Computational Linguistics in the Netherlands Journal 12 (2022) 207-233 Submitted 12/2022; Published 12/2022

Responsibility Framing under the Magnifying Lens of NLP: The Case of Gender-based Violence and Traffic Danger

Gosse Minnema[a] g.f.minnema@rug.nl
Gaetana Ruggiero[b] gaetana.ruggiero@unito.it
Marion Bartl[e] marion.bartl@insight-centre.org
Sara Gemelli[d] sara.gemelli01@universitadipavia.it
Tommaso Caselli[a] t.caselli@rug.nl
Chiara Zanchi[d] chiara.zanchi01@unipv.it
Viviana Patti[b] viviana.patti@unito.it
Marco te Brömmelstroet[c] m.c.g.tebrommelstroet@uva.nl
Malvina Nissim[a] m.nissim@rug.nl

[a] *University of Groningen, The Netherlands* [b] *University of Turin, Italy*

[c] *University of Amsterdam, The Netherlands* [d] *University of Pavia, Italy*

[e] *Insight SFI Research Centre, University College Dublin, Ireland*

Abstract

We introduce a framework for the computational analysis of how responsibility is framed in the reporting of two types of socially relevant events: gender-based violence (specifically, femicides in the Italian press), and traffic danger (specifically, traffic crashes in Dutch and Flemish news reports). We advocate for the parallel analysis of these two phenomena under the same theoretical framework, which draws on Frame Semantics, Critical Discourse Analysis and Natural Language Processing. Reusing two existing event-text datasets we show how computational experiments and the resulting analyses can be run. This work supports the testing and development of tools for NLP practitioners, as well as large-scale linguistic analyses for activists and journalists, in the context of socially impacting events.

1. Introduction

Reporting an event almost always implies taking a perspective on it. This can be done with full awareness, but it can also happen subconsciously. Prime cases of intentional perspective-taking or framing (Iyengar 1994, Entman 1993) can be observed in the political arena, where different newspapers will describe the same event in widely varying ways, or in sport events, where outlets associated to opposing teams might describe the same competition in rather different terms (Semetko and Valkenburg 2000, Matthes 2012).

Unconscious, and thus more subtle perspectives are often embedded in the communities' cultural and social aspects. Linguistically, such perspectives may be conveyed by community-established discourse practices, including lexical choices, which are only one of the many ways that could represent the same event. The ideological power of discourse precisely emerges from representation alternatives: "ideology is made possible by the choices a language allows for representing the same material situation in different ways" (Haynes 1989).

The analysis of how language is used to talk about things that happen can shed light on the specific perspective(s) that is adopted to frame these events and reflect on their effects on the society. While manual critical discourse analysis can provide valuable insights, it suffers from limited power and from a margin of subjectivity. The biggest shortcoming of manual analysis comes from the finite amount of data that can be analyzed. This could also introduce additional bias in the selection and subjective reading of the data. Indeed, to overcome such limitations, practitioners in the field of Critical Discourse Analysis (Van Dijk 2015), have recently emphasized the need to open the field toward multiple means of inquiry, including corpus linguistics, cognitive linguistics, and experimental perception studies (Hart 2018a).

Computational Social Science (CSS) has gained popularity as a method that applies Natural Language Processing (NLP) tools to run large-scale analyses of language use and language variation, especially in English (Conte et al. 2012, Baumer et al. 2015, Radford and Joseph 2020, Kabbach and Herbelot 2021, Mendelsohn et al. 2021).

In this work we propose a framework where NLP tools are used to study *responsibility framing*. It is known that the way a piece of news is written, especially in terms of perspective-taking, heavily influences the way readers perceive attribution of responsibility in the events described (Iyengar 1994). We focus on two phenomena where there is a potential imbalance of power between the actors involved, namely violence against women, as reported in Italian news, and traffic crashes, as reported in Dutch news. We introduce a single NLP-based methodological framework to treat these two phenomena in parallel, thereby using NLP tools as a magnifying lens to better look at how responsibility is framed in news reports of such events, and where biases emerge. We argue that specific NLP tools, grounded in specific linguistic theories, can support large-scale language-based analyses on these phenomena.

Contributions This work offers several contributions.

First, we suggest to unify the computational treatment of two socially relevant phenomena, underlying their similarities and differences, thereby opening up the opportunity to extend our framework to yet other phenomena that involve an imbalance of power. The computational framework we propose innovatively draws from a combination of Frame Semantics, Critical Discourse Analysis, and Natural Language Processing. With this work, we also introduce two datasets (`RAI-F` and "TheCrashes.Org" Dataset) to the NLP community by running computational analyses on them for the first time.

Second, we test a recent end-to-end system for frame semantic parsing (LOME, Xia et al. 2021) on new datasets for languages for which no LOME-evaluation exists as of yet (Italian and Dutch), both in a zero-shot setting as well as with additional language-specific fine-tuning; contextually, we highlight the difficulties of testing frame semantic models due to differences across annotation strategies and propose a solid evaluation framework.

Third, we develop and make available SocioFillmore, an online research tool for identifying responsibility-backgrounding frames and constructions in texts given a database of linked text-event information.[1]

Impact The impact of this project is twofold. On the social side, it can support activists in finding evidence of misrepresentation of specific events and social groups in the news in a systematic way, and on a large scale; also, ongoing extensions of the work presented in this paper will make it possible to help journalists gain awareness of the way in which they are framing a situation in their news reports, possibly not consciously, and that there could be alternative - and more neutral - ways of describing the very same event. On the NLP side, we provide testbeds for evaluating existing methods and tools in real-life situations and with new data; we provide methodological blueprints for running similar analyses and evaluations on yet new socially-relevant phenomena; and we contribute to the ongoing tool and resource development for languages other than English, thereby also assessing the potential and validity of multilingual approaches.

2. Phenomena and Approach

We consider two phenomena, violence against women and traffic danger, that are acknowledged to have a strong social relevance and, precisely due to their impact on society, to be subject to perspective-taking in news reports (Tranchese and Zollo 2013, Santaemilia and Maruenda 2014, Busso et al. 2020, Ralph et al. 2019, Goddard et al. 2019, Te Brömmelstroet 2020). We outline here the characteristics of these event types as well as their similarities and differences to show why the approach chosen is appropriate and what we expect to observe and obtain from our analysis.

1. `https://demo.let.rug.nl/gossminn/sociofillmore/explore`. Access may be subject to authorization due to the nature of the data.

Violence against women, and more specifically femicide, is worryingly common and therefore regularly reported in the news. The last briefing of femicide published in November 2012 by the European parliament[2] reports that the United Nations Office on Drugs and Crime (UNODC) estimates that 87,000 women were intentionally killed in 2017, which is the last global estimate from reliable sources. The same report underlines that Italy is one of the European countries with the lowest number of femicides.[3] Nevertheless, in this paper we focus on Italian news due to data availability, while keeping in mind that the phenomenon is obviously world-spread and specific cultural factors always influence the way in which such events are reported in the news. Focusing on the incidence of femicide in Italy, according to a report by ISTAT,[4] in Italy in 2019, out of 315 intentional killings perpetrated, 111 qualify as femicides.

The last 25 years have seen a constant number of intentional killing of women (from 0.6 in 1982 to 0.4 every 100,000 women in 2017). The number of victims dramatically increases if one considers gender-based violence in all its forms. Discouragingly enough, a report from November 2018 by the Italian National News Agency ANSA points out that the stereotype of a shared responsibility between the violence victim and its perpetrator is still widespread among young generations: "56.8% of boys and 38.8% of girls believe that the female is at least partly responsible for the violence she has suffered".[5]

Sexual power asymmetries are also discussed in the "Gender Gap" report, which ranks countries according to the gap calculated between how women and men are treated in key social areas (health, education, economy, and politics), to compare the extent of gender equality across the world. In 2021, Italy ranked at 63th position: in the European region only Greece, Malta, and Cyprus ranked lower.[6] Given that women are arguably a less privileged social group with respect to men, they are also very likely to undergo misrepresentation in media discourse, especially — but by no means limited to — when involved in facts in which men also are involved. This misrepresentation was precisely observed by critical discourse analysts in the studies cited above.

The other socially relevant phenomenon we study is traffic crashes. Worldwide, more than 1.3 million individuals are killed in traffic annually, while an approximated ten- to twentyfold are severely injured while being underway (Culver 2018, p. 153). According to the World Health Organization (2018), traffic crashes are the leading cause of death for young people. As such, it forms a unique, and ever-present (but often taken for granted) threat to life and limb in contemporary society. Although it is a form of violence that we unintentionally impose on each other, traffic crashes are highly complicated phenomena, due to a large variety of parties involved and highly complex cause-effect (or victim-culprit) relations due to dispersed responsibility and liability among those parties: think only of the role that traffic rules play in disciplining road users and how rule-breaking can be seen as both a cause and a context variable of a crash (Te Brömmelstroet 2020). While we can assume a relative straightforward relation in crashes between car drivers and pedestrians, this quickly dissolves when we for instance include crashes between car drivers. Again, data availability drives our language choice, so that we work on Dutch/Flemish news, but the computational analyses that we run can be extended to any other language for which data exist.

The two phenomena share similarities which made us conceive a single unified framework to treat them. Such similarities consist in (i) the presence of violence, with at least one of the actors getting injured, or dying; (ii) the (potential) imbalance of power between the actors with the associated higher vulnerability of one party; (iii) default or accepted state-of-affairs at the cultural level. To clarify this latter point, consider the following: the members of a society, including gender-based violence perpetrators and victims as well as parties involved in traffic

2. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/653655/EXPO_BRI(2021)653655_EN.pdf
3. In Italy and in the other EU countries, the legal category of femicide/feminicide does not exist. In the International Classification of Crime for Statistical Purposes 1.0 (https://unstats.un.org/unsd/statcom/doc15/BG-ICCS-UNODC.pdf), drawn by the UNODC in 2015, femicide is characterized as "[. . .] the intentional killing of a woman for misogynous or gender-based reasons" (p. 32).
4. ISTAT - Italian National Institute for Statistics https://www.istat.it/it/violenza-sulle-donne/il-fenomeno/omicidi-di-donne
5. http://www.ansa.it/canale_saluteebenessere/notizie/stili_di_vita/2018/11/30/violenza-donne-per-4-giovani-su-10-dipende-anche-da-lei_b834f656-fdf2-4a0c-8de5-2d6c6dfcfc82.html
6. http://www3.weforum.org/docs/WEF_GGGR_2021.pdf

crashes, can be categorized as members of the dominant social group (*ingroup*) or as members of weaker social groups (*outgroup*). The categorization of social actors is connected with ideology, a way of culturally interpreting reality: according to Van Dijk (1998), ideologies involve an ingroup vs. outgroup polarization, together with positive attitudes toward members of the ingroup and negative attitudes toward members of the outgroups. Ideologies also imply naturalized assumptions with respect to how events must take place in reality and the relative roles and responsibilities of event participants.

It is important to stress that the two phenomena are also different, at least in two respects: (i) intentionality before the event often exists in femicides, while crashes normally happen without any of the parties intentionally causing them; and (ii) the responsibility post-event: while femicides are more of a unidirectional event where it is usually clear where the guilt of the crime lies, traffic crashes can be conceived as more bi- or multi-directional, where responsibility can be shared by both/multiple parties. In spite of such differences, we argue that the aspects they share make them suitable for a unified framework of analysis. This allows for greater portability to yet other phenomena, to increase awareness of standardized reporting practices that might tend to blame the weaker party, and for any shared findings to be stronger both in terms of analysis as well as in testing theories and NLP methods.

3. Theoretical Framework: Computational-Critical Frame Semantics

Our theoretical background aims to unite different linguistic subfields, which together provide the conceptual and methodological tools for automatically analyzing responsibility analysis: (i) Fillmorian Frame Semantics (FFS, Fillmore 2006)[7] as a branch of Cognitive Linguistics (CogLing), for linking language to conceptual structures; (ii) Critical Discourse Analysis (CDA, Van Dijk 1995), for linking language to ideology; and (iii) Computational Linguistics (CompLing), specifically the FrameNet system of lexical representation and related parsing tools, for automatically mapping language to conceptual and structural representations.

Combining these fields requires several bridging steps. First, in §3.1, we will review previous literature that shows how a combination of CogLing and CDA can provide powerful tools for analyzing how specific linguistic constructions can reflect and reinforce ideologies. Then, in §3.2, we take this a step further, arguing that FFS, a specific branch of CogLing, is particularly suitable for systematically analyzing ideologically-biased social frames. Finally, in §3.3, we discuss how computational tools derived from FFS can be used to put this systematic analysis into practice.

3.1 Background: Combining Cognitive Linguistics and Critical Discourse Analysis

How Social is Cognitive Linguistics? The main aim of this study is to find systematic ways of linking specific linguistic structure to ideological ways of framing responsibility for particular events. CogLing promotes a vision of the use of natural language from a perspective of general cognition. CogLing researchers have proposed various kinds of theoretical devices such as metaphors (Lakoff and Johnson 1980), constructions (Goldberg 1995), construals (Langacker 1991), and frames (Fillmore 2006), each of which links linguistic forms to (often perceptually grounded) cognitive structures. In particular, in order to be able to describe responsibility framing, it is crucial to analyze from what *perspective* an event is being conceptualized: a perspective highlighting a conscious agent's intentional action is much more likely to be perceived as attributing responsibility to that agent than a perspective abstracting away from the participants or focusing on another participant's (e.g. the victim's) perspective. Perspective-taking is a phenomenon that explicitly or implicitly plays an important role in many cognitive linguistic theories. For example, Construction Grammar research has shown how different constructional choices can give rise to different viewpoints to the same event (Goldberg 1995); in Cognitive Grammar *construals* are defined as "our ability to conceive and portray the same situation in alternate ways" (Langacker 2019).

However, while providing many powerful tools for analyzing perspective-taking, CogLing does not generally consider the *social context* that influences the way that language perspectivizes

7. Throughout this paper, we will consistently use the term *Fillmorian Frame Semantics* to distinguish 'Fillmore-style' frames from other notions of frames elsewhere in cognitive linguistics and social science.

events. Although cognitive linguists recognized early on that the theoretical devices they proposed have a social significance – e.g., Lakoff and Johnson (1980, p. 156) stated that "metaphors create realities for us, especially social realities" – the interplay between discourse and society is not a central concern of cognitive linguistics, and as such, notions of perspective in cognitive linguistics are generally socially and politically 'neutral'.

CDA: Understanding and Challenging Society A research tradition that, on the other hand, can help to better connect discourse and society is CDA (Fairclough 2010, Van Dijk 2015). As Fairclough (2010) points out, CDA has two basic properties that can help us in better understanding perspectives and the connections between language and society. First, it is a relational form of research whose primary focus is on societal relations rather than on entities or individuals; second, it is dialectical as the study of the societal relations - of which discourse is an expression - requires an understanding the continuous connections between the internal and the external relations between the objects that compose society.

For CDA, the socially constructive effects of discourse are a primary concern. However, while the world is socially construed, "which construals come to have socially constructive effects depends upon a range of conditions which include for instance power relations [...]." (Fairclough 2010, p. 5). The "critical" element of CDA brings a critique (i.e., values, views of what is good) on the basis of which a society can be evaluated (i.e., what is wrong and how this can be improved). This critical component puts the effects of power relations, and in particular of the discoursive aspects of power relations, and inequalities on the spotlight and helps to unveil the reproduction of social structure in discourse (Halliday 1978, Fairclough 2010). CDA is not necessarily defined by any particular method: it is a cross-disciplinary tradition drawing on various methods from the humanities and social sciences. In some ways, CDA is as much a social movement as it is a theoretical framework, whose adherents are researchers who "take an explicit position and thus want to understand, expose, and ultimately challenge social inequality" (Van Dijk 2015, p. 465).

CDA and CogLing have in common basic assumptions concerning the relationship of language and cognition: (i) linguistic structures are integrated in other domains of cognition and are grounded in extralinguistic experience; (ii) cognitive models, which are mirrored in our linguistic expressions and which we take for granted, contribute to construing our surrounding reality and are socio-culturally specific (see also Section 1). According to its original social afflatus, the theory of conceptual metaphor (Lakoff and Johnson 1980) has been the earliest and, up to date, principal appropriation of CogLing in CDA (Chilton 1996). Sporadically, theoretical tools of CogLing different from conceptual metaphors have been employed by critical discourse analysts: prototype theory (O'Halloran 2003), force-dynamics (Hart 2011a), certain aspects of Cognitive Grammar (Marín Arrese 2011) and Mental Space Theory (Chilton 2004).

Despite said link between language and cognition assumed by CogLing and CDA, studies in which the latter approaches are paired with experimental methodologies are relatively rare up to date, with some exceptions (Hart 2013, Fuoli and Hart 2018, Hart 2018a, Hart 2018b, Fuoli and Hart 2018, Hart and Fuoli 2020, Hart 2021). However, concerning argument structure syntactic constructions, Henley et al. (1995) investigate how passive voice is used and perceived in US media reports by combining a corpus- and a perception study. Bohner (2001) also focuses on passive voice investigating its effects on the perception of gender-based violence reports among German native speakers. Finally, Hart (2018a) compares English native speakers' perception of transitive and reciprocal constructions in reports of political protests and how the selection of the latter affects blame apportionment.

3.2 From Fillmorian Frames to Social Frames

In this paper, we adopt a particular cognitive linguistic theory: Fillmorian Frame Semantics. Historically, FFS has its origin in Case Grammar (Fillmore 1968), an early theory of semantic roles and predicate-argument structures within the framework of transformational-generative grammar (Chomsky 1957, Chomsky 1965). Over the years, Fillmore's approach to semantic roles shifted away from its initial focus on syntax and became increasingly semantic and cognitively oriented until, by (Fillmore 1985), frame semantics had become a "theory of language understanding" (or "U-semantics"), which is also the primary perspective on frames that we will take in this paper.

Within the context of language understanding, a *interpretive frame* is defined as a unit that represents the conceptual background knowledge needed to semantically process a particular linguistic expression. Frames have both a semantic and pragmatic function: they help interpreters "give [a text] a maximally rich interpretation [...] by having at hand, for each conventional linguistic form in it, an implicit answer to the question: '[w]hy does the language have the category which the form represents?' and [...] '[w]hy did the speaker select this form in this context?"' (Fillmore 1985, p. 234).

Frames are everywhere: every word needs a background to be interpretable at all. Some classical examples where this is especially obvious include family terms ("sister" and "father" only have a meaning within the context of networks of family relations), calendric units ("Saturday" would be meaningless without a concept of "week"), and geometry ("hypothenuse" cannot be interpreted without knowledge about rectangular triangles). However, frames do not simply make linguistic expressions interpretable, but also *selectively encode* parts of the reality surrounding us, and can provide different perspectives ("alternative 'ways of seeing things"' (Fillmore 1985, p. 229)) on the same part of reality. A classical example of this is the distinction between "land" and "ground", which have the same referential denotation but different connotations, as they are interpreted from the perspective of air travel and sea travel, respectively.

Frames can also encode aspects of social reality: a well-known example (introduced as early as Fillmore (1971)) is the framing of commercial transactions: "buying", "selling", and "paying" refer to the same type of event, but encode it from different participants' perspectives by (syntactically) foregrounding certain roles and backgrounding others: "buying" foregrounds the buyer and the thing that is bought but backgrounds the seller; "paying" foregrounds the buyer and the seller but backgrounds the the thing that is bought. We believe that this is exactly the same type of 'framing' that we observe in the news coverage of femicides and traffic crashes: particular linguistic expressions are often used that enforce the encoding of certain participants (the victims) over others (the stronger parties).

Modern frame semantics, mostly carried out within the context of the FrameNet project (Baker et al. 2003), has a strong focus on systematically documenting sets of lexical units (predicates[8]) sharing a set of semantic roles that are expressed through certain syntactic patterns. Within this framework, 'frames' are defined as structured objects that link together lexical units and semantic roles, and carry the conceptual background shared by them. While it also contains frames encoding objects and abstract concepts, an important part of FrameNet is dedicated to frames describing many different kinds of real-world events and situations. Such frames differ from each other not only in terms of what kinds of events they describe, but also what aspects of and which participants in these events they emphasize, both conceptually and linguistically.

Table 1 gives some examples of frames that are relevant for describing violence. As can be seen from the example sentences, these frames can be used to describe the same (or similar) situations in reality, but conceptualize these situations in very different ways: while the first three frames encode the fact that someone dies (or is dead), while CAUSE_HARM only encodes the fact that someone gets hurt (not whether they survived) and EVENT does not specify the event in any way at all. Meanwhile, even though the first three frames are similar in the kinds of situation they describe, they differ in how much of an 'agentive' perspective they take: KILLING and DEATH both describe the event in a dynamic way, but the former attributes responsibility for the event to an active agent, whereas the latter focuses on the person who dies and leaves completely open who or what caused the death. Finally, DEAD_OR_ALIVE removes the dynamicity of the event, and instead conceptualizes it as a static situation.

From our perspective, it is just a small step from merely documenting frames and their usage in corpora from a purely linguistic point of view, to critically questioning them: giving a social twist to (Fillmore 1971)'s formulation, we can ask why a speaker, in a given situation, consciously or not used any particular frame over another one to refer to an event, foregrounding certain participants and backgrounding others, and perhaps we can even question why a language has these frames at all. However, to our knowledge, this social-critical potential of FFS has barely

8. Throughout this paper, we will use 'predicates' in a very broad sense, understanding them to be any linguistic unit capable of assigning semantic roles in the FrameNet sense. In this sense, predicates are prototypically verbs, but can also be nouns, adjectives, or even prepositions.

Frame	Description	Example
KILLING	an agent *(Killer)* actively causes the DEATH of a patient *(Victim)*	[The man] **killed** [his wife]
DEATH	event in which someone *(Protagonist)* dies	[The woman] **died**
DEAD_OR_ALIVE	state of someone *(Protagonist)* being dead or alive	[She] was found **dead**
CAUSE_HARM	an agent *(Agent)* actively causes a patient *(Victim)* to be hurt	[He] **stabbed** [his girlfriend]
EVENT	an unspecified event *(Event)* happens	[The dramatic events] **happened** last week

Table 1: Examples of FrameNet frames relevant for describing (gender-based) violence. Semantic role names indicated in *italics*, lexical units indicated in **bold**.

been noticed in the literature. An important potential exception is George Lakoff's *Don't Think of an Elephant* (Lakoff 2014), which does not explicitly reference Fillmore's work but does use a very similar notion of framing and applies it to political discourse. However, the only work that we know of that explicitly links FFS and FrameNet to social-political notions of framing is (Ziem et al. 2018), who propose frame-based semantic role annotation as a tool for analyzing media frames, and perform a corpus analysis of a single FrameNet frame in order to analyze framing of information gathering during the Edward Snowden affair. The results of this analysis show that certain semantic roles are more frequently expressed in texts describing the social events under analysis compared to a reference corpus of similar texts, from which the authors draw tentative conclusions about how the media 'frame' certain event participants.

3.3 Towards a Computational-Critical Frame Semantics of Responsibility Framing

In the previous subsections, we have seen that, used together, Cognitive Linguistics and Critical Discourse Analysis provide powerful tools for analyzing discourse in society (§3.1) and that there is an unexplored potential for using Fillmorian Frame Semantics (§3.2), a particular branch of Cognitive Linguistics, for critically analyzing perspective-taking on events. In this subsection, we will make the case for using frame semantics, and the ecosystem of computational tools and resources developed around the FrameNet project, for computationally analyzing responsibility framing as a special case of perspective-taking. From a conceptual point of view, our proposal is a direct descendant from Pinelli and Zanchi (2021)'s Construction Grammar approach (CxG) approach to analyzing responsibility framing of gender-based violence, and extends this approach not only from a theoretical perspective, by taking into account semantic frames in addition to grammatical constructions, but even more importantly from a practical point of view, by enabling large-scale automatic analysis in addition to manual analysis.

Computational Frame Semantics Computational work on FFS can be divided into two main categories: on the one hand, starting with Baker et al. (2003), there has been much work on assembling large-scale lexical databases (Berkeley FrameNet for English and similar projects for other languages) that document frames, semantic roles, and lexical units, and on the other hand, starting with Gildea and Jurafsky (2002) there has been much work on frame semantic parsing: automatically analyzing texts in terms of semantic frames and their associated semantic roles. For an impression of the current state of FrameNet and related projects, see Torrent et al. (2020), and for a fairly recent overview of the state of frame semantic parsing, see Kabbach (2019).

For our purposes here, one family of computational approaches to FFS is relevant in particular: recent work, starting with Vossen et al. (2020), has proposed performing computational semantic frame analysis within the context of large-scale *data-to-text* datasets (Vossen et al. 2018b) linking structured information about real-events to texts referencing these events. The central idea in this line of work is that, if it is known which texts refer to which specific real-world events, it becomes possible to study different ways in which particular event types, or even particular event instances, are conceptualized in texts using frames. A crucial concept for doing this kind of analysis is that of *typical frames* (Postma et al. 2020, Remijnse and Minnema 2020, Remijnse et al. 2021), which are frames that are pre-determined to be conceptually important to a particular event type. Once defined, these typical frames can be used for performing a top-down, event-driven analysis of a text: instead of annotating all possible mentions of all possible semantic frames, only the relevant frames for a particular event are analyzed.

Extending Pinelli & Zanchi 2021 The central idea of Pinelli and Zanchi (2021) is that, in descriptions of violence against women, the use of syntactic constructions with varying levels of transitivity – from transitive active constructions on one side of the spectrum, via passives and anticausatives to nominalization constructions on the other side – corresponds to various degrees of responsibility attributed to the (male) perpetrator. For example, while "he killed her" (active/transitive) makes the involvement of an active agent fully explicit, "she was killed (by him)" (passive) shifts attention away from the agent and to the patient, and "the murder" or even "the event" (nominal construction) moves both participants to the background. Based on this idea, the authors performed a manual CxG-based analysis of constructions describing violence in a small corpus of Italian news articles reporting on femicides, and found that responsibility-backgrounding constructions are very prevalent. They furthermore argue that this is problematic, given that studies such as Bohner (2001) and Hart (2018a) have shown the substantial impact of transitivity on the perception of responsibility and blame.

The approach that we will propose in this paper aims at reproducing (or approximating) Pinelli and Zanchi (2021)'s findings, but using different conceptual and practical tools, and apply it to different datasets and to the additional social domain of traffic crashes. Conceptually, we propose to analyze descriptions of violence in terms of not just syntactic constructions, but also of semantic (FrameNet) frames. This will make the analysis richer because it allows taking into consideration semantic factors (e.g. the distinction between KILLING, DEATH, and DEAD_OR_ALIVE) that contribute to 'de-agentivizing' descriptions of violence. Additionally, using semantic frames allows us to make other conceptual distinctions between different descriptions of violence (e.g., KILLING versus CAUSE_HARM).

From a practical point of view, we will be able to work on a much larger scale by, on one hand, making use of existing data-to-text datasets which enable us to consider only texts about specific (types of) events, and, on the other hand, by making use of various NLP tools for syntactic and frame semantic analysis to automate the annotation process.

4. Datasets

An essential feature that contribute to the success of investigating perspectives, and thus for this project too, is the availability of lots of data from different sources concerning the same specific event mention or happening. The collection of such dataset from scratch is burdensome and non-trivial. A solution that alleviates this aspect comes from the application of the *data-to-text* approach (henceforth, `D2T`) (Vossen et al. 2018b, Vossen et al. 2020). The approach exploits existing event repositories (either structured or semi-structured) in which specific events (e.g., femicides, car crashes, terrorist attacks, among others) are defined *a priori* and texts (i.e., mentions) are collected and mapped to these events, thus making the aggregation of sources with different perspectives on the same event easier. The following paragraphs introduce the two datasets we have used in this paper, which are both based on the `D2T`method.

RAI Femicide Dataset The RAI Femicide dataset (henceforth, `RAI-F`) has been compiled by the CRITS research unit at RAI (Radiotelevisione Italiana), the Italian national public broadcasting organization, in collaboration with a team of sociologists. It is composed of news articles in Italian from 31 different news sources (see Table A. 1 in Appendix A.1 for a detailed list of the sources.) The dataset consists of 2,734 documents covering a total of 937 unique femicides that occurred in Italy between January 2015 and November 2017. The same event is reported by different news outlets, and at different points in time after the event occurred.

Each event is annotated with structured metadata about the event itself and the participants involved in it, categorized into about thirty different variables. In particular, the annotations contain information about the date and location (geographical and semantic) of the event, the category of the event (e.g. aggression, abuse and/or homicide), the weapon used and the motive. Moreover, the dataset includes details about the victim and the attacker, such as age, gender, birth place, nationality, occupation, and number of children. It is also specified whether or not the attackers committed suicide, escaped, or turned themselves in, and whether they were affected by psychological issues. Additionally, the events are annotated with information regarding the investigation process and respective court rulings. Alesiani and Metta (2021) provide more information about the collection and annotation process.

Although we did not use all of the text sources to conduct our experiments, the many different dimensions present in the dataset allow the analysis to possibly expand in multiple directions.

"TheCrashes.Org" Dataset The website `https://thecrashes.org` was launched in January 2019 to collect news articles reporting on traffic crashes. It offers a platform that allows volunteers to add reports from their online (local) newspaper. By uploading the news link, the platform collects all relevant metadata and then asks the volunteer to specify a number of aspects about the reported crash (i.e., traffic modes, injury/deaths, involvement of alcohol/children or hit-and-run). Volunteers who have a moderator role can then put the report on the website (and/or merge them with other reports of the same event) where it becomes public.

To date, 297 volunteers have added 9,424 reports of 8,038 traffic crashes. Most of these occurred in the Netherlands and are reported in Dutch, but the website increasingly draws volunteers from other countries. Te Brömmelstroet (2020) provides additional details regarding the collection and the resulting dataset.

5. Computational Analysis: Frame Semantic Parsing for Responsibility Analysis

As a frame semantic parsing model, we propose using LOME (Xia et al. 2021), an Information Extraction (IE) system which includes a frame semantic parser. LOME is one of the very few true end-to-end models (i.e. designed to take raw text as input and to output frame structures found in this text),[9] that reports excellent performance on English frame identification. Furthermore, the use of the multilingual XLM-R language model (Conneau et al. 2020) as its encoder, makes it suitable for analyzing text in other languages in a zero-shot approach.

However, there are three important issues that need to be addressed before we can apply LOME to our problem domain:

(i) there is only very limited published data on the model's performance. The frame semantic parser is part of a larger IE system. Even though the reported evaluation on English and our own initial observations from using the system's demo version[10] all made a very favorable impression, we have very limited quantitative information on how well the model actually works, and none at all for languages other than English;

(ii) while LOME is trained on English data only, there exists (albeit limited) training data available for our target languages as well. This suggests that we could use language specific or adapted version of the frame parser in alternative to a zero-shot approach;

(iii) we do not know how well LOME (or any other model) works for our specific domains: it is known from previous work that frame semantic parsing models have difficulties adapting to new domains (Hartmann et al. 2017).

To address each of these issues, we propose a multilingual evaluation benchmark for end-to-end frame semantic parsing (§5.1); we then compare the performance of several strategies for using Italian and Dutch training data to that of the original system (§5.2); finally, we take initial steps towards testing real-world applicability by comparing the two top-ranked LOME models for Italian against each other on a small manually annotated sample of the `RAI-F` dataset (§5.3).

5.1 Evaluating LOME

Traditionally, frame semantic parsing has been conceived as a pipeline consisting of a predicate identification module (`targetID`), a semantic frame classification module (`frameID`), and a semantic role labeling module (`argID`). Most work in the literature so far has focused on individual pipeline components, particularly `argID`, rather than on the 'end-to-end' task. We discuss the available

9. See Minnema and Nissim (2021) for a discussion of frame semantic parsing as an end-to-end task.
10. See `https://nlp.jhu.edu/demos/lome/`

datasets in English, Italian, and Dutch, and their suitability for evaluating end-to-end performance, and introduce our strategy for benchmarking LOME models for these languages.

Evaluation data In the context of frame semantic parsing, it is important to distinguish between two kinds of datasets: *fulltext* and *exemplar* data. In *fulltext* corpora, entire documents have been selected for annotation, and within these documents, it has been attempted to annotate every possible lexical unit and frame structure in the text. In such corpora, a single sentence can contain several lexical units and frames. By contrast, in *exemplar* corpora, specific sentences have been chosen (from a larger corpus) to illustrate the use of particular lexical units. In such corpora, each sentence usually has no more than one annotated lexical unit.

Awareness of the dataset one is dealing with is crucial when evaluating end-to-end performance of frame semantic parsing models. An important challenge for accurately measuring precision comes from 'missing' annotations: if no gold annotation exists for a particular target in the sentence, but a parsing model does predict a frame structure for that target, it is possible that the model is over-generating, but it is also possible that the model was in fact correct and is 'filling' a gap in the gold annotations. This problem is particularly significant in cases when the only available evaluation data consists of exemplars, because these by definition contain very many gaps. However, the same problem may be encountered when dealing with fulltext data: frame semantic annotation is a complex task and it is possible that some targets are missed during annotation.

Datasets For English, we use the standard Berkeley FrameNet 1.7 (henceforth, BFN) evaluation set (Baker et al. 2003). BFN is also the dataset that the original LOME model was trained on. BFN contains both fulltext and exemplar data, but in this study, we only use the fulltext data. BFN contains texts from several corpora (including parts of the American National Corpus, the Nuclear Threat Initiative website, and other sources), in significant part consisting of news, history and political texts. The training and evaluation datasets come from different random selections from the same set of texts.

For Italian, we use the dataset from the EVALITA-2011 Shared Task on Frame Labeling over Italian Texts (Basili et al. 2013). It consists of training data from two pre-existing Italian FrameNet-based corpora (Tonelli and Pianta 2008, Lenci et al. 2012) including text from the EuroParl corpus (Koehn 2005) as well as from newspapers and periodicals, and a test set (another portion of the EuroParl corpus) specifically created for the task. An important limitation of this datasets is that they are exemplar-like data: only one predicate per sentence was selected for annotation. The training set covers 1,255 sentences and predicates, covering 38 different frames, while the test set has 318 sentences covering 36 of these frames.[11] Since no development data was included in the official data release, we held off a randomly selected 10% portion of the training set for this purpose.

For Dutch, we use the Open Dutch FrameNet corpus (Vossen et al. 2018a), which was created starting from an existing dataset of PropBank annotations (De Clercq et al. 2012) covering part of the SoNaR corpus (Oostdijk et al. 2008), including Dutch texts from various genres. All predicates covered by the PropBank annotations were further annotated with BFN frame labels. This can be seen as a 'partial fulltext' approach: more than one predicate per sentence was annotated, but the PropBank predicates do not cover everything that would be seen as a predicate according to FrameNet standards (e.g., nouns and adjectives are predicates in FrameNet but not in PropBank). All texts in the corpus were annotated by two annotators, with 47-51% agreement (depending on the strictness of evaluation) on frame labels and 79% agreement on role labels. The released dataset[12] provides annotations separately for both of the annotators (i.e., without a final consensus version); for this reason, we repeated our training and evaluation procedures for each annotator.

Table 2 provides an overview of all available data in the three languages, split by annotation type.[13] The numbers show that English has by far the largest annotation dataset, especially when also considering the exemplar sentences. Dutch is not very far off in terms of the number of

11. Note that this data has been annotated using the 'iFrameNet' lexicon, which has some modifications relative to BFN. In particular, the training and test sets contain 109 and 75 instances, respectively, of frames that are not present in BFN.
12. `https://github.com/cltl/FrameNet_annotations_on_SoNaR/`
13. Sentence and frame counts have been calculated after preprocessing of the corpora with the `bert-for-framenet` toolkit (`https://gitlab.com/gosseminnema/bert-for-framenet`). Different counting methods might produce

		English	Italian	Dutch
fulltext	sentences	5,093	0	3,287
	frames	29,359	0	5,091
exemplar	sentences	163,801	1,569	0
	frames	169,473	1,569	0
total	sentences	168,894	1,569	3,287
	frames	198,832	1,569	5,091

Table 2: Comparison of sentences and annotations in the datasets of each language

fulltext sentences, but with a far lower ratio of annotations per sentence. The Italian dataset is even smaller, with a very limited number of sentences and annotations.

Benchmark For evaluating end-to-end performance of LOME (and potentially other frame semantic parsing systems) on the three datasets described above, we make use of the *SeqLabel* metric introduced in (Minnema and Nissim 2021), which compares, on a token-by-token basis, the predicted frame and role labels to the gold standard. This metric is particularly suited for end-to-end evaluation since it does not require target predicates as input. However, since this metric was originally designed for fulltext-only data, using it as-is will result in very low precision scores: any predictions that do not match the gold annotations will be counted as erroneous, even if they are in fact correct but simply not included in the annotations. For this reason, we deploy a modified version of the metric, *gold_pred* (Minnema 2021): only frame and role predictions corresponding to predicates annotated in the gold dataset are evaluated. Wit this, we can get a reliable estimate of models' performance on exemplar datasets as well as partial fulltext datasets such as Dutch. Note, however, that an inherent limitation of such datasets unfortunately remains: it is not possible to evaluate in any way a model's predictions for predicates that are never annotated in the evaluation set. This means that the gold_pred scores should be interpreted with some caution: while we have no indication that this is the case for our models, it is in principle possible for a model suffering from widespread 'hallucination', i.e., predicting frames for target words that are not predicates, still achieve high precision scores.

5.2 Adapting LOME to Italian and Dutch

Based in part on earlier work on adapting LOME to a domain-specific FrameNet dataset (Minnema 2021), we propose three strategies for making optimal use of the available training data for the two target languages:

(i) ***Simple:*** train LOME using the same settings as the original model, but use only the target language's own training data (i.e., instead of the English training data);

(ii) ***Concat:*** train LOME using standard settings, but on the concatenation of the training data for the target language and the English training data;

(iii) ***Berkeley:*** first train LOME for English using standard settings, and then use this model's encoder (XLM-R) parameters to initialize the encoder of a second LOME model. Next, train the second model using only the target language's own training data. The intuition behind this technique is that encoder representations that are already 'made useful' for doing frame semantic parsing would help make the training process for the target language more efficient.

LOME was trained by extracting the source code from the Docker image provided by the authors,[14] converting the training annotations to the appropriate format, and modifying the configuration files. Apart from the encoder model (in the Berkeley strategy), we kept all of the original settings: models are trained for up to 128 epochs with early stopping based on development set scores; Adam (Kingma and Ba 2017) is used for optimization, with the initial learning rate set to $1e^{-3}$.[15]

slightly different results, e.g. due to different sentence tokenization choices. Counts for Dutch are based on annotator A1's annotations.

14. `https://hub.docker.com/r/hltcoe/lome`

15. The models were trained on the Peregrine HPC cluster,[16] using a single Nvidia V100 GPU. Training a single model took between one and five hours depending on the setup and training set sizes.

lang	model	conf	frames raw R	P	F	frames gold_pred R	P	F	roles raw R	P	F	roles gold_pred R	P	F
EN	***FN-LOME***	*0.95*	0.70	0.50	0.58	0.70	0.89	0.78	0.59	0.41	0.49	0.59	0.69	0.64
	Sesame	-	-0.12	0.12	0.02	-0.12	-0.08	-0.11	-0.22	0.00	-0.09	-0.21	-0.16	-0.20
IT	***FN-LOME***	*0.9*	0.52	0.07	0.13	0.52	0.63	0.57	0.54	0.15	0.23	0.50	0.63	0.56
	Simple	*0.95*	0.14	0.00	0.00	0.14	-0.14	-0.01	0.10	0.07	0.10	0.16	-0.14	0.00
	Concat	*1.0*	0.14	0.07	0.11	0.14	0.21	0.17	0.05	0.10	0.11	0.08	0.10	0.09
	Berkeley	*0.9*	0.17	-0.02	-0.02	0.17	-0.07	0.05	0.17	0.04	0.07	0.12	0.04	0.09
NL (A1)	***FN-LOME***	*0.95*	0.29	0.08	0.12	0.29	0.46	0.35	0.28	0.15	0.20	0.25	0.41	0.31
	Simple	*0.95*	0.03	-0.03	-0.04	0.03	-0.08	-0.01	0.06	0.03	0.04	-0.01	0.04	0.01
	Concat	*0.9*	-0.02	0.03	0.03	-0.02	0.03	-0.01	-0.05	0.22	0.09	-0.08	0.16	-0.05
	Berkeley	*0.95*	0.03	-0.03	-0.04	0.03	-0.03	0.01	0.07	0.04	0.05	-0.01	0.05	0.01

Table 3: SeqLabel scores (precision/recall/F1-score) for the LOME experiments. Baseline scores are in blue, improvements/decreases in performance are in green/red. The 'conf' value indicates the (empirically chosen) confidence score cutoff. *FN-LOME* refers to the original LOME. *Sesame* refers to Open-SESAME (Swayamdipta et al., 2017), a previous state-of-the-art English-only model. For Dutch, only results based on the data from annotator A1 are displayed.

Before evaluation, predictions were filtered according to the confidence scores of the LOME model, removing all predictions with a score below a threshold in $\{0, 0.9, 0.95, 1.00\}$. The same thresholds were used for both frame and role labels; if a certain frame was removed, all of its associated roles were automatically removed as well. Optimal confidence thresholds were selected on the development set for each model individually, in order to achieve the best possible balance between recall and precision on gold_pred frame F1-scores.

Table 3 shows the SeqLabel scores for each of the models for English, Italian, and Dutch. On English, our results are in line with the evaluation by (Xia et al. 2021): when compared on the same metrics, LOME shows an improvement over Open-SESAME on frame prediction (by 2 F1 points on unadjusted scores, and 11 F1 points on gold-predicate scores), and also on role prediction (by 9 and 20 F1 points, respectively). On Italian and Dutch, it is harder to interpret the results as there are no previous end-to-end models to compare against. However, LOME achieves moderate recall (52%, around 20 points lower than English) on frame prediction for Italian, and performs worse on Dutch (29%). Raw (unadjusted) precision scores are very low for both languages, which is to be expected at least for Italian, given that only exemplar data is available, but less so for Dutch. Looking at gold_pred scores, precision improves considerably, reaching somewhat more acceptable levels for both languages. Role prediction scores follow mostly the same patterns.

Our language-specific training strategies seem to have mixed success: on Italian they lead to considerable improvements of up to 17 F1 points on frame predictions and 9 F1 points on role prediction. On Dutch, there is not much improvement overall, with frame scores seeing not much change and role scores seeing improvements on precision but not on the final (gold_pred) F1 scores. For Italian, the best overall model is clearly ***Concat***: on the test set we get a gold_pred F1 score of 62% for frames and 55% for roles, which seems quite good given the small quantity of training data available. On the other hand, for Dutch, it seems that using FN-LOME in a zero-shot setup gives the best, although still quite bad, results.

The disappointing results on Dutch are quite unclear to us: we see no straightforward explanation for the significantly worse performance of the zero-shot model on Dutch compared to Italian, nor for the fact that using language-specific training data, despite the relatively large set of available annotations, yields barely any improvements. This result is especially puzzling given that we do see substantial improvement on Italian even with much less available data, and given that similarly promising results have been found in (Minnema 2021) for training a domain-specific FrameNet model with limited exemplar data.

	frames	
	all	relevant
predictions		
FN-LOME	63,714	4,444
IT-Concat	31,237	2,899
either	68,679	4,982
differences		
#	48,143	2,910
%	70.1%	58.4%

(a) Frame predictions differences

	differences	
	any	no nulls
total	48,143	859
by predicate		
scomparsa	153	56
morta	131	100
omicidio	122	50
by frame		
KILLING	705	124
EMOTION_D.	429	62
DEATH	374	245

(b) Typical frame differences

	best prediction			
	FN	IT	both	none
overall	0.51	0.12	0.12	0.25
by frame				
KILLING	0.70	0.19	0.11	0.00
EMOTION_D.	0.77	0.05	0.05	0.14
DEATH	0.33	0.05	0.19	0.42

(c) Scoring the differences

Table 4: Comparing Italian LOME models: FN-LOME ('zero-shot') vs. IT-Concat

5.3 EVALITA-LOME: Estimating Real-World Performance

Having found that the ***Concat*** model achieves best performance on Italian on the EVALITA evaluation set, the next step is to test the real-world performance of this model compared to the FN-LOME baseline. In order to do this, we used both models to produce predictions on the development portion of the `RAI-F` dataset, automatically extracted all the differences in frame predictions, and set up a small-scale annotation. A single annotator blindly compared the predictions from the two models for a randomized sample of 150 predictions.[17] We believe that this approach leads to a realistic estimate of the relative performance of the two models, avoiding the effort of fully manual annotation. Furthermore, we evaluated only on frame prediction, since it is known to be more domain-sensitive than role prediction (Hartmann et al. 2017). The annotation was limited to only cover differences in predictions for the *typical frames* targeted by our critical framing analysis (see below). During the annotation, only the relevant sentence, target predicate, and the two frame predictions were visible, without any information about by which model each prediction was produced. To further reduce the risk of bias, the order of presentation of the two models' predictions was shuffled for every example.

The results of the comparison between the models is given in Table 4. Sub-table (a) gives a general overview of the predictions by the two models, and shows a striking difference: FN-LOME produces more than double the number of frame instances predicted by ***Concat***. When looking only at typical frames, the difference becomes smaller but is still very considerable. The gap becomes even larger for the difference rates: in over 70% of the cases in which at least one of the models predicts a frame instance for a particular predicate, the other model produces a different frame or no frame at all. Again, the picture is slightly better when we look at only typical frames, but even then, the difference rate is almost 60%. In sub-table (b), we zoom in on the differences for the typical frame predictions. This reveals that the vast majority of differences arise in cases where one model makes a prediction whereas the other model predicts nothing: there are only 859 cases in which both models predict different frames. Interestingly, of these cases, more than a third can be attributed to a disagreement over two frames in particular: KILLING and DEATH.

What these numbers demonstrate is that IT-Concat 'misses' many of the predicates for which FN-LOME predicts a frame, leading to the suspicion that the former model might be suffering from poor recall. While it is possible that FN-LOME is over-generating (as discussed above, this is a problem that would not have been visible from the EVALITA evaluation scores due to the nature of the dataset), the results in sub-table (c) make it unlikely: in 51% of differences, the FN-LOME is marked as producing the most correct prediction, against only 12% of cases where ***Concat*** is better. In another 12% of cases, both models are judged as producing equally good predictions, and in 25% of cases both models are judged to be bad. Looking at the most frequently confused frames, LOME-FN is overwhelmingly on KILLING and EMOTION_DIRECTED, but not on DEATH, for which both models' predictions are rejected in 42% of cases.

While our annotation experiment was limited both in scope and in number of annotations, the preliminary conclusion is clear: FN-LOME, while performing largely worse that ***Concat*** on

17. The annotator is author S.G.

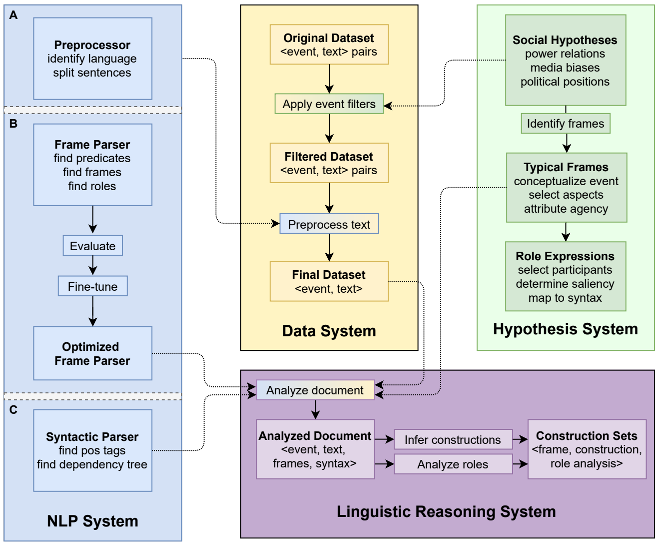


Figure 1: Conceptual overview of SocioFillmore

the EVALITA evaluation sets, is clearly more robust, producing more and better predictions in a real-world setting, at least for our particular domain. Hence, in the remainder of this paper, our analysis will be based on the predictions from the FN-LOME model for both Italian and Dutch.

5.4 Introducing SocioFillmore

Having proposed and evaluated frame semantic parsing models for our target languages, we can now build on these models to introduce a first attempt at critical frame semantics. In this subsection, we introduce SocioFillmore, a tool for identifying responsibility-backgrounding constructions in texts given a database of linked text-event information.

Figure 1 shows a schematic overview of how the tool works and divides it into conceptual 'systems'. Everything starts with the Hypothesis System (green): the researcher starts with hypotheses about the analyzed event type and its conceptualization in the media; based on this, they define relevant semantic frames for conceptualizing relevant aspects of the events and relations linking the roles associated with these frames to even participants. Next to this, there are the Data System (yellow), consisting of the dataset itself, and the NLP System (blue), consisting of several tools, including a basic preprocessing system (A), the frame semantic parser (B), and a syntactic parser (C). All of these systems come together in the Linguistic Reasoning System (purple): the NLP tools are applied to the dataset to produce a syntactic and semantic analysis of the frames specified in the Hypothesis System. The Linguistic Reasoning System then combines this information and applies a set of rules to extract meaningful syntactic structures and semantic role expression labels.

Hypothesis System For each of our two target domains, we defined a set of typical frames based on a preliminary manual analysis of a number of example sentences from (Pinelli and Zanchi 2021) (for femicides) and (Te Brömmelstroet 2020) (for crashes) that we believed to be representative of the linguistic phenomena analyzed in these papers, i.e. the level of agentivity and responsibility ascribed to the parties involved in femicides and unbalanced crashes, respectively. Additional frames were selected by inspecting the overall frequency distribution of the output of LOME on the two corpora and selecting frames that were both conceptually related to the two event domains and predicted frequently enough to perform a meaningful quantitative analysis.

Femicides

Frame	Target	Killer	Victim	Object/Concept
Picchiata brutalmente dal fidanzato all'alba del primo dell'anno. '[she was] brutally hit by [her] boyfriend at the dawn of the first day of the year' (P&Z'21, p. 131, ex. 3)				
CAUSE_ HARM	*picchiata* 'hit'	*dal fidanzato* 'by [her] boyfriend'	(implicit)	-
La furia omicida si è scatenata nel pomeriggio in un appartamento di Pietra Ligure. 'the murderous fury broke out in the afternoon in an apartment in Pietra Ligure.' (P&Z'21, p.134, ex. 10)				
KILLING	*omicida* 'murderous'	(implicit)	(implicit)	-
EMOTION_ DIRECTED	*furia* 'fury'	-	-	*omicida* 'murderous'
EVENT	*scatenata* 'broke out'	-	-	*furia omicida* 'murderous fury'

Unbalanced crashes

Frame	Target	Driver	Victim	Object/Concept
Bestuurder auto rijdt door na aanrijding met 9-jarig meisje 'car driver drives away after crash with a 9-year old girl' (B'20, p. 100109, ex. 7/2)				
OPERATE_ VEHICLE	rijdt 'drives'	*bestuurder* 'driver'	-	*auto* 'car'
CAUSE_ IMPACT	*aanrijding* 'crash'	*bestuurder auto* 'car driver'	9-jarig meisje '9-year old girl'	-
Fietser gewond geraakt door botsing in Lunteren 'cyclist hurt after collision in Lunteren' (B'20, p. 100109, ex. 7/5)				
CAUSE_ HARM	*gewond geraakt* 'hurt'	-	*fietser* 'cyclist'	*botsing in Lunteren* 'collision in Lunteren'
IMPACT	*botsing* 'collision'	(implicit)	*fietser* 'cyclist'	-

Table 5: Annotated examples used for frame selection from P&Z'21 (Pinelli & Zanchi 2021) and B'20 (Te Brömmelstroet 2020)

Event	Participant	Group/Frame							
Femicides	*sub-event*	*murder*	*abuse*	*shooting*	*circumstances*				
	Agent	KILLING	CAUSE_HARM	ATTACK	ABUSING	RAPE	USE_FIREARM	HIT_TARGET	CAUSATION
	Patient	DEATH	DEAD_OR_ALIVE	EXPERIENCE_BODILY_HARM					
	Event	EVENT							
Unbalanced crashes	*sub-event*	*crash*	*consequences*	*circumstances*					
	Agent	CAUSE_IMPACT	CAUSE_HARM	KILLING	CAUSATION				
	Patient	EXPERIENCE_BODILY_HARM		DEATH	DEAD_OR_ALIVE	CATCH_FIRE	EMOTION_DIRECTED		
	Event	IMPACT	EVENT						

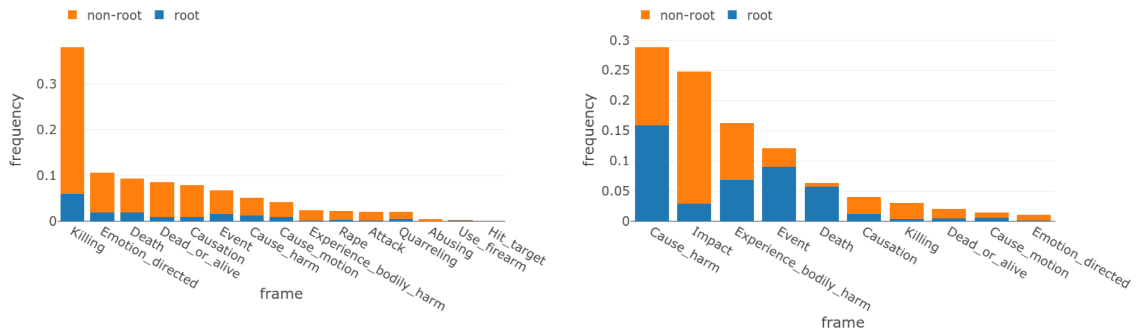
Table 6: Selected frames by agentivity and sub-event

Table 5 shows some of the annotated sentences from which typical frames were extracted. For each sentence, we manually assigned BFN frames to every word in the sentence denoting an event or state.[18] Even from this small set of examples, an interesting set of frames emerges that appear to be useful for capturing the responsibility framing-related phenomena that we would like to analyze. For example, for femicides, we find KILLING and EVENT; both of these can be used to conceptualize the murder event itself, but whereas the former evokes a situation involving both an agent and a victim, the latter describes the event as a 'bare happening' without any participants. Furthermore, we find two frames describing relevant secondary events: CAUSE_HARM denotes events where an agent physically harms a patient (potentially describing the way in which the murder was executed, or describing another act of aggression in the context of which the femicide took place), and EMOTION_DIRECTED describes emotions in relation to an experiencer (possibly the murderer or the victim), which potentially describe a (perceived) cause of the murder.

As shown in Table 6, the selected frames can be categorized along two dimensions: the part of the event that is denoted, and the level of agentivity that is encoded by the frame. For example, for femicides, KILLING denotes the murder itself and implies the presence of an agent (and a patient), whereas EXPERIENCE_BODILY_HARM denotes an event in which someone is harmed, from the perspective of the harmed person. On the other hand, for unbalanced crashes, the crash event itself could be denoted by IMPACT (without a specific focus on one of the parties) or CAUSE_IMPACT (implying a specific cause or agent), while frames such as CAUSE_HARM, DEATH, and CATCH_FIRE describe potential outcomes of the crash event from different perspectives.

The selected frames can be interpreted relative to a specific hypothesis about how social context influences how a particular event type is portrayed in public discourse. In our case studies, based on the previous literature, we are working on the hypothesis that media portrayal of femicides and car crashes are influenced by the difference in social status between a male perpetrator and a female victim, and between a car driver and a vulnerable road user, respectively, resulting in the frequent use of lexical and syntactic choices that de-emphasize the role of the more powerful party as an active agent, and portray the event as spontaneous or caused by an inanimate entity.

18. In order to restrict the complexity of the analysis, we excluded entity-denoting frames, even where they might have been relevant, e.g. WEAPON or VEHICLE.



(a) Femicides

(b) Imbalanced crashes

Figure 2: Frame frequencies (relative to the total number of typical frame instances), split by root status

Given this hypothesis and the frames that we defined, we can now use SocioFillmore to test the relative frequencies of frames with different levels of agentivity (and, within each frame, the relative frequencies of syntactic constructions implying different levels of agentivity) against our expectations, as well as to inspect annotations of specific sentences and use these as a basis for further qualitative analysis.

Data Filtering For each target domain, we also defined a set of *event filters* and *document filters* that are relevant for responsibility framing. One of these filters combines both event and document information, and is shared between the two domains: the number of days between the event itself and the publication date of the article. Also common to both domains is a filter that selects articles by their news sources and by the geographical scope (national, regional, local) of these news sources. Apart from these, each domain has its own set of event filters (e.g., for femicides, there are filters for the nationality of the victim and type of location where the crime was committed; for crashes, there are filters for which types of road users were involved and/or got injured in the event), and document filters (e.g., for femicides, there is a filter for the religious orientation of the news source; for crashes, there is a filter for whether the news source is Dutch or Flemish). For crashes, one event filter is particularly relevant: the 'imbalanced' filter, which picks out only those events involving at least one vehicle, injuring or killing at least one pedestrian or cyclist, and not injuring people in vehicles.

Frame Semantic and Syntactic Parsing Given the availability of an end-to-end frame semantic parser optimized as much as possible for the target language and dataset, actually performing the frame analysis was straightforward: we first split every article to be analyzed into sentences using the `Punkt` tokenizer in NLTK (Bird et al. 2009), and then ran the sentences through LOME one-by-one. Finally, we extracted all of the predicted instances of typical frames from the output. On the other hand, for the syntactic analysis, we used the spaCy package (particularly, the `it_core_news_md` and `nl_core_news_md` models), which provides a high-quality Universal Dependencies parsing system as well as a part-of-speech tagger for both target languages (Honnibal et al. 2020).[19]

Linguistic Reasoning Once each sentence has been parsed semantically and syntactically, the next step was to combine these two annotation layers in order to get an analysis of constructions and role expressions. We defined five construction types: *nonverbal*, *verbal:active*, *verbal:passive*, *verbal:unaccusative*[20], *verbal:impersonal*, and *verbal:reflexive*. Frame structures are assigned to one of these types on the basis of a set of heuristics taking into account part-of-speech tags (e.g.,

19. The only (minor) technical hurdle for deploying the parser was the fact that spaCy and LOME use different word tokenization systems; this was solved by automatically defining a token mapping for every sentence.
20. *Unaccusative* refers to intransitive constructions with a single patient-like argument.

	femicides				imbalanced crashes		
frame	target	pos	freq	frame	target	pos	freq
KILLING	uccisa 'killed [fem.]'	v:part	0.16	CAUSE_HARM	gewond 'wounded'	v:part	0.52
	omicidio 'homicide'	n	0.13		geraakt 'hit'*	v:part	0.26
	ucciso 'killed [masc.]'	v:part	0.09		aangereden 'hit (by vehicle)'	v:part	0.10
	l'omicidio 'the homicide'	n	0.06		geschept 'run over/hit'	v:part	0.04
	uccidere 'kill'	v:inf	0.05		zwaargewond 'gravely injured'	adj	0.03
EMOTION_DIRECTED	dolore 'pain'	n	0.10	IMPACT	aanrijding 'collision'	n	0.62
	depressione 'depression'	n	0.07		botsing 'collision/crash'	n	0.22
	lutto 'grief'	n	0.03		aangereden 'hit (by vehicle)'	v:part	0.05
	sofferenza 'suffering'	n	0.03		botste 'collided [past sg]'	v:fin	0.03
	rabbia 'anger'	n	0.02		botsten 'collided [past pl'	v:fin	0.01

Table 7: Top-5 targets for the top-2 most frequent typical frames. Frequencies are relative to the total number of instances of each frame. (N.B.: Dutch *geraakt* 'hit' is marked with '*' to indicate that it often occurs as a support verb for *gewond* 'wounded', but is erroneously tagged as evoking a frame instance of its own.)

nominal and adjectival predicates are always nonverbal), dependency information (e.g., participle verbs with 'have' auxiliaries are always active constructions), and frame information (e.g. a finite verb expressing KILLING is active, but a finite verb expressing DEATH is unaccusative).

For analyzing semantic role expressions, we try to find a path through the dependency tree from the predicate to each of its semantic roles. Once a path is found, it is classified in two different ways: by *depth*, and by *dependency label*. Depth is simply a measure of how many steps are needed; the dependency label is a single label summarizing paths of two steps are shorter. There are five possible label types: *SELF* (for zero-length paths), *arc*↓ (go one step down in the tree along *arc*), *arc*↑ (go one step up in the tree along *arc*), ↓-*arc*↓ (go one step down, then go another step down along *arc*), and ↑-*arc*↓ (go one step up, then go one step down along *arc*). See the next subsection for examples of these labels.

Web Application Finally, in order to aid our own exploratory analysis of the datasets as well as to make our system available to researchers without a technical background, we built our analysis system in the form of a web application. This application consists of an HTTP API (built with Flask[21]) for performing analyses on the corpus on request and a front-end application (built with JavaScript and jQuery[22]) that can query the API and display the results in a user-friendly way. A screenshot of the front-end interface is given in Figure A.1 in the Appendix. The interface consists of several input panels that allow the user to specify the target dataset, select frames, and select filtering options, as well as two output types: the system can compute various *descriptive statistics* on the selected events and documents, and display these in several types of figures and tables; alternatively, the user can request the system to analyze sentences from the selected part of the corpus, either by selecting particular documents, or by asking for a random sample of sentences with particular properties (e.g. particular frames, constructions, or semantic roles).

5.5 Reproducing and extending Pinelli & Zanchi

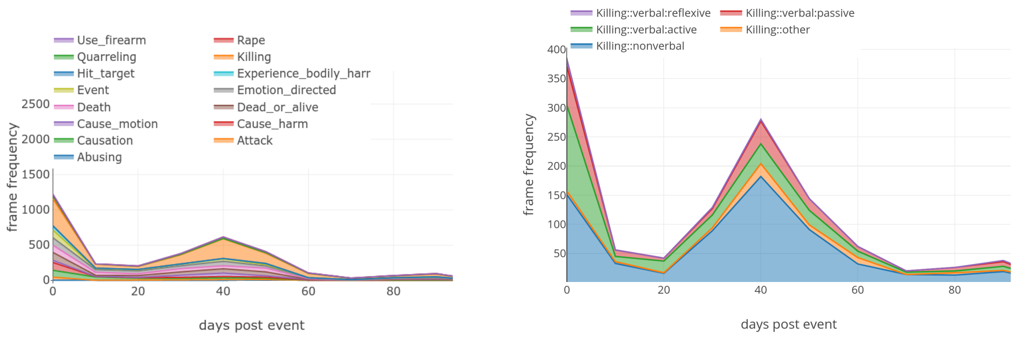
In this subsection, we apply the SOCIOFILLMORE system to produce an initial analysis of responsibility framing in our two target datasets:[23] which types of backgrounding constructions are found in the dataset, and how common are they?

Frames, Targets and Constructions The first step in our analysis is to see how frequent each of the typical frames occurs, and how they are expressed in the sentence. Figure 2 shows that, for femicides, the frame that is by far the most common is KILLING, followed at a significant distance by EMOTION_DIRECTED, DEATH, DEAD_OR_ALIVE, and CAUSATION. We also find

21. `https://flask.palletsprojects.com/en/2.0.x/`
22. `https://jquery.com/`
23. All analyses being reported on are done on the 'development' portion of the dataset.

		femicides					imbalanced crashes		
frame	pos-tag	freq	construction	freq	frame	pos-tag	freq	construction	freq
KILLING	n	0.45	nonverbal	0.53	CAUSE_HARM	v:part	0.50	nonverbal	0.48
	v:part	0.26	verbal:active	0.24		adj	0.48	verbal:passive-unacc.	0.48
	v:inf	0.08	verbal:passive-unacc.	0.15		v:fin	0.01	verbal:active	0.04
EMOTION DIRECTED	n	0.58	nonverbal	0.79	IMPACT	noun	0.86	nonverbal	0.86
	adj	0.22	verbal:passive-unacc.	0.09		v:fin	0.06	verbal:active	0.07
	v:part	0.11	verbal:active	0.05		v:part	0.06	verbal:passive-unacc.	0.06
DEATH	n	0.60	nonverbal	0.64	EXPERIENCE_BODILY_HARM	adj	0.42	nonverbal	0.60
	v:part	0.25	verbal:unaccusative	0.33		v:part	0.22	verbal:passive-unacc.	0.22
	v:fin	0.06	verbal:active	0.02		n	0.19	verbal:active	0.17
DEAD_OR_ALIVE	v:part	0.49	verbal:unaccusative	0.51	EVENT	v:fin	0.67	verbal:impersonal	0.84
	n	0.22	nonverbal	0.46		v:part	0.21	verbal:active	0.10
	adj	0.20	verbal:passive	0.02		v:fin	0.08	nonverbal	0.05
CAUSATION	v:part	0.42	verbal:active	0.58	DEATH	v:part	0.67	verbal:unaccusative	0.96
	n	0.13	nonverbal	0.16		v:fin	0.28	nonverbal	0.02
	other	0.12	other	0.12		n	0.02	other	0.01

Table 8: Syntactic expression of the most frequent typical frames



(a) typical frame frequency over time (b) KILLING over time, split by constructions

Figure 3: Femicides: frame usage over time (0-90 days post event, grouped by 10-day periods)

that, for each of these frames, the large majority of instances are expressed by non-root predicates. On the other hand, for imbalanced crashes, the most frequent frames are CAUSE_HARM, IMPACT, EXPERIENCE_BODILY_HARM, and EVENT. Interestingly, here, there are large differences with respect to root status, with some frames overwhelmingly being expressed by root predicates (CAUSE_HARM, EVENT) but not others (IMPACT). However, in general, the typical frames in the crashes domain seem to be overall much more foregrounded with respect to root status compared to their counterparts in the femicides domain.

Taking a more concrete look at the predicates expressing the typical frames, in Table 7, we find that the two most common femicide frames are evoked by a wide variety of targets (e.g., the most common predicate for KILLING accounts for only 16% of instances of this frame), whereas the two most frequent crashes frames are each dominated by a single predicate. Zooming out, in Table 8, we find that most of the typical frames in both domains are dominated by nominal, adjectival, and participial predicates, which is also reflected by the fact that nonverbal, passive, and unaccusative constructions are most common, indicating a high level of backgrounding. Strikingly, the only frequent frame for which finite verbs are most frequent is EVENT in the crashes domain, but these usually express impersonal constructions (“it happened on a Tuesday . . . ”) which also background the event participants. However, there are a few frames with a substantial majority of active constructions, most notably KILLING in the femicides domain, and CAUSATION even has a majority of active constructions.

Finally, it is interesting to take a look at how frames are expressed differently over time. Whereas in the crashes dataset, we tend to have only one or a few articles covering each event,

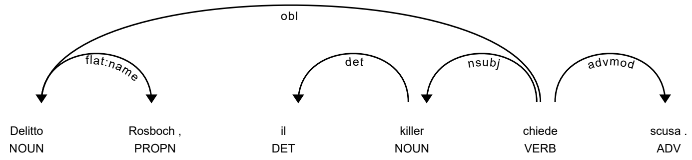
	femicides					imbalanced crashes			
frame (cx)	**agent/cause**	**freq**	**victim**	**freq**	**frame (cx)**	**agent/cause**	**freq**	**victim**	**freq**
	Killer	*0.20*	*Victim*	*0.40*	CAUSE_	*Agent*	*0.00*	*Victim*	*1.00*
KILLING	dep:SELF	0.10	dep:UNK	0.20	HARM			dep:UNK	0.74
(nonverbal)	dep:UNK	0.05	dep:nmod↓	0.10	(nonverbal)			dep:nsubj↓	0.22
	dep:amod↓	0.01	dep:amod↑	0.07				dep:amod↑	0.01
	Killer	*0.71*	*Victim*	*0.87*	CAUSE_	*Agent*	*0.14*	*Victim*	*1.00*
KILLING	dep:UNK	0.30	dep:obj↓	0.65	HARM	dep:obl:agent↓	0.11	dep:nsubj↓	0.52
(active)	dep:nsubj↓	0.26	dep:SELF	0.06	(passive-	dep:obl↓	0.01	dep:nsubj:pass↓	0.38
	dep:aux↓	0.03	dep:UNK	0.05	unacc.)	dep:↓-nmod↓	0.01	dep:↓-nsubj-↓	0.04
	Explanation	0.03	*Protagonist*	0.70	EXPERIENCE	*Injuring_entity*	*0.02*	*Experiencer*	*0.86*
DEATH	dep:nmod↓	0.01	dep:nmod↓	0.55	BODILY_	dep:UNK	0.02	dep:UNK	0.62
(nonverbal)	dep:SELF	0.01	dep:UNK	0.07	HARM			dep:nsubj↓	0.15
	dep:UNK	0.01	dep:amod↑	0.02	(nonverbal)			dep:nmod:poss↓	0.04
	Explanation	0.00	*Protagonist*	0.94		*Explanation*	0.49	*Protagonist*	1.00
DEATH			dep:nsubj↓	0.38	DEATH	dep:obl↓	0.49	dep:nsubj↓	0.90
(unaccusative)			dep:acl↑	0.20	(unaccusative)			dep:UNK	0.04
			dep:UNK	0.17				dep:↓-appos↓	0.02

Table 9: Frequency of Agent-, Cause-, and Victim-like role labels, and their syntactic expression using UD labels, for the most frequent typical frames and constructions. All frequencies are expressed relative to the total number of instances of each frame.

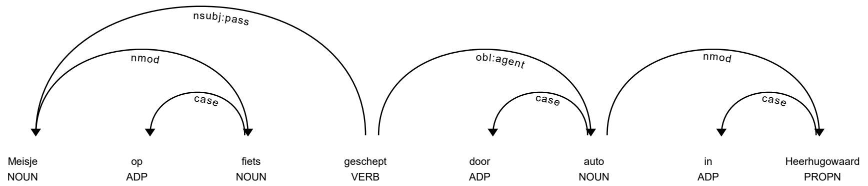
usually published shortly after the event itself, for femicides there are an average of 25 articles for each event, often spread out over a long period of time (covering the murder itself, various phases of the criminal investigation, the trial, etc.). Looking at the evaluation of typical frames over time is important because all of our selected typical frames pertain to the murder event itself, and can be expected to be expressed in a more foregrounded way shortly after the event but less so afterwards. This is indeed what we find: in Figure 3, subfigure (a) shows that, overall, most typical frame instances occur in articles published in the first week after the murder, followed by a fast decline and then a second peak 40 days later. Interestingly, in both of these peaks, the KILLING frame dominates, whereas its relative frequency decreases in the period in between the two peaks. In subfigures (b), we get a clear idea of how backgrounding changes over time: active constructions are relevatively much more dominant in the first peak (accounting for almost 40% of instances during the first 10-day period), and are much rarer in the second peak (12% active constructions).

Frames and Semantic Roles Shifting our focus from frames and predicates towards semantic roles, in Table 9 we observe that frames and constructions differ greatly in the extent to which they explicitly express core event participants. In general, roles corresponding to victims are expressed much more frequently than those corresponding to agents or causes, and, as expected, verbal constructions more frequently express participants than nonverbal ones. For example, for KILLING in the femicides domain, nonverbal constructions have a Victim role 40% of the time, but have a Killer only 20% of the time. If we look at active constructions, these numbers rise to 87% and 71%, respectively, but they still greatly 'favor' the victim. On the other hand, moving to the crashes domain, we find very few instances of constructions expressing overt agents or causes, with the interesting exception of DEATH, which has an Explanation role half of the time. However, in these cases, the attributed 'cause' does not always refer directly to the crash, e.g., we find constructions like "he later died from his wounds". The only frequent frame with an Agent role is CAUSE_HARM, but it is important to note that this role can be filled either by a vehicle ("hit by a car") or by a person ("hit by a car driver"), leaving room for a different type of backgrounding that involves conceptual metonymy: the Agent is not evoked directly, but by mentioning another entity belonging to their same conceptual domain.

Looking at the dependencies with which semantic roles are expressed, we find a lot of variation across frames and constructions, but a general observation is that the UNK dependency, indicating that there is no path of two steps or shorter between the predicate and the role, is fairly common; this too can be seen as expressing backgrounding of participants. Also, for KILLING in femicides, agents of nonverbal constructions are frequently realized with the SELF dependency,



(a) Example of "dep:SELF" for a KILLING instance: *"Delitto Rosboch, il killer chiede scusa"* ("Rosboch crime, the killer apologizes"). Frame target: *"killer"*, Killer: *"killer"*.



(b) Example of "dep:nsubj:pass↓" and "dep:obl:agent↓" for a CAUSE_HARM instance: *"Meisje op fiets geschept door auto in Heerhugowaard"* ("Girl on bike hit by a car in Heerhugowaard [city]"). Frame target: *"geschept"*, Agent: *"door een auto"*, Victim: *"meisje op fiets"*.

Figure 4: Examples of dependency relations between frames and targets

i.e., by agentive nominals (e.g. *assassino* 'murderer'). While such constructions might be taken to be foregrounding the perpetrator, they paradoxically background him at the same time because they leave little room for the overt expression of a more concrete description of the killer. Otherwise, the most frequent dependencies generally follow the expected syntactic patterns: agent-like arguments of active constructions and patient-like arguments of unaccusative and passive constructions tend to be expressed as subjects, patients of active constructions tend to be objects, and roles of nonverbal constructions are often expressed by nominal (nmod) or adjectival (amod) modifiers.

Some examples of expressions of core participants are shown in Figure 4: in subfigure (a), we find a instance of the KILLING frame expressed by an agentive nominal (*killer*), with the predicate and the Killer role being expressed by the same token. On the other hand, subfigure (b) shows an instance of CAUSE_HARM expressed by a passive construction. Both core participants are expressed: the victim is expressed as a subject, whereas the 'agent' (conceptualized as a vehicle rather than as the driver) is expressed by an oblique role.

Discussion In this subsection, we have used SOCIOFILLMORE to perform a preliminary statistical analysis of agentivity backgrounding in media reports of femicides and imbalanced crashes. A natural question is what we can conclude from this analysis: do the data match the expectations that we had based on (Pinelli and Zanchi 2021) and (Te Brömmelstroet 2020)? At first glance, it seems that they largely do: while, for both femicides and unbalanced crashes, the overall most frequent frame is agent-focused, agent-backgrounding constructions and role configurations are the majority within each frame. Thus, a large proportion of frame instances for both event types is analyzed by SOCIOFILLMORE as exhibiting some degree of agent backgrounding. However, from our preliminary analysis it is difficult to draw any hard conclusions as several important questions are still unanswered: for example, when can we say that a particular type of frame or construction is 'frequent' or 'dominant'? Should we consider the analysis on a specific corpus on its own or relative to some baseline (e.g. framing of femicides vs. framing of other kinds of murder)? How do we weigh the contributions of different dimensions of backgrounding (frames, constructions, role configurations, dependencies, etc.) against each other? There might not be definitive answers to all of these questions, but future work using SOCIOFILLMORE is likely to benefit from trying to make the hypotheses to test as precise as possible.

References

Alesiani, Giulia and Sabino Metta (2021), Il racconto del femminicidio in Italia. Un approccio data-driven, *in* Belluati, Marinella, editor, *Femminicidio. Una lettura tra realtà e rappresentazione*, Biblioteca di testi e studi, Carocci editore, chapter 4, pp. 77–93.

Baker, Collin F., Charles J. Fillmore, and Beau Cronin (2003), The structure of the FrameNet database, *International Journal of Lexicography* **16** (3), pp. 281–296, Oxford University Press.

Basili, Roberto, Diego De Cao, Alessandro Lenci, Alessandro Moschitti, and Giulia Venturi (2013), Evalita 2011: The frame labelingover italian texts task, *in* Magnini, Bernardo, Francesco Cutugno, Mauro Falcone, and Emanuele Pianta, editors, *Evaluation of Natural Language and Speech Tools for Italian*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 195–204.

Baumer, Eric, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay (2015), Testing and comparing computational approaches for identifying the language of framing in political news, *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Denver, Colorado, pp. 1472–1482. https://aclanthology.org/N15-1171.

Bird, Steven, Ewan Klein, and Edward Loper (2009), *Natural language processing with Python: analyzing text with the natural language toolkit*, " O'Reilly Media, Inc.".

Bohner, Gerd (2001), Writing about rape: Use of the passive voice and other distancing text features as an expression of perceived responsibility of the victim, *British Journal of Social Psychology* **40** (4), pp. 515–529, Wiley Online Library.

Busso, Lucia, Claudia Roberta Combei, and Ottavia Tordini (2020), Narrating gender violence a corpus-based study on the representation of gender-based violence in italian media, *in* Giusti, Giuliana and Gabriele Iannàccaro, editors, *Language, Gender and Hate Speech*, Language, Gender and Hate Speech A Multidisciplinary Approach, Fondazione Università Ca' Foscari. `https://publications.aston.ac.uk/id/eprint/42283/`.

Chilton, Paul (2004), *Analysing political discourse: Theory and practice*, Routledge:London.

Chilton, Paul Anthony (1996), *Security metaphors: Cold war discourse from containment to common house*, Vol. 2, Peter Lang Pub Incorporated.

Chomsky, Noam (1957), *Syntactic Structures*, Mouton & Co., The Hague.

Chomsky, Noam (1965), *Aspects of the theory of syntax*, MIT Press, Cambridge, Massachusetts.

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov (2020), Unsupervised cross-lingual representation learning at scale, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, pp. 8440–8451. https://aclanthology.org/2020.acl-main.747.

Conte, Rosaria, Nigel Gilbert, Giulia Bonelli, Claudio Cioffi-Revilla, Guillaume Deffuant, Janos Kertesz, Vittorio Loreto, Suzy Moat, J-P Nadal, Anxo Sanchez, et al. (2012), Manifesto of computational social science, *The European Physical Journal Special Topics* **214** (1), pp. 325–346, Springer.

Culver, Gregg (2018), Death and the car: On (auto) mobility, violence, and injustice, *ACME: An International Journal for Critical Geographies* **17** (1), pp. 144–170.

De Clercq, Orphée, Veronique Hoste, and Paola Monachesi (2012), Evaluating automatic cross-domain Dutch semantic role annotation, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, pp. 88–93. http://www.lrec-conf.org/proceedings/lrec2012/pdf/680_Paper.pdf.

Entman, Robert M (1993), Framing: Towards clarification of a fractured paradigm, *Journal of Communication*.

Fairclough, Norman (2010), *Critical Discourse Analysis: The Critical Study of Language (Second Edition)*, Routledge.

Fillmore, Charles J. (1968), The case for case, *in* Bach, E. and R. T. Harms, editors, *Universals in linguistic theory*, Holt, Rhinehart, and Winston, New York, pp. 1–88.

Fillmore, Charles J. (1971), Subjects, speakers, and roles, *Synthese*, Springer.

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

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

Fuoli, Matteo and Christopher Hart (2018), Trust-building strategies in corporate discourse: An experimental study, *Discourse & Society* **29** (5), pp. 514–552, SAGE Publications Sage UK: London, England.

Gildea, Daniel and Daniel Jurafsky (2002), Automatic labeling of semantic roles, *Computational Linguistics* **28** (3), pp. 245–288. `https://doi.org/10.1162/089120102760275983`.

Goddard, Tara, Kelcie Ralph, Calvin G Thigpen, and Evan Iacobucci (2019), Does news coverage of traffic crashes affect perceived blame and preferred solutions? evidence from an experiment, *Transportation research interdisciplinary perspectives* **3**, pp. 100073, Elsevier.

Goldberg, Adele E. (1995), *Constructions: A construction grammar approach to argument structure*, University of Chicago Press, Chicago.

Halliday, Michael Alexander Kirkwood (1978), *Language as Social Semiotic*, Edward Arnold.

Hart, Christopher (2010), *Critical discourse analysis and cognitive science: New perspectives on immigration discourse*, Springer.

Hart, Christopher (2011a), Force-interactive patterns in immigration discourse: A cognitive linguistic approach to cda, *Discourse & Society* **22** (3), pp. 269–286, SAGE Publications Sage UK: London, England.

Hart, Christopher (2011b), Moving beyond metaphor in the cognitive linguistic approach to CDA, *Critical discourse studies in context and cognition* **43**, pp. 171–192, Palgrave Macmillan London, England.

Hart, Christopher (2013), Event-construal in press reports of violence in two recent political protests: A cognitive linguistic approach to cda, *Journal of Language and Politics* **12** (3), pp. 400–423, John Benjamins.

Hart, Christopher (2015), Viewpoint in linguistic discourse: Space and evaluation in news reports of political protests, *Critical Discourse Studies* **12** (3), pp. 238–260, Taylor & Francis.

Hart, Christopher (2018a), Event-frames affect blame assignment and perception of aggression in discourse on political protests: An experimental case study in critical discourse analysis, *Applied Linguistics* **39** (3), pp. 400–421, Oxford University Press.

Hart, Christopher (2018b), 'Riots engulfed the city': An experimental study investigating the legitimating effects of fire metaphors in discourses of disorder, *Discourse & Society* **29** (3), pp. 279–298, SAGE Publications Sage UK: London, England.

Hart, Christopher (2021), Animals vs. armies: Resistance to extreme metaphors in anti-immigration discourse, *Journal of Language and Politics* **20** (2), pp. 226–253, John Benjamins Publishing Company Amsterdam/Philadelphia.

Hart, Christopher and Matteo Fuoli (2020), Objectification strategies outperform subjectification strategies in military interventionist discourses, *Journal of Pragmatics* **162**, pp. 17–28, Elsevier.

Hartmann, Silvana, Ilia Kuznetsov, Teresa Martin, and Iryna Gurevych (2017), Out-of-domain FrameNet semantic role labeling, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Association for Computational Linguistics, Valencia, Spain, pp. 471–482. https://aclanthology.org/E17-1045.

Haynes, John (1989), *Introducing stylistics*, Allen & Unwin Australia.

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

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd (2020), spaCy: Industrial-strength Natural Language Processing in Python. `https://doi.org/10.5281/zenodo.1212303`.

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

Kabbach, Alexandre (2019), Debugging frame semantic role labeling, *CoRR*. http://arxiv.org/abs/1901.07475.

Kabbach, Alexandre and Aurélie Herbelot (2021), Avoiding conflict: When speaker coordination does not require conceptual agreement, *Frontiers in Artificial Intelligence* **3**, pp. 95. https://www.frontiersin.org/article/10.3389/frai.2020.523920.

Kingma, Diederik P. and Jimmy Ba (2017), Adam: A method for stochastic optimization.

Koehn, Philipp (2005), Europarl: A parallel corpus for statistical machine translation, *Proceedings of the MT Summit*.

Lakoff, George (2014), *The all new don't think of an elephant!: Know your values and frame the debate*, Chelsea Green Publishing.

Lakoff, George and Mark Johnson (1980), *Metaphors We Live By*, University of Chicago Press, Chicago.

Langacker, Robert W. (1991), *Foundations of Cognitive Grammar, vol. II: Descriptive application*, Stanford University Press, Stanford.

Langacker, Ronald W. (2019), Chapter 6: Construal:, *in* Dąbrowska, Ewa and Dagmar Divjak, editors, *Cognitive Linguistics - Foundations of Language*, De Gruyter Mouton, pp. 140–166. `https://doi.org/10.1515/9783110626476-007`.

Lenci, Alessandro, Simonetta Montemagni, Giulia Venturi, and Maria Grazia Cutrullà (2012), Enriching the ISST-TANL corpus with semantic frames, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, pp. 3719–3726. http://www.lrec-conf.org/proceedings/lrec2012/pdf/986_Paper.pdf.

Marín Arrese, Juana (2011), Effective vs. epistemic stance and subjectivity in political discourse: Legitimising strategies and mystification of responsibility, *Critical discourse studies in context and cognition. Amsterdam: John Benjamins* pp. 193–224.

Matthes, Jörg (2012), Framing politics: An integrative approach, *American behavioral scientist* **56** (3), pp. 247–259, Sage Publications Sage CA: Los Angeles, CA.

Mendelsohn, Julia, Ceren Budak, and David Jurgens (2021), Modeling framing in immigration discourse on social media, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Online, pp. 2219–2263. https://aclanthology.org/2021.naacl-main.179.

Minnema, Gosse (2021), Kicktionary-LOME: A domain-specific multilingual frame semantic parsing model for football language, *CoRR*. `https://arxiv.org/abs/2108.05575`.

Minnema, Gosse and Malvina Nissim (2021), Breeding Fillmore's chickens and hatching the eggs: Recombining frames and roles in frame-semantic parsing, *Proceedings of the 14th International Conference on Computational Semantics*. `https://iwcs2021.github.io/proceedings/iwcs/pdf/2021.iwcs-1.15.pdf`.

O'Halloran, Kieran (2003), *Critical discourse analysis and language cognition*, Edinburgh University Press.

Oostdijk, Nelleke, Martin Reynaert, Paola Monachesi, Gertjan Van Noord, Roeland Ordelman, Ineke Schuurman, and Vincent Vandeghinste (2008), From D-coi to SoNaR: a reference corpus for Dutch, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco. http://www.lrec-conf.org/proceedings/lrec2008/pdf/365_paper.pdf.

Pinelli, Erica 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* p. 117, Springer.

Postma, Marten, Levi Remijnse, Filip Ilievski, Antske Fokkens, Sam Titarsolej, and Piek Vossen (2020), Combining conceptual and referential annotation to study variation in framing, *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, European Language Resources Association, Marseille, France, pp. 31–40. https://aclanthology.org/2020.framenet-1.5.

Radford, Jason and Kenneth Joseph (2020), Theory in, theory out: The uses of social theory in machine learning for social science, *Frontiers in Big Data* **3**, pp. 18. https://www.frontiersin.org/article/10.3389/fdata.2020.00018.

Ralph, Kelcie, Evan Iacobucci, Calvin G Thigpen, and Tara Goddard (2019), Editorial patterns in bicyclist and pedestrian crash reporting, *Transportation research record* **2673** (2), pp. 663–671, SAGE Publications Sage CA: Los Angeles, CA.

Remijnse, Levi and Gosse Minnema (2020), Towards reference-aware FrameNet annotation, *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, European Language Resources Association, Marseille, France, pp. 13–22. https://aclanthology.org/2020.framenet-1.3.

Remijnse, Levi, Marten Postma, and Piek Vossen (2021), Variation in framing as a function of temporal reporting distance, *Proceedings of the 14th International Conference on Computational Semantics*. `https://iwcs2021.github.io/proceedings/iwcs/pdf/2021.iwcs-1.22.pdf`.

Santaemilia, José and Sergio Maruenda (2014), The linguistic representation of gender violence in (written) media discourse: The term 'woman'in spanish contemporary newspapers, *Journal of Language Aggression and Conflict* **2** (2), pp. 249–273, John Benjamins.

Semetko, Holli A and Patti M Valkenburg (2000), Framing european politics: A content analysis of press and television news, *Journal of communication* **50** (2), pp. 93–109, Wiley Online Library.

Swayamdipta, Swabha, Sam Thomson, Chris Dyer, and Noah A. Smith (2017), Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold, *CoRR*. `http://arxiv.org/abs/1706.09528`.

Te Brömmelstroet, Marco (2020), Framing systemic traffic violence: Media coverage of dutch traffic crashes, *Transportation research interdisciplinary perspectives*, Elsevier.

Tonelli, Sara and Emanuele Pianta (2008), Frame information transfer from English to Italian, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco. http://www.lrec-conf.org/proceedings/lrec2008/pdf/567_paper.pdf.

Torrent, Tiago T., Collin F. Baker, Oliver Czulo, Kyoko Ohara, and Miriam R. L. Petruck, editors (2020), *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, European Language Resources Association, Marseille, France. https://aclanthology.org/2020.framenet-1.0.

Tranchese, Alessia and Sole Alba Zollo (2013), The construction of gender-based violence in the british printed and broadcast media., *Critical Approaches to Discourse Analysis Across Disciplines*.

Van Dijk, Teun A. (1995), Discourse analysis as ideology analysis, *in* Schäffner, C. and A.I. Wenden, editors, *Language and Peace*, Harwood Academic Publishers, Amsterdam.

Van Dijk, Teun A. (1998), *Ideology*, SAGE, London.

Van Dijk, Teun A. (2015), Critical discourse analysis, *in* Tannen, Deborah, Heidi E. Hamilton, and Deborah Schiffrin, editors, *The Handbook of Discourse Analysis*, Wiley-Blackwell, Chichester, West Sussex, pp. 466–485.

Vossen, Piek, Antske Fokkens, Isa Maks, and Chantal van Son (2018a), Towards an Open Dutch FrameNet lexicon and corpus.

Vossen, Piek, Filip Ilievski, Marten Postma, and Roxane Segers (2018b), Don't annotate, but validate: a data-to-text method for capturing event data, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, European Language Resources Association (ELRA), Miyazaki, Japan. https://aclanthology.org/L18-1480.

Vossen, Piek, Filip Ilievski, Marten Postma, Antske Fokkens, Gosse Minnema, and Levi Remijnse (2020), Large-scale cross-lingual language resources for referencing and framing, *Proceedings of the 12th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, pp. 3162–3171. https://aclanthology.org/2020.lrec-1.387.

World Health Organization (2018), Global status on road safety. https://www.who.int/publications/i/item/9789241565684.

Xia, Patrick, 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, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, Online, pp. 149–159. https://aclanthology.org/2021.eacl-demos.19.

Source	Main		Development	
	# Docs	**# Unique Events**	**# Docs**	**# Unique Events**
AGI	76	36	22	6
ANSA.it	502	127	145	19
Affaritaliani.it	10	5	6	2
Agenzia Stampa Italpress	1	1	–	–
America Oggi	4	3	2	2
Avvenire	6	4	1	1
Corriere del Mezzogiorno	52	19	13	3
Direattanews.it	2	2	–	–
Il Corriere della Sera	179	69	45	11
Il Fatto Quotidiano	25	13	8	5
Il Giornale	7	4	2	1
Il Mattino	43	13	20	3
Il Messaggero	32	8	2	2
Il Secolo XIX	98	45	29	5
Italian-News.it	1	1	–	–
L'Unione Sarda	21	17	6	2
La Gazzetta del Mezzogiorno	5	4	2	1
La Gazzetta dello Sport	1	1	–	–
La Repubblica	379	119	150	18
La Stampa	132	64	93	5
La Voce	21	15	4	4
Leggo.it	–	–	1	1
NewNotizie.it	37	22	6	2
OMNIMILANO	3	3	1	1
OMNIROMA	12	4 –	–	
RAI.it	94	30	27	5
Rassegna.it	2	2	–	–
SKY.it	36	21	10	4
TGCOM	271	110	52	16
TM News	6	3	1	1
Tiscali	24	14	4	2
total	2,082	815	652	122

Figure A.1: `RAI-F`- Corpus overview.

Ziem, Alexander, Christian Pentzold, and Claudia Fraas (2018), Medien-Frames als semantische Frames: Aspekte ihrer methodischen und analytischen Verschränkung am Beispiel der 'Snowdon-Affäre', *in* Alexander Ziem, Detmer Wulf, Lars Inderelst, editor, *Frames interdisziplinär: Modelle, Anwendungsfelder, Methoden*, DUP, Düsseldorf, pp. 155–184.

Appendix A. Appendix

A.1 `RAI-F`: Source overview and distribution

Full list of the sources for the `RAI-F`, including number of documents and number of unique events per split (Main and Development).

frame	**role:perpetrator_like**	**role:victim_like**	**role:cause_like**
Abusing	Abuser	Victim	-
Attack	Assailant	Victim	-
Causation	Causer	Affected	Cause
Cause_harm	Agent	Victim	Cause
Cause_motion	-	-	-
Dead_or_alive	-	Protagonist	Explanation
Death	-	Protagonist	Cause
Emotion_directed	-	-	-
Event	-	-	-
Experience_bodily_harm	Experiencer|Body_part	-	-
Hit_target	Agent	Target	-
Killing	Killer	Victim	Cause
Quarreling	-	-	-
Rape	Perpetrator	Victim	-
Use_firearm	Agent	Goal	-

Figure A.2: Femicides: mapping frames, participants, and roles

frame	**role:perpetrator_like**	**role:victim_like**	**role:cause_like**
Catch_fire	-	-	-
Causation	Causer	Affected	Cause
Cause_harm	Agent	Victim	Cause
Cause_motion	-	-	-
Dead_or_alive	-	Protagonist	Explanation
Death	-	Protagonist	Cause
Emotion_directed	-	-	-
Event	-	-	-
Experience_bodily_harm	Experiencer|Body_part	-	-
Impact	Impactor	Impactee	-
Killing	Killer	Victim	Cause

Figure A.3: Crashes: mapping frames, participants, and roles

A.2 SocioFillmore Hypothesis System: mapping frames and roles

Full list of relevant frames and mappings between frame elements (semantic roles) and core event participants.

A.3 SocioFillmore Web Interface

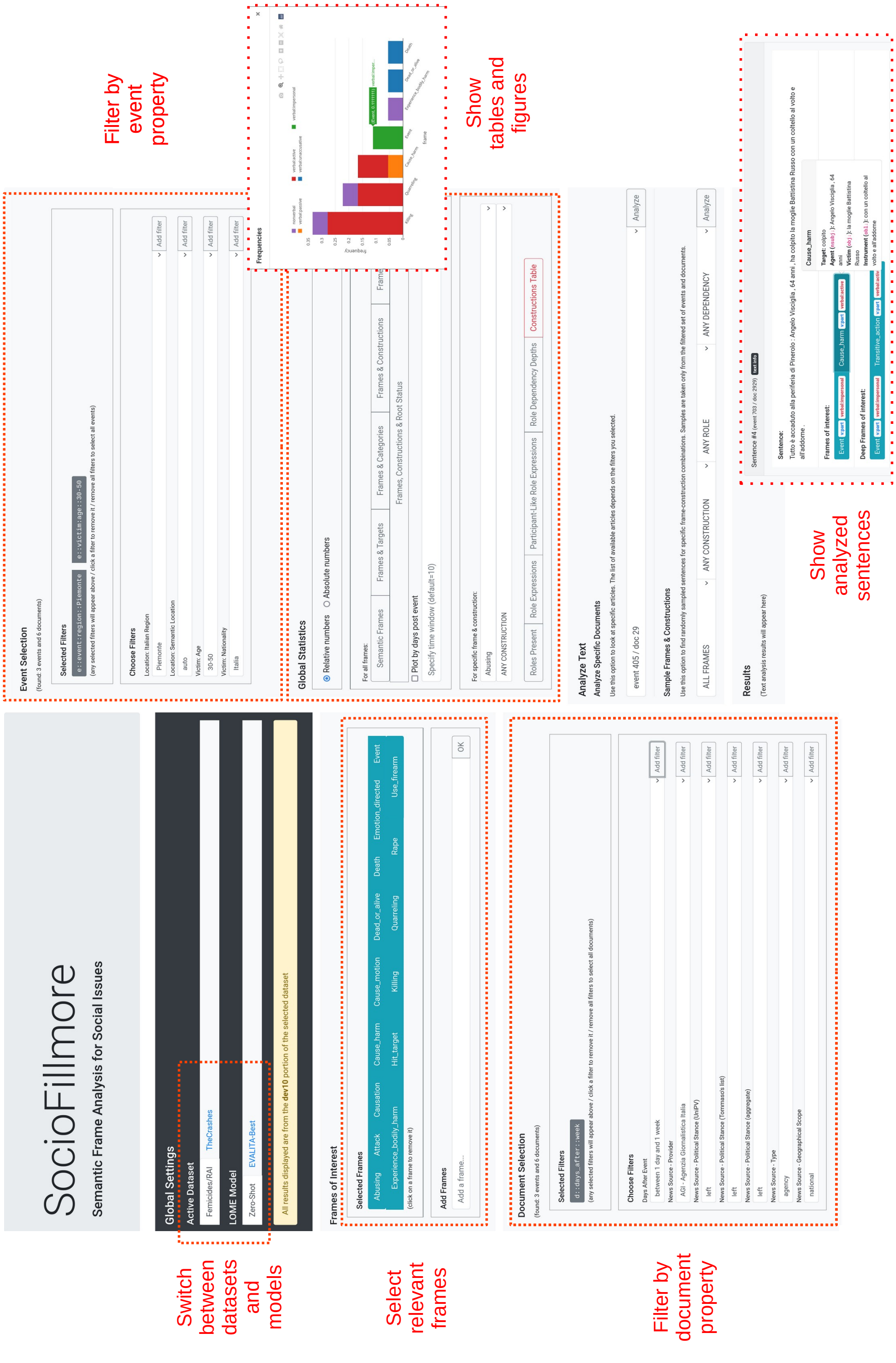


Figure A.1: SocioFillmore Screenshot