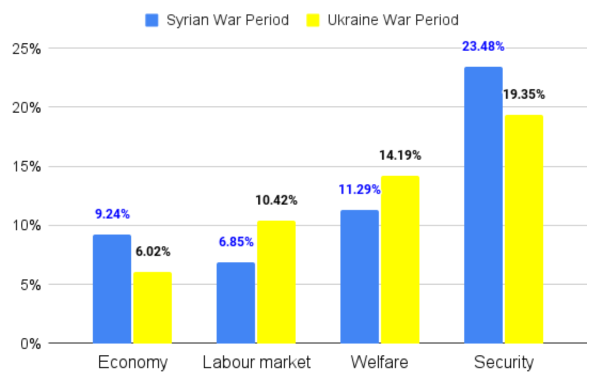


Figure 4: Final Zero-Shot prediction results on a complete Slovene migration news corpus obtained with the best model from the fine-tuning results.

5. Conclusion

We tested several approaches for news frame classification on the REMINDER multilingual migration news corpus. We discovered that employing a binary relevance problem transformation approach combined with the XLM-Roberta-base pre-trained model yields the most effective results. The model was then evaluated on a small Slovene sample, where zero-shot performance is around 0.37 regarding classification accuracy.

Although our model performance is clearly suboptimal, and individual labels and absolute percentages will be errorful, we can still draw some conclusions from relative distributions and comparisons across settings. Given this, we take our results as tentative support for the hypothesis that the portrayal of migration by the news media varied across the two periods in its focus on the economy, labour market, welfare, and security. Moreover, our analysis suggests that economic and security issues were more prominent in media reports on migration during the Middle East conflict than during the Ukraine war. Likewise, it is apparent that labour market and welfare concerns received more emphasis in discussions of migration during the period of the Ukraine war.

Overall, it is also interesting to see that security framing, in line with the REMINDER results, remains the most prominent news frame detected in the Slovenian corpus, independent of context and type of migration (Eberl and Galyga, 2021).

In our pursuit of enhancing the classification model, which is crucial for a more reliable interpretation of results, future efforts will focus on several key areas of improvement. Firstly, we plan to pre-train the models further using the Slovene migration corpus, specifically targeting the masked language modelling task. This approach aims to deepen the models' understanding of context and nuances regarding migration within the Slovene language. Secondly, to mitigate the effects of possible truncation, which can lead to the loss of vital information in longer texts, we intend to explore the use of models designed for handling extended sequences, such as ToBERT (Pappagari et al., 2019), Longformer (Beltagy et al., 2020), Big Bird (Zaheer et al., 2021). Lastly, we want to incorporate contrastive learning techniques tailored for Few-Shot scenarios (Reiter-Haas et al., 2023; Liao et al., 2023). This innovative approach could enhance the model's ability to learn from a limited number of examples, thereby improving its performance in classifying new, unseen data with minimal additional input. Additionally, we plan to investigate the use of generative model approaches, not only to improve classification accuracy potentially but also to enrich the training corpus.

6. Acknowledgements

This work was supported by the Slovenian Research Agency grants via the core research programmes Knowledge Technologies (P2-0103) and the projects Computer-assisted multilingual news discourse analysis with contextual embeddings (J6-2581), Embeddings-based techniques for Media Monitoring Applications (L2-50070) and Hate Speech in Contemporary Conceptualizations of Nationalism, Racism, Gender and Migration (J5-3102).

7. Limitations

The classification performance is above the baselines but still far from optimal. Consequently, any analysis and resultant conclusions regarding framing within the Slovene corpus must be approached cautiously and can only be considered tentative support for the hypotheses.

8. References

8.1. Bibliographical References

Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Veronika Bajt. 2019. The schengen border and the criminalization of migration in slovenia. *Südosteuropa*, 67(3):304–327.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer.

Björn Burscher, Daan Odijk, Rens Vliegenthart, Maarten de Rijke, and Claes H. de Vreese. 2014. Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8(3):190–206.

Aleš Bučar Ručman. 2022. Slovenija in begunske migracije: Primerjava odnosa do beguncev iz Bosne in Hercegovine, Sirije in Ukrajine [Slovenia and refugee migration: Comparison of attitudes towards refugees from bosnia and herzegovina, syria and ukraine]. In *8. Nacionalna konferenca o varnosti v lokalnih skupnostih: Cilji trajnostnega razvoja in varnost v lokalnih skupnostih*. Univerza v Mariboru, Univerzitetna založba.

Jaya Caporusso, Damar Hoogland, Mojca Berglez, Boshko Koloski, Matthew Purver, and Pollak Senja. 2024. A computational analysis of the dehumanisation of migrants from syria and ukraine in slovene news media. In *Proceedings of LREC-Coling 2024*. Forthcoming.

Ming-Wei Chang, Lev Ratinov, Dan Roth, and Vivek Srikumar. 2008. Importance of semantic representation: Dataless classification. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, page 830–835. AAAI Press.

Lilie Chouliaraki and Rafal Zaborowski. 2017. Voice and community in the 2015 refugee crisis: A content analysis of news coverage in eight european countries. *International Communication Gazette*, 79(6-7):613–635.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Claes de Vreese. 2005. News framing: Theory and typology. *Information Design Journal*, 13:51–62.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Jakob-Moritz Eberl and Sebastian Galyga. 2021. Mapping media coverage of migration within and into europe. In *Media and Public Attitudes Toward Migration in Europe: A Comparative Approach*, pages 105–122. Routledge.

Jakob-Moritz Eberl, Christine Meltzer, Tobias Heidenreich, Beatriz Herrero, Nora Theorin, Fabienne Lind, Rosa Berganza, Hajo Boomgaarden, Christian Schemer, and Jesper Strömbäck. 2018. The european media discourse on immigration and its effects: a literature review. *Annals of the International Communication Association*, 42.

Olga Eisele, Tobias Heidenreich, Olga Litvyak, and Hajo Boomgaarden. 2023. Capturing a news frame – comparing machine-learning approaches to frame analysis with different degrees of supervision. *Communication Methods and Measures*, 17:1–23.

Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

Robert M. Entman. 2003. Cascading activation: Contesting the white house's frame after 9/11. *Political Communication*, 20(4):415–432.

Dhatri Ganda and Rachana Buch. 2018. A survey on multi label classification. *Recent Trends in Programming Languages*, 5(1):19–23.

Neža Kogovšek Šalamon and Veronika Bajt, editors. 2016. *Razor-Wired: Reflections on Migration Movements through Slovenia in 2015*. Peace Institute, Ljubljana.

Arun Kundnani. 2015. *The Muslims are Coming: Islamophobia, Extremism, and the Domestic War on Terror*. Verso.

Taja Kuzman, Igor Mozetič, and Nikola Ljubešić. 2023. Chatgpt: Beginning of an end of manual linguistic data annotation? use case of automatic genre identification.

Hugo Larochelle, Dumitru Erhan, and Yoshua Bengio. 2008. Zero-data learning of new tasks. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, page 646–651. AAAI Press.

Qisheng Liao, Meiting Lai, and Preslav Nakov. 2023. MarsEclipse at SemEval-2023 task 3: Multi-lingual and multi-label framing detection with contrastive learning. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 83–87, Toronto, Canada. Association for Computational Linguistics.

Fabienne Lind, Tobias Heidenreich, Christoph Kralj, and Hajo Boomgaarden. 2021. Greasing the wheels for comparative communication research: Supervised text classification for multilingual corpora. *Computational Communication Research*, 3(3):1–30. Publisher Copyright: © Fabienne Lind, Tobias Heidenreich, Christoph Kralj, Hajo G. Boomgaarden.

Gjorgji Madjarov, Dragi Kocev, Dejan Gjorgjevikj, and Sašo Džeroski. 2012. An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, 45(9):3084–3104. Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).

Stephynie C. Perkins Nataliya Roman, Anna Young. 2020. Displaced and invisible: Ukrainian refugee crisis coverage in the us, uk, ukrainian, and russian newspapers. *Negotiation and Conflict Management Research*, 14.

Arne Niemann and Natascha Zaun. 2023. EU external migration policy and EU migration governance: Introduction. *Journal of Ethnic and Migration Studies*, 49(12):2965–2985.

Mojca Pajnik and Marko Ribać. 2021. Medijski populizem in afektivno novinarstvo. *Javnost*, 28(sup.1):S103–S121.

Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. 2009. Zero-shot learning with semantic output codes. In *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc.

Salvatore Palidda, editor. 2011. *A Review of the Principal European Countries. Racial Criminalization of Migrants in 21st Century*. Farnham.

Raghavendra Pappagari, Piotr Żelasko, Jesús Villalba, Yishay Carmiel, and Najim Dehak. 2019. Hierarchical transformers for long document classification.

Andraž Pelicon, Marko Pranjić, Dragana Miljković, Blaž Škrlj, and Senja Pollak. 2020. Zero-shot learning for cross-lingual news sentiment classification. *Applied Sciences*, 10(17).

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert?

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

Markus Reiter-Haas, Alexander Ertl, Kevin Innerhofer, and Elisabeth Lex. 2023. mCPT at SemEval-2023 task 3: Multilingual label-aware contrastive pre-training of transformers for few- and zero-shot framing detection. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 941–949, Toronto, Canada. Association for Computational Linguistics.

Rahul Sambaraju and Natasha Shrikant. 2023. 'All of a sudden for no reason they've been displaced': Constructing the 'contingent refugee' in early media reports on the ukrainian refugees. *British Journal of Social Psychology*, 62(4):1715–1732.

Daniel Thiele, Mojca Pajnik, Birgit Sauer, and Iztok Šori. 2023. Borderless fear? How right-wing populism aligns in affectively framing migration as a security threat in Austria and Slovenia. *Journal of Language and Politics*.

Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. 2010. *Mining Multi-label Data*, pages 667–685. Springer US, Boston, MA.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in*

Natural Language Processing: System Demonstrations, pages 38–45, Online. Association for Computational Linguistics.

Ben Wu, Olesya Razuvayevskaya, Freddy Heppell, João A. Leite, Carolina Scarton, Kalina Bontcheva, and Xingyi Song. 2023. SheffieldVeraAI at SemEval-2023 task 3: Mono and multilingual approaches for news genre, topic and persuasion technique classification. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1995–2008, Toronto, Canada. Association for Computational Linguistics.

Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, and Amr Ahmed. 2021. Big bird: Transformers for longer sequences.

8.2. Language Resource References

Lind, Fabienne and Heidenreich, Tobias and Eberl, Jakob-Moritz and Galyga, Sebastian and Edie, Rachel and Herrero-Jiménez, Beatriz and Gómez Montero, Eva Luisa and Berganza, Rosa and Boomgaarden, Hajo G. 2020. *REMINDER: Historical Media Analysis on Migration 2003-2017 (OA edition)*. AUSSDA. PID https://doi.org/10.11587/IEGQ1B.

9. Appendices

A. Formulas

In this section, we use N for the number of samples, L for the number of labels, y_i for an individual example label set, and $\hat{y}_i$ for the predicted example label set.

For the example-based evaluation, we used

- Subset accuracy:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} I(y_i = \hat{y}_i)$$

where $I(true) = 1$ and $I(false) = 0$

- Hamming loss:

$$Hamming\text{-}Loss = \frac{1}{N \cdot L}\sum_{i=1}^{N}\sum_{j=1}^{L} xor(y_i, \hat{y}_i)$$

For the label-based macro-averaged evaluation, we used:

$$Precision_{macro} = \frac{1}{L}\sum_{i=1}^{L}\frac{TP_i}{TP_i + FP_i}$$

$$Recall_{macro} = \frac{1}{L}\sum_{i=1}^{L}\frac{TP_i}{TP_i + FN_i}$$

For the label-based micro-averaged evaluation, we used:

$$Precision_{micro} = \frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L}(TP_i + FP_i)}$$

$$Recall_{micro} = \frac{\sum_{i=1}^{L} TP_i}{\sum_{i=1}^{L}(TP_i + FN_i)}$$

In both cases the $F_1\text{-}score$ can be computed as follows:

$$F_1\text{-}score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

B. Micro-averaged Fine-Tuning Results

Model	Method	Field	Micro F1	Precision	Recall
XLMRb	BinRel	B	$\mathbf{0.721} \pm 0.020$	$\mathbf{0.723} \pm 0.026$	$\mathbf{0.720} \pm 0.028$
XLMRb	BinRel	T+B	0.719 ± 0.016	0.717 ± 0.030	0.722 ± 0.023
XLMRb	LPSet	B	0.702 ± 0.021	0.719 ± 0.030	0.686 ± 0.024
XLMRb	LPSet	T+B	0.702 ± 0.018	0.714 ± 0.028	0.692 ± 0.025
BERTmc	BinRel	B	0.690 ± 0.017	0.703 ± 0.018	0.678 ± 0.027
BERTmc	BinRel	T+B	0.690 ± 0.016	0.696 ± 0.023	0.685 ± 0.021
BERTmc	LPSet	B	0.669 ± 0.015	0.572 ± 0.027	0.667 ± 0.031
BERTmc	LPSet	T+B	0.663 ± 0.017	0.661 ± 0.026	0.531 ± 0.032
			Baseline		
Majority L_1			$\mathbf{0.309}$	$\mathbf{0.300}$	$\mathbf{0.319}$
Random			0.247	0.246	0.248

Table 7: Label-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, micro F1 score, micro precision and micro recall.

C. Micro-averaged Zero-Shot Results

Model	Method	Fields	micro F1	micro P	micro R
XLMRb	BinRel	B	0.457	$\mathbf{0.662}$	0.350
XLMRb	BinRel	T+B	0.471	$\mathbf{0.662}$	0.366
XLMRb	LPSet	B	$\mathbf{0.490}$	0.658	0.390
XLMRb	LPSet	T+B	$\mathbf{0.490}$	0.636	$\mathbf{0.398}$
			Baseline models		
Majority L_1			$\mathbf{0.511}$	$\mathbf{0.570}$	$\mathbf{0.463}$
Random			0.410	0.422	0.398

Table 8: Label-based Zero-Shot classifier performance - shows the pre-trained model used, followed by a problem transformation method, selected article fields for training, micro F1 score, micro precision and micro recall

D. Slovene corpus media outlets

id	name	url
10	24ur.com	https://www.24ur.com/
20	Celje.info	https://www.celje.info/
30	Delo.si	https://www.delo.si/
40	Demokracija.si	https://demokracija.si/
50	Dnevnik.si	https://www.dnevnik.si/
60	Dolenjskilist.si	https://www.dolenjskilist.si/
70	Domovina.je	https://www.domovina.je/
80	Druzina.si	https://www.druzina.si/
90	Ekipa.svet24.si	https://ekipa.svet24.si/
100	Gorenjskiglas.si	https://www.gorenjskiglas.si/
110	Kozjansko.info	https://kozjansko.info/
120	Lokalec.si	https://lokalec.si/
130	Mladina.si	https://www.mladina.si/
140	N1info.si	https://n1info.si
150	Necenzurirano.si	https://necenzurirano.si/
160	Nova24tv.si	https://nova24tv.si/
170	Novice.svet24.si	https://novice.svet24.si/
180	Novitednik.si	https://www.novitednik.si/
190	Politikis.si	http://www.politikis.si/
200	Primorske.si	https://www.primorske.si/
210	Primorski.eu	https://www.primorski.eu/
220	Prlekija-on.net	https://www.prlekija-on.net/
230	Reporter.si	https://reporter.si/
240	Rtvslo.si	https://www.rtvslo.si/
250	Siol.net	https://siol.net/
260	Slovenskenovice.si	https://www.slovenskenovice.si/
270	Vecer.com	https://www.vecer.com/
280	Vestnik.si	https://vestnik.si/
290	Zurnal24.si	https://www.zurnal24.si/

Table 9: Slovene news media sources. Showing corpus media identifier, media source name, and media source URL

Manosphrames: exploring an Italian incel community through the lens of NLP and Frame Semantics

Sara Gemelli[a] and **Gosse Minnema**[b]
University of Pavia[a]; University of Bergamo[a]; University of Groningen[b]
sara.gemelli@unibg.it; gosseminnema@gmail.com

Abstract

We introduce a large corpus of comments extracted from an Italian online *incel* ('involuntary celibate') forum, a community of men who build a collective identity and anti-feminist ideology centered around their inability to find a sexual or romantic partner and who frequently use explicitly misogynistic language. Our corpus consists of 2.4K comments that have been manually collected, analyzed and annotated with topic labels, and a further 32K threads (700K comments) that have been automatically scraped and automatically annotated with FrameNet. We show how large-scale frame semantic analysis can shed a light on what is discussed in the community, and introduce incel topic classification as a new NLP task and benchmark.

Keywords: Incels, Typical frames

1. Introduction

Among the many communities that prosper on the Internet, the so-called *manosphere* seems to be one of the most obscure yet active of the last years. More precisely, the manosphere is better described as a *network* of forums, groups, and blogs that are heterogeneous with respect to their approaches to society, but united by the commitment to expose the oppression that men supposedly suffer at the hands of women (Ging, 2019). Within this landscape, the incel community is probably the best-known. The term is a portmanteau for *involuntary celibate*, and it is used by the users to describe themselves; the men who populate incel forums, in fact, argue that their inability to find a partner is due to their appearance or their personality not meeting the standards imposed by the society in general and by women in particular.

The language used within these groups is often explicitly derogatory towards women, who are seen both as active oppressors and object of desire, whereas the users of the forums conceive themselves as passive victims who are doomed to suffer their condition of *inceldom*. Their linguistic repertoire is characterised by what Jane (2014) calls *e-bile*, a term that encapsulates a variety of offensive expressions typical of online communication, often misogynistic and homophobic; it is also rich in neologisms and acronyms, that reflect the users' linguistic creativity and often disguise disparaging expressions. These communities are often conceived as hermetic environments, exceptional and somehow remote, but in fact the very nature of the Web makes it easy for their theories to disperse in the mainstream areas of the online world, with language as the vehicle. Their discursive choices and linguistic creativity contribute to the strengthening of their sense of belonging and the creation of an in-group (the users of the forum and incels in general, depending on the discussion) and of an out-group (women, and sometimes people that are considered 'not ugly') very rigidly delimited. The misogynistic ideologies perpetuated within these echo chambers are strictly linked to the patriarchal rules that permeate society, and reflect the sense of entitlement (Manne, 2017) that men who live by those rules often feel towards women; these sexist positions become more explicit in the forums through the users' discourse, working as a 'social glue' and justifying their sense of misery and failure, recurring motifs within the community.

We believe that developing computational linguistic analyses of incel discourse is urgently necessary: a better understanding of how the manosphere works is vital for addressing 'modern' forms of misogyny, and computational tools are needed for quantitatively analyzing the enormous amounts of online content that are produced every day. Moreover, incel discourse sometimes leads to extremist violence;[1] this increases the need for real-time automatic monitoring tools.

So far, there has been only a small amount of NLP research about inceldom, all of it very recent and mostly investigating Anglophone contexts (see §2). In this paper, we build on this previous work by analyzing incel language through the lens of frame semantics (Fillmore, 1985) as a tool for systematically analyzing *what incels say* about themselves and others. In doing so, we take a non-anglocentric perspective, analyzing an Italian incel forum: *Il Forum dei Brutti* ("The Forum of the

[1] E.g., see the well-known 2018 attack in Toronto; https://www.theguardian.com/world/2022/jun/13/toronto-van-murders-court-victim-2018-attack (archived version at https://archive.is/QurUY).

First Workshop on Reference, Framing, and Perspective, pages 28–39

Ugly").[2]

Contributions Our main contributions are:

- A new dataset comprising over 700K comments from the forum, 2.4K of which have been manually annotated with topic labels;[3]
- A novel frame semantics-based analysis of the corpus;
- A benchmark of preliminary machine learning experiments for predicting topic labels.

2. Background

In recent years, there has been a growing interest in exploring the manosphere and its complexity from various perspectives. Scholars investigating this network primarily examine anglophone groups or communities and come from various disciplines, including gender studies, sociology, psychology, communication sciences, economics, and linguistics. These analyses cover several important aspects, such as the representation of masculinity within different communities (Schmitz and Kazyak, 2016; Van Valkenburgh, 2021), the presence of misogyny within these groups (Farrell et al., 2019), discourses and perceptions of violence within the communities (Bryan and Warren, 2023; Lounela and Murphy, 2023), and the manosphere's relationship with the neoliberal economic model (Banet-Weiser and Bratich, 2019).

In the linguistic field, discourse within the manosphere, and especially incel communities, has been often studied with corpus-based Critical Discourse Analysis and Cognitive Linguistics approaches (Heritage and Koller, 2020; Maxwell et al., 2020; Tranchese and Sugiura, 2021), but also by means of Computational Linguistics techniques. In particular, Jaki et al. (2019) worked on the description of the language used in incel forums in order to facilitate the identification of gendered hate speech; Jelodar and Frank (2021) analyse comments from an incel forum from a semantic perspective, through tasks such as topic modeling and opinion mining; Yoder et al. (2023) investigated the terminology used by users to construct their own identity in incel forum.

Despite the limited literature in the Italian context, researchers' interest in this type of online phenomenon has been increasing over the past few years. The sociologists Farci and Righetti (2019) have focused on analyzing communities of men's rights activists and their emergence as a response to feminism, particularly online. The International Journal of Gender Studies *About Gender* dedicated an issue (vol. 10, No. 19, 2021: *Doing masculinities online: defining and studying the manosphere*) to exploring the manosphere in both Italian and international contexts. In this monographic issue, in particular, Dordoni and Magaraggia (2021) explore the representations of gender identities and masculinity within Italian Red Pill and incel communities; De Gasperis (2021) aims to analyze interactions within the Italian incel forum Il Forum dei Brutti (also subject to analysis in this contribution) to explore the intertwining of gender identity representations and the Italian literary imagination, particularly by considering threads where users compare themselves to the poet Giacomo Leopardi. The linguists Nodari and Fiorentini (2023) propose a first description of the language used within various communities of the Italian manosphere, analysing data extracted from blogs and webpages. In particular, the authors observe how narratives within the Italian manosphere resemble those in the Anglophone context, where individuals are rigidly classified based on gender definitions. Additionally, they highlight that users primarily discuss gender relations in terms of economic transactions and employ a language that echoes scientific discourse in presenting quantitative data and para-scientific evidence in support of their ideology. Lastly, from a cross-lingual perspective, Gajo et al. (2023) collected and analysed data from English- and Italian-language incel forums. They performed tasks of automatic identification of hate speech, with a specific focus on misogyny and racism, and attempted to forecast the extent to which the posts would trigger more hateful content.

We are not aware of any previous working using frame semantics to analyze manosphere discourse.

3. Datasets

In order to examine the communication dynamics within the Italian forum from multiple perspectives, we collected two different datasets using two different approaches. It is important to note that we only accessed the parts of the forum that are public. While there are private sections of the forum that are only accessible to users who sign up, we believe that the threads that anyone can freely access are representative enough of the discourse

[2]This is the same forum that Gajo et al. (2023) used; since their paper had not yet been published at the time that we conducted our study, we collected our dataset independently of their work. However, it is likely that their is considerable overlap between our and their dataset.

[3]Our scraping and analysis code as well as annotations on a stand-off basis are available on `https://gitlab.com/sociofillmore/manosphrames`. Due to privacy and copyright concerns, we cannot publicly release the full corpus, but it is available upon request for interested researchers.

within incel communities; in addition, we expect these threads to have a stronger impact on the expansion of these ideologies and linguistic expressions outside these spaces. In section 3.1, we discuss the collection of the first corpus and the process of manual annotation that was performed on it; section 3.2 presents the second corpus, which was scraped automatically.

3.1. Manually Collected Corpus

The first dataset includes threads and posts that were collected manually between November 2020 and October 2021. We accessed the forum monthly and extracted random threads and posts that were actively being commented on at the time of access. This resulted in a total of 60 threads, 2,406 posts and approximately 94,000 words.

Manual Annotation One human annotator classified comments in the dataset by topic, in order to better understand the most common subjects of discussion in the forum and to obtain different subcorpora based on topics. Assigning only one topic to each comment was often too reductive, given that many comments were complex and quite long. Therefore, two topics were assigned to each comment when necessary. The set of topics used in the annotation was based on both literature regarding incel communities (Dordoni and Magaraggia, 2021; Heritage and Koller, 2020; Tranchese and Sugiura, 2021) and the experience frequenting the forum during the year in which the data were collected. The corpus was annotated through a process of *open coding* (Khandkar, 2009). The annotator read the corpus in order to identify a first set of recurring topics; subsequently, each comment was annotated by topic, assigning one or two labels according to the topics discussed. During the classification, every new comment was compared to the previously annotated ones, in order to determine if the label was fitting and if the set of labels was sufficiently representative; if not, new labels were added to the set, or the criteria to assign the labels were adjusted. This resulted in a definite set of labels that was suitable to annotate the whole corpus.

Annotation Set We define the following topic labels: D (*donne*, 'women'), AF (*aspetto fisico*, 'physical aspect'), IN (*incels*), SS (*sé stessi*, 'themselves' – the users), U (*uomini* 'men'), AL (*altri utenti*, 'other users'), FDB (*Forum dei Brutti*), SOC (*società*, 'society'), RP (*Red Pill*, a theory shared within the manosphere), M (*mondo esterno*, 'outside world'), POL (*politica*, 'politics'), IRR (*irrilevante*, 'irrelevant'). See Table 1 for the descriptions of the labels and examples[4].

Inter-Annotator Agreement For validating the annotations, a second annotator annotated a subset of the corpus (157 comments[5]). Up to two topic labels could be assigned per comment; in order to compute set similarity between the annotations for each comment, we experimented with both Jaccard Distance and the MASI Distance (Passonneau, 2006). We obtained a Cohen's Kappa score of $\kappa = 0.59$ using Jaccard Distance, and a more conservative $\kappa = 0.52$ using MASI Distance.

3.2. Scraped Corpus

In addition to the manually collected corpus, we also automatically scraped the entire history of threads on the *Una vita da brutto* ('a life as a ugly person') subforum, containing 32,560 threads posted between April 2010 and May 2023, containing 706,086 posts in total (number of posts/thread ranging from 1–2,165; median 13.0) amounting to 31.8 million words (between 1–58,815 words per thread; median 478 words).

4. Automatic Annotation

We performed a small-scale set of machine-learning experiments aiming at automatically predicting topic labels for forum comments in a multi-label setting. We tried two different approaches: on one hand, we trained a linear SVM model using as input either text-based features (raw unigram counts and TF-IDF weighted unigram counts) or frame-based features (a count vector of automatically tagged FrameNet frames); on the other hand, we used ChatGPT[6] with several zero-shot and few-shot prompts: a prompt including only the label definitions, a prompt including one hand-picked prototypical example for each label, and a prompt including both examples and definitions.

The annotated dataset was split into 70% training samples (n=1684), 20% test samples (n=481) and 10% development samples (n=241). For the ChatGPT experiments, we used the `gpt-3.5-turbo` model from the OpenAI API[7], with the default generation settings and the default system parameter.

[4] Some words in the examples provided in the table were slightly atlered, in order to preserve the anonymity of the users.

[5] For choosing this subset, we randomly selected threads from the corpus until the total number of selected comments exceeded 150, or about 6% of the total corpus.

[6] https://openai.com/blog/chatgpt

[7] The experiments were done in May 2023.

Code	Description	Example		Typical frames (FFICF)	Freq.
D	Comments about women (in general, about specific women, or about relationships)	*La maggior parte delle donne, a causa del ciclo mestruale, sono di umore instabile e capricciose.*	"The majority of women, because of their menstrual cycle, are unstable and moody"	PEOPLE (1.0), PERSONAL_REL. (0.94), DESIRING (0.79)	863
AL	Comments about other forum users	*Sembra che tu ti senta giudicato o attaccato anche quando non hai fatto niente di male.*	"It looks like you feel judged and attacked even if you haven't done anything bad."	STATEMENT (1.0), AWARENESS (0.99), DESIRABILITY (0.96)	697
AF	Comments about physical appearance	*Per avere gli addominali bisogna mangiare troppo poco, per come la vedo io.*	"In order to have nice abs one has to eat too little, in my opinion."	BODY_PARTS (1.0), AESTHETICS (0.69), BODY_DESCRIPTION_HOLISTIC (0.62)	622
SS	Comments by users talking about themselves	*Ma cosa state dicendo? Io prima d'ora non avevo neanche mai avvicinato una ragazza.*	"What are you talking about? Until now I had never even approached a girl"	AWARENESS (1.0), CALENDRIC_UNIT (0.97), EMOTION_DIRECTED (0.90)	433
IN	Comments talking about incels (as individuals or as a community)	*Tutto questo è assurdo, noi al giorno d'oggi siamo perseguitati come gli ebrei... esagero, ma avete capito.*	"All this is crazy, we are persecuted just like the Jews... I am overstating, but you get what I mean."	STATEMENT (1.0), PEOPLE (0.87), INCREMENT (0.78)	219
M	Comments about external people				212
IRR	Irrelevant / not possible to assign a label				188
FDB	Comments about the forum itself				153
SOC	Comments about society and social issues				83
U	Comments about the position of men in society				71
RP	Comments related to the redpill theory				22
POL	Comments related to political issues and ideologies				13

Table 1: Annotated topic labels in the manually collected forum, with examples and typical frames (see §5) for the top-5 most frequent topics. N.B.: a comment can have up to two topic labels.

Table 2 lists the results of our experiments. Interestingly, the best overall model are the two word count-based SVM models. ChatGPT performs substiantially worse, the best-performing setup being the zero-shot one. While our results are from a single run of the model, the fact that scores are very consistent between the development and test sets suggests that the model's performance is stable across different runs.

5. Frames

We build on recent work that applies Fillmorean frame semantics to societal issues (Minnema et al., 2022a,b,c). Fillmorean frames (Fillmore, 1985, 2006), catalogued in lexical databases such as Berkeley FrameNet (Baker et al., 1998), are pieces of conceptual information, grounded in human experience and cognition, that pick out a particular event, situation or object in the world around us. By looking at which frames are used in a text, we gain information about *what* is said about the world, but also about *how* and *from whose perspective* it is said. For example, in English, "buying" and "selling" pick out the same type of real-world event, but do so from a different event participant's point of view. In the context of sociopolitically loaded events, studying why one frame is used over another can be used as a tool for Critical Discourse Analysis (CDA): for example, when

Input representation	Model	dev P	dev R	dev F1	test P	test R	test F1
Frame count vectors	Linear SVM ($C = 4$)	0.40	0.35	0.37	0.36	0.34	0.35
Bag-of-words (raw) vectors	Linear SVM ($C = 0.5$)	0.63	0.50	0.56	0.58	0.54	0.51
Bag-of-words (tf-idf) vectors	Linear SVM ($C = 2$)	0.72	0.51	0.59	0.68	0.45	0.54
Zero-shot prompt (only definitions)	ChatGPT	0.42	0.41	0.41	0.43	0.43	0.43
Few-shot prompt (only examples)	ChatGPT	0.22	0.14	0.17	0.21	0.14	0.17
Few-shot prompt (examples + definitions)	ChatGPT	0.42	0.32	0.37	0.45	0.34	0.39

Table 2: Classification experiment results (P, R, and F1 are micro-averaged across topic labels)

discussing traffic incidents, the headlines "cyclist *dies* in traffic" or "driver hits and *kills* a cyclist" could both be factually correct ways of describing the same event, but convey different perspectives on the event and could imply different ideological positions (in this case about the place of cars and pedestrians in urban planning) (Minnema et al., 2022c). This type of variation in framing has also been linked to differences in event perception, e.g. with respect to who is to blame (Minnema et al., 2022a).

In the present study, we are interested in studying how users of incel forums conceptualize the world, especially when relating to gender relations. In this section, we perform two types of analysis as preliminary steps to better understanding the conceptual world of the community: (i) analyzing which frame types are most representative for the different topics discussed in the corpus; (ii) analyzing which semantic frames are used to talk about men versus women.

For both analyses, we use LOME (Xia et al., 2021) to automatically annotate our automatically-scraped corpus with FrameNet frames. LOME has been trained only on the English-language Berkeley FrameNet, but, since it has a multilingual encoder model (XLM-R, Conneau et al. 2020), can be applied to other languages in a zero-shot cross-lingual transfer setting. To our knowledge, LOME is the only recent model to have been tested on an Italian FrameNet benchmark; in Minnema et al. (2022c), we showed that it has acceptable overall performance for Italian on a standard benchmark. We also evaluated LOME's predictions on a dataset of Italian news articles about gender-based violence and showed that applying the (original) version of LOME that was trained on English data and tested directly (zero-shot) to Italian outperformed a (new) version of LOME that was trained on both English and Italian data.

The use of automated FrameNet-based analysis has several limitations. First of all, frame semantic parsers make errors and performance may vary across different types of texts. Due to the high complexity and cost of manually annotating high-quality evaluation data, we were unable to systematically test LOME's performance on our corpus for this study. However, based on a preliminary check of the automatic annotations, we found that there is at least one serious domain adaptation issue, namely that words relating to (consensual) sex are frequently mistagged. In particular, the verb *scopare* ("fuck") was frequently mistagged as evoking violence-related frames (KILLING, RAPE); we therefore decided to exclude all instances of *scopare* from our frame analysis. We speculate that these errors may originate from the nature of Berkeley FrameNet; the original FrameNet corpus contains many descriptions of violence (e.g., in the context of geopolitics) and frames corresponding to this, but few descriptions of and frames relating to (consensual) sex. Apart from affecting parsing performance, the semantic coverage of Berkeley FrameNet also forms a limitation for our analysis by itself: we are likely to miss out on many important aspects of incel discourse due to lacking frames. For example, while there are frames related to emotion in general (e.g., EMOTION_DIRECTED) there are no frames specifically for capturing expressions of hate, or for analyzing (misogynistic) emotion descriptions such as "unstable" and "moody" (see the first example in Table 1), which are frequently found in the corpus. In the future, it could be interesting to look at expanding FrameNet's coverage for our specific domain, as has also been proposed, for example, for the domain of gender-based violence (Dutra et al., 2023).

5.1. Typical Frames

We adopt the notion of *typical frames* proposed in Vossen et al. (2020) and Remijnse et al. (2021): a set of FrameNet frames that is most representative for a particular subcorpus within a larger corpus, and that can be automatically detected using *FF-ICF*, a modified version of the *TF-IDF* metric:[8]

$$\text{FF-ICF}_i = \frac{t_i}{f_i} \times \log \frac{m}{\sum_j^n t_j}$$

[8]Specificaly, it is a derivative of C-TFIDF (Grootendorst, 2022).

where i is a subcorpus, t_i is the frequency of frame t in i, f_i is the total number of frame instances in i, and m is the total number of documents across all subcorpora (Remijnse et al., 2021, p. 233). This results in each frame being assigned a score in $[0, 1]$ for each subcorpus, with the highest-ranked frames in each subcorpus being most informative for distinguishing between subcorpora.

Table 1 shows the highest-ranking frames for each of the most frequent topic labels in the manually annotated corpus. For example, for topic D ("women"), we find PEOPLE (frequently triggered by words such as *ragazza* "girl"), PERSONAL_RELATIONSHIP (triggered by words such as *amica* "[female] friend", *fidanzamento* "engagement"), or DESIRING (triggered by words such as *volere* "to want"). In this case, the frames correspond quite closely to aspects of the manually definition of the topic. However, we also find more specific information: for example, when looking at different instances of DESIRING, we find that forum users frequently talk about desires (romantic or otherwise) both from their own (male) perspective (e.g. *non la voglio*, "I don't want her") but also from the perspective of women (e.g. *loro vogliono il brivido di far eccitare i maschi e sentirsi desiderate*, "they [women] want the thrill of getting boys turned on and they want to feel desired"). Similarly, for other topics we also find frames closely corresponding to the topic definition — e.g. for topic AF ("physical appearance") we find BODY_PARTS, triggered in phrases such as *poteva farsi un trapianto di capelli*, "[he] could get a hair transplant" or *bei lineamenti e bonus occhi, 6*, "nice features and a bonus [for her] eyes, [she gets a] 6" — but also less expected, but still informative frames: e.g., for topic SS ("talking about themselves"), we find CALENDRIC_UNIT, triggered by words like *ieri* "yesterday", *stamattina* "this morning", which is often an indicator of stories about the users' personal lives, e.g. *prima di ieri non avevo neanche baciato* "until yesterday, I had never even kissed". In addition, the presence of the frame EMOTION_DIRECTED as one of the most frequent in comments about themselves, evoked by words like *tristezza* "sadness", *imbarazzo* "embarrassment", *ansia* "anxiety", suggests that the users frequently present their experiences adopting emotional narratives, choosing to share their feelings (often negative) about themselves or their life with the rest of the community.

5.2. Gender and Semantic Roles

In the typical frame analysis, we used semantic frames essentially as a way to group together related lexical units: words that express the same concept. Here, we go a step further and exploit the ability of semantic frames to relate concepts to *frame elements*: semantic roles that express the prototypical participants of an event or situation type (e.g. in "Chiara sold a book to Tommaso", "Chiara" fills the Seller role of the COMMERCE_SELL frame whereas "a book" fills the Goods role and "Tommaso" fills the Buyer role). By analyzing the contents of semantic role spans, we can get an insight into what is said about event participants: in which frames does a given participant appear as a role filler, and which roles does that participant fill? Here, we are interested in men vs. women: what kind of conceptual information do forum users typically convey about members of each gender? We are particularly interested in *agentive frames*: semantic frames that describe the main participant as actively doing something. In the literature about language and gender, two (seemingly) conflicting patterns have been observed relating agentivity: on the one hand, there seems to be a general tendency in several languages that in active sentences, men are more often expressed as syntactic subjects than as objects, while women are more often expressed as syntactic objects (Kotek et al., 2021; da Cunha and Abeillé, 2022). On the other hand, linguists and feminist scholars have identified patterns of language use where actions by men are described in a *de-agentivized* way. For example, according to Penelope (1990), expressions without an explicit agent such as "it is widely *understood* that..." or "the only reason for *ordering* a war ..." are frequently found in contexts in which the only plausible implicit agent is a man or a group of men, and omitting the agent in such cases can contribute to presenting men's experiences as 'universal' or to present male actions as inevitable and to absolve the agents of responsibility. This latter pattern has also been observed in a number of recent empirical (cognitive, corpus-based and computational) studies on the reporting of gender-based violence, where journalists and other writers frequently use agent-removing constructions, which has been shown to decrease the level of blame that readers attribute to the perpetrator (Henley et al., 1995; Bohner, 2002; Pinelli and Zanchi, 2021; Meluzzi et al., 2021; Minnema et al., 2022a). In the light of this, it is interesting to observe agentivity patterns in the language use of the incel community.

We implement our analysis as follows: first, we perform a keyword search in all automatically detected semantic roles in the 32K threads scraped from the forum for words referring to men or women, respectively.[9] Next, we identify agen-

[9]We used the following keywords: *donna, donne, ragazza, ragazze, lei, np, co* ("woman", "women", "np", "co", "girl", "girls", "she/her"; *np* and *co* are abbreviations of pejorative terms frequently used in the corpus to mean "woman") for women, and *uomo, uomini, ragazzo, ragazzi, lui* ("man", "men", "boy", "boys", "he/him").

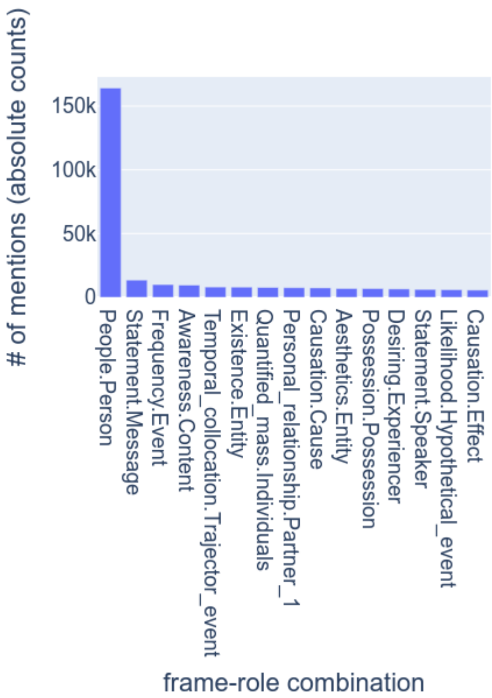


Figure 1: Roles mentioning women

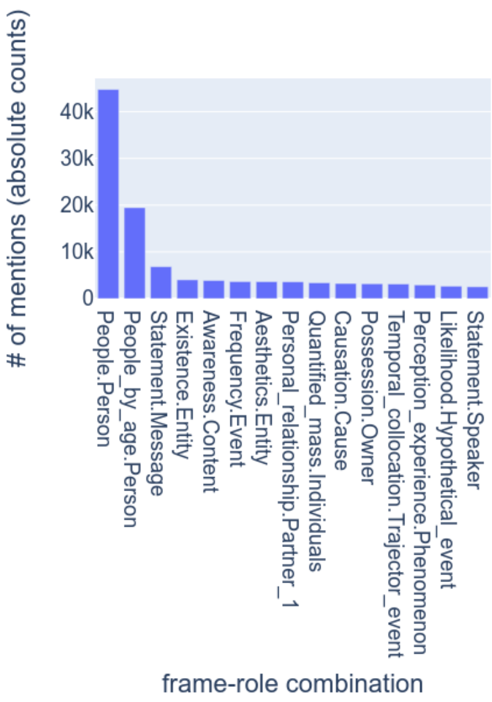


Figure 2: Roles mentioning men

tive frames by checking for each frame if it (indirectly) inherits from either TRANSITIVE_ACTION or INTENTIONALLY_ACT, and identifying which of the frame's roles expresses an agent.[10] Note that both

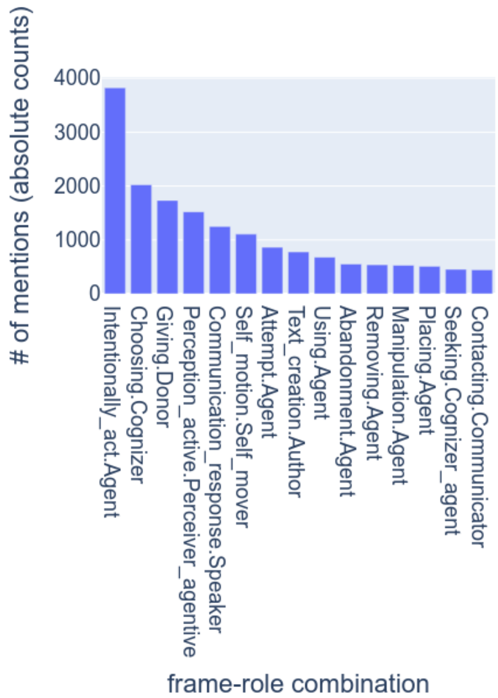


Figure 3: Agentive roles mentioning women

of these steps have limited recall: our keyword search does not include all possible words for referring to men and women, and also ignores all cases where a participant is expressed using a proper name or anaphorically as a pro-drop subject. Moreover, the FrameNet hierarchy is incomplete, and not all frames that are semantically agentive can be detected as such using the inheritance hierarchy.

Figures 1 and 2 show the top-15 most frequent semantic roles across all frames that match one of the keywords. The first observation is that women are mentioned much more frequently than men. For both genders, "man/men" or "woman/women" are the most commonly matched keywords (we find 605K matches in total for the female keywords, of which *donna/donne* accounts for 317K; for men, *uomo/uomini* account for 126K out of 257K total matches). In both cases, the by far most frequent frame-role combination is PEOPLE.Person, which is expected as any and all mentions of "man" or "woman" (and variants/synonyms of those words) trigger this frame. The next-most frequent frames are more interesting: for example, we find 13.5K instances of women matching STATEMENT.Message role, i.e., being mentioned as the content of some-

[10]FrameNet has a rich and complex graph structure of relating frames and frame elements to each other. Here, we only use the *inheritance* frame-to-frame relation. We allow for both direct and indirect inheritance; e.g., KILLING inherits directly from TRANSITIVE_ACTION (where the KILLING.Killer role is mapped to TRANSITIVE_ACTION.Agent; on the other hand, COOKING_CREATION inherits from INTENTIONALLY_CREATE (mapping the Cook role to Creator), which in turn inherits from INTENTIONALLY_ACT (mapping Creator to Agent).

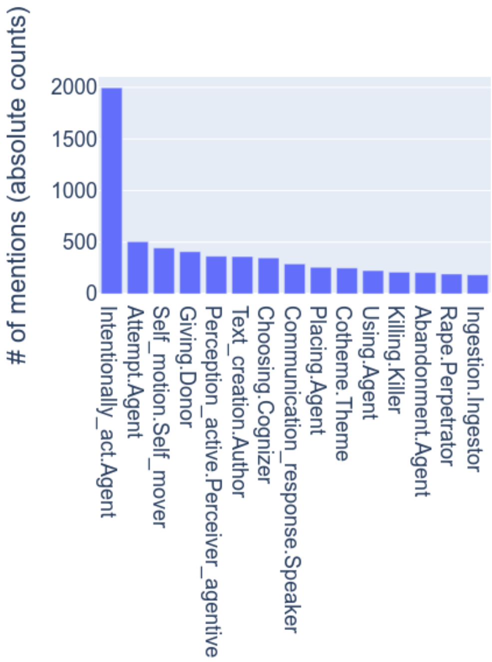


Figure 4: Agentive roles mentioning men

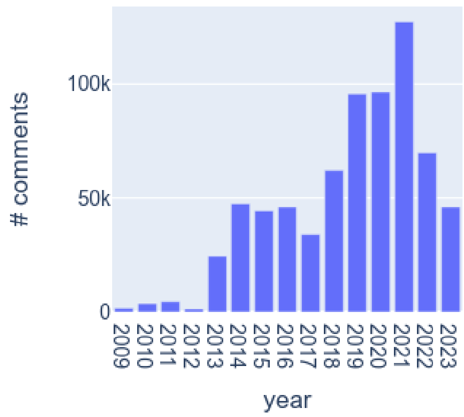


Figure 5: Comments by year

thing that someone says (e.g., in *ma almeno non potete* ***dire*** *[che nessuna ragazza starebbe mai con voi]* "but at least you cannot **say** [that no girl will ever be with you]")[11]. Similarly, we find 9.5K mentions of women matching Awareness.Content, i.e. being mentioned in the content of a (stated) knowledge or belief (e.g. *io non* ***so*** *[se le donne cerchino un uomo per riprodursi]* "I don't **know** [if women are looking for a man to reproduce]"). Interestingly, these same frame/role pairs are also in the top-5 for mentions of men.

Moving to agentive frames (Figures 3 and 4), Intentionally_act.Agent is the most frequent role for both genders; the majority of these instances correspond to the subject of the verb *fare* "do/make/act" (e.g. *Io non ho mai visto [ragazze]* ***fare*** *così*, "I have never seen [women] **acting** like this"). In the rest of the top-5, for women we find Choosing.Cognizer (e.g. *Tanto è la donna che sceglie* "It's the woman who chooses, anyway"), Giving.Donor, and Perception_active.Perceiver_agentive (e.g., "seeing", "hearing"). By contrast, for men, we find Rape.Perpetrator, Attempt.Agent, Self_motion.Mover ("going"), and Choosing.Donor.

5.3. Diachronic Analysis

Since our corpus spans more than a decade's worth of posts (April 2009–May 2023), we were also able to start exploring the question of how incel discourses changes over time. Figure 5 shows the number of comments scraped from *Il Forum dei Brutti*. The annual number of comments steadily increases after 2012, rising to over 127,000 posts in 2021. This trend ended in 2022, which saw substantially fewer comments; our data for 2023 are incomplete. Note that a decrease in comments in our corpus does not necessarily imply a decrease in overall activity: certain sections of the forum are private, and we are unable to monitor the activity trends there. Figures 6 and 7 show the top-7 of most frequent role mentions (collectively accounting for at least 30% of the total number of matching role mentions in each year) that match women-related and men-related keywords, respectively. The clearest visible pattern is the stability over time of the most frequent frames: the top frames are mostly the same ones for each year. However, an interesting development is the decline in frequency of Aesthetics.Entity (e.g., sentences like *Ho ritenuto da sempre [le donne indiane] le più* ***belle*** *del mondo*, "I have always considered [Indian women] the most **beautiful** in the world"), for both women and men: relating to women, the relative frequency of this frame-role pair peaked at 2.5% in 2011, and then entered a steep decline, falling to 1.4% in 2016 and 0.9% in 2021; relating to men, there is a similar pattern, but with a peak in 2013 (3.1%), falling to 1.0% in 2021. While it is hard to draw strong conclusions from this, it could be an indicator that the importance of discussing physical attractiveness is declining or increasingly expressed in a different way over time.

[11] Frame trigger highlighted in boldface, semantic role instance between square brackets.

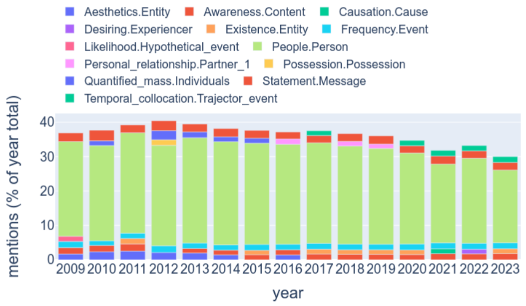


Figure 6: Roles mentioning women by year

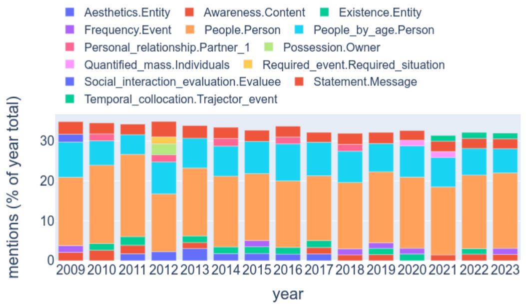


Figure 7: Roles mentioning men by year

6. Conclusion

This paper introduced a large corpus of comments extracted from *Il Forum dei Brutti*, an Italian online incel community. Our corpus consists of 2.4K comments that have been manually collected, analyzed and annotated with topic labels, and a further 32K that have been automatically scraped and automatically annotated with FrameNet annotations. We also provided a benchmark with basic machine learning experiments for automatically predicting topic labels. Our experiments yielded mixed results: while simple SVM-based approaches work surprisingly well, ChatGPT performed surprisingly poorly. Finally, we performed an automatic FrameNet-based analysis of the contents of the corpus. In the first step of our analysis, we showed the usefulness of typical frame detection (Vossen et al., 2020; Remijnse et al., 2021) for analyzing topic-based subcorpora. In the second step of the analysis, we showed that forum users talk twice as frequently frequently (explicitly) about women than about men — this is true both across all frames and when only including frame instances where a man or woman is described in an agentive way — and we found interesting parallels and differences in the patterns of semantic roles in which men and women appear most commonly. We also took a first peek at how incel discourse changes over time; clearly, there is a lot of room for expansion here. In particular, Baele et al. (2023) observed a general increase of violent extremist discourse in incel forums in recent years; it could be interesting to investigate this through the lens of violence-related frames (e.g. KILLING, CAUSE_HARM, etc.).

7. Acknowledgements

Author G.M.'s contribution was funded by the Dutch National Science organisation (NWO) through the project *Framing situations in the Dutch language*, `VC.GW17.083/6215` (see `www.dutchframenet.nl`). We would like to thank the anonymous reviewers for their extensive comments and insights.

8. Limitations

Some of the peculiarities of the language used by the users of the forum can produce various challenges in computational tasks like automatic frame analysis and annotation. In particular, we came across three issues that we tried solve, but that partially still represent areas for improvement: the first two have to do with the way the users refer to women and to themselves, whereas the last one concerns lexical opacity typical of these type of communities.

To identify semantic roles mentioning women we used a list of keywords that seemed to cover the majority of cases (cfr. note 2), but in some comments the users refer to women by only using the plural personal pronoun *loro* ("they/them"), or even just by relying on verbal morphology, expressing the verb in the third-person plural form, as italian is a pro-drop language (e.g. *pens-ano*, "[they] think"; *dic-ono*, "[they] say"). These strategies contribute to the idea that women - seen as a homogeneous whole - are the out-group with respect to the users of the forum. At the same time, the keywords used for the men were not always sufficient, as they often talk about themselves and the community in general using the personal pronoun *noi* ("we/us"), sometimes the adverb *qui* ("here" [in the forum]), or, similarly to the case of women, they simply use the first-person singular form of the verb (e.g. *sembr-iamo*, "[we] seem"), marking themselves as in-group. Lastly, the presence of neologisms and acronyms constitutes an obstacle for the automatic exploration of the corpus. If not all the words are understood by the models, it is harder to obtain a correct classification of the comments; similarly, analyzing semantic frames and roles is more difficult if there are opaque terms, e.g. *CO* for *cessa obesa* ("ugly fat [woman]"), *zerbinare* (lit. "doormatting", the act of submitting completely to a woman in hopes of being noticed and, ultimately, loved by her).

9. Bibliographical References

Stephane Baele, Lewys Brace, and Debbie Ging. 2023. A diachronic cross-platforms analysis of violent extremist language in the incel online ecosystem. *Terrorism and Political Violence*.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 1*, COLING '98, pages 86–90.

Sarah Banet-Weiser and Jack Bratich. 2019. From pick-up artists to incels: Con (fidence) games, networked misogyny, and the failure of neoliberalism. *International Journal of Communication*, 13:1.

Gerd Bohner. 2002. Writing about rape: Use of the passive voice and other distancing features as an expression of perceived responsibility of the victim. *British Journal of Social Psychology*, 40:515–529.

Elliott Bryan and Shane Warren. 2023. The manosphere under the microscope: a critical discourse analysis of the news media reporting of rowan baxter's murder of hannah clarke and her family. *Feminist Media Studies*, pages 1–18.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Édouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451.

Yanis da Cunha and Anne Abeillé. 2022. Objectifying women? a syntactic bias in French and English corpora. In *Proceedings of the First Workshop on Language Technology and Resources for a Fair, Inclusive, and Safe Society within the 13th Language Resources and Evaluation Conference*, pages 8–16, Marseille, France. European Language Resources Association.

Arianna De Gasperis. 2021. "giacomino uno di noi". letteratura italiana e pratiche di maschilità nel forum dei brutti. *AG About Gender-International Journal of Gender Studies*, 10(19).

Annalisa Dordoni and Sveva Magaraggia. 2021. Modelli di mascolinità nei gruppi online incel e red pill: Narrazione vittimistica di sé, deumanizzazione e violenza contro le donne. *AG About Gender-International Journal of Gender Studies*, 10(19).

Lívia Dutra, Arthur Lorenzi, Lorena Larré, Frederico Belcavello, Ely Matos, Amanda Pestana, Kenneth Brown, Mariana Gonçalves, Victor Herbst, Sofia Reinach, Renato Teixeira, Pedro de Paula, Alessandra Pellini, Cibele Sequeira, Ester Sabino, Fábio Leal, Mônica Conde, Regina Grespan, and Tiago Torrent. 2023. Building a frame-semantic model of the healthcare domain: Towards the identification of gender-based violence in public health data. In *Jornada de Descrição do Português: Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*.

Manolo Farci and Nicola Righetti. 2019. Italian men's rights activism and online backlash against feminism. *Rassegna Italiana di Sociologia*, 60(4):765–781.

Tracie Farrell, Miriam Fernandez, Jakub Novotny, and Harith Alani. 2019. Exploring misogyny across the manosphere in reddit. In *Proceedings of the 10th ACM conference on web science*, pages 87–96.

Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di semantica*, 6(2):222–254.

Charles J. Fillmore. 2006. Frame semantics. In D. Geeraerts, editor, *Cognitive Linguistics: Basic Readings*, pages 373–400. De Gruyter Mouton, Berlin, Boston. Originally published in 1982.

Paolo Gajo, Arianna Muti, Katerina Korre, Silvia Bernardini, and Alberto Barrón-Cedeño. 2023. On the identification and forecasting of hate speech in inceldom. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 373–384, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Debbie Ging. 2019. Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and masculinities*, 22(4):638–657.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure.

Nancy M. Henley, Michelle Miller, and Jo Anne Beazley. 1995. Syntax, semantics, and sexual violence: Agency and passive voice. *Journal of Language and Social Psychology*, 14(1-2):60–84.

Frazer Heritage and Veronika Koller. 2020. Incels, in-groups, and ideologies: The representation of gendered social actors in a sexuality-based online community. *Journal of Language and Sexuality*, 9(2):152–178.

Sylvia Jaki, Tom De Smedt, Maja Gwóźdź, Rudresh Panchal, Alexander Rossa, and Guy De Pauw. 2019. Online hatred of women<? br?> in the incels. me forum: Linguistic analysis and automatic detection. *Journal of Language Aggression and Conflict*, 7(2):240–268.

Emma A. Jane. 2014. “your a ugly, whorish, slut” understanding e-bile. *Feminist Media Studies*, 14(4):531–546.

Hamed Jelodar and Richard Frank. 2021. Semantic knowledge discovery and discussion mining of incel online community: Topic modeling. *arXiv preprint arXiv:2104.09586*.

Shahedul Huq Khandkar. 2009. Open coding. *University of Calgary*, 23:2009.

Hadas Kotek, Rokker Dockum, Sarah Babinksi, and Christopher Geissler. 2021. Gender bias and stereotypes in linguistic example sentences. *Language*, 97(4):653–677.

Emilia Lounela and Shane Murphy. 2023. Incel violence and victimhood: Negotiating inceldom in online discussions of the plymouth shooting. *Terrorism and Political Violence*, pages 1–22.

Kate Manne. 2017. *Down girl: The logic of misogyny*. Oxford University Press.

December Maxwell, Sarah R Robinson, Jessica R Williams, and Craig Keaton. 2020. “a short story of a lonely guy”: A qualitative thematic analysis of involuntary celibacy using reddit. *Sexuality & Culture*, 24(6):1852–1874.

Chiara Meluzzi, Erica Pinelli, Elena Valvason, and Chiara Zanchi. 2021. Responsibility attribution in gender-based domestic violence: A study bridging corpus-assisted discourse analysis and readers’ perception. *Journal of pragmatics*, 185:73–92.

Gosse Minnema, Sara Gemelli, Chiara Zanchi, Tommaso Caselli, and Malvina Nissim. 2022a. Dead or murdered? Predicting responsibility perception in femicide news reports. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1078–1090. Association for Computational Linguistics.

Gosse Minnema, Sara Gemelli, Chiara Zanchi, Tommaso Caselli, and Malvina Nissim. 2022b. SocioFillmore: A tool for discovering perspectives. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 240–250, Dublin, Ireland. Association for Computational Linguistics.

Gosse Minnema, Gaetana Ruggiero, Marion Bartl, Sara Gemelli, Tommaso Caselli, Chiara Zanchi, Marco te Brömmelstroet, and Malvina Nissim. 2022c. Responsibility framing under the magnifying lens of NLP. *Computational Linguistics in the Netherlands*, 12:207–233.

Rosalba Nodari and Ilaria Fiorentini. 2023. Gli alpha e i beta. la lingua della uomosfera nelle comunità online italiane. In V. Caruso and M. Maffia, editors, *Vecchie e nuove forme di comunicazione diseguale: canali, strutture e modelli*, pages 307–321. Officinaventuno.

Rebecca Passonneau. 2006. Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*, Genoa, Italy. European Language Resources Association (ELRA).

Julia Penelope. 1990. *Speaking freely: Unlearning the lies of the fathers’ tongues*. The Athene Series. Permagmon Press, Elmsford, New York.

Erica Pinelli and Chiara Zanchi. 2021. Gender-based violence in italian local newspapers: How argument structure constructions can diminish a perpetrator’s responsibility. *Discourse Processes between Reason and Emotion: A Post-disciplinary Perspective*, page 117.

Levi Remijnse, Marten Postma, and Piek Vossen. 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, Groningen, The Netherlands (online). Association for Computational Linguistics.

Rachel M Schmitz and Emily Kazyak. 2016. Masculinities in cyberspace: An analysis of portrayals of manhood in men’s rights activist websites. *Social Sciences*, 5(2):18.

Alessia Tranchese and Lisa Sugiura. 2021. “i don’t hate all women, just those stuck-up bitches”: How incels and mainstream pornography speak the same extreme language of misogyny. *Violence against women*, 27(14):2709–2734.

Shawn P Van Valkenburgh. 2021. Digesting the red pill: Masculinity and neoliberalism in the manosphere. *Men and masculinities*, 24(1):84–103.

Piek Vossen, Filip Ilievski, Marten Postma, Antske Fokkens, Gosse Minnema, and Levi Remijnse.

2020. Large-scale cross-lingual language resources for referencing and framing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3162–3171, Marseille, France. European Language Resources Association.

Patrick Xia, Guanghui Qin, Siddharth Vashishtha, Yunmo Chen, Tongfei Chen, Chandler May, Craig Harman, Kyle Rawlins, Aaron Steven White, and Benjamin Van Durme. 2021. LOME: Large ontology multilingual extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 149–159, Online. Association for Computational Linguistics.

Michael Yoder, Chloe Perry, David Brown, Kathleen Carley, and Meredith Pruden. 2023. Identity construction in a misogynist incels forum. In *The 7th Workshop on Online Abuse and Harms (WOAH)*, pages 1–13, Toronto, Canada. Association for Computational Linguistics.

10. Language Resource References

Baker, Collin F. and Fillmore, Charles J. and Lowe, John B. 1998. *FrameNet*. Berkeley International Computer Science Institute, version 1.7. PID https://framenet.icsi.berkeley.edu/.

Broadening the coverage of computational representations of metaphor through Dynamic Metaphor Theory

Xiaojuan Tan♣ and Jelke Bloem♢♠

♣ Department of Language, Literature and Communication, Vrije Universiteit Amsterdam
♢ Institute for Logic, Language and Computation, University of Amsterdam
♠ Data Science Centre, University of Amsterdam
tanxiaojuan01@gmail.com, j.bloem@uva.nl

Abstract

Current approaches to computational metaphor processing typically incorporate static representations of metaphor. We aim to show that this limits the coverage of such systems. We take insights from dynamic metaphor theory and discuss how existing computational models of metaphor might benefit from representing the dynamics of metaphor when applied to the analysis of conflicting discourse. We propose that a frame-based approach to metaphor representation based on the model of YinYang Dynamics of Metaphoricity (YYDM) would pave the way to more comprehensive modeling of metaphor. In particular, the metaphoricity cues of the YYDM model could be used to address the task of dynamic metaphor identification. Frame-based modeling of dynamic metaphor would facilitate the computational analysis of perspectives in conflicting discourse, with potential applications in analyzing political discourse.

Keywords: Dynamic metaphor theory, conflicting discourse, metaphoricity, extended metaphor

1. Introduction

The ubiquity and power of metaphor in discourse have served as an impetus for computational linguists to develop ways to automatically identify and process metaphors. Although various computational approaches to metaphor have been developed, the problem is far from solved partially due to incomplete coverage of the nature and mechanism of metaphor. Computational work typically focuses on representing the surface realizations of metaphor (called 'linguistic metaphor' in the typology of Shutova, 2015) or the metaphorical mappings underlying them as an inventory of mappings (called 'conceptual metaphor'), but these representations are static.

Conversely, cognitive linguists have made breakthroughs in developing dynamic metaphor models, but these models are only applied in manual semantic and/or pragmatic analysis. This article aims to bring together these two lines of work, by connecting recent developments in metaphor theory to recent computational approaches. In particular, we take insights from dynamic metaphor theory and discuss how existing computational models of metaphor might benefit from representing the dynamics of metaphor, particularly in the case of conflicting discourse. With this theoretical contribution, we aim to outline a theoretically informed path towards computational representations of metaphors that go beyond static metaphors and to introduce cognitive linguists to the possibility of the computational modeling of dynamic metaphors.

2. Metaphors are Dynamic

The most well-known metaphor theory is Lakoff and Johnson's (1980) Conceptual Metaphor Theory (CMT). According to CMT, the essence of metaphor is understanding and experiencing one kind of thing in terms of another. Accordingly, metaphor structures a cross-domain mapping of thought, from a relatively concrete target domain to an abstract source domain. As an example, in everyday life we often come across expressions that reason the target domain of 'love' in terms of the source domain of JOURNEY:

(1) We're *at* a *crossroads*.
We can't *move forward*.
I don't think our relationship is *going anywhere*.

The italicized metaphors here are called linguistic metaphors, while the mapping between the source domain and target domain (LOVE IS JOURNEY) is termed a conceptual metaphor.

Dynamic metaphor theory is a key recent development in the field of metaphor studies. A commonly held assumption in many earlier and contemporary metaphor theories is that metaphor is static. From the static viewpoint, linguistic metaphors are either dead (conventional) or alive (novel). For instance, Black (1979) regards conventional metaphorical expressions as dead and only novel metaphorical expressions as alive. In *Metaphors We Live By*, Lakoff and Johnson (1980) imply that the category of 'live' metaphor is much larger than generally assumed and should encompass the conventional metaphors of ordinary lan-

First Workshop on Reference, Framing, and Perspective, pages 40–50

guage. However, their perspective on metaphors is still static in that metaphors are restricted to a fixed cognitive structure of thought. Lakoff (1993, p. 210) characterizes the mapping of two frames as universally "fixed patterns of ontological correspondences" between two conceptual domains. Such a static view of metaphor has been attacked by discourse analysts because Lakoffian works take conceptual metaphors as highly stabilized conceptual mappings across speech communities. Rejecting the static view, Müller (2008) was the first to argue that the property of metaphor has the potential for activation and thus metaphor is dynamic. She claims that in certain discourse contexts, the source domains and conventionalized linguistic metaphors may be active for a given speaker at a given moment in time. This argument can be illustrated via the following two examples.

(2) We have to do these things to make America great again. Because we can't *lose* almost $800 billion on the start of the trade dispute, like has been done for many years. (Donald Trump's speech, 24/01/2019; italicization added)

(3) We were *losing* all our cases in the World Trade Organization. Almost every case, we were *lost*, *lost*, *lost*. (Donald Trump's speech, 13/08/2019; italicization added)

In example 2, the conventional linguistic metaphor 'lose' indicates a possible mapping between the source domain of Competition and the target domain of 'trade dispute'. The metaphoricity is static since the metaphor occurs only once without any other semantic elaboration. However, in example 3, 'lose' becomes a more salient linguistic metaphor and Competition becomes a more salient source domain in Trump's trade speech, through a strategy that foregrounds metaphor use – the repetition of the lexeme 'lose' in different verb tenses. This is what we call a metaphoricity cue. Through this cue, the metaphoricity of 'lose' is activated by former president Trump and/or the speech writers at that moment. Therefore, 'lose' is not static but dynamic, as it is no longer strictly restricted to the rigid category of *conventional* linguistic metaphor. Its metaphoricity achieves a higher degree of activation in example 3 than in example 2.

Challenging the static perspective of metaphor which has been taken for granted for decades, pivotal dynamic metaphor scholars have put forth different usage-based models (e.g., Cameron, 2010; Jensen and Cuffari, 2014; Müller, 2008). However, these dynamic perspectives are limited due to their focus on either the change of linguistic metaphor or the change of source domain. For instance, Cameron (2010) focuses on patterns of development of metaphorical expressions, while Kyratzis (1997) focuses on the chains of the source domains. Until recently, there has been little attention paid to the mechanisms by which changes in both source domains and/or target domains can activate metaphoricity.

The recently proposed YinYang Dynamics of Metaphoricity (YYDM, Tan, 2023; Tan and Cienki, 2024) addresses this theoretical gap. This usage-based model, created for metaphor analysis in texts, emphasizes how change within and between source and/or target domains can activate metaphoricity. As this model defines dynamicity in terms of explicit metaphoricity cues that are textually expressed, we consider YYDM a promising theoretical framework for the computational modeling of dynamic metaphors. It puts forward that metaphor develops with the emotions and attitudes of discourse participants, which outlines a way to empirically reconstruct the inner mechanism of dynamic metaphors, the motivation behind their use, and therefore their effect on society.

2.1. YinYang Dynamics of Metaphoricity

The model of YinYang Dynamics of Metaphoricity (YYDM) assumes that metaphorical expressions range from Yin-inactive metaphors to Yang-active metaphors; there is no strict boundary between them. Yin-inactive metaphors have a low degree of metaphoricity because they are not surrounded by any metaphoricity cues (cf. example 2). On the contrary, Yang-active metaphors have a high degree of metaphoricity because they are surrounded by metaphoricity cues (cf. repetition in example 3). The same metaphorical expression can be inactive in one context and active in another. The degree of activation of metaphoricity can be documented through Tan and Cienki's (2024) metaphoricity cues. In this section, we show a limited sample of different types of metaphoricity cues through examples from conflicting discourse.[1]

2.1.1. Cues highlighting the same source domain

Clustering of metaphorical expressions in the source domain This is exemplified by the clustered metaphorical expressions ('capitulation', 'submission', and 'retreat') that all highlight the War source domain in the following example: "On the question of foreign trade, previous leaders were guided by a shameful policy of capitulation, submission, and retreat."

[1]Examples taken from the Trump subcorpus and Xi subcorpus of Tan and Cienki's (2023) corpus on US-China trade conflict.

Explicit mapping Presenting the source domain explicitly. Consider the explicit mapping of CATASTROPHE (WTO IS CATASTROPHE): "World Trade Organization is a catastrophe."

Marking devices are used (Goatly, 1997, pp. 262-263; Cameron and Deignan, 2003) to mark the source domain, e.g., 'sort of', 'like', 'kind of', 'really', 'imagine', 'so to speak', 'actually', 'literally', 'if you like', 'in a way', 'as it were'. For instance, using "as" to mark the source domain LEVERAGE: "...the previous administration was unwilling to use our huge trade deficit with China as leverage..."

Repetition of the same linguistic metaphor with the same source domain. In this example, the cue is repeating 'stole' within the source domain of CRIME: "...other countries stole our factories, stole our plants, stole our wealth, and stole our jobs."

2.1.2. Cues indicating the change of source domain, but non-change of target domain

Diversification Using different source domains to refer to the same target domain is a metaphoricity cue. Consider the diversified source domains (POISON; GOOD PRESCRIPTION) in the following example, i.e., "Trade protectionism is a poison rather than a good prescription."

Novelization Using novel linguistic metaphors to refer to the new source domain is a metaphoricity cue. Considering the novel linguistic metaphor 'top student' to refer to the new source domain (TOP STUDENTS): "China is a top student among the members of the World Trade Organization."

2.1.3. Cue indicating the change of target domain, but non-change of source domain

Multivalency Using the same source domain to refer to different target domains exemplifies this metaphoricity cue. Consider the repeated source domain PILLAR and different target domains 'trade policy' and 'trade regulation' in the following example, i.e., "In addition to trade policy, trade regulation is also the pillar of our economic development"

2.1.4. Cue indicating the change of both source domain and target domain

Mixing different source domains mapped to different target domains illustrates this cue. Consider the mixed mappings (CHINA IS A TOP STUDENT; WTO IS CONTAINER) in the following example, i.e., "China has been a top student since its entry into the World Trade Organization".

2.2. Exemplification of YYDM through data on conflicting discourse

Generally, the more semantic representation of a source domain is present in a discourse, the more the source domain is foregrounded and a higher degree of activated metaphoricity is achieved. The more metaphoricity cues that point to a linguistic metaphor, the more the linguistic metaphor is highlighted and the higher the degree of activation (cf. Tan, 2023). The following examples of 'war' metaphors from discourse on the recent U.S.-China trade war will illustrate more clearly how the dynamic model functions. These examples are selected because they use commonly studied frames that also exist in frame repositories such as FrameNet (Baker et al., 1998) and these frames can show how the divergent opinions of the Chinese and American governments evolve in the process of trade negotiation.

(4) Q Talking about a trade *war*? PRESIDENT TRUMP: I don't think you'll have a trade *war*. Q No trade *war*? PRESIDENT TRUMP: I don't think so. I don't think you're going to have trade *war*, no. (Remarks, 05/03/2018; italicization added)

(5) Q On the tariffs, the President tweeted that trade *wars* are good, easy to *win*. Can you explain what he meant by that? MS. SANDERS: Look, the President, I think, is very confident that if that's where we ended up, we certainly would *win*. But that's not the *goal*. The *goal* is to get free, fair, and reciprocal trade, and hope that other countries will join in. (Press Briefing of Press Secretary, 05/03/2018; italicization added)

From example 4 to example 5, the Yin-inactive metaphor 'war', framing the target domain of 'trade negotiation', became a Yang-active metaphor on March 5th, 2018 through different metaphoricity cues. In example 4, the *repetition* of 'war' activated both the metaphorical expression 'war' and the source domain of 'WAR'. It shows former president Trump's position in the morning that he could threaten China to make concessions in trade negotiations without launching a trade war. However, in example 5, 'war' and the source domain of WAR are further highlighted in a press briefing.

The journalist first activates 'war' and WAR through *a cluster of* WAR metaphors ('war', 'win'). Then the Press Secretary foregrounds them further through the *repetition* of 'win', *the change of the source domain* (the change from WAR to JOURNEY via the change from 'war' to 'goal'), and *the change of both the source domain and target domain* (from TRADE IS JOURNEY to GET FREE, FAIR, AND RECIPROCAL TRADE IS GOAL). With the activation of the metaphoricity, Trump's administration changed their attitude and triggered a nationalist sentiment of winning the war, i.e., from the non-necessity of a trade war to the determination to get free trade through a trade war. This kind of trade discourse from the Trump administration was attacked by the Chinese government, which can be shown through the activation of Chinese dynamic metaphors below.

(6) 历史已经证明，贸易战没有赢家，中国不愿意打贸易战。'The history has proved that there is no *winner* in the trade *war*. China is not willing to *fight* a trade *war*' (Reports of Leaders' Activities, 06/03/2018; italicization added)

(7) Q 中方目前的态度比较克制，但并不代表没有好牌... 贸易战会发展到什么程度，要看美国走到哪一步。中国要反击这场贸易战的 "牌" 有不少，从大豆到汽车、飞机，可以打出组合拳来回击，这些商品的可替代性都比较强。对于美方挑起的贸易战，我们完全有底气采取强有力措施精准还击。'China's current attitude is relatively restrained, but it does not mean that there is no *good card*...the extent of the trade *war* depends on the procedure taken by the U.S. China has many *"cards"* to *fight back* against this trade *war*. From soybean to car and airplane, it can *hit back* with a *combination combo*. These commodities are highly replaceable. For the trade *war* provoked/*shouldered* by the U.S., We have the confidence to take strong measures to *fight back accurately*.' (China Daily, 26/03/2018; italicization added)

From examples 6 to 7, the source domain WAR becomes more and more salient, and 战 'war' changes from a Yin-inactive metaphor to a Yang-active metaphor within a month. In example 6, the repetition of 战 'war' and the clustered metaphors of WAR (打 'fight', 战 'war', 赢家 'winner') activate 战 'war' within the same source domain of WAR. Built on the activation of metaphor, the Chinese government conveyed its stance that China was unwilling to go to war on March 6th, 2018, which replies to America's decision to launch a trade war on March 5th, 2018. On March 26th, the Chinese attitude evolved to counterattack, which aroused a nationalist sentiment of protecting China through a trade war. This was shown by a higher activation of 战 'war' and WAR through many metaphoricity cues across sentences in example 7.

At the beginning of example 7, China is framed as a card player having a set of good cards in the CARDS game and then is *reframed* as a defender fighting back the U.S. aggression with a series of WAR weapons. With the *change of source domain* (from CARDS to WAR), the cards of soybeans, cars, and airplanes are *reframed* as weapons. As the news report continues, the *collocated* metaphors (打出 'hit' and 组合拳 'combination combo') introduce a *new* source domain for trade (BOXING COMBO). That is, the actions of playing cards are reconstructed as blows in a BOXING COMBO, which reconstructs China as a boxer hitting back the U.S. through the combo, and foregrounds WAR and 战 'war' even further through another *reframing*. In the next sentences, WAR and 战 'war' are even more foregrounded through *aggregated* metaphoricity cues. Namely, a *new* reframing reconstructs trade war as a PHYSICAL OBJECT through 挑起 'shoulder'. With the *change of source domain and target domain*, China attributes the guilt of starting the trade war to the U.S. The following *clustered metaphors* of WAR (e.g., 反击/反制 'counterattack', 精准还击 'fight back accurately') as well as the *repetition* of 战 'war' and 精准 'accurate(ly)', portray China's strong skills in counterattacking and its confidence in winning the trade war.

Applying the dynamic model (YYDM) to authentic data gathered from discourse on trade conflicts, these examples reveal that metaphors can be activated and become dynamic through additional semantic representations of the source domain, and additional changes of the source domain and/or target domains. Dynamic metaphors can connect various thoughts and participants over stretches of texts and even an entire large-scale corpus. With the development of metaphors, the sentiments and attitudes on the China-U.S. trade war also changed. This shows discourse is not a matter of detached meaning construction but instead a dynamic system intertwined with intersected levels (e.g., linguistic, conceptual, sociopolitical) which needs to be understood as processes, flows, or movements (Larsen-Freeman and Cameron, 2008). Since the dynamics in the micro level of language use function all the way up to the discourse dynamics at the social group level (Tan et al., 2024), automatic identification of metaphoricity cues at the micro level can lead to the prediction of different changes driving the production and reproduction of conflicting discourses.

3. Computational representations

Next, we examine the extent to which the dynamics of metaphor might be represented in the field of NLP. In recent years, several excellent surveys on the state of computational metaphor processing have been written (Rai and Chakraverty, 2020; Tong et al., 2021; Ge et al., 2023), which we will not reiterate here. We instead aim to survey computational work that constructs detailed or extended computational representations of metaphor, which may cover aspects of the dynamics of metaphor. Broadly, computational work on metaphor has centered around two tasks: automated metaphor identification and automated interpretation. Typically, identification is operationalized as a sequence labeling task, and interpretation is operationalized as a paraphrasing task.

For many years, computational metaphor processing relied mainly on hand-crafted resources such as MetaNet (Dodge et al., 2015). MetaNet is a multilingual repository of conceptual metaphors that is linked to FrameNet (Baker et al., 1998), enabling computational representation of conceptual metaphors in terms of source and target domains, and theoretically grounding those domains in terms of frame semantics (Fillmore, 1976). For metaphor identification, metaphoric expressions can be linked to conceptual metaphors as listed in MetaNet. However, such approaches have limited coverage, with few possibilities to generalize beyond the hand-crafted metaphor inventory.

More recent approaches rely on the use of dense vector representations as features for predicting metaphoricity labels. Typically, the data and labeling from the VU Amsterdam (VUA) Metaphor Corpus (Steen et al., 2010) are used. This corpus contains token-level binary annotation indicating metaphoric or non-metaphoric use. These tools identify a wider range of metaphoric expressions than those relying on metaphor repositories, but lack explanatory power and theoretical grounding in metaphor theory. Such tools do not tell us why an expression is metaphoric, e.g. by performing conceptual mapping, identifying it as an instance of a particular conceptual metaphor with a particular source and target domain. An example of this approach is Gong et al.'s (2020) RoBERTA-based system.

A particularly accessible example of this approach is Mao et al.'s (2023) MetaPro 2.0, an end-to-end metaphor processing system incorporating the tasks of identification and interpretation with state of the art performance on standard benchmarks. The identification module is trained on the VUA corpus, and the interpretation paraphrasing is done by having RoBERTa mask a metaphorically used word and predict a synonym or hypernym of that word in its place. The approach is limited to substituting metaphorically used words with a synonym or hypernym that fits the context literally, excluding more creative metaphoric uses. At the time of writing, this system is available as a functioning online demo[2]. It is thus a good way for cognitive linguists to assess the state of the art in computational metaphor processing.

[2]https://metapro.ruimao.tech/

3.1. Conceptual mapping

A few computational studies do address the conceptual mapping task in addition to metaphor identification. Firstly, the aforementioned MetaNet (Dodge et al., 2015) was used to perform conceptual mapping by identifying candidate metaphoric expressions through grammatical patterns and then matching the words in the source domain slot and target domain slot to frames using MetaNet, FrameNet, Wordnet or Wiktionary. If those frames have metaphoric mappings in the hand-coded MetaNet repository, it is identified as metaphoric. This formalizes connections between different instances of a conceptual metaphor but does not generalize to novel mappings. Although this is not discussed in the paper, due to the use of a standardized resource, different instances of the same conceptual metaphor occurring throughout various discourses can easily be connected using this approach.

Shutova et al. (2017, p. 79) emphasize the importance of conceptual metaphoric mappings, stating that "one needs to address conceptual properties of metaphor, along with the surface ones". They use semi-supervised clustering to create source and target domains based on seed expressions (e.g. "grasp theory", "ideology embraces"). These expressions are used to learn how to map these domains, allowing the detection of mappings between other expressions within these domains that are not in the seed set. This is extended to an unsupervised method where the same clusters are automatically organized hierarchically by inferring connections at a hypernym level which can represent the kind of broad conceptual mappings that metaphors consist of. However, the mappings are static and limited to ones based on two-word verb-subject and verb-object relations without context. Domains also remain unlabeled and not linked to a word sense, frame or metaphor repository.

Ge et al. (2022) take conceptual mapping one step further by explicitly defining concepts as WordNet hypernyms, with the goal of increasing the explainability of metaphor identification methods. They use an algorithm to determine the level of hypernymy in the WordNet hierarchy that sufficiently covers most senses of a noun without being too abstract. This approach is also incorporated in the aforementioned MetaPro 2.0 (Mao et al., 2023) system to map to the source domain. For a given metaphor, paraphrases from MetaPro's interpretation module are regarded as labels of the metaphor's target domain. The concept resulting from the application of Ge et al.'s (2022) algorithm is regarded as the metaphor's source domain, yielding a mapping. However, the analysis is limited to decontextualized pairs of dependent words (verb-noun or adjective-noun), and the mapping to WordNet only exists for the source domain, not the target domain.

Wachowiak and Gromann (2023) predict source domains from GPT-3 given a sentence and a target domain in a one-shot text completion task. This form of metaphor mapping is fairly flexible by not being connected to pre-defined domains and by drawing upon a huge amount of training data. However, it presupposes that a specific linguistic metaphor statically maps to a source domain regardless of discourse context and the approach does not consider other aspects of dynamic metaphor theory such as the amount of activation of the metaphor. It may also yield non-standard source domain descriptions that do not map to metaphor inventories.

3.2. Extended metaphors

While no computational work directly addresses the dynamicity of metaphor, the related concept of extended metaphor is discussed by a few authors. However, these works diverge on what extended metaphors are. The aforementioned surveys of computational metaphor illustrate that "broader tasks of identifying conceptual metaphors, extended metaphors, and metaphoric framing, have been largely ignored" (Tong et al., 2021, p. 4679) and "identifying other types of metaphors, such as extended metaphors or MWEs, has yet to be well solved" (Ge et al., 2023, p. 1857).

Klebanov and Beigman (2010) probably comes closest to describing the dynamics of metaphor, discussing the case of bargaining in political communication. They describe an extended metaphor of the European Union as a train. This metaphor received various politically charged positive and negative extensions over time and was the subject of counter-metaphors in European political discourse. The authors note the extensive attempts at bargaining over the same metaphor rather than re-framing the discussion with a novel metaphor, and aim to explain this bargaining with a game-theoretical model. Klebanov and Beigman's (2010) model represents extended metaphor as an abstract set of frames that can be negotiated about, and this is the representation of extended metaphor used throughout the work. However, the discussion proceeds in terms of game-theoretic negotiation about these frames rather than in terms of cognitive metaphor theory – there is no representation of source and target domains. The work does illustrate the importance of acknowledging diverging perspectives on a metaphor in the domain of political communication and the need for a model that can represent this. The authors note the difficulty of automatic detection of extended metaphor due to a lack of sufficient training data for particular metaphors.

Subsequently, Shutova (2015, p. 585) states that "a computational method for identification and interpretation of extended metaphor in real-world discourse is yet to be proposed. A discourse-level metaphor processing system would need to identify a chain of metaphorical expressions in a text, which indicates a systematic association of the text topic with a particular domain. These chains would then demonstrate how continuous scenarios can be transferred across domains". This line of work regards extended metaphors as a group of linguistic metaphors elaborating on the same conceptual metaphor, which is close to the type of dynamic metaphor in section 2.1.1. However, they do not encompass other types of dynamic metaphors in which there is a change of frame (sections 2.1.2-2.1.4). Dynamic metaphor theory can provide a framework for operationalizing this.

Dankers et al. (2020) address another aspect of dynamic metaphor by emphasizing the importance of discourse context in the identification task. However, the task itself remains a binary identification of linguistic metaphor, just with a larger context window for the sequence prediction task. The analysis does not show that the model recognizes elements such as clusters of linguistic metaphors that refer to a particular conceptual metaphor, and it would not explicitly identify conceptual metaphors anyway as there is no conceptual mapping. The main improvements using this method come from having more information on the topic of the text and implicit coreference resolution.

We are aware of only one study that computationally models extended metaphor occurring across sentences. Jang (2017) argues that metaphor should be defined in terms of frames in order to be able to process metaphor at the discourse level. Metaphor is then defined as a phenomenon that occurs when "a speaker brings one frame into a context/situation governed by another, and explicitly relates parts of each". Extended metaphor is described as metaphors that can be around related metaphors. Specifically, in a dissertation chapter not published as a paper, Jang (2017, Ch. 8) applies template induction to find frame elements in a connected discourse, using this frame information for metaphor detection. In this context, the concept of extended metaphor is defined as a switch of source frames in discourse (corresponding to section 2.1.2 in our theoretical description): "metaphor performs social functions through the switching of frames" Jang (2017, p. 79). Other types of dynamic metaphor (sections 2.1.3-2.1.4) are not addressed, and the concept of dynamic metaphor is not mentioned.

In this work, the frame elements are extracted from lexico-grammatical patterns in a seeded but unsupervised way. The template induction then identifies more frame elements for the target frame, which may occur in the vicinity of the candidate linguistic metaphor. The final task is once again metaphor identification, thus the identified frame elements are only used as features to solve this task, and explicit representation of clusters of metaphoric expressions with the same source frame (as in Section 2.1.1) is not demonstrated or evaluated directly. Although the notions of frames and frame elements ('frame facets') and their linguistic instances ('slot instances') are used, they are not formalized. The frames are not connected to repositories such as FrameNet and it is not clear where the frame seed words come from.

4. Dynamic metaphors and frames

After comparing state-of-the-art theory on dynamic metaphor with the state of the art in computational metaphor processing, we identified several gaps that would need to be addressed in order to computationally represent the dynamics of metaphor.

4.1. Mapping clustered and repeated references to a metaphor

In the computational literature, instances of metaphor are largely viewed in isolation, and when they are not, surrounding instances within a discourse are mainly viewed as features to aid in the identification of a targeted instance. In the new dynamic model (YYDM), linguistic metaphors are connected and may partially instantiate different elements of the metaphor's source or target domain. These connected linguistic metaphors indicate the dynamic activation of the metaphor. A computational representation of dynamic metaphor should be able to formalize these connections, for example, by linking them to a common metaphoric mapping as in MetaNet. At the same time, it should represent the fact that some linguistic metaphors serve as cues marking the activation of a certain linguistic/conceptual metaphor, which is similar to the frame elements of Jang (2017). This should be possible even when the conceptual metaphor changes due to changes in source and/or target domains (cf. section 2.1.2- 2.1.4), unlike in static MetaNet representations.

4.2. Representing changing domains

An important aspect of dynamic metaphors is that a metaphor's source domain, target domain or both can change between instances. Firstly, this requires a computational model of metaphor to have explicit representations of a source and target domain, which requires conceptual mapping. This was only performed by a few studies until now, as we saw in section 3.1. These studies either map to unsupervised clusters, WordNet hypernym levels, or FrameNet frames.

Secondly, mappings should either be able to have different source and target domains, or should be a set of related mappings that map different source and target domains as part of a dynamic metaphor. This could be operationalized through something like FrameNet's inheritance relation or MetaNet's related metaphor property. Mappings can change even across longer spans of discourse, as a metaphor can be used dynamically throughout a political discourse across time. Therefore, it seems useful to have standardized frame and metaphor identifiers that instances of a metaphor can be linked to (e.g. Klebanov and Beigman's (2010) EU train metaphor).

Next, these frame changes are often used in conflicting discourse to represent or emphasize divergent perspectives as in our example 7 on the US-China trade war. This could be represented using something like FrameNet's perspectivization relation, where frames describing the same situation from different perspectives are linked. Metaphoric mappings might also be considered perspectivized in this way.

Lastly, in the YYDM model, metaphors have a degree of activation based on the amount of metaphoricity cues in their context. A computational model of dynamic metaphor would also have to account for this. This possibility is already addressed to some extent by Jang (2017) – the frame elements they detect in the vicinity of the frame involved in the metaphoric mapping can be considered as metaphoricity cues. A computational representation of a metaphor that can be active or inactive following the YYDM model could include the number of metaphoricity cues found to indicate whether it is a Yin-inactive metaphor or a Yang-active metaphor.

4.3. Choosing a representation

We propose that frame-based approaches to metaphor representation are the best choice for modelling dynamic metaphors computationally. These approaches can incorporate the necessary elements from the YYDM model in order to represent detailed aspects of metaphor that may change dynamically throughout a discourse (e.g., different frame-evoking words denoting different frame elements). Frame-based approaches can enable their alignment between source and target domains to facilitate analysis. They can also incorporate perspectivization, which is important if computational metaphor processing is to be applied to the study of conflicting discourse.

Existing approaches that get closer to representing dynamic aspects of metaphor, such as the conceptual mapping of MetaNet or the frame elements of Jang (2017), already draw ideas from Frame Semantics (Fillmore, 1985). Jang (2017, p. 92) also argues that "modeling metaphor through the lens of frame theory could be the first step in detecting extended metaphor". Frame-based approaches are well grounded in theory, particularly in cognitive linguistics, which is a framework that aligns well with the current data-driven and distributional paradigm in the field of natural language processing (Levshina and Heylen, 2014; van Trijp, 2017; Rambelli et al., 2019). Frame-based representations could also handle multi-word units, a weakness of most current computational approaches.

In previous work, frame-based approaches such as MetaNet have shown a lack of scalability due to their dependence on hand-crafted linguistic resources. However, with the recent increase of interest in grounding elements of large language models in linguistic theory, we are starting to see efforts to induce even complex linguistic representations such as frames from data. Yamada et al. (2021) perform semantic frame induction from contextual word embeddings, paving the way for automatic frame construction. They show that clusters of contextualized word representations can be used to distinguish the difference between multiple frames invoked by the same verb. Furthermore, recent work has shown that evidence for Construction Grammar constructions, another branch of linguistic theory related to cognitive linguistics, can be found in transformer-based sentence embedding models. Li et al. (2022) observed that argument structure constructions get clustered by their construction type (e.g. ditransitive, caused-motion) rather than by their verb, and Veenboer and Bloem (2023) note that constructions in embedding space are surrounded by nearest neighbors with similar constructional semantics, also generalizing to instances containing verbs not seen in example constructions.

Therefore, it may be possible to create frame-based representations of dynamic metaphors in an unsupervised way in the future, especially given some seed set of frames, metaphoric mappings, frame elements or annotated metaphoricity cues.

5. Towards identification

Besides the representation of dynamic metaphors, the YYDM model can also be operationalized to aid in the popular task of metaphor identification. We propose that the metaphoricity cues discussed in section 2.1 can be used not only to characterize, but also to identify dynamic metaphor use.

Specifically, cues such as **repetition** and **clustering** of metaphoric expressions (section 2.1.1) could be detected by searching for multiple frame elements in the context of an expression, as done by Jang (2017, Ch. 8). **Marking devices** indicating metaphoricity such as 'so to speak' could be used directly as identification features. This was backed with empirical corpus data by Cameron and Deignan (2003) ('tuning devices'), but surprisingly this work has never been cited in the field of NLP. **Explicit mapping** can also be detected if a model is able to perform conceptual mapping or draws upon a repository of metaphoric mappings - it would only require a small step to match the conceptual metaphor of CATASTROPHE (WTO IS CATASTROPHE) to the lexemes in "World Trade Organization is a catastrophe.", especially when a frame-semantic parser can be used.

The cues indicating frame changes are more abstract, as they require frame representations that may be beyond the capabilities of current frame-semantic parsers. We tried our example of the **diversification** cue, "Trade protectionism is a poison rather than a good prescription" with the Frame Semantic Transformer parser (Chanin, 2023). It does detect both source frames: the Toxic_substance frame, with *poison* filling the TOXIC_SUBSTANCE FE, and Usefulness frame, with *prescription* filling the ENTITY FE. However, a frame related to 'trade protectionism' is not detected, as this is a concept rather specific to the domain of trade war. As for existing metaphor representations, TRADE PROTECTIONISM IS POISON can be categorized as a subcase of the NEGATIVELY EVALUATED CONDITIONS ARE HARMFUL AGENTS mapping in MetaNet. The latter does exist, but the former is not in the repository, and neither is something corresponding to the domain-specific TRADE PROTECTIONISM IS PRESCRIPTION. However, missing frames could be induced (Yamada et al., 2021), substituted by domains defined as WordNet hypernyms or clusters of related words, as was done in work discussed in section 3.1.

With such frame-based representations of metaphor, metaphoricity cues indicating frame changes can be detected in theory. **Diversification** could be detected by identifying the various source frames in the discourse, and checking whether they map to the same target domain in an available inventory of metaphoric mappings. Detecting **multivalency** would be the inverse of this, identifying target domain frames instead. **Novelization** is more difficult to detect as novel metaphors would not be in an inventory of conventional metaphors. The novelization cue might be found by detecting linguistic patterns that look like metaphors (e.g. from induced templates as in Jang, 2017), where the target domain has been used before in the discourse and the source domain is unknown in the local context of that target domain. The **mixing** of source and target domains appears to be a difficult cue to detect, but if we take metaphor to be dynamic over a larger discourse, we could find them by restricting the search space to only metaphors that have already been used in the discourse. Instances of mixing of the source and target domains of these previously used metaphors can then be found by detecting novel combinations of the source and target domain frames within a limited context window.

Such work on cue-based identification of dynamic metaphor or induction of missing MetaNet mappings may be aided by an annotated metaphor corpus that includes annotation of metaphoricity cues. Tan and Cienki (2023) annotated a 6M word corpus of texts relating to US-China trade conflicts with detailed features of the YYDM model, including metaphoricity cues. This labeled data could be used to train a classifier that can use the metaphoricity cues as features for the task of dynamic metaphor identification. It could also be used to extend the computational task of static conceptual mapping to the task of dynamic mapping, where multiple different but related metaphoric mappings may exist within the same discourse.

6. Discussion

We have sketched a proposal for more comprehensive computational representations of metaphor. Using the model of YinYang Dynamics of Metaphoricity as a theoretical framework, we demonstrated that metaphors are dynamic rather than static. Next, we surveyed the state of the art in computational metaphor representation. We found that, although dynamic metaphor theory was never explicitly addressed computationally, some of its ingredients, such as conceptual mappings, are represented. We then proposed ways to incorporate the main elements from the YYDM model into computational representations of metaphor using frame representations. Lastly, we discussed how the metaphoricity cues of the YYDM model could be used to address the task of dynamic metaphor identification. Overall, we hope to have shown that approaches based on Frame Semantics, such as FrameNet, with the addition of an inventory of metaphoric mappings, such as MetaNet, provide the necessary ingredients for computational representation of dynamic metaphor. The main weakness of this approach is limited coverage, but combining recent work on frame induction with an annotated corpus of dynamic metaphor may help to address this.

The importance of representing metaphor dynamically in the computational domain lies in the increasing importance of representing different perspectives on events and issues. This is true in NLP where the real-world application of large language models has shown that aggregating all data points into a single distribution or ground truth label erases minority perspectives (Cabitza et al., 2023). In political discourse analysis, metaphor researchers holding the static view fail to demonstrate how metaphors can develop together with political perspectives. Political discourse is a dynamic system where metaphors developing at the micro-level of language use are dynamically intertwined with the hidden political interests and power at the macro-level of discourse context which influences the evolution of political perspectives.

Dynamic metaphors evolve in discourse over time and can be sustained over many years. Having computational representations thereof would open up the possibility of performing diachronic metaphor analysis by comparing diachronic representations. Research on lexical semantic change using diachronic word embeddings has been quite successful (Tahmasebi et al., 2021), but similar approaches have not been developed for metaphor.

Representations of dynamic metaphors may also have benefits for downstream NLP tasks. Metaphoric expressions in conflicting discourse are often used to express polarized sentiment, and detecting this could contribute to better sentiment analysis. When metaphors are explicitly resisted (van Poppel and Pilgram, 2023), they may carry negative sentiments and conflicting perspectives.

Event detection is another possible application area – dynamic use of metaphor can involve many mentions of a particular event from various perspectives, each adding more information about the event. Our analysis of example 4 and 5 shows that the former American president and the Press Secretary make multiple metaphoric references to a trade war in two statements on the same day. This points toward the possibility of detecting the evolution of big political conflicts, which complements the existing computational techniques that focus on the detection of detached events.

In a nutshell, this study aims to bring cognitive linguists and computational linguists together, by showing recent developments in metaphor theory as well as a path towards computational application. Contrary to the static view of cognitive linguists and computational linguists, this paper argues that the cognitive dimension (frames), affective dimension (sentiments), and social-political dimension (perspectives) are constantly interacting. This inherent variability of the discourse system has implications for experts from both fields. Future computational operationalizations of this new dynamic model applied to different datasets could have impactful applications in analyzing political discourse in general and in analyzing conflicting discourse in particular.

7. Bibliographical References

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The Berkeley FrameNet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Max Black. 1979. More about metaphor. In A. Ortony, editor, *Metaphor and thought*, pages 19–43. Cambridge University Press.

Federico Cabitza, Andrea Campagner, and Valerio Basile. 2023. Toward a perspectivist turn in ground truthing for predictive computing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 6860–6868.

Lynne Cameron. 2010. The discourse dynamics framework for metaphor. In R. Maslen L. Cameron, editor, *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities*, pages 77–94. Equinox London.

Lynne Cameron and Alice Deignan. 2003. Combining large and small corpora to investigate tuning devices around metaphor in spoken discourse. *Metaphor and symbol*, 18(3):149–160.

David Chanin. 2023. Open-source frame semantic parsing. *arXiv preprint arXiv:2303.12788*.

Verna Dankers, Karan Malhotra, Gaurav Kudva, Volodymyr Medentsiy, and Ekaterina Shutova. 2020. Being neighbourly: Neural metaphor identification in discourse. In *Proceedings of the second workshop on figurative language processing*, pages 227–234.

Ellen K Dodge, Jisup Hong, and Elise Stickles. 2015. Metanet: Deep semantic automatic metaphor analysis. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 40–49.

Charles Fillmore. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32.

Charles Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di semantica*, 6:222–254.

Mengshi Ge, Rui Mao, and Erik Cambria. 2022. Explainable metaphor identification inspired by conceptual metaphor theory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10681–10689.

Mengshi Ge, Rui Mao, and Erik Cambria. 2023. A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artificial Intelligence Review*, 56(Suppl 2):1829–1895.

Andrew Goatly. 1997. *The language of metaphors*. Routledge.

Hongyu Gong, Kshitij Gupta, Akriti Jain, and Suma Bhat. 2020. Illinimet: Illinois system for metaphor detection with contextual and linguistic information. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 146–153.

Hyeju Jang. 2017. *Computational Modeling of Metaphor in Discourse*. Ph.D. thesis, Carnegie Mellon University.

Thomas Wiben Jensen and Elena Cuffari. 2014. Doubleness in experience: toward a distributed enactive approach to metaphoricity. *Metaphor and Symbol*, 29(4):278–297.

Beata Beigman Klebanov and Eyal Beigman. 2010. A game-theoretic model of metaphorical bargaining. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics*, pages 698–709.

Sakis Kyratzis. 1997. *Metaphorically speaking, sex, politics, and the Greeks*. Ph.D. thesis, Lancaster University.

George Lakoff. 1993. The contemporary theory of metaphor. In A. Ortony, editor, *Metaphor and thought*, page 202–251. Cambridge University Press.

George Lakoff and Mark Johnson. 1980. *Metaphors we live by*. University of Chicago Press.

Diane Larsen-Freeman and Lynne Cameron. 2008. *Complex systems and applied linguistics*. Oxford University Press.

Natalia Levshina and Kris Heylen. 2014. A radically data-driven construction grammar: Experiments with Dutch causative constructions. *Extending the scope of Construction Grammar*, 54:17.

Bai Li, Zining Zhu, Guillaume Thomas, Frank Rudzicz, and Yang Xu. 2022. Neural reality of argument structure constructions. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7410–7423.

Rui Mao, Xiao Li, Kai He, Mengshi Ge, and Erik Cambria. 2023. MetaPro Online: A computational metaphor processing online system. In *Proceedings of the 61st Annual Meeting of the*

Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 127–135.

Cornelia Müller. 2008. *Metaphors dead and alive, sleeping and waking: A dynamic view*. University of Chicago Press.

Sunny Rai and Shampa Chakraverty. 2020. A survey on computational metaphor processing. *ACM Computing Surveys (CSUR)*, 53(2):1–37.

Giulia Rambelli, Emmanuele Chersoni, Philippe Blache, Chu-Ren Huang, and Alessandro Lenci. 2019. Distributional semantics meets Construction Grammar. Towards a unified usage-based model of grammar and meaning. In *First International Workshop on Designing Meaning Representations (DMR 2019)*.

Ekaterina Shutova. 2015. Design and evaluation of metaphor processing systems. *Computational Linguistics*, 41(4):579–623.

Ekaterina Shutova, Lin Sun, Elkin Darío Gutiérrez, Patricia Lichtenstein, and Srini Narayanan. 2017. Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning. *Computational Linguistics*, 43(1):71–123.

Gerard Steen, Aletta G Dorst, J Berenike Herrmann, Anna Kaal, Tina Krennmayr, Trijntje Pasma, et al. 2010. A method for linguistic metaphor identification. *Amsterdam: Benjamins*.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2021. Survey of computational approaches to lexical semantic change detection. *Computational approaches to semantic change*, 6(1).

Xiaojuan Tan. 2023. *Static and dynamic metaphoricity in US-China trade discourse: A transdisciplinary perspective*. Ph.D. thesis, Vrije Universiteit Amsterdam.

Xiaojuan Tan and Alan Cienki. 2023. Dynamic metaphor theory.

Xiaojuan Tan and Alan Cienki. 2024. A dynamic metaphor perspective on Trump and Xi's trade negotiation in governmental discourse. *Language & Communication*, 96:42–53.

Xiaojuan Tan, Alan Cienki, and Bertie Kaal. 2024. The diachronic and cross-linguistic use of trade metaphors in us-china governmental discourse: A socio-cognitive approach. *Metaphor and the Social World*, 14:130–153.

Xiaoyu Tong, Ekaterina Shutova, and Martha Lewis. 2021. Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4673–4686.

Lotte van Poppel and Roosmaryn Pilgram. 2023. Types of resistance to metaphor. *Metaphor and Symbol*, 38(4):311–328.

Remi van Trijp. 2017. A computational construction grammar for English. In *2017 AAAI Spring Symposium Series*.

Tim Veenboer and Jelke Bloem. 2023. Using collostructional analysis to evaluate bert's representation of linguistic constructions. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12937–12951.

Lennart Wachowiak and Dagmar Gromann. 2023. Does GPT-3 grasp metaphors? Identifying metaphor mappings with generative language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1018–1032.

Kosuke Yamada, Ryohei Sasano, and Koichi Takeda. 2021. Verb sense clustering using contextualized word representations for semantic frame induction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4353–4362.

Author Index