Language Analysis on Dutch Traffic Accident News

...

Jonas van Oenen

A thesis presented for the degree of
Bachelor of Science

UNIVERSITY OF AMSTERDAM

University of Amsterdam
2019

Contents

Chapter 1

Introduction

1.1 Introduction

There are numerous sources which signify the societal relevance of traffic accidents in the Netherlands. There is the general discussion in the news on how traffic accidents should be viewed [Verkade, a, Verkade and Brömmelstroet, , Verkade, b, Verkade, 2018].

Furthermore, there are the official reports on traffic accidents in the Netherlands which show the scale on which accidents occur[Kemler and den Hertog, 2009]. In addition, there has also been discussion in the Dutch government on how to prevent these accidents and the viability of different approaches[Aarts et al., 2014].There is also a report on the different data available on traffic accidents, with a score on how viable the information is [Tros,].

1.2 What

1.2.1 Research question

What kind of information and language is used in the reporting of traffic accidents?

1.3 Why (relevancy)

1.4 How

For this research, a dataset of news articles will be collected via the Lexis-Nexis service. The specifics of this dataset are to be determined. However, most likely, the data will consist of a certain time period and be collected through the use of a specified set of search terms. This dataset will be analysed through two consecutive approaches. The first approach is retrieving the information that is being reported and try to classify these automatically. Meaning, the article analysis will try to extract at much information about the subject of the news article.

The second approach is more oriented towards specific language use in the article. This approach will delve deeply into the way language is used and what kind of language is used in the news articles. Some of the sub-questions we will be looking at are described below.

Chapter 2

Theoretical foundation

2.1   Research context

This research relies heavily on theoretical knowledge of language and its relations with the forming of knowledge. There are different aspects which need to be addressed in order to give a complete overview of the subject. Firstly, a general understanding of the relation between power and knowledge. This relation between power and knowledge is needed in order to grasp the idea of discourse. And more importantly, how discourse and the language within discourse is affected by power relations. Understanding the role of discourse and its role in the forming knowledge is important because only then can we understand the forms in which the power relations in discourse present themselves.

There is, however, another important piece of the puzzle. News on traffic accidents cannot be simply defined as the language discourse on traffic accidents. There is of course the way we talk about traffic accidents -and traffic in general- in the news. However, traffic is also discussed in the sciences, politics and governmental reports. To suggest that news on traffic accidents represent the general discourse on traffic would be wrong. However, we can say something about the way in which specifically news talks about traffic accidents. In this context of news, there is a key factor. In scientific literature it is mostly referred to as framing. Framing refers to way in which an article is presented. Newspapers want their articles to be read, they are made appealing in various ways. For instance, a lot of news will be linked with other stories in order for them to be more relevant. This factor needs

to be kept in mind when talking about the discourse. And at times we must try to make a sharp distinction between what can be seen as discourse and what is result of framing.

2.1.1 Power and knowledge

...

2.2 Discourse and Framing

There are two concepts closely related to the subject at hand. There is the concept of *discourse*, originally used by Foucault, as that through which a subject is constituted and generates knowledge . It entails far more than this short description. However, in the context of this paper one aspect of this idea of discourse is of great importance. The feature of language as a means through which knowledge on the subject is instantiated. Language cannot be interpreted as a objective means through which information is simply communicated. Language itself forms the knowledge and the information of a subject. The terms, concepts and descriptions used in a specific context are paramount to the way we -as a people- understand a subject. This idea of language as forming knowledge also gives rise to a complication, the intertwinement with power. Stating that language is forming for knowledge also means that language can be used to alter the discussion simply by changing the way we talk about a subject. The power to define the language which is used to discuss a specific subject therefore becomes important. These power relations are not simply on display for everyone to see. They are deeply embedded in the concepts and terms we use, with no real perpetrator to be singled out. The task at hand is to uncover these relations, not too appoint blame, but to make clear the limitations and biases on which the language of a subject is based.

The second concept is not so much a philosophical concept but more used in sociological and psychological areas. However, it is of great importance in our approach to news articles. The concept of *framing*, the context in which a news article is presented.

referentie foucault

2.2.1 Critical Discourse Analysis

There is some very specific research on how to approach critical analysis of news discourse[Xie, 2018].

And some more general ideas on discourse and ideology in the news [Fowler, 1991]

2.2.2 News framing

There is various research on the concept on framing and the significance it has in the realm of news articles[Kühne and Schemer, 2015, Entman, 2007, Weaver, 2007, Scheufele and Tewksbury, 2006].

2.3 Research on traffic accidents in the news

There is various research on traffic accidents in the news. There is research on the impact news reporting has on the public image of traffic accidents[Mejia, 2017]. And this article also elaborates on the agenda setting importance of news.

Then there has also been substantial research on the coverage of traffic accidents in the news in relation to official accident databases[De Ceunynck et al., 2015, Daniels et al., 2010]. This research can be used to identify some of the main features that occur in the reporting of traffic accidents.

Lastly, there has also been some specific research on traffic discourse in newspapers [MacRitchie and Seedat, 2008]. However, this research was done in South Africa and there are quite a number of differences with traffic situation in the Netherlands. Still, it gives meaningful insights in the way discourse can affect the news.

2.4 Text analysis research

This research relies heavily on natural language processing. A number of approaches have been considered to extract the needed information from news articles. A -very recent- general overview of some text summarising techniques are given in this article [Nenkova and McKeown, 2012]. Furthermore, a lot of research has been done on more specific text mining

[Allahyari et al., 2017]. In this general overview on text mining, numerous approaches to our problem can be found.

There is the options of clustering articles, to find patterns which exist between the articles. Specifically, topic modeling has been met with a lot of success in creating a model for text documents. The two main algorithms are *Probabilistic Latent Semantic Analysis*(pLSA) and *Latent Dirichlet Allocation* (LDA). The use of topic modeling might enable us to attribute topics to articles and find specific information on the kind of accident that happened.

The second approach is more generally called the information extraction approach. There are four important techniques: Named Entity Recognition, Hidden Markov Models, Conditional Random Fields and Relation Extraction. These techniques are all interesting in relation to our problem and give us slightly different information. The most promising approach seems to be Relation Extraction, which takes an in depth look at sentences and the meaning of the relation between words in a sentence.

Lastly, there is also the approach of machine learning. Somehow labelling the data (maybe through some of the previous algorithms) and then classifying a large number of articles.

Chapter 3

Research Method

3.1 Data

3.1.1 Text preprocessing

Tokenization

Filtering

Lemmatization

Stemming

3.1.2 Lexis Nexis

Bibliography

[Aarts et al., 2014] Aarts, L., Eenink, R., Weijermans, W., Knapper, A., and Schagen, I. (2014). Soms moet er iets gebeuren voor er iets gebeurt.

[Allahyari et al., 2017] Allahyari, M., Pouriyeh, S. A., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., and Kochut, K. J. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *CoRR*, abs/1707.0.

[Daniels et al., 2010] Daniels, S., Brijs, T., and Keunen, D. (2010). Official reporting and newspaper coverage of road crashes: A case study. *Safety Science*, 48(10):1469–1476.

[De Ceunynck et al., 2015] De Ceunynck, T., De Smedt, J., Daniels, S., Wouters, R., and Baets, M. (2015). "Crashing the gates" - Selection criteria for television news reporting of traffic crashes. *Accident Analysis and Prevention*, 80:142–152.

[Entman, 2007] Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173.

[Fowler, 1991] Fowler, R. (1991). *Language in the news : discourse and ideology in the press.* Routledge.

[Kemler and den Hertog, 2009] Kemler, H. J. and den Hertog, P. (2009). Ongevallen in Nederland: Overzicht en risicogroepen per ongevalscategorie. *Tijdschrift voor gezondheidswetenschappen*, 87(8):360–365.

[Kühne and Schemer, 2015] Kühne, R. and Schemer, C. (2015). The emotional effects of news frames on information processing and opinion formation. *Communication Research*, 42(3):387–407.

[MacRitchie and Seedat, 2008] MacRitchie, V. and Seedat, M. (2008). Headlines and discourses in newspaper reports on traffic accidents. *South African Journal of Psychology*, 38(2):337–354.

[Mejia, 2017] Mejia, P. (2017). Beyond the traffic report: The news about road safety and vision zero in San Francisco. Technical report, Berkeley Media Studies Group.

[Nenkova and McKeown, 2012] Nenkova, A. and McKeown, K. (2012). A Survey of Text Summarization Techniques. In *Mining Text Data*.

[Scheufele and Tewksbury, 2006] Scheufele, D. A. and Tewksbury, D. (2006). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication*, 57(1):9–20.

[Tros,] Tros, m. d. i. M. E. M. De beschikbaarheid en kwaliteit van informatie over verkeersongevallen.

[Verkade, a] Verkade, T. Door deze fietsprofessor kijk je voor altijd anders tegen het fileprobleem aan.

[Verkade, b] Verkade, T. Waarom meer asfalt het fileprobleem niet oplost (en zelfs erger maakt).

[Verkade, 2018] Verkade, T. (2018). Waarom Nederland het grootste fietsland ter wereld is: daar is keihard voor gestreden.

[Verkade and Brömmelstroet,] Verkade, T. and Brömmelstroet, M. t. Honderden doden, duizenden gewonden, tienduizenden trauma's: verkeersleed is akelig dichtbij.

[Weaver, 2007] Weaver, D. H. (2007). Thoughts on agenda setting, framing, and priming. *Journal of communication*, 57(1):142–147.

[Xie, 2018] Xie, Q. (2018). Critical discourse analysis of News Discourse. *Theory and Practice in Language Studies*, 8(4):399–403.