One Event, Different Stories

Towards the Automatic Identification of Framing Differences

Upgrade Report

Martino Mensio

Supervisors:
Harith Alani
Alistair Willis

The Open University
United Kingdom
June 2020

Abstract

News media is the way that we get informed about information every day. We read the news to know what happens and when we do that, not only do we receive the facts, but also the interpretation of the writer. This phenomenon is called *framing* and different works have analysed it.

With this PhD, we want to tackle instances of framing that are difficult to recognise. These can manifest subtly through word choices and selection of what to include or exclude from the narrative line.

We present a method that could help both human readers and automated approaches to find more easily when such framing techniques occur in the news. We want to provide an analysis of articles, by exploiting the differences in how information is presented by different sources. We can then analyse how different sources in the media landscape are using framing in different ways.

Contents

List of Figures

1 Introduction

Thousands of articles are published every single day about the latest events happening. The usage of specific language, the selection of details and how the narrative is presented, are all different aspects that are unique of each news outlet and author. And these peculiarities, at the same time, can influence what the reader perceives about the events described. This whole set of information, may it be subtle or explicit, extends the raw facts that happen and is usually called *framing* (Gamson & Modigliani 1989, Scheufele 1999). Framing can be created with any subjective statements that are mixed with the description of the event narrated, but does not stop there. Even selecting which details or features to report makes a big difference in the message sent to the reader. For example, taking two sentences "*the black man was shot*" vs "*the man was shot*", they have a different framing because, although the man shot was black, it is a judgement of the reporter whether this detail needs to be emphasised or not, considering the ethnicity of the person who was shot to be important. Framing can have big impacts on the way readers perceive the content and relevance of the news (Cohen 2015).

Spotting the occurrence of framing is therefore a very difficult task, even for humans (Morstatter et al. 2018). Something that could help in this situation is looking at different sources and analyse how they present the same event with different framing. By seeing "the other sides" of the story we, as readers, could create a more complete picture and spot the differences at the macro (the perspective of the overall article) and micro-level (specific linguistic cues) (Gamson & Modigliani 1989). The main problem of this technique is that it requires a lot of time, and many people only read news articles superficially (Pennycook & Rand 2019). At the moment, we can see a gap in the tools available to provide this functionality automatically. Some technologies analyse parts of the problem, e.g. by grouping articles together by events (news aggregators), or anti-plagiarism tools that spot sentences occurring in multiple documents, or theoretical studies and conceptualisations analysing framing under different aspects. But in our knowledge, none of them is bringing together different stories to highlight the framing differences in an automatic way.

This PhD aims at creating a methodology to extract and characterise framing differences among news articles. This includes on one side revealing the choices done by the authors, and bring to the light the types of techniques they use to stand their point of view (e.g., selection and emphasis of details, addition of subjective content). And on the other side, to study the information flow between sources and see which relationships exist between them. Given this aim, we target two Research Questions:

- RQ1: How can we automatically reveal the framing differences in articles presenting the same event?
- RQ2: How do news sources with different characteristics change stories over time?

The document is structured as follows. In Chapter 2 we are providing a re-

view of the literature available about framing and the comparison of multiple documents. Then, after revisiting and expanding the Research Questions in Chapter 3, in Chapter 4 we describe the methodology we plan to develop to tackle the research questions. In Chapter 5 we provide a timeline for the project, describing early achievements and a plan for the next two years of research.

2 Literature Review

Given the aims identified in the introduction, this chapter investigates the existing work that can help us understand the *framing* phenomenon, and how it is detected by automatic approaches, usually performed on single articles. But, as mentioned previously, framing can be very subtle when it manifests through techniques such as focusing on certain aspects, or selecting which details to include or omit, or even selecting a specific order inside a story (Morstatter et al. 2018). For these specific cases, it is very complicated to spot the signals of framing taking place, and we argue that both human and automated analysis would benefit from comparing different articles that talk about the same event but have different perspectives.

This multi-article knowledge requires introducing a second family of methods that aims at finding *similarities* between articles, by using document and sentence representations, and to find which pieces appear in multiple documents. In this way the common ground can be identified, as well as the diversity in the information reported. With this second family of methods, we can see that for a certain event there are several articles which overlap in part, and some of them have unique pieces or specific words that differ from the others. But this analysis needs an interpretation framework that can describe the framing role of the choices done by each of the sources.

Exploring these two different fields of research, we can see that there is an intersection area that is not much explored, that is the analysis of framing differences. Given the differences identified by this second group of methods, we need to expand the framing analysis to work with multiple documents and understand why modifications in the text happen. Some works analyse this specific field, but using manual handcrafted analysis, that for example shows how events are described from sources belonging to different political leanings. We want to target this specific area with automated analysis.

For these reasons, this chapter is structured in three parts: framing analysis of single documents (Section 2.1), similarities between multiple documents (Section 2.2) and finally the point of contact between the two (Section 2.3), as our target niche.

2.1 Framing

As we are using the term *framing*, we need to define what we mean with it, because multiple definitions do exist. The broadest definition comes from Goffman (1974) who defines frames as the cognitive maps or patterns of interpretation that people use to organise their understandings of reality. Then we have the Frame Semantics (Fillmore 2006) that instead defines a frame as "*any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits*". This second definition goes into the direction of associating sentences with units of meaning (semantics) that are evoked by specific words, e.g. FrameNet (Baker et al. 1998).

But the definition that we rely on, instead, focuses on the power of mass media to present facts under a specific light: instead of focusing on *what* is contained in a text (semantics), we focus on *how* it is described. This facet of the term can be seen in the work of Entman (1993) that defines framing as "*select[ing] some aspects of a perceived reality and make them more salient in a communicating text*". Or from Goffman (1974) that describes framing as "*how a certain story is presented to shape mass opinion*". Using this connotation, framing becomes an additional layer on top of the factual information focusing on the presentation side, giving a structure (identifying problems, causes, resolutions) where each of the semantic elements is given a role to create a coherent story (Pan & Kosicki 1993).

Framing is used every day to influence the readers and make some messages pass more than others, and becomes very important in the political sphere where it is used as a propaganda technique to promote or demote candidates and ideologies by creating and sustaining narrative lines that coherently align with a specific communication goal. Although this is a very common use of framing, in a day by day communication there is a lot of framing that can influence the readers, even though there is not a clear propaganda target.

2.1.1 Framing techniques

Media framing, under the theoretical perspective, is composed of several concepts and processes (Scheufele & Tewksbury 2007). *Frame building* is the process of creating a narrative in which the articles can fit. It corresponds to defining the objectives of the communication in which several articles can be fitted coherently. Under this point of view, the frames are the output of the societal, organisational, ideological, political and cultural contexts (Scheufele 1999). Then there is the process of *frame setting* that instead takes the frames as inputs and describes the transfer from the writer to the readers. How the document has been written creates the *audience frames*: the reader is placed in a certain mental perspective. For this to happen, a big role is played by the selection of media coverage (which issues to cover or not, how to use emphasis) (Iyengar 1994) and the specific use of language that conveys how to think and feel about an issue (Bryant et al. 2012). Inside this process we can place the quite known *agenda-setting* theory (McCombs & Shaw 1972) that describes the ability of news media to influence the importance of the topics, by selecting the information for the reader (what to think about, not how): the number of articles that are published about each event is an indicator of this process, which can have an avalanche effect as lots of outlets want to have their piece on the same event too. But the agenda-setting can be seen also at a more fine-grained level: stories are made of details and different sources can emphasise them with different weights.

Framing is usually defined at macro and micro levels. For the macro-level, usually, each article is associated with a set of media framing packages, which correspond to a central idea with a surrounding interpretive structure that gives meaning to an issue (Gamson & Modigliani 1989). And then the micro-level analysis focuses on seeing how the defined frames come out from specific words, used as *framing devices* (e.g., word choice, metaphors, catchphrases, use of contrast, quantification) and *reasoning devices* (e.g., problem definition, cause, consequence, solution, action). The macro-level labelling of framing (e.g., saying that a certain article is belonging

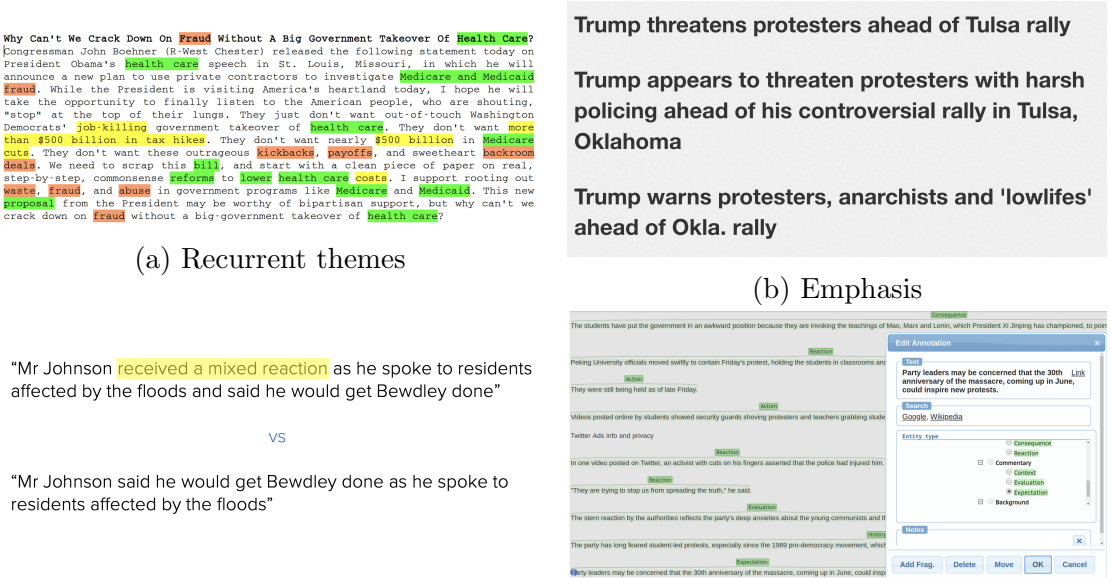
to a specific narrative) is very difficult for automatic approaches because of different reasons. Firstly, it needs to be specific for every topic and story, it cannot encode all the possible positions for all the possible issues. And it requires contextual knowledge of the events. For these reasons, we want to focus more on the micro-level that instead targets the detection of *techniques* that are used transversely to the specific topics. These framing techniques need to be simple to describe and capture punctual details.

From the literature, many works describe and categorise framing techniques that can be seen on the micro-level. Here we have a list of the most important.

- **Recurrent themes**: this technique is achieved by repeating and reinforcing certain themes in the articles, making the readers perceive certain topics more important than others because they appear more in the news. This is the direct manifestation of the agenda-setting theory, where the press has the power to influence the importance of the issues. This can be done on broad granularity, as analysed by different studies (Tsur et al. 2015, Card et al. 2015), where each document is associated with very wide media frames, not focused on any specific topic (e.g., economics, politics, health, ...). Or other studies that investigate more on a more fine-grained level using hierarchical topic modelling to identify both issue-specific and generic frames (Boydstun 2013). This, however, is still limiting the ability to evidence how a certain opinion is supported inside the articles. For this reason, some studies identify a set of specific themes that exist around a certain issue and analyse the polarity of the article with respect to them. For example, the work by Morstatter et al. (2018) defines a set of 10 themes concerning the narrow topic of "ballistic missile defence system in Europe". They define framing analysis as the extraction of the presence and polarity to the 10 themes (e.g., Collective Security, Political Tension, Threat to Russia).
- **Emphasis**: this second technique instead uses language to selectively emphasise on certain details instead of others. Some studies focus on the use of specific language which evokes framing (Baumer et al. 2015), evidencing some constructs that are recurring when the articles want to give a "spin". This type of studies is related to the analysis of subjectivity because, when the authors bring their opinion, some signs can be observed, e.g., loaded terms and specific adjectives that go beyond the impartiality, and contain an implicit sentiment (Greene & Resnik 2009). But sometimes the manifestation of emphasis is more subtle and is done with a selection of synonyms that have a slightly different connotation (Schuldt et al. 2011, Rugg 1941, Tversky & Kahneman 1981). Sometimes the different emphasis can be seen by simply looking at the titles of the articles, which can indicate the point of view of the authors (Liu et al. 2019): for example on articles about gun violence, there can be an emphasis on gun control or the 2nd amendment or mental health. The emphasis gives more importance to some details and demotes others that are just mentioned as secondary.
- **Selection of details**: another technique found in the literature is the strategic selection of the details that may be supporting the communication objectives or the omission of some information that could make the readers less convinced of a certain opinion. This is different from the recurrent themes because the details are not directly part of a narrative line that is being repeated over and

over. Some studies have focused on analysing quoting behaviours from public speeches, observing how different news outlets select different parts and omit others (Niculae et al. 2015). This technique requires a knowledge external to the article to be seen.

- **Pre-packaging**: in this technique, the narrative of the article identifies structural roles of the details narrated, defining what is the problem, the relationships with other problems, who are the people linked, and what is the suggested solution (Entman 1993, Bell 1991, Zahid et al. 2019). This technique is very effective in providing readers with a way to interpret the different details mentioned. Sometimes the interpretive framework is packaged by using different components such as "keywords, stock phrases, stereotype images, sources of information" (Entman 1993) and "metaphors, exemplars, catchphrases, depictions, [...] visual images" (Gamson & Modigliani 1989). The role of these specific components is to provide an interpretive lens that packages the issues described.
- **Propaganda techniques**: this set is borrowed from the specific case of propaganda analysis, and includes a set of logical fallacies and other oversimplification methods that are very commonly used (Da San Martino et al. 2019) (e.g., name-calling, repetitions, exaggerations, whataboutism, straw man). Some of these techniques overlap with the previous ones.



(a) Recurrent themes

(b) Emphasis

(c) Selection of details

(d) Pre-packaging

Figure 2.1: Examples of framing techniques

While these studies in political, social communication and linguistics fields focus on defining these techniques, it is unclear how important it is for the users to have available news articles from other sources that show the same story under a different perspective. It is an intuitive concept that, to see better how someone is trying to convince us to have a certain point of view, it would be helpful to compare his persuasion with ideas coming from different sides.

In our opinion, some of these framing techniques would need to have external sources of knowledge to be detected more easily: how could we know that an article is showing an incomplete picture if we do not have a ground truth available, or at

least some other sources with a different opinion to compare? And some others techniques that, even if they can be seen on articles on their own, would anyway benefit from a comparison: we can see that an adjective is emphasising the importance of something, but if we see another sentence that is using a different adjective maybe we can identify more easily how the author is trying to persuade in a specific way.

Using external sources is already suggested in other fields that deal with the process of information, for example in guides to spot misinformation: media literacy refers to searching for external sources to find corroboration of the claims,[1] but we do not see the same for more subtle influences on the opinion of the readers, that happen when framing occurs.

2.1.2 Automated framing detection

Automated framing detection is defined as a task where articles need to be labelled, in their whole or sub-parts, with categories that relate to some of the framing techniques identified previously (Morstatter et al. 2018, Liu et al. 2019).

The main motivation to develop and use automated framing detection is to analyse at scale and reveal the type and magnitude of influence that the articles are trying to push to the readers. The hypothesis that drawing the attention to this type of language may help mitigate framing effects has been already stated by other researchers (Baumer et al. 2015).

Different strategies have been used to perform this type of detection. A first group can be identified by using specific lexicon-based approaches, which work by counting the occurrences of terms in the analysed texts. This has been used on large scale identification of recurring themes, which are identified by specific words that have been selected by putting together the manual analysis of annotators (Field et al. 2018). Similarly, the automatic inference of specific words of the recurring themes has been done by using LDA analysis (Tsur et al. 2015). Another work instead compared different computational approaches to find out that the most important contribution to detecting the language of framing is given by analysing the grammatical structure, seconded by lexical features (n-grams) (Baumer et al. 2015).

More recent works instead make use of complex models based on neural networks. With the advance of language models like BERT (Devlin et al. 2018), which have shown to achieve the state of the art in many different tasks, more models are making use of them for almost every problem in NLP. The work of Da San Martino et al. (2019), which is focused on propaganda techniques detection, makes use of a model based on BERT. The architecture first uses this language model to provide contextualised representations of the words, and then uses these representations to classify each word with one of the labels available (one label for each of the propaganda techniques, plus one extra label to indicate the absence of propaganda). In this way, their model needs to train only the weights of the added layer for the specific task, reusing the rich encoded knowledge of the language owned by the pre-trained BERT. In Figure 2.2, we can see the analysis performed by this tool, which evidences the usage of several propaganda techniques.[2]

[1] http://www.spencerauthor.com/fake-news-is-a-real-problem-heres-how-students-can-solve-it/

[2] https://www.tanbih.org/propaganda/223/BREXIT/

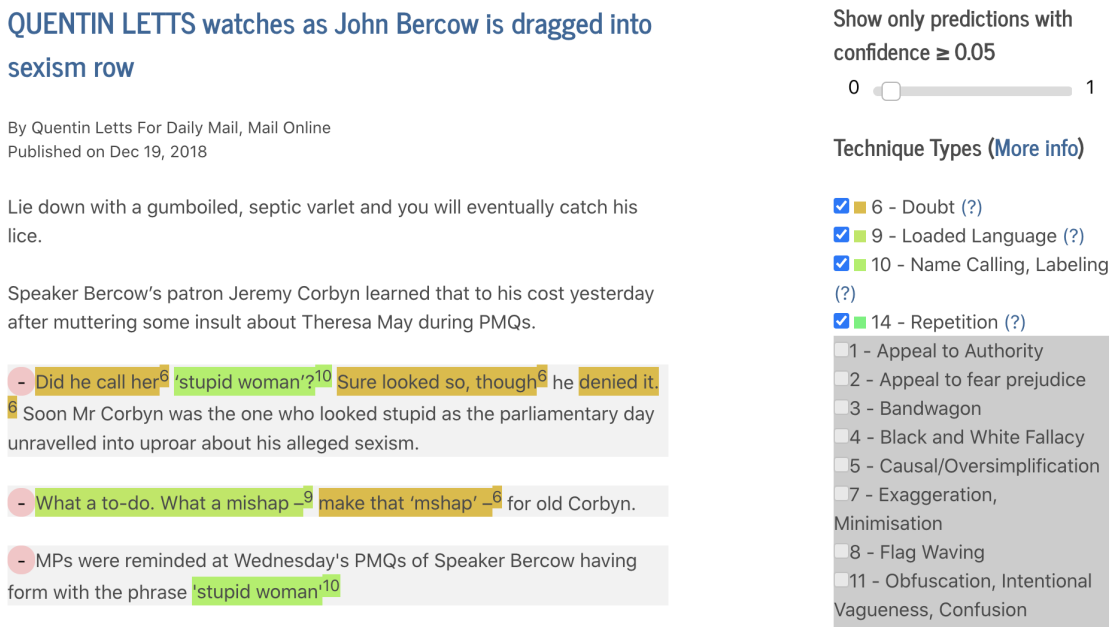


Figure 2.2: Propaganda detection on an article about a political event.

The methods described focus on detecting how framing looks like, by highlighting occurrences which are quite visible. But, as for human readers, also automated methods will probably benefit by comparing different articles that present the same issue. For example, if any detail is missing, methods like the ones presented here would not be able to see the omission. External knowledge is required. Related research done on verification of facts is already making use of comparison with external documents (Yin et al. 2008, Karadzhov et al. 2017). We would like to bring a similar analysis but with a different goal: instead of chasing the truth, we want to bring together different sides of each story and compare their use of framing.

2.1.3 Limitations

To summarise here we present the limitations found in the framing literature.

As we have seen, on the theoretical side different framing techniques have been defined. Some of them, to be seen, require external knowledge that could be provided by using external articles. These techniques include the omission of facts that appear on other sources, or adding details that other sources did not report because of different relevance criteria. Or also using specific language choices that other sources do differently. For these cases, narrowing the analysis to articles in isolation is likely to be a limiting factor (both for human reader and for automated detection).

It is very unlikely that how the news is communicated is independent of political forces and also articles that may present 100% factual information may result in driving the opinions in a certain way. Given that every news source has a different ideological alignment, we aim at using stories from multiple sources, written by authors with different perspectives. In this way we can use the variety of news to our advantage, revealing even more framing occurrences.

2.2 Similarity between articles

From the motivations of the previous section, here we want to expand on exploring relationships between different articles, with the objective to use multiple sources at the same time to help to face the limitations of framing analysis.

There are different possible types of relationships between news articles, such as similarity (covering the same information), referencing (one article is citing another one), and temporal proximity. They can be performed at the document level (e.g., the whole article is similar to another one) or at the sentence level (e.g., the same sentence is corroborated by a sentence in another article) (Bountouridis et al. 2018). Since we are interested in finding articles discussing the same information, we focus on similarity relationships.

Here in this section, we first explore what we mean with similarity, then we look at how it is used and computed, concluding with some limitations of this branch of studies. Our overall goal is to find the tools that would allow us to enrich the framing analysis by bringing information from other articles. And the articles need to relate to the same events.

2.2.1 Similarity definition

The concept of similarity is widely defined as "*the fact that people or things look or are the same*"[3]. In our case, we want to focus on the specific **semantic textual similarity**, that expresses how two different pieces of text are similar considering their semantics. This narrower definition is used in the NLP community, and for several years a specific SentEval task has been targeted to evaluate the capacity of evaluating the level of similarity between sentences (Conneau & Kiela 2018).[4] The similarity can match with how similar the corresponding strings of the text are (character-wise similarity) but has also to account for entities, relationships and concepts that can be expressed by using different words with the same meaning. For example, if we consider two synonyms, their semantic similarity needs to be high. And another requirement is that the similarity depends on the context of the words (Miller & Charles 1991). The same word can have different meanings depending on the sentence it is used in.

2.2.2 Applications of similarity

Similarity is used in many NLP applications that deal with documents and need to find how close they are to each other. This includes both applications that work on whole documents, like document clustering, and more detailed analyses of overlap which aim at finding pieces occurring in multiple documents (paragraphs, sentences, words).

Document clustering is widely used in the Information Retrieval community, to discover topics on different levels, to find similar items and to find possible duplicates. Given big quantities of articles, different methods can form groups based on wider or narrower topics.

[3] `https://dictionary.cambridge.org/dictionary/english/similarity`
[4] `http://nlpprogress.com/english/semantic_textual_similarity.html`

The literature is quite rich in approaches to news clustering (Carpineto et al. 2009, Andrews & Fox 2007) that work in many different ways. Some clustering techniques requires specifying the number of clusters wanted as an input (K-Means, Mean-Shift, GMM), others instead are more flexible, discovering the number of clusters based on other geometric properties (DBSCAN). And some others have even nicer properties. For example, Agglomerative Hierarchical Clustering allows cutting at different stages of the algorithm which represent a threshold distance to merge clusters.

Then there are applications of similarity at a more fine-grained level. The overall objective of this group of works is to study how many parts of a document also appear in other documents. The objectives for this analysis may vary from plagiarism detection (Potthast et al. 2010) where the occurrence of text pieces too similar is an indication of non-originality. Or in the fact-checking perspectives, when external confirmation is searched about a specific detail to be able to corroborate and establish if the content is true (Karadzhov et al. 2017). Or to compute how complete is the reporting from different sources, analysing the overlap between different articles.

In one recent work (Bountouridis et al. 2018), groups of similar articles are found, then broken down to pieces of information and analysed to find if these details are *corroborated* (occurring in multiple documents) or *omitted* (occurring in other documents of the same group, but not the current one). We aim to use this idea of applying similarity hierarchically to both article-level and sentence-level, extending it even to the word-level. By doing so, not only we might be able to recognise which sentences appear in multiple documents (with different degrees of similarity) but also we would be able to identify the specific words that have been changed.

Going into a more fine-grained comparison, we can shift from analysing the similarity of meaning, which is required to relate different articles, to the similarity of the linguistic surface, which is our target since we want to analyse *how* the events are presented in different ways.

2.2.3 Similarity computation

To be able to compute a similarity between different texts, there are two steps:

1. representing the texts with a numeric representation;
2. use similarity metrics between these representations.

The first step is the one that is more interesting and so many different methods exist. This group of methods aims to find a condensed representation that contains the meaning of the text, independently from its length.

Starting with very naive methods, we have approaches based on *bag-of-words* that simply take into consideration which words are present in the considered text. They run with an initial scan of all the available documents to create a dictionary, and then each document is represented with a one-hot-vector that is saying for each word in the dictionary if it belongs to the current document. Many refinements to this basic approach exist, e.g., by removing words that do not carry meaning (stopwords) or by limiting the dictionary based on term frequencies (only the top k words overall are considered, or only words that appear at least n times). This group of methods culminates with a model that is called TF-IDF (Jones 1972) which is based on two components: Term Frequency and Inverse Document Frequency. In simple words, the first component gives higher weights to terms that appear more in the considered document, while the second measures how important a term is by promoting rare

terms (Jones 1972).

Then we have another group of models that relies on explicit semantic representation. This is the field of knowledge extraction that is based on recognising the entities mentioned in a text (entity recognition and linking) and extracting the relationships expressed between them. Using these models, a text is translated into a knowledge graph (Gangemi 2013). The main limitations of this set of models it that they may work accurately enough just when entities that are known are mentioned and need the supervision of rich knowledge bases.

And finally, we have a whole line of research that is based on distributional semantics. The assumption is that the words take their meaning from the context around them. For example, two synonyms can be recognised as similar because one can be switched for the other and therefore they have similar contexts. This unsupervised learning strategy builds words representation by observing their context. A milestone in this area is given by the word embedding models Word2Vec (Mikolov et al. 2013) and GloVe (Pennington et al. 2014) which have shown to achieve better performance metrics in a big variety of tasks. They can work with syntax and semantics and even solve analogy tests (Mikolov et al. 2013) with great performances.

Then in recent years, new models have arisen in the language representation that are extremely powerful to encode entire documents and capture their meaning. It is the era of Language Model that can provide document representation, like BERT (Devlin et al. 2018), XLnet (Yang et al. 2019), or even more oriented towards the similarity task: Universal Sentence Encoder (Cer et al. 2018). And all these models can be used directly without the need to be trained, thanks to pre-trained models that perform already well out-of-the-box.

The main strength of Language models is that they can relate documents that talk about the same events, even if they use different linguistic surface, from articles that may use the same subset of words but talk about different events. It is a semantic matching more than a word-based matching which accounts for the specific order of words (a problem of word vectors). The semantics they are able to express is shallower with respect to the explicit semantics based on Entity Linking, but it is more resistant to changes in the linguistic surface and works without having knowledge bases encoding the reality.

Once these models synthesise documents of different length into numeric representations, different geometrical measures can be used to compute the distance or similarity. The most commonly used distance measure is euclidean, although with word and sentence embedding usually the cosine similarity is more common.

And we can compare therefore representations for single words, for sentences or even for complete documents. The computation of similarity allows targeting all the different applications that we saw previously, exploring commonalities and differences at multiple granularity levels.

2.2.4 Limitations

However, this set of approaches is limited to bringing to the attention of the reader the linked information pieces with a measure of similarity, without characterising the differences. The reader would then need to evaluate the impact of each of the multiple versions of the texts, and the framing that they imply. Different documents may express the same set of details, but give them a different role (reporting

an action, commenting, contextualising, doing a digression, identifying causes and consequences) and use different words that are semantically similar but may imply a different framing perspective. And also the fact that a certain detail is present or not in certain articles within a group where other sources mention it, needs to be analysed with framing theories.

Another problem of using only this group of works for our objectives is that, while we can analyse the overlap of information, we do not consider the influence between the sources (one could have taken the content from the other and modified it slightly). This problem is currently tackled by works on plagiarism analysis, but this is not studied jointly with framing analysis. Furthermore, plagiarism detection is done by having one questionable article and a set of original works, while here we do not have this distinction. We would like to analyse how the same information is recycled and slightly changed, by tracking it across several news sources.

2.3 Comparing different perspectives

Although these two previous research areas have done significant progress, they have evolved independently without many points of contact. We need to bring together the works on framing and similarities to spot more subtle examples of framing.

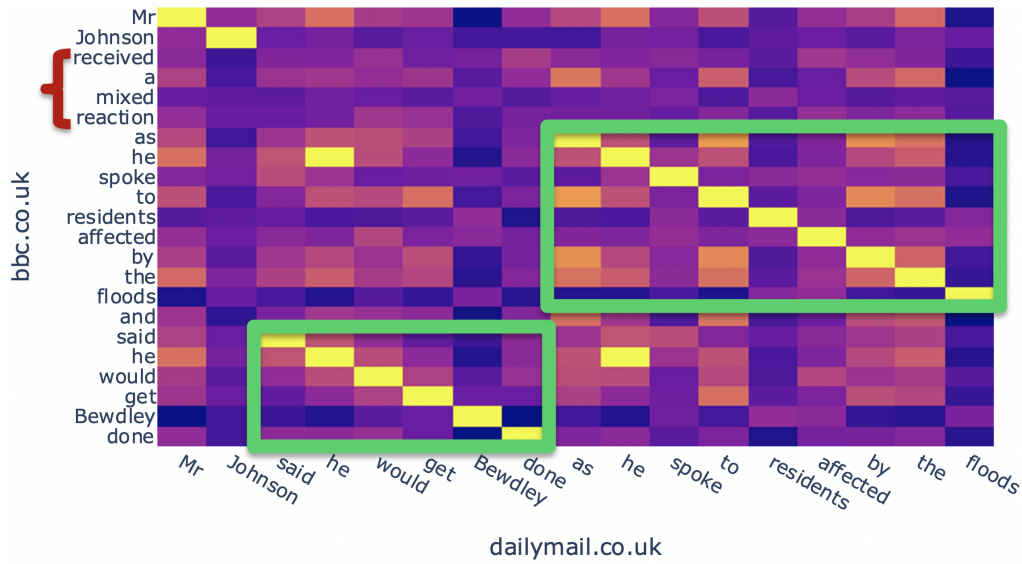


Figure 2.3: An example where omitting a detail provides a different framing.

If we take the example of Figure 2.3, we have two sentences coming from two different news outlets (on the vertical axis from BBC[5] while on the horizontal from Daily Mail[6]) and compare the words they contain by using a similarity metric (Cer et al. 2018). From this figure, we can see that in the green boxes we have information reported from both sources, while in the red curly brace we have a detail that has been reported only by the BBC. This specific detail, describing a "mixed reaction", gets omitted from the sentence from Daily Mail that is known to be right-leaning.[7]

[5] https://www.bbc.co.uk/news/uk-england-hereford-worcester-51791346

[6] https://www.dailymail.co.uk/news/article-8088805/Britons-facing-heavy-downpours-four-inches-rain-50mph-winds-set-batter-UK.html

[7] https://mediabiasfactcheck.com/daily-mail/

This example presents one specific manifestation of framing that current methods do not focus on detecting, that is the selection of details. And the main reason is that such techniques require a method that goes beyond the analysis of one single article at a time. We cannot know when something has been omitted if we don't have or a ground truth or different stories to corroborate with, using the richness of information available from multiple points of view. There are facets of framing that have not been tackled by automated approaches because they cannot be pointed out by considering a single article, and we need to investigate which ones we can define more rigorously and how important they are.

There exist some efforts in this joint area, that try to provide multiple points of view of the same events by considering sources that come from different political leanings. For example, AllSides[8] is a platform that collects articles from different news sources providing a hand-curated set of stories where articles with opposing political bias are put together, with also a brief textual introduction of the framing differences.

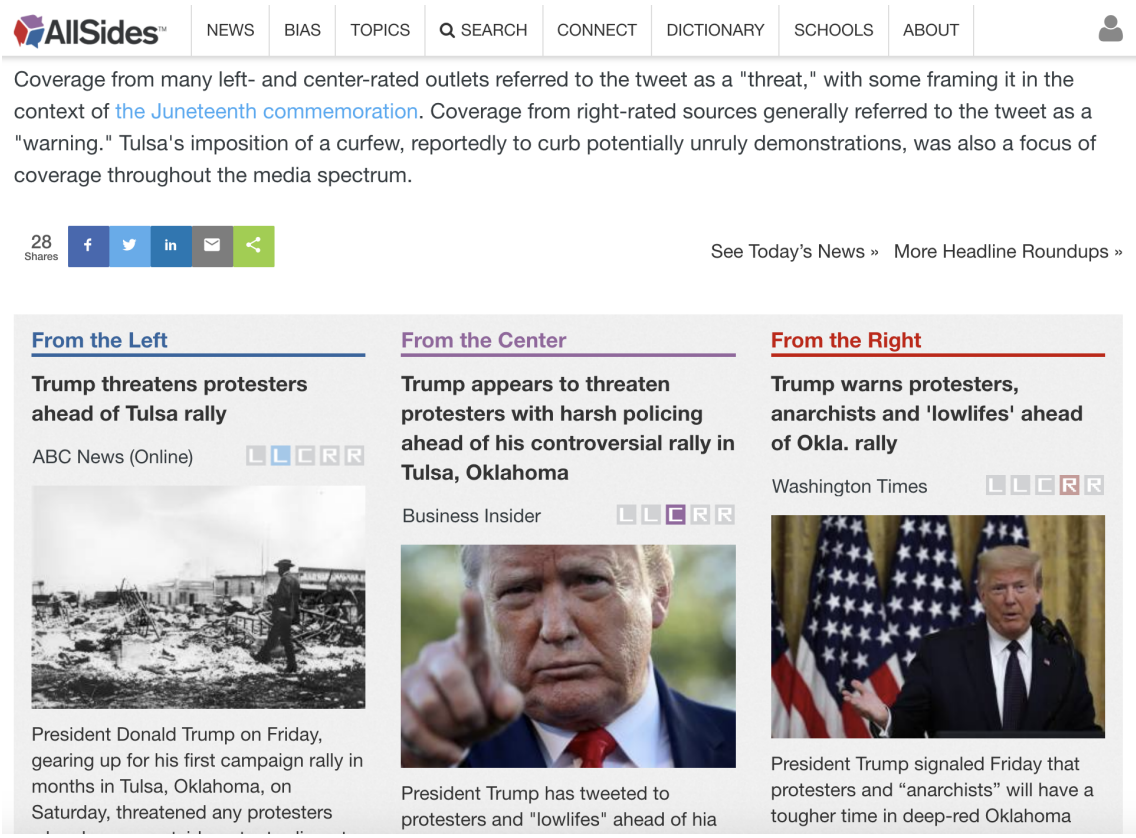


Figure 2.4: Example of analysis from AllSides: `https://www.allsides.com/story/trump-tweets-about-protesters-ahead-saturday-tulsa-rally`

In Figure 2.4 we can see an example that is displaying how a source from the left, one from the centre and one from the right frame differently Trump tweeting about his rally in Tulsa. Sources from the left use the term "threat" while sources from the right use "warning", which can be seen in the specific words used in the headlines. This *selection of terms* gives a specific framing which can be seen as

[8] `https://www.allsides.com/story/admin`

the emphasis from the left on the strength of the tweet versus an attenuation from the right, plus the centre instead is keeping in an intermediate position. But if we look at the articles, this term is not the only difference. The sources from the left give more space to the narrative about Juneteenth[9] and the past history of the city, which emerges also in the image of the article from ABC News.

We would like to provide a similar analysis, but with an automated analysis. The main limitations of AllSides come from the analysis being handmade and therefore they publish just a few stories per week and only focusing on the US. With this work, we want to create an automated method, which will need to rely on existing works of framing analysis and document similarities.

We argue that by the integration of these two fields, a deeper analysis of framing can be achieved, with the goal to understand better how, from the same event, multiple stories are generated.

With respect to AllSides or other tools (Blue Feed, Red Feed[10]) that just focus on the left/right political bias (mainly from the US with liberal vs conservative), we want to conduct an analysis that compares several characteristics of the news outlets, such as the newsgroup they belong to, or how they are rated by multiple news rating organisations (factuality, bias), or their ideology in general. We want to have an idea of the correlation between certain features of the sources and the type of framing that they use. This goes beyond the left/right spectrum because not all the news is focused on political issues.

Furthermore, seeing that the overlap of information between the different sources is quite high, we want to analyse more deeply how the content is reused by different sources (as evidenced in 2.2.4). The main focus is not in evidencing plagiarism (as other studies have done), but instead to find how the news is modified considering the time of publishing. We need to investigate which sources use contents from others, and how they slightly modify the information contained to provide a different perspective.

[9] `https://en.wikipedia.org/wiki/Juneteenth`
[10] `https://graphics.wsj.com/blue-feed-red-feed/`

3 Research Questions

Given the gaps identified in the Literature Review, this PhD has the following Research Questions.

RQ1: How can we automatically reveal the framing differences in articles presenting the same event?

This question focuses on how to develop a cross-article framing analysis that would overcome some of the limitations evidenced. We can break it down in two sub-questions, one on the reader-side and the other on the technological side:

- **RQ1.1: What is the contribution of comparing multiple articles to see framing signals?** We need to understand from the point of view of the reader how much important it is to have multiple articles available and compare them to be aware of the several framing techniques used.
- **RQ1.2: To what extent can we automatically recognise framing signals using multiple documents?** We want to investigate the impact of having multiple articles available on the ability of an automated method to identify and label framing differences.

RQ2: How do news sources with different characteristics change stories over time?

This second question instead moves the emphasis on the sources where articles are published. We can break it down in two subquestions:

- **RQ2.1: Which features of a news source influence the framing techniques used?** News sources can be described by using their political alignment, the news group they belong to, their location, and we want to see if some of them are related to the amount and type of framing techniques that occur in their articles.
- **RQ2.2: To what extent can we identify the evolution of the framing of a certain event?** We aim at extracting temporised chains that track a specific event across news sources, from where it is seen the first time and how it evolves through time and sources. Can we identify some patterns that describe how the information flows between news sources?

4 Research Proposal

In this chapter, we present an overall description of the framework we want to build to highlight the different framing techniques and compare the articles that different sources create around a single event. As we can see in Figure 4.1, we are taking the research questions and sub-questions from the previous chapter and, for each one of them, we have a process that from the hypothesis leads to some outputs using a separate and distinct analysis.

The structure of this chapter follows the columns that represent the different research questions.

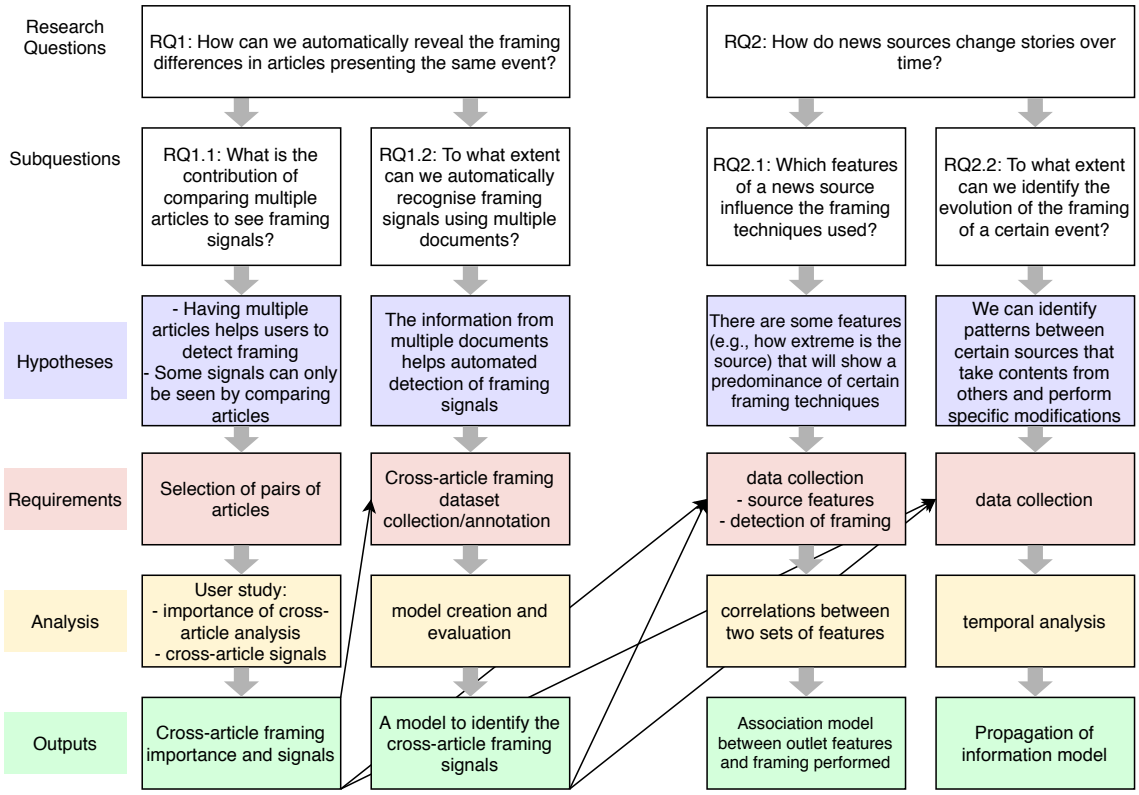


Figure 4.1: The relationships between the research questions and the evaluations.

4.1 Cross-article framing analysis (RQ1)

To address the first Research Question **(How can we automatically reveal the framing differences in articles presenting the same event?)**, we aim to resolve two sub-questions separately. The first one analyses the *importance* for the reader of comparing multiple articles to understand the framing that is being applied to the news through the revelation of specific framing signals. The second one instead

explores *how* we can build a tool to identify these framing signals automatically. In other words, while the first sub-question deals with the importance for the readers, the second question instead targets automated detection. The dependency between the two sub-questions underlines that the second one needs the outputs of the first one to know how to label the framing differences.

4.1.1 The usefulness of multiple articles (RQ1.1)

The first column represents a line of research that is needed to understand how useful is for readers to confront the information available in multiple articles to spot the framing techniques used. To describe and characterise the framing, we started from the literature review to gather concrete manifestations of framing, both in theoretical works and in detection methods. As we have seen in section 2.1, different techniques have been theorised and studied, both in theoretical and practical works. We name the manifestation of such techniques *signals* because they are signalling the presence of the framing, which is an abstract concept, but they are tangible and can be seen as concrete features. For example, loaded language (technique) emerges from specific words that have a strong sentiment (signals) and evokes frames of strength, violence, threat and similar. Or the strategic omission or inclusion (technique) of certain details (signals) that have the role of shifting the opinion of the reader. What we want to investigate is whether we can ease their detection from readers by comparing multiple articles. Therefore, the specific sub-question is **What is the contribution of comparing multiple articles to see framing signals?**

We hypothesise that having multiple articles would allow users to more easily detect the framing signals that they are faced with, especially when some of them are particularly subtle to reveal when just looking at a single article. As we described in our position paper (Mensio et al. 2020), such signals vary, from differences in emphasis (e.g., different articles focus their headlines on a different detail, to presenting parts of the story with more weight), to the selection of details (e.g., some articles may decide to omit some parts of the story that is reported by others, or include unnecessary details to push for a certain narrative), or also to different term choices (e.g., being stronger in the language, or evoke specific feelings or related events). For this reason, we want to test how important it is to see multiple articles that treat the same event. By comparing them, the difference in their point of view would emerge through specific differences of the texts.

To validate our hypothesis, we will run a *user study* where participants will be shown different articles that present the same event with some differences. To test the hypothesis of framing being easier to spot when multiple articles are available, we plan to first present to the user just one article and ask to find the framing signals. Then we show also the second article on the side with different details and ask again. In both the steps, the input asked from the user is to highlight parts of the article(s) and to say which framing techniques it is representing, showing a certain view angle, as one of many labels decided before (one extra label "other" which participants can use to describe something different). To obtain these annotations, users will first be shown with the framing techniques identified in the literature, to understand what they are looking for. Then they will be shown with a single article and asked to annotate it. And the second phase will ask them to compare two articles and to point at differences in terms, themes, and details that are representative of framing

techniques. We will then compare the number and type of annotations coming from the two different situations (one single article or multiple articles). It is important to divide the annotators of a certain pair of articles in two sub-groups to divide the effect of "seeing again the same article twice" (unwanted) from "seeing the article compared to another one" (wanted). For this reason, if we consider two articles A1 and A2 and two users U1 and U2, user U1 will first see A1 then A1+A2, while U2 will first see A2 then A1+A2. In this way, we can sum the sets of annotations done in the first stage, when seeing just one article, and in the second stage, when the reader has already analysed one article and is presented with another one. In the second stage, we sum the sets of annotations given to the corresponding new articles (A2 for user U1 and A1 for user U2). By comparing the number and types of annotations in the two settings, we can see the effect of having already knowledge from another point of view, and be able to compare with the new article (see Figure 4.2).

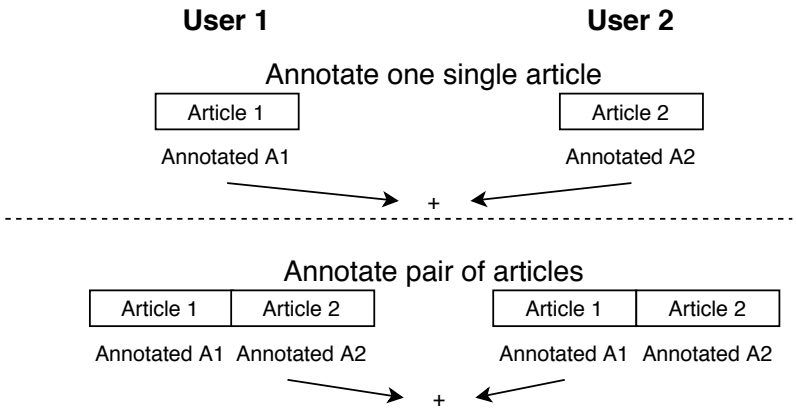


Figure 4.2: How to compare the user annotations to exclude effects of secondary-reading.

The pairs of articles to be shown will come from a selected subset from clusters of documents that are presenting the same detail, with some differences in the terms used. For this reason, they need to be revised and checked. The clusters will come from articles suggested by AllSides and other news aggregators like Google News, testing both differences from opposite political views as well as articles without any difference in the political bias. To identify the pairs of documents we plan to use the help from similarity models that, as will be seen in Chapter 5, have been investigated and used to cluster documents and sentences.

The annotations will be manually examined with two goals. First, to see if the annotators (using inter-annotator agreement measures) are able to spot more signals when they have different articles available. This will prove or disprove our hypothesis that framing can be revealed easier in this setting, especially for some subtle framing techniques. And second, the annotations will be revised manually to manage the cases where participants have expressed the need for other types of signal. For this reason it is important to design the labels well enough during the preparation period of the study, making use of a small initial pilot to gather feedback. The labels need to account for the framing technique used, which serves as link between the documents, together with a role of the participating articles. For example, if the technique *selection of details* is used, one article will participate with the role

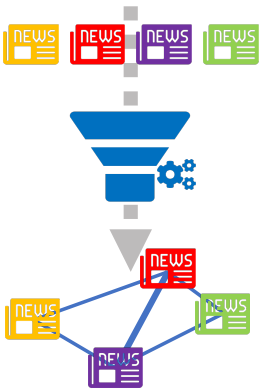
omitting and the other *adding*. Or for the *emphasis* technique, one article can have the role of having a *stronger term* while the other a *weaker term*.

This user study will provide, from the perspective of the readers, which differences are important in relation to the framing phenomenon. From seeing multiple differences in the articles, the participants will make a decision of which ones are providing a different connotation to the events reported, ignoring other modifications that are not meaningful for them.

4.1.2 Automated framing detection (RQ1.2)

The second part of the first Research Question requires building and evaluating an automated method to extract the signals identified in the first user study. Unlike traditional methods of frame detection that mainly rely on extracting features from single articles to build a classification model, we want to use external features that come from the comparison of multiple articles. Such features could be the uniqueness of certain details (sentences or words) when comparing to other articles describing the same event, or including similar words that are used as substitutes in other articles. The sub-question we address is the following: **To what extent can we automatically recognise framing signals using multiple documents?**

To implement a method that exploits the differences between multiple articles, we need to be able to use the similarities and differences between the considered documents. For this reason, we plan to rely on the approaches described in the Literature Review 2.2 where we analysed some methods that can find the similarities at different granularities, and expose the differences of the details provided. The step of feature extraction will be based on a pipeline that will process the documents in several steps as shown in Figure 4.3.



Stage	Description	Inputs	Outputs
retrieval	Articles need to be retrieved from multiple sources	Public datasets, news feeds, news aggregators	Articles (headline, body, source, URL, publication date)
processing	NLP pipeline: tokenisation, parsing, embedding	Articles	Documents with: - Parsing, segmenting - Embedded representation
linking	Similar articles are linked as they cover the same event, and similar sentences are recognised across documents	Processed articles	Similarity links between - Articles - sentences

Figure 4.3: The processing pipeline that retrieves, processes and links together different articles.

The first step, *retrieval*, keeps collecting articles from multiple sources and stores them with a unified schema that includes a set of useful fields: headline and body that are needed by linguistic analysis, the source and publication date which are needed for RQ2. Then the *processing* stage will use NLP libraries with two main objectives: dividing the articles into sentences and compute the embedded semantic representation of both the articles and the sentences. The articles and sentences embeddings are then used in the last stage, *linking*, which performs clustering at both

article and sentence granularity. Furthermore, corroborated and omitted sentences are identified (Bountouridis et al. 2018) and this is added to the features. In this way, articles that cover the same event are brought together and are ready to be compared in their overlap and in their differences.

What is available after this initial processing, is a set of differences between articles. But since not all the differences are related to framing, we need a model that, given as input one specific difference, estimates whether this difference is *meaningful* in the framing perspective or not. There are differences that are not conveying any framing, and therefore should be discarded. In Figure 4.4 we can see that from a pair of details that are different but belong to similar contexts (e.g., a word that has been changed with one other, or a detail that is present in one article and not in the other), we need a model that learns to recognise which differences are framing the reported events differently. In order to do so, there may be multiple features, explicit (e.g., sentiment strength, proxies of emphasis, POS) or implicit (e.g., the representations coming from a Language Model) that the model can use to recognise when this difference is related to framing, and which type of signal it represents.

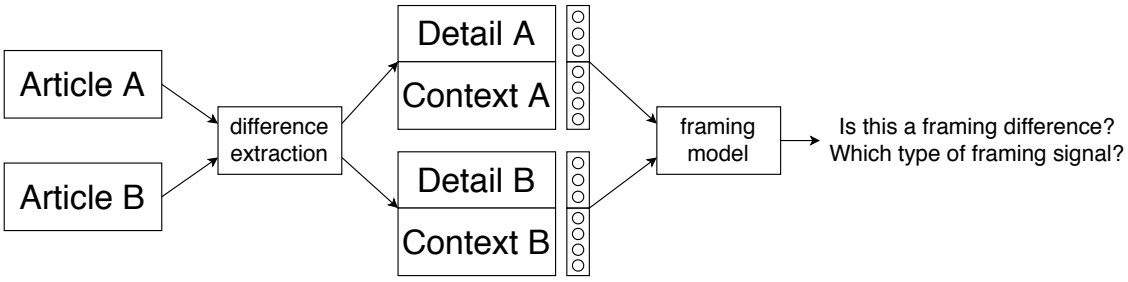


Figure 4.4: The general structure of the model that estimates the framing differences.

The general structure of this model is inspired from a similar one that is used on articles in isolation to detect propaganda techniques (Da San Martino et al. 2019). The commonality is that, on top of the representation layer (BERT in the compared work), we have an additional layer which is responsible to recognise when and which specific techniques are occurring. The difference is that we will work with a dual network with a siamese structure (Bromley et al. 1994). Siamese networks have shown wide application in comparison scenarios, and are made of two identical parts (here represented by the encoding of the details changed and their contexts) that are merged into a final comparison layer. This final layer will try to learn when the differences provided are conveying a different framing perspective, and assign one of the many labels identified in the user study. We need to experiment with this type of architecture and see whether it can achieve the task assigned.

As a simple example, the model should learn that when the two details are using a different strength in the language, this needs to be categorised as a signal of emphasis which is probably conveying a different framing. Other signals may require to take into consideration also the context of the differences, for example when an adjective is giving an additional connotation to an entity in one article, the type of entity may guide the model to understand how to interpret the situation.

We hypothesise that using the differences between different articles would help an automated model to find more signals of framing, with respect to models that work just by considering single articles in isolation. As the first sub-question analyses

the importance for human readers, this analysis wants to see if the same applies to automated detection. We also hypothesise that some types of framing signals are only detectable by using the external knowledge coming from other articles, which is not used by the existing methods.

To validate or refute our hypothesis, the model will be evaluated by looking at how accurately the spans containing framing differences are identified. For example, the complete model will be compared with another version of the model that will take as inputs raw articles without using the difference extraction. And we will also use models found in the literature that focus on specific types of framing signals (e.g., emphasis detection, loaded terms) to see how the detection results compare with our difference-based analysis.

To evaluate the model proposed, we will need to expand the dataset annotated in the user study. Manual annotators will be given the refined task description and will annotate on document pairs the framing differences according to the schema refined with the first user study. The pairs of documents will be extracted with the help of the pipeline which will provide articles that are very similar in the content but at the same time have enough differences to be annotated. The annotations will contain the type of signal and also the role of the two documents (e.g. in an omission signal, one article is *including* and the other is *excluding*, or instead in a term strength variation, one article is *stronger* and the other is *weaker*). The annotations can be then used both in the multi-article scenario and in the single-article scenario, because different spans will be annotated on the pairs of documents. For example, if we have an annotation that is highlighting a term that is used with a difference in strength, the fact that one term is stronger is available even without having the information of the other article. This dataset will be made available as a resource paper to promote the research on this task.

Once this method has been analysed and validated, the detection model will be made available as a tool to help users to see the framing differences between multiple articles. This tool will provide highlights on the current article, by evidencing where it is showing the usage of framing techniques. Users will be able to see how articles from other sources have presented the same story framed in different ways.

4.2 The role of news sources (RQ2)

The purpose of the second research question is to understand the link between the features of the sources and how they frame the news, changing details and taking information from other sources. While the first research question focused on how to reveal the framing, here we want to investigate how the sources interact with framing. For this purpose, we have two sub-questions that investigate *i)* the relationships between the features of a source and the framing that it uses, and *ii)* how stories are modified and re-framed over time when sources reuse contents from other articles.

4.2.1 Source features vs framing (RQ2.1)

News sources are different and there is a wide variety of them in the media landscape. A source can have a certain political alignment, be part of a certain newsgroup, have an affiliation to certain movements and ideals, or reside in a certain country, and we

want to analyse the framing techniques across all these features. Lots of tools and evaluation methods just rely on the provenance of the news, looking at which source published it. The reputation of the sources is a big theme in the credibility domain. Here we want to analyse the relationship between the features of a news source and how it applies framing differently from others. The sub-question is: **Which features of a news source influence the framing techniques used?** We want to analyse the relationships between the framing choices done by the authors and the news sources where the articles are published. In other words, we want to understand the relationships between the ideological alignment of the news outlets (political alignment, bias, news media groups) and the effective differences contained in the articles.

The hypothesis that we have is that certain features, such as the political alignment, have bigger effects than others on how the news is presented, especially when the news topic is about politics. Or considering the magnitude of political leaning of a source, more biased sources will be using more loaded terms and report information more incompletely than balanced sources. We want to understand what is the effect of other features as well, and discover which ones influence mostly the framing of the articles written.

To test this hypothesis, we need two sets of information: the features for each of the sources and the framing analysis. In the first group, we can use many features:

- **Political alignment**: different ratings exist describing the political leaning of the sources. AllSides, besides comparing articles from different news sources, also provides the bias rating for more than 600 media outlets.[1] Also, Media Bias/Fact Check[2] provides the political leaning for more than 3200 news sources. This is probably one of the most important features when analysing the framing about political issues, as their leaning would show slighter support or distancing from public actors.
- **News media group/company**: this feature groups together media sources that are part of the same economical group, and therefore would have similar economical reasons in presenting the news about certain topics.[3]
- **Geographical information**: knowing the country and region of a news outlet can provide an analysis of framing across countries to analyse how the same issue is framed differently depending on the position.
- **Factuality and credibility ratings**: there exist multiple ratings that say how factual is the reporting, or how transparent the organisation is in different aspects (Media Bias/Fact Check, NewsGuard[4]). All these ratings can instead drive the analysis in seeing how the reputation of a news outlet makes emerge different usage of framing.

Instead, to provide the framing analysis, the model from the first Research Question will provide how much each source uses each of the framing techniques, by doing an aggregation on the source level of the analysis of the articles.

The analysis will be done by comparing the two sets of features and extracting the most important correlations. We can apply correlation analysis techniques that will tell us which features relate mostly with which framing technique.

[1] https://www.allsides.com/media-bias/media-bias-ratings
[2] https://mediabiasfactcheck.com/
[3] https://en.wikipedia.org/wiki/Category:Newspaper_companies_by_country
[4] https://www.newsguardtech.com/

This correlation analysis will tell us if there are any characteristics of the news sources that make them heavily use specific framing techniques. One example that we may discover is that sources heavily biased tend to perform omissions (as seen in Bountouridis et al. (2018)) or use loaded language.

The outputs of this analysis will hopefully be able to describe the news sources and how they present information. This would be a step further in making the readers of news more conscious about whom they are reading from and guide them, for example, to sources that report the information more completely.

4.2.2 Framing evolution (RQ2.2)

The last branch of research instead takes the direction of analysing the evolution over time of an event in how it is presented in the media landscape. We want to track pieces of information and see how different news sources describe them by using different framing techniques and reuse contents from other sources with different modification types. Our sub-question is the following: **To what extent can we identify the evolution of the framing of a certain event?**

We want to analyse this phenomenon to explore how the information gets reused and presented under different shapes. The analysis of framing differences that we described before does not look at temporal information and therefore has no way of understanding, when two details are similar, which description came first and which one instead is reusing the content from the other. Although news articles usually cite their sources, there is not much work that explores how the information is modified when reused.

We hypothesise that we will be able to identify patterns between certain sources that take contents from others and perform specific modifications. It is very likely that, especially small and secondary news outlets, sometimes recycle material from other bigger sources, with citations, but they re-adapt parts for example by changing some wording. Or on the other side, major headlines that reuse contents from local news outlets may remove details that are too dense for their reporting needs.

To do this analysis, we need to have precise timing information about the publication of the articles involved. The metadata of the articles usually contain the publication date and sometimes also the precise minute. Relying on this information (that sometimes could be imperfect because of further edits after the first publication), we have the challenge that many articles may indicate the date only and therefore any precise time-dependency could be undetectable.

From the pipeline that gives us similar articles and characterises which parts have been changed, and with the framing analysis that serves to understand which type of modification has been performed, we will conduct this further analysis by ordering documents inside a cluster by publication time. Once the documents are sorted, we can then identify the different details, and analyse where do they appear first and how they evolve with time. Following these details we extract information flows that will give us an understanding of which sources influence others, creating an influence network. Furthermore, we can add to this directed network labels on the edges that correspond to the types of framing modifications that are done. To see whether there exist patterns we will take the statistics of the framing modifications article-wise and aggregate them to the source level. We will count between each pair of sources how many times they reused details from others and which types of

framing they used.

In this way, we will show, for example, how a certain source evolves the narrative starting from other sources, merging new elements on its way, emphasising some details, and dropping others. This is important for the readers to understand the provenance of the information they are consuming and how it has been modified over time by different sources.

5 Timeline

This final chapter of this report describes the timeline, considering both the work that has been done up to the current date and the plan for the next two years of research. While the previous chapter described the methodology and evaluation that is planned, here we describe when each part of the project is expected to be addressed.

5.1 Work to date

During the first year, several activities have been done. Some have been done as initial explorations in the field of research, understanding what other researchers have done, "getting the hands dirty" with data and NLP tools. And their function within the PhD project has been to lead to the Research Questions that we described earlier. Some other activities have been done to kick start the future experimentation, for example doing data collection of news articles or by beginning to implement some of the stages of the processing pipeline that will be used.

5.1.1 Paper reproduction (Bountouridis et al. 2018)

We started by analysing the paper from Bountouridis et al. (2018) which presents a methodology to analyse how much information overlaps between different similar documents, identifying points of information that are corroborated or omitted. We wanted to analyse this resource because it seems to be going in the broad direction of our project, analysing different presentations in the news of the same event.

The paper has been analysed and reproduced, to get a deep understanding of how it works. The implementation started with the code publicly provided by the authors,[1] but its incompleteness in some stages of the processing (e.g., the creation of the article-level cliques, and all the specific hyperparameters of the algorithms used) required integrating the codebase.[2]

With the experiments reproduced, we have been able to inspect the cliques of documents and sentences identified by the model, seeing the following limitations:

- The *document clustering* sometimes splits similar articles over different clusters, or in some cases has different stories that talk about a different detail within the same cluster (e.g., when a news story re-emerges because further details are discovered). This can be a consequence of having TF-IDF as the underlying method to represent the documents. This method is fast and efficient for coarse topic detection, because it is based on bag-of-words which works well with specific terms that distinguish the topics. But when we need to have a finer-grained clustering such in this case, the limitation of such method

[1] `https://github.com/dbountouridis/InCredible`
[2] `https://github.com/MartinoMensio/InCredible`

may surface because the terms of two political events with the same entities mentioned result in having similar feature vectors.

- The *sentence clustering* method provided just says that two sentences are similar, but does not point to which specific words are responsible for the similarities and differences. This would require a framing analysis that in this paper is not included. Furthermore, sentences in a clique are very similar and no big differences have been observed because the similarity metrics is based again on TF-IDF. This method is not robust enough to the usage of synonyms and other variations on the linguistic surface, while at the same time is unable to distinguish two sentences that use the same words but have different meanings because of the sentence structure, so it makes selecting a threshold value very difficult.
- The *clique algorithms* are not the best choice for doing clustering when we have the information of how much similar two items are (a real-value is available from the similarity metric). The approach considers an unweighted version of the similarity graph by using a threshold (weights are just used to select the most appropriate clique during their creation), but instead dealing with the original weighted graph would allow better clustering techniques.

This experiment helped to see the limitation of this type of work, that belongs to the *similarity* area of research (Section 2.2). This paper provides a great way to analyse the overlap between articles and extracts pieces that have been omitted or that are corroborated, but does not investigate further in the reason behind the selection of what to include or not. This opened up for further work to:

1. investigate how to better represent the documents to provide meaningful similarity metrics;
2. investigate the works that analyse framing theories and detection;
3. experiment with document and sentence clustering to bring up differences;
4. collect data from more recent articles that would be more relevant and interesting.

Apart from these limitations, we are building our processing pipeline on top of this type of analysis, that links together news articles at different granularity levels (documents, sentences, words). This gives us the opportunity to use features from multiple articles for the next stage of automated detection of framing techniques.

5.1.2 Models for similarity analysis

Moving to the problem of representing documents and sentences in a way that captures more the semantic similarity, we decided to analyse closer the existing works, including word embeddings and language models. We wanted to see in practice how the usage of different representation models would affect the measurements of similarity, experimenting with a small set of articles. Having a solid base for computing the distances between articles and sentences is a pillar for comparing different articles. The applications of similarity range from document clustering to identification of omitted pieces of information in a cluster, therefore it is very important to use a proper method that is not fooled by the usage of synonyms and other linguistic variations in communicating the same information. To study the differences in the language of framing we first need to be able to tell whether two pieces of text are discussing the same information, and distinguish properly degrees of similarity.

We set up a small benchmark where the goal is to find the most similar pairs of sentences coming from selected pairs of news articles which cover the same event. Each model candidate has to tell which 10 most similar pairs of sentences has found, one from one article and one from the other. The pairs of articles have been chosen manually, by considering three constraints: *i)* description of the same event, *ii)* from different news outlets, *iii)* published near in time, with a maximum distance of one day. Each model would extract the most similar 10 pairs, and we then compare the pairs provided and their relative order.

The selected models used in the benchmark are the following:

- **TF-IDF**: with a feature size of 2000, with a preprocessing made of lowercasing and tokenizing, without lemmatisation;
- **GloVe-average**: considering GloVe word embeddings trained on the CommonCrawl dataset, and doing an average of the vectors over the sentence;[3]
- **BERT**: using the most popular embeddings provided by Google Research (Devlin et al. 2018) with the base uncased pre-trained weights;[4]
- **USE**: using sentence embeddings coming from Universal Sentence Encoder (Cer et al. 2018) which has been specifically trained for sentence similarity[5]

In all the cases, the representations from these models have been compared with the cosine similarity. For each pair of sentences that was provided by any of the models, we listed by manual analysis which differences were contained, in terms of details that changed, or different words used.

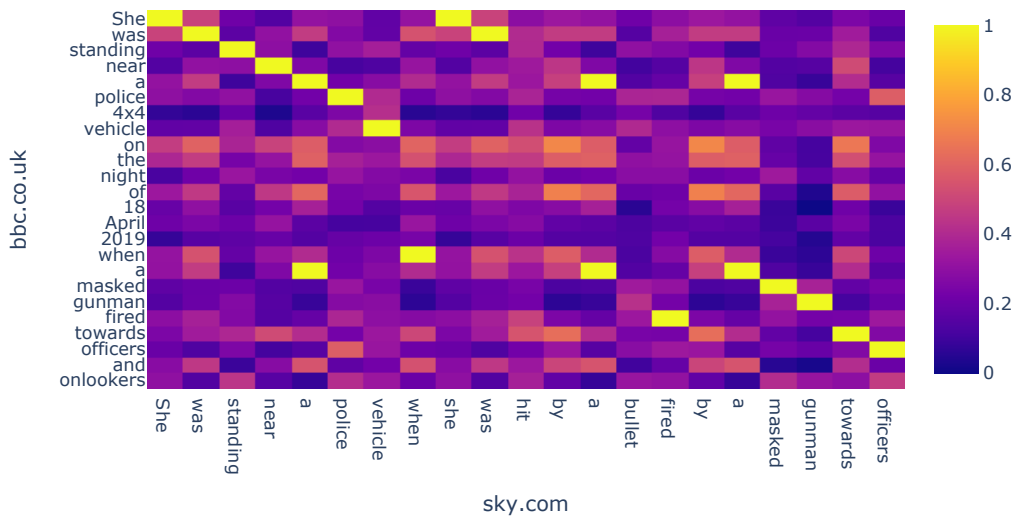


Figure 5.1: The comparison between two sentences, one from the BBC and the other from Sky News, where multiple differences exist.

An example can be seen in Figure 5.1 that shows a sentence from the BBC[6] and one from Sky News,[7] where the differences are the following:

- the detail "4x4" just appears on the BBC article;

[3] https://spacy.io/models/en#en_core_web_lg
[4] https://spacy.io/models/en-starters#en_trf_bertbaseuncased_lg
[5] https://tfhub.dev/google/universal-sentence-encoder/
[6] https://www.bbc.co.uk/news/uk-england-hereford-worcester-51791346
[7] https://www.dailymail.co.uk/news/article-8088805/Britons-facing-heavy-downpours-four-inches-rain-50mph-winds-set-batter-UK.html

- the detail "on the night of 18 April 2019" just appears on the BBC article;
- Active vs passive sentence "a masked gunman fired" vs "she was hit by a bullet fired by a masked gunman";
- the detail "a bullet" just appears in the Sky article;
- "Towards officers and onlookers" vs "towards officers": in the second, the targets of the gunman are just the officers.

With this kind of information on the number, type and magnitude of changes contained in different pairs of sentences, we can get a qualitative idea of how the measures of similarity coming from the different models are representative of the effective differences. If a model scores more similar a pair of sentences that appear to us to be less related than another pair, that is a negative sign for that model.

The observations that we have for the TF-IDF model is that the feature size changes the results a lot. It requires to be computed on a set of documents all together (which also changes which features are selected), and it is not possible to encode an additional document without changing the representation of the already encoded documents. We also see that the type of pre-processing affects the results: without a lemmatization step to the pipeline, it is sufficient to change the verb tense to have a different term. Given these limitations, we find that sentences that are very similar in the meaning but have some differences in the linguistic surface see a drop in their similarity with this model.

Instead considering GloVe-average, we observe in some cases that the measure of similarity provided does not capture substantial changes in the meaning. The problem is that, while it can use wordwise similarity quite well, the sentence structure is not accounted for its representation. The representation is a simple average of the word vectors (e.g. "Luke insulted John" results in being equal to "John insulted Luke").

For the language models (BERT and USE models), we see a big improvement in the pairs of sentences that come as more similar. We observe that USE provides values less skewed to the higher end, distributing the similarity values more evenly. The numbers make more sense without any re-scaling technique, and therefore the heatmaps shown in this document come from this model. It is also the only model considered that is purposely trained on a semantic similarity task, while the other models can provide similarity measures just because of how they represent language.

This experiment shows the need for a similarity model that accounts for the semantics more than the linguistic surface. And given the continuous progress of language models, we need to be able to switch our choice relatively easily. For example, by looking at the Semantic Textual Similarity benchmark,[8] at the moment the best model available is XLNet (Yang et al. 2019) so we will use it for our future experiments.

The purpose of this experiment is to have a good observation of how different types of models can be effective or not, and to experiment with them to drive the implementation of the processing pipeline.

5.1.3 Sentence clustering and extraction of differences

The next experimentation that we have done regards the usage of the similarity values to group together sentences describing the same details and at the same time

[8] `http://nlpprogress.com/english/semantic_textual_similarity.html`

study the uniqueness of the words used. We have seen with the reproduction of the model from Bountouridis et al. (2018) that one big limitation of using cliquing techniques over unweighted graphs is that they do not exploit the full power of the distances available, which resulted in having fragmented clusters due to a choice of the "similar-enough threshold" that is very sensible.

With this idea, we retrieved some groups of articles that relate to the same event from Google Headlines, which aggregates and clusters together news articles from multiple sources.[9] These documents are processed with the SpaCy NLP Python library[10] to split the documents into sentences and have available different NLP functions (e.g., tokenisation, POS tagging).

Each of the sentences is then passed through a language model which creates a sentence embedding, in this case using the Universal Sentence Encoder because it showed to distribute the similarity values more evenly and is specifically trained for sentence similarity.

We then use agglomerative hierarchical clustering for different reasons:

- it does not require the specification of the number of clusters wanted, we want to be flexible;
- we can truncate the clustering when we reach a certain level of distance between the clusters, or a certain number of clusters;
- We can see the evolution of many different features (e.g., number of clusters, size, internal cohesion) while performing the clustering step by step;
- we have a graphical representation (dendrogram) which helps to inspect and understand what is happening;
- it has widely been used for similar tasks (e.g., finding related claims (Almeida & Santos 2020))

As we can see in Figure 5.2, we can explore what is the distance required to have different sentences inside the same cluster, and select a certain threshold more consistently.

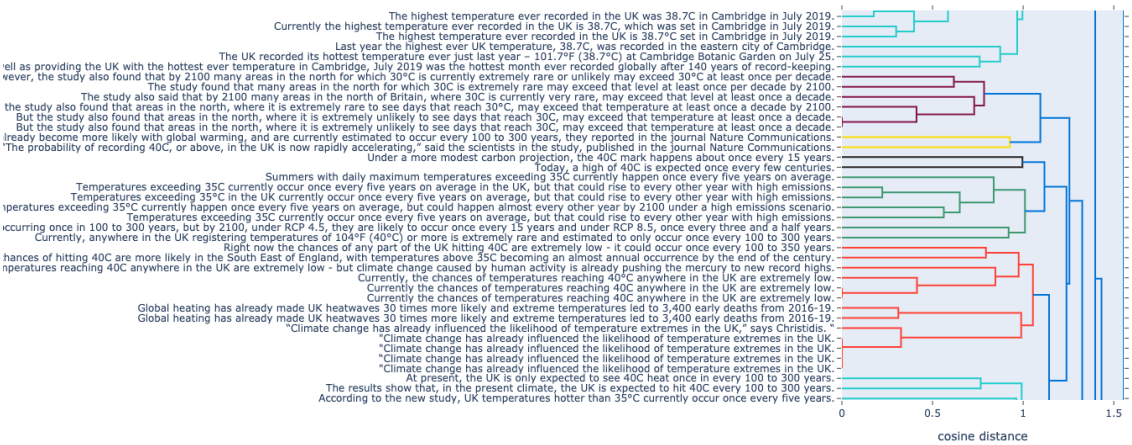


Figure 5.2: A portion of the dendrogram that shows how different sentences are merged in the clusters by increasing distance values.

Using this type of inspection, we can decide specific threshold values more con-

[9] https://www.blog.google/products/news/new-google-news-ai-meets-human-intelligence/

[10] https://spacy.io/

sistently: we can see what is the required similarity to make two pairs of sentences be in the same cluster.

With this method, we create sentence clusters that are very similar in their semantic content, but at the same time have linguistic changes. To facilitate an analysis of the differences, we experimented with different methods of highlighting the uniqueness of the words in a cluster. This can be seen in Figure 5.3 where we score each of the words with a scale of uniqueness that is defined as:

$$u_w = 1 - \frac{|\{ s_i \mid s_i \in S \wedge w \in s_i \}|}{|S|}$$

where S is the set of sentences considered, w is the word for which to compute the index. The expression compares the sentences of the current cluster where w appears (numerator) with respect to the cluster size. A value close to 1 means that the word is used in just a few sentences in the cluster. This gives a higher value of uniqueness to the word "deceased" that just appears in one over three sentences ($u = 2/3$), while "surgery" has $u = 0$.

0: Three officers are deceased , two are in surgery , and three are in critical condition .

1: Two officers underwent surgery , and three are in critical condition , according to the police .

2: Two officers were in surgery Thursday night and three others were in critical condition , according to a police statement .

Figure 5.3: A sentence cluster example, where the uniqueness of words in the cluster is highlighted with different shades of green.

This methodology can be used in our framework to help the preparation of the dataset in different steps. First of all, during the creation of article couples that have to be compared (using at the article-level the same clustering methodology). In this case, we will need to use a specific threshold that cuts out unrelated articles but at the same time keeps a considerable number of differences on the document level (not too similar because identical articles, that are a lot, are not useful). While for the first user study we can exploit hand-curated groups of articles, when doing a larger creation of the dataset we will need to rely on automated techniques. This methodology could also be used as support for the annotators, to see both which sentences are most similar, and the words that differ inside. This would make the annotation process faster and easier.

This experiment needs to be completed, to choose some parameters with better criteria. We want to identify good intervals for the thresholds and parameters (e.g. with euclidean distance around 0.6-1.0 sentences start to have linguistic variations but still very related to the same concepts).

This experiment evidences the need to explore more on the interpretation of the differences, pushing for the user study of RQ1.1.

It serves the RQ1.2 as a first implementation of the processing pipeline by making available different articles and their document and sentence-wise relationships. Building on top of these features, we can then develop the methodology for doing the cross-article framing analysis.

5.1.4 Data collection

We also started collecting, early this year, several types of data that will be useful for the analysis planned. When looking for data, we are interested in different features. First of all, a wide set of articles is needed, dense in time and from a wide variety of news outlets. We need substantial overlap between articles and the more sources we can include, the better we can observe variations of the framing phenomena. Another very important feature is to have a good pre-clustered set of articles to help the curation of a dataset, especially for the first Research Question, by knowing that the articles are well related. Then when we will have the document clustering in action, this feature is not anymore required, but still can serve as a benchmark for that stage. And another desirable feature would be to have articles that come from sources with a different opinion, that would be beneficial to create examples especially for the user study, where we want to maximise the occurrence of framing techniques.

Given these requirements and after exploring different news aggregators, we found that Google Headlines Full Coverage feature[11] would fit the requirements of covering a big number of news sources (the `en-GB` version contains articles from more than 10k domains) and being very dense (an average of 9k new articles each day). The data comes divided by topics (Latest, United Kingdom, World, Business, Technology, Entertainment, Sports, Science, Health) and inside each topic the articles are grouped in "stories". Each story has articles from the most relevant sources ("Top coverage") and then lists also articles from other less important sources, for an average of 21 articles inside each story. The stories are created automatically by Google News and this allows it to be always updated and be so diverse in the sources included. We have captured from mid-march every day the published set of stories, and managed to retrieve more than 700k articles (as of the end of June, and not all the articles listed can be retrieved because of paywalls or other filtering techniques by the publishers).

Another data source that we actively retrieve is AllSides which provides a curated set of "headlines"[12] where three articles with a different political alignment are put together and compared in their difference. The curators describe how the story gets framed by the considered sources, using natural language. This description usually contains the usage of terms or themes that get mentioned. At the end of June, we have available 4764 headlines, with 13979 articles linked. Differently from Google Headlines which has different versions for each country, this data is US-focused being curated in the US and therefore has a much more limited scope. Also, the discussion of bias and framing is mainly focused on political issues, while we want to focus also on other types of differences of opinion. This data, although the description of the differences is not directly parseable, will be used to feed the user study and understand the role of comparing different sides.

The data collection done in these current months will continue across the PhD, and will be used in different stages of the analysis. It will serve as the seed to create the labelled dataset for the first Research Question and also provide a wide set of articles from different news sources to empower the studies of the second Research Question. Having articles from so many different news sources, we can on one side

[11] `https://www.blog.google/products/news/new-google-news-ai-meets-human-intelligence/`

[12] `https://www.allsides.com/story/admin`

provide some indication of framing for news sources that are not usually targeted by manual framing studies because not enough "important", and on the other side be more confident to observe some phenomenon of information re-usage that is the underlying hypothesis for the last sub-question.

5.1.5 Formalisation and dissemination

The last type of activity carried out during this year has been the presentation of this work to other researchers throughout different events, both internally to the Open University and externally. The motivation of this activity has been to get some feedback from both people working inside the same research space and also from other fields. From the first group, we wanted to get an expert opinion and mainly understand if we are missing some related research work that could be helpful. Instead, from the more general-audience group, we wanted to understand if this type of research makes sense to them and try to explain more motivationally and in an easier format.

Belonging to the first group, this work has been presented firstly to an internal seminar in KMi on the 25th March, then to the Text2Story workshop part of the ECIR conference as a position paper which was presented in April[13]. This position paper (Mensio et al. 2020) focuses on describing some proposed cross-article signals that would show differences in how stories are narrated. The proposal described completely in this report has also been presented in the CRC PhD Conference, that is an internal conference for PhD students of KMi and C&C schools.

For the more wide audience instead, during June we submitted a poster to the OU PhD Poster Competition which, involving a more general audience, focused more on being simple to understand.

These documents can be found at the end of this file.

5.2 Plan

In Figure 5.4 we show the time plan, which mirrors the structure of the research questions.

The first phase will target RQ1.1, by doing the user study proposed before. This will require different stages, from the preparation of the data to be annotated and guidelines for the participants, to a small pilot to correct potential errors in the procedure, to the real user study and analysis of the results.

The outputs of the first phase will be needed by the second one, to create and annotate a larger dataset with the specific labels refined with the first study. The implementation of the pipeline is independent of this requirement so it can be done in parallel. The implementation of the detection model will sit on top of the pipeline.

After a month of buffer time, the analysis of the second research question will be faced, with two cycles of data preparation, implementation and analysis.

We plan for each of the groups to use the results of the analysis in papers that will be submitted to upcoming conferences.

The last six months will be dedicated exclusively to writing the thesis.

[13] `http://text2story20.inesctec.pt/`

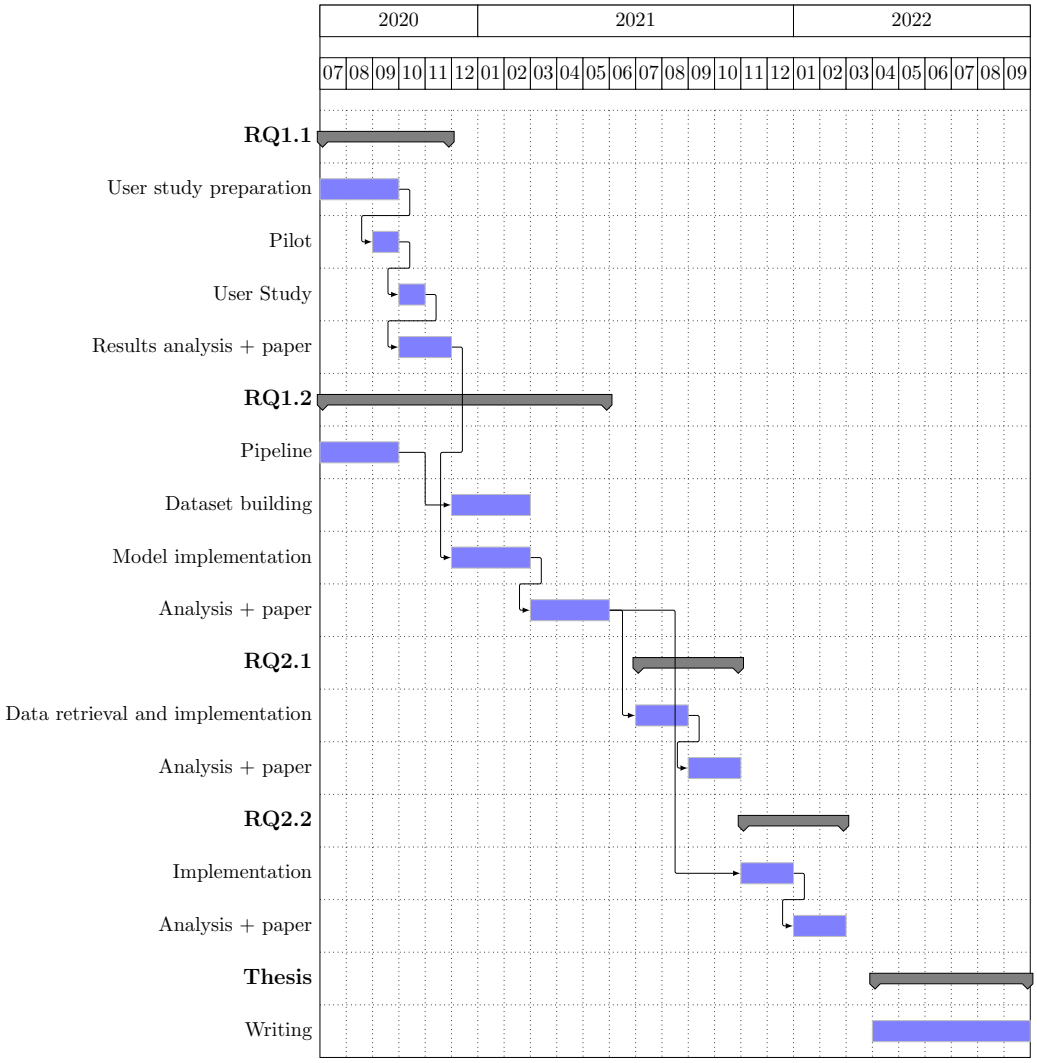


Figure 5.4: Gantt chart showing the time plan

Bibliography

Almeida, C. & Santos, D. (2020), 'Text similarity using word embeddings to classify misinformation', *arXiv preprint arXiv:2003.06634* .

Andrews, N. O. & Fox, E. A. (2007), Recent developments in document clustering, Technical report, Department of Computer Science, Virginia Polytechnic Institute & State ....

Baker, C. F., Fillmore, C. J. & Lowe, J. B. (1998), The berkeley framenet project, *in* 'Proceedings of the 17th international conference on Computational linguistics-Volume 1', Association for Computational Linguistics, pp. 86–90.

Baumer, E., Elovic, E., Qin, Y., Polletta, F. & Gay, G. (2015), Testing and comparing computational approaches for identifying the language of framing in political news, *in* 'Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies', pp. 1472–1482.

Bell, A. (1991), *The language of news media*, Blackwell Oxford.

Bountouridis, D., Marrero, M., Tintarev, N. & Hauff, C. (2018), Explaining credibility in news articles using cross-referencing, *in* 'SIGIR workshop on ExplainAble Recommendation and Search (EARS)'.

Boydstun, A. E. (2013), *Making the news: Politics, the media, and agenda setting*, University of Chicago Press.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E. & Shah, R. (1994), Signature verification using a" siamese" time delay neural network, *in* 'Advances in neural information processing systems', pp. 737–744.

Bryant, J., Thompson, S. & Finklea, B. W. (2012), *Fundamentals of media effects*, Waveland Press.

Card, D., Boydstun, A., Gross, J. H., Resnik, P. & Smith, N. A. (2015), The media frames corpus: Annotations of frames across issues, *in* 'Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)', pp. 438–444.

Carpineto, C., Osiński, S., Romano, G. & Weiss, D. (2009), 'A survey of web clustering engines', *ACM Computing Surveys (CSUR)* **41**(3), 1–38.

Cer, D., Yang, Y., yi Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B. & Kurzweil, R. (2018), 'Universal sentence encoder', *arXiv preprint arXiv:1803.11175* .

Cohen, B. C. (2015), *Press and foreign policy*, princeton university press.

Conneau, A. & Kiela, D. (2018), SentEval: An evaluation toolkit for universal sentence representations, *in* 'Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)', European Language Resources Association (ELRA), Miyazaki, Japan.
URL: *https://www.aclweb.org/anthology/L18-1269*

Da San Martino, G., Yu, S., Barrón-Cedeno, A., Petrov, R. & Nakov, P. (2019), Fine-grained analysis of propaganda in news article, *in* 'Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)', pp. 5640–5650.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018), 'Bert: Pre-training of deep bidirectional transformers for language understanding', *arXiv preprint arXiv:1810.04805* .

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

Field, A., Kliger, D., Wintner, S., Pan, J., Jurafsky, D. & Tsvetkov, Y. (2018), 'Framing and agenda-setting in russian news: a computational analysis of intricate political strategies', *arXiv preprint arXiv:1808.09386* .

Fillmore, C. J. (2006), 'Frame semantics', *Cognitive linguistics: Basic readings* **34**, 373–400.

Gamson, W. A. & Modigliani, A. (1989), 'Media discourse and public opinion on nuclear power: A constructionist approach', *American journal of sociology* **95**(1), 1–37.

Gangemi, A. (2013), A comparison of knowledge extraction tools for the semantic web, *in* 'Extended semantic web conference', Springer, pp. 351–366.

Goffman, E. (1974), *Frame analysis: An essay on the organization of experience.*, Harvard University Press.

Greene, S. & Resnik, P. (2009), More than words: Syntactic packaging and implicit sentiment, *in* 'Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics', pp. 503–511.

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

Jones, K. S. (1972), 'A statistical interpretation of term specificity and its application in retrieval', *Journal of documentation* .

Karadzhov, G., Nakov, P., Màrquez, L., Barrón-Cedeño, A. & Koychev, I. (2017), 'Fully automated fact checking using external sources', *arXiv preprint arXiv:1710.00341* .

Liu, S., Guo, L., Mays, K., Betke, M. & Wijaya, D. T. (2019), Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence, *in* 'Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)', pp. 504–514.

McCombs, M. E. & Shaw, D. L. (1972), 'The agenda-setting function of mass media', *Public opinion quarterly* **36**(2), 176–187.

Mensio, M., Alani, H. & Willis, A. (2020), Towards a cross-article narrative comparison of news, *in* 'Proceedings of the Text2Story'20 Workshop', CEUR WS.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013), 'Efficient estimation of word representations in vector space', *arXiv preprint arXiv:1301.3781* .

Miller, G. A. & Charles, W. G. (1991), 'Contextual correlates of semantic similarity', *Language and cognitive processes* **6**(1), 1–28.

Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. R. & Liu, H. (2018), 'Identifying framing bias in online news', *ACM Transactions on Social Computing* **1**(2), 1–18.

Niculae, V., Suen, C., Zhang, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. (2015), Quotus: The structure of political media coverage as revealed by quoting patterns, *in* 'Proceedings of the 24th International Conference on World Wide Web', pp. 798–808.

Pan, Z. & Kosicki, G. M. (1993), 'Framing analysis: An approach to news discourse', *Political communication* **10**(1), 55–75.

Pennington, J., Socher, R. & Manning, C. D. (2014), Glove: Global vectors for word representation, *in* 'Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)', pp. 1532–1543.

Pennycook, G. & Rand, D. G. (2019), 'Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning', *Cognition* **188**, 39–50.

Potthast, M., Stein, B., Barrón-Cedeño, A. & Rosso, P. (2010), An evaluation framework for plagiarism detection, *in* 'Coling 2010: Posters', pp. 997–1005.

Rugg, D. (1941), 'Experiments in wording questions: Ii.', *Public opinion quarterly* .

Scheufele, D. A. (1999), 'Framing as a theory of media effects', *Journal of communication* **49**(1), 103–122.

Scheufele, D. A. & Tewksbury, D. (2007), 'Framing, agenda setting, and priming: The evolution of three media effects models', *Journal of communication* **57**(1), 9–20.

Schuldt, J. P., Konrath, S. H. & Schwarz, N. (2011), '"global warming" or "climate change"? whether the planet is warming depends on question wording', *Public opinion quarterly* **75**(1), 115–124.

Tsur, O., Calacci, D. & Lazer, D. (2015), A frame of mind: Using statistical models for detection of framing and agenda setting campaigns, *in* 'Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)', pp. 1629–1638.

Tversky, A. & Kahneman, D. (1981), 'The framing of decisions and the psychology of choice', *science* **211**(4481), 453–458.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R. & Le, Q. V. (2019), Xlnet: Generalized autoregressive pretraining for language understanding, *in* 'Advances in neural information processing systems', pp. 5754–5764.

Yin, X., Han, J. & Philip, S. Y. (2008), 'Truth discovery with multiple conflicting information providers on the web', *IEEE Transactions on Knowledge and Data Engineering* **20**(6), 796–808.

Zahid, I., Zhang, H., Boons, F. & Batista-Navarro, R. (2019), Towards the automatic analysis of the structure of news stories., *in* 'Text2Story@ ECIR', pp. 71–79.

Towards a Cross-article Narrative Comparison of News

Martino Mensio
martino.mensio@open.ac.uk

Harith Alani
h.alani@open.ac.uk

Alistair Willis
alistair.willis@open.ac.uk

The Open University, United Kingdom

Abstract

In the world of public misinformation, there are many cases where the information is not false or fabricated, but rather has been manipulated using more subtle techniques such as word replacements, selection of details, omissions and argument distortion. These techniques can have the effect of influencing the reader's frame of mind towards the events reported. We currently lack the necessary tools to uncover such manipulations automatically. In this position paper, we propose an integrated analysis framework and pipeline to identify various narrative signals in news articles; such as structural roles, framing, and subjectivity. By comparing these at the document level and sentence level, it will be possible to highlight differences of narrative techniques used to report the same news events.

1 Introduction

Narrative analysis refers to the processing of a piece of text to understand and characterise its structure [Rie93]. Such an analysis could help to distinguish between event reports based on their narrative structure. These are usually reflected through linguistic signals that can be more or less explicit, such as emphasising certain aspects, changing the order in which certain information is presented, or using specific terminology to impose or stress a certain opinion.

In the specific case of news articles, their narrative structure usually follows a complex non-chronological sequence, which tends to differ from other kinds of narrative that proceed more linearly [ZZBBN19]. It is a choice that is made to "*get a good story*" [Bel05], and can be exploited to emphasise or introduce non-objective statements or causality relationships between events [Dah10].

To avoid being manipulated, one solution suggested in the literature is to gather information from multiple sources [ABS14, GAR97], and to cross-compare them in order to get a broader view of the event. The same information, for example, may be presented by some sources and omitted by others, or the sequence of events be presented differently to emphasise different aspects. Therefore, we believe that readers should be made more aware of the narrative and framing embedded in the piece of news they are consuming, and how they compare with those in other articles reporting the same event. Currently, there are hardly any automated tools that offer such functionality: the best readers can do is to use news aggregators that show articles grouped by events, but they have to do such comparison on their own.

In this position paper, we suggest a framework to automatically highlight the differences in how the same story is presented by different articles, by cross-comparing their narratives. To this end, the contributions of this paper are: *i)* integrating several signals characterising the narrative of news; *ii)* presenting a processing pipeline to link together similar articles at the document and sentence level, integrating the signals identified; and *iii)* introducing a set of cross-article signals that aim to highlight the difference of narrative techniques applied.

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia (eds.): Proceedings of the Text2Story'20 Workshop, Lisbon, Portugal, 14-April-2020, published at http://ceur-ws.org

2 Related work

In this section, we provide an overview of previous studies in two areas of research. First, the investigation on relationships between news articles which aims to find documents that cover the same information. Second, the detection of narrative linguistic signals, which investigates and characterises several aspects of structure, framing, and subjectivity. For both of them, we gather a set of techniques that enable our approach described in the next Section 3.

2.1 Relationships between news articles

There are different possible types of relationships between news articles, such as similarity (covering the same information), referencing (one is citing another one), and temporal proximity. They can be performed at the document level (e.g., the whole article is similar to another one) or at the sentence level (e.g., the same sentence is corroborated by a sentence in another article [BMTH18]) or even at the paragraph level. Since we are interested in finding articles discussing the same information, we focus on similarity relationships. Other relationships could add interesting features, such as the order of publication which would help to identify which of the articles might have taken inspiration from the other. For the time being, we focus on studying and understanding the role of similarity.

At the article-level, there is a wide variety of work that investigates article clustering, and the methods mostly used are Latent Dirichlet Allocation (LDA) or document embedding. LDA [BNJ03] is the most used technique for topic modelling, as it allows the discovery of topics and to group articles accordingly using word distributions. Another technique for grouping articles together is to compute a similarity measure (e.g., cosine similarity) between numeric representations of the documents (TF-IDF [Jon72] or Language Models [DCLT18, CYyK$^+$18, YDY$^+$19]). We plan to study these models in order to select the one that can efficiently discriminate articles that talk about the same events, even if they use different linguistics, from articles that may use the same subset of words but talk about different events.

Furthermore, there are works that not only link the articles at a document level, but also investigate in more detail the connections between sentences. In one recent work [BMTH18], groups of similar articles are found, then broken down to pieces of information and analysed to find if these details are *corroborated* (occurring in multiple documents) or *omitted* (occurring in other documents of the same group, but not the current one). We aim to use this idea of applying similarity to both article-level and sentence-level, extending it even to the word-level. By doing so, not only we might be able to recognise which sentences appear in multiple documents (with different degrees of similarity) but also we would be able to identify the specific words that have been changed.

However, this set of approaches are limited to bringing to the attention of the reader the linked information pieces with a measure of similarity, without characterising the differences. The reader would then need to evaluate the differences in the role of the sentence, the framing that it implies and how it compares with other sentences in terms of subjectivity. Different documents may express the same set of details, but give them a different role (reporting an action, commenting, contextualising, doing a digression, identifying causes and consequences) and use different words that are semantically similar but may imply a different framing perspective. For this reason, the next subsection presents a set of narrative linguistic signals that could provide us with the missing features.

2.2 Narrative linguistic signals

There is much research on exposing the narrative using linguistic signals [ZZBBN19], with specific words that indicate the *structural role*, *framing* and *subjectivity* of the part of text they belong to. One limitation is that most of such works are applied to single articles, with little comparison between them.

On one hand, some research considers the *structural role* of a sentence in the document (e.g., is it providing some background, the main event, an evaluation). Different structural roles have been defined in the literature, such as news schema [Bel91], which identifies hierarchical categories (e.g., action, reaction, consequence, context, history), narrative structure [Bel05] (e.g., abstract, orientation, evaluation, complication, resolution), or linguistic signals [ZZBBN19, Mar00]. Such signals could be used to identify the differences between similar sentences with regards to their structural roles in the articles.

On the other hand, there is much literature on *framing*, defined as how a certain story is presented to shape mass opinion [Gof74], the addition to the underlying facts that reflects the sociocultural context and acts as an underlying force to persuade the reader. The work by [GM89] describes a set of *framing packages*, made

of *framing devices* (e.g., word choice, metaphors, catchphrases, use of contrast, quantification) and *reasoning devices* (e.g., problem definition, cause, consequence, solution, action). Additionally, the Frame Semantics Theory [Fil06] can be used to recognise lexical units of known frames. By extracting these linguistic signals, we could represent the framing behind a certain piece of text, and there exist different approaches to extract the listed features [MGB+17, GCCZ18, Asg16, STDS17].

In addition to these two characterisations, we can add other signals derived from studies on *subjectivity*. As found by recent research, in contemporary journalism the line between opinion and facts is blurring more and more [JWJ+19]. For this reason, having signals of subjectivity on the document and paragraph-level would be very useful [Liu10]. In this way, each article and each paragraph can be characterised with an indication of subjectivity.

All these features have been used in previous research, but as mentioned above, they are mainly applied to single-article analysis. Extending this kind of analysis by taking into consideration the relationships both at the article level and the sentence level would bring a big contribution by providing contrastive signals that would not come up otherwise.

3 Cross-article comparison framework

In this section, we propose a description of our comparison framework. We plan to use methods coming from both the research areas identified (document linking and linguistic signals) as a starting point. In order to do so, we propose the following processing pipeline:

- **preprocessing**: documents are retrieved, cleaned up and fragmented into paragraphs and sentences;
- **narrative features** are attached to each document, paragraph and sentence belonging to three main types: *structural role* using and highlighting the linguistic devices provided by [ZZBBN19]; *framing features* are extracted (framing and reasoning devices) finding some linguistic representatives from [GM89, Fil06]; *subjectivity* is computed, and strong word choices are highlighted [Liu10];
- **linking**: *similar articles* are found by using document-level similarity measures: in this way it would be possible to find groups of documents that describe the same events; *similar sentences and paragraphs* are found by sentence-level similarity measures, inside each group of documents: corroborated and omitted sentences are identified [BMTH18].

Figure 1 shows the result of such processing over two articles, where we have several features attached to the sentences, with similar paragraphs across the two articles linked together using a similarity measure [CYyK+18].

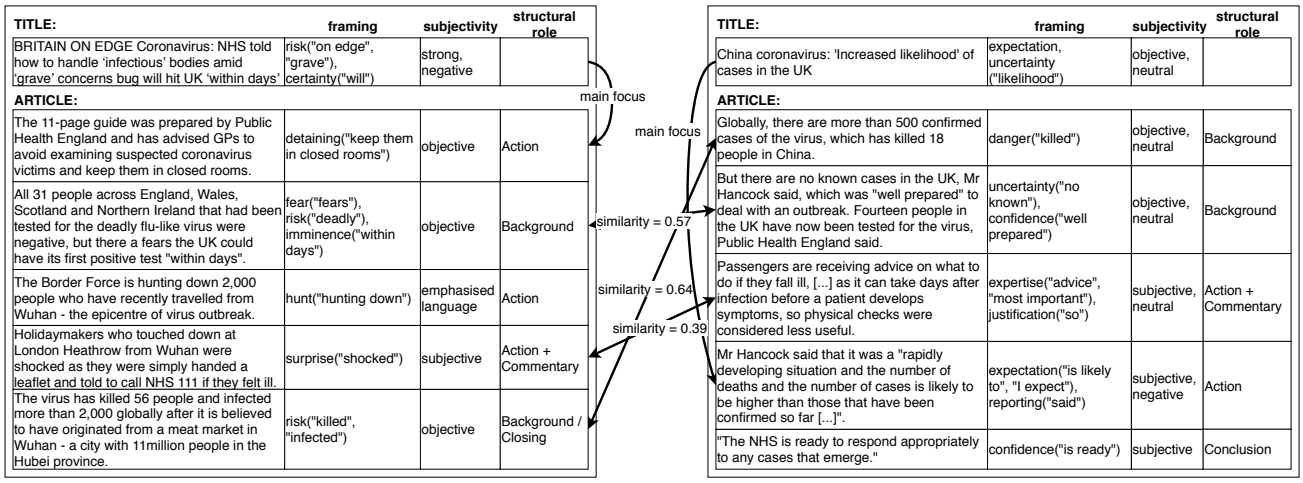


Figure 1: An example of analysis between two news articles that both talk about the risk of coronavirus spread in the UK. The first one (from The Sun) emphasises the risks from the virus, while the second article (from BBC) is more focused on presenting the UK as ready to face the problem. Each paragraph is characterised with framing, subjectivity and structural signals, and the links between the articles represent the most similar pairs of sentences.

This is is the starting point to identify the differences, with a contrastive analysis. We propose here a set of *cross-article comparative signals* that can bring the narrative analysis a step further:

- The **main focus** of the compared articles is on a different part or detail of the story: this means that while they are both describing the same broad event, they are trying to emphasise or prioritise two different aspects. This signal can be computed by looking at the most similar sentence to the article title (proxy of the emphasis), and seeing how it is represented in other documents.
- **Ordering**: the compared articles present the same details, but in a different order. Re-ordering events tends to be an efficient way of creating implicit cause-effect relationships. To do this comparison, it is sufficient to find the crossovers in the sentence-level connections.
- **Selection of details**: One article is *omitting* certain details that have been reported by other articles, or is describing events that are *corroborated* by other sources, or has *unique parts* that do not occur in other articles [BMTH18]. In addition to seeing which parts are selected or omitted, the narrative analysis can help us to find some insights about them (e.g., the article is omitting subjective statements reported by others, or is describing a background event that others did not include).
- The articles are **framing** the narrative in different ways from each other. This manifests through comparing linked sentences to observe the differences in terms of framing features: the considered articles are describing the same events but with different framing and reasoning. One concrete example is the usage of *causality*: one article may contain causality signposting between a pair of sentences that is absent elsewhere. Or as another example, the usage of *specific words* can reveal a specific framing: talking about the same detail or entity, the usage of verbs or adjectives may change. For detecting such peculiarities, features as Named Entities and subjectivity may be combined.
- The comparison can be also done on the **subjectivity** of the article: both at the document level (saying that this is an opinion piece, while a similar one is more factual) or at the sentence level, by interweaving this signal with the ones proposed before.

From the signals in Figure 1, we can see that the first article pushes the narrative towards `risk` and other negative frames, to sustain the idea presented in the title "Britain on Edge". The second article, even though it has a lot of information in common with the first one, is more confident on the preparedness of the National Health Service to face the virus (e.g., `confidence`, `expertise`). The extraction of these cross-article signals is the first step to finding possible cases of manipulation.

4 Evaluation

The evaluation of this framework needs to be performed at different levels. Firstly, we need to find a similarity model that performs well both at the article and sentence levels, going beyond the linguistic surface and being able to relate pieces of text that may use different terms for describing the same events. The evaluation of the similarity measure will be done at the article level using data coming from tools that aggregate articles talking about the same events, such as Google News Headlines[1] and AllSides[2] as well as research datasets such as NewsAggregator.[3] Instead, for the sentence-level similarity, user feedback will be needed to understand when and why a sentence is considered to describe the same detail while we are dealing with manipulations that can be significant.

Following that, we would also need to evaluate the whole framework with user studies to understand the relevance, quality and usefulness of the indicators proposed. Currently, and to the best of our knowledge, there are no similar approaches to the task we are addressing in this paper, and hence we are unable to establish comparisons with other baseline approaches from the literature.

5 Discussion

Much research exists that address the problem of misinformation. However, the vast majority of such research focuses on distinguishing what is true from what is false, and hence mainly applies to a small subsection of the

[1] `https://news.google.com/`
[2] `https://www.allsides.com/story/admin`
[3] `http://archive.ics.uci.edu/ml/datasets/News+Aggregator`

misinformation ecosystem.[4] There is a lack of research on identifying *misleading content*, *false connection* and *false context*. To this end, there is an immediate need for technological solutions to address such cases, where the information is manipulated in a subtle fashion, and thus cannot be easily dismissed as false. We want to reveal the differences in reporting, without declaring that one article contains true or false information, but rather to provide a tool that exposes such diversities.

In this paper we proposed a comparative approach that aims at bringing into light the differences in the narratives of news articles, using a set of cross-article narrative signals. These signals only exist when multiple documents are compared, in contrast to single-article ones that already exist. With this method, we aim to reveal the framing intentions of the writers, and making them more evident and comparable.

This analysis may be useful for empowering users to form a critical view of pieces of news they are consuming, to find missing pieces that have been omitted and to see the same information presented with a different framing by different articles and sources.

Acknowledgements

This work is partially supported by EU H2020 Project Co-Inform (grant no. 770302).

References

[ABS14] Øistein Anmarkrud, Ivar Bråten, and Helge I Strømsø. Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, 30:64–76, 2014.

[Asg16] Nabiha Asghar. Automatic extraction of causal relations from natural language texts: a comprehensive survey. *arXiv preprint arXiv:1605.07895*, 2016.

[Bel91] Allan Bell. *The language of news media.* Blackwell Oxford, 1991.

[Bel05] Allan Bell. News stories as narratives. *The language of time: a reader*, page 397, 2005.

[BMTH18] Dimitrios Bountouridis, Mónica Marrero, Nava Tintarev, and Claudia Hauff. Explaining credibility in news articles using cross-referencing. In *SIGIR workshop on ExplainAble Recommendation and Search (EARS)*, 2018.

[BNJ03] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[CYyK+18] Daniel Cer, Yinfei Yang, Sheng yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*, 2018.

[Dah10] Michael F Dahlstrom. The role of causality in information acceptance in narratives: An example from science communication. *Communication Research*, 37(6):857–875, 2010.

[DCLT18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Fil06] Charles J Fillmore. Frame semantics. *Cognitive linguistics: Basic readings*, 34:373–400, 2006.

[GAR97] John T Guthrie, Solomon Alao, and Jennifer M Rinehart. Literacy issues in focus: Engagement in reading for young adolescents. *Journal of Adolescent & Adult Literacy*, 40(6):438–446, 1997.

[GCCZ18] Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. Neural metaphor detection in context. *arXiv preprint arXiv:1808.09653*, 2018.

[GM89] William A Gamson and Andre Modigliani. Media discourse and public opinion on nuclear power: A constructionist approach. *American journal of sociology*, 95(1):1–37, 1989.

[4] `https://firstdraftnews.org/latest/fake-news-complicated/`

[Gof74] Erving Goffman. *Frame analysis: An essay on the organization of experience.* Harvard University Press, 1974.

[Jon72] Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 1972.

[JWJ+19] Kavanagh Jennifer, Marcellino William, Blake Jonathan, Smith Shawn, Davenport Steven, and Tebeka Mahlet Gizaw. *News in a Digital Age: Comparing the Presentation of News Information over Time and Across Media Platforms.* Rand Corporation, 2019.

[Liu10] Bing Liu. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition*, pages 627–666. CRC Press Book, 2010.

[Mar00] Daniel Marcu. *The theory and practice of discourse parsing and summarization.* MIT press, 2000.

[MGB+17] Arpan Mandal, Kripabandhu Ghosh, Arnab Bhattacharya, Arindam Pal, and Saptarshi Ghosh. Overview of the fire 2017 irled track: Information retrieval from legal documents. In *FIRE (Working Notes)*, pages 63–68, 2017.

[Rie93] Catherine Kohler Riessman. *Narrative analysis*, volume 30. Sage, 1993.

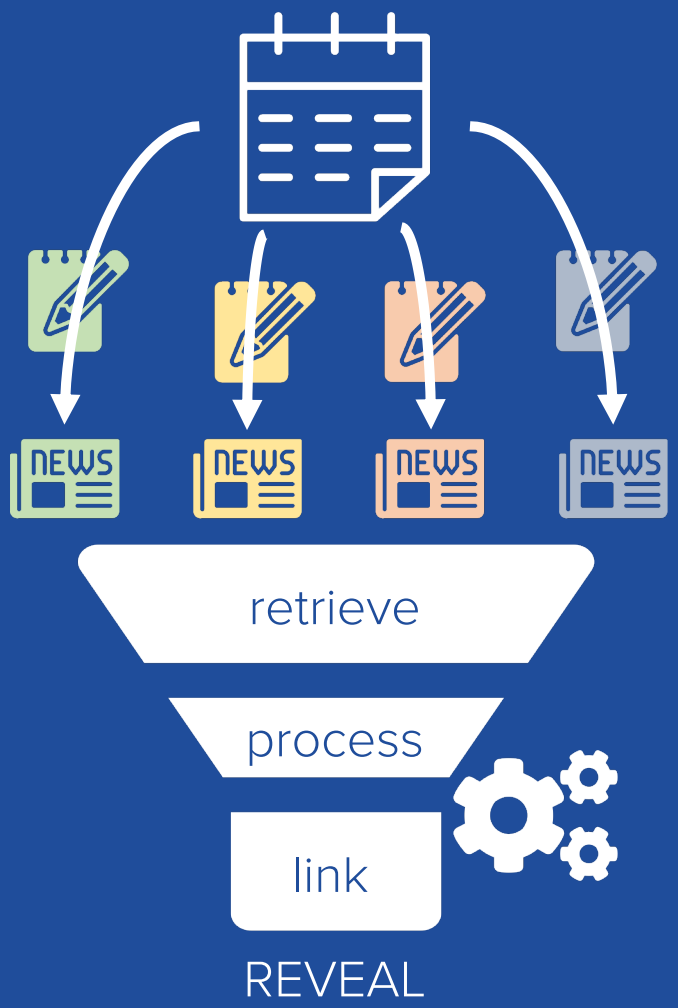
[STDS17] Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold. *arXiv preprint arXiv:1706.09528*, 2017.

[YDY+19] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pages 5754–5764, 2019.

[ZZBBN19] Iqra Zahid, Hao Zhang, Frank Boons, and Riza Batista-Navarro. Towards the automatic analysis of the structure of news stories. In *Text2Story@ ECIR*, pages 71–79, 2019.

One Event, Different Stories

Martino Mensio, Harith Alani, Alistair Willis
Knowledge Media Institute & Computing and Communications



How do different sources describe the same event?

Different factors influence the writing:

- Telling a "good story"
- Political leanings / personal views
- Intention of communication
- Target audience

Possible problems

- Incomplete picture of the event
- Addition of subjective content
- Mix of facts and opinions
- Manipulation of details
- Different contexts ("framing")

Research Questions

1. How to automatically reveal the differences in stories about an event?
2. How to identify patterns of information flow between news sources?

Differences in emphasis

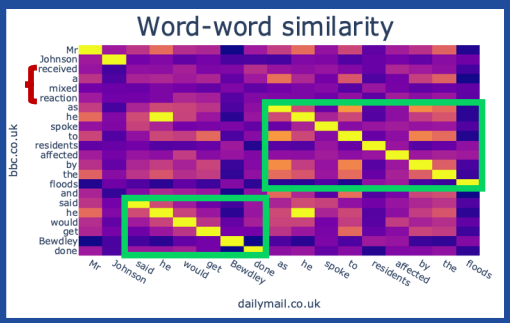
- Main focus
- Ordering events
- Loaded language

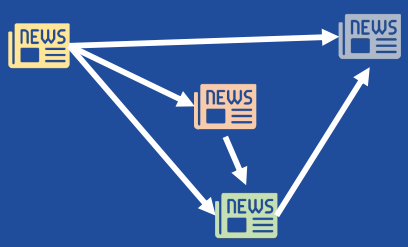
Selection of details

- The common pieces
- Adding subjective phrases
- Omitting unwanted details

Information flow

- How sources reuse contents from others
- Small changes





More details?

Position paper http://ceur-ws.org/Vol-2593/paper11.pdf