Studying Muslim Stereotyping through Microportrait Extraction

Antske Fokkens♣, Nel Ruigrok♡, Camiel Beukeboom♢,
Sarah Gagestein♠ and Wouter van Atteveldt♢
♣ Computational Lexicology and Terminology Lab. Vrije Universiteit Amsterdam, the Netherlands.
♠ LJS Nieuwsmonitor. Amsterdam, the Netherlands
♢ Department of Communication Science. Vrije Universiteit Amsterdam, the Netherlands
♡ Taalstrategie. Amsterdam, the Netherlands.
{antske.fokkens,c.j.beukeboom,w.van.atteveldt}@vu.nl,nelruigrok@nieuwsmonitor.org,sarah@taalstrategie.nl

Abstract

Research from communication science has shown that stereotypical ideas are often reflected in language use. Media coverage of different groups in society influences the perception people have about these groups and even increases distrust and polarization among different groups. Investigating the forms of (especially subtle) stereotyping can raise awareness to journalists and help prevent reinforcing oppositions between groups in society. Conducting large-scale, deep investigations to determine whether we are faced with stereotyping is time-consuming and costly. We propose to tackle this challenges through the means of *microportraits*: an impression of a target group or individual conveyed in a single text. We introduce the first system implementation for Dutch and show that microportraits allow social scientists to explore various dimensions of stereotyping. We explore the possibilities provided by microportraits by investigating stereotyping of Muslims in the Dutch media. Our (preliminary) results show that microportraits provide more detailed insights into stereotyping compared to more basic models such as word clouds.

Keywords: stereotyping, digital social science, text analysis

1. Introduction

Research in social sciences and computational linguistics shows that language use plays a crucial role in creating stereotypes about social categories (like Muslims). Several fields within Social Sciences have studied how stereotypic beliefs are reflected in language and consequently become shared knowledge (Ruscher et al., 2005; Klein et al., 2008; Beukeboom, 2014, among others). These studies show that people (unconsciously) express stereotypic beliefs they hold about described individuals or groups in both linguistic content (what information is provided) and form. Computational linguistic studies show that language models learned from corpora reflect such human bias, including stereotypes (Caliskan et al., 2017, e.g.).

Insights from computational linguistics and communication science are complementary. Distributional semantic models applying purely mathematical models to large corpora of text reveal that biases are present in the texts. Yet, these abstract language models do not provide means to reveal how bias is expressed in language exactly. Social science studies fill this gap, for example by studying the *frames* that media use in their coverage (Entman, 1993). These studies, however, only cover limited sets of data as they often rely on experimentally manipulated sentences or manual annotation to study biased language use. Moreover, the need for double-blind annotations on each new set under investigation makes it challenging to study the expression of stereotypes over time and across a high variety of sources. Automatic approaches that can help tackle this problem can therefore greatly enhance the possibility of investigating stereotyping in natural language. Previous work has shown that NLP can be used to identify various forms of framing (Card et al., 2016). These general classifiers perform deeper insight than pure distributional models, but still do not provide the means to look at the more subtle aspects of stereotyping.

This paper introduces a method that allows for stereotype detection in natural language by means of **microportraits**. A microportrait is the collection of descriptions that a single text provides on a given entity or event. We can investigate stereotyping by targeting information about individuals or groups. A microportrait of a person, for instance, combines the labels used to refer to them, the characteristics assigned to them and the roles they played in various events. We provide the first implementation of microportrait extraction for Dutch and use this to investigate stereotyping of Muslims in the Dutch media.

The contributions of this paper are the following:

1. We introduce the notion of microportraits and outline how they can be used to investigate expressions of stereotypes in texts.
2. We provide the first implementation for 'vanilla' microportrait extraction for Dutch. This implementation is open source.
3. We analyze a corpus of microportraits around the Dutch election time and share the first observations from this study

Even though results on the quality of the microportraits themselves are preliminary, the outcome of our study indicates that microportraits can lead to deeper insight into stereotypical descriptions compared to more basic methods such as LDA. We therefore conclude that microportrait extraction is a valuable new task in NLP with large potential in interdisciplinary research.

This paper is structured as follows. Section 2 provides background information on stereotyping in general, recent advances in stereotyping of Muslims and related work in NLP. We introduce microportraits and outline the proposed method for studying stereotyping in Section 3. In Section 4, we describe the pilot study. The initial evaluation and first steps towards a proper evaluation set are described in Section 5. Sections 6 and 7 provide our discussion and conclusion.

2. Background and Related Work

In this section, we outline related work. We first provide an overview of the ways in which stereotypes can be expressed in language in Subsection 21. We then address related work in NLP in Subsection 22. This section ends with an overview of previous findings concerning stereotyping of Muslims in the media.

2.1. Linguistic cues for stereotyping

There is a wide range of literature on stereotyping and biased language use, hence a full overview is beyond the scope of this paper. Instead, we provide an overview of the principle observations and dimensions as outlined by Beukeboom and Burgers (2017).

Following Beukeboom and Burgers (2017)'s Social Categories and Stereotypes Communication (SCSC) framework, we distinguish aspects of stereotyping along two dimensions: the first dimension looks at **labels** that are used to refer to groups or individuals as well as **characteristics** and **behaviors** of these groups and individuals. Second, for both of these aspects it holds that stereotyping can be expressed through the **content** and through the **form** of an expression or, in other words, through the **frames** used.

2.1.1. Bias in labels

The bias in label content is reflected in terminological choices which can be neutral terms (e.g., 'refugees') or terms with a negative connotation (e.g., 'fortune-seekers', 'aliens'). Bias in label form is observed in at least two ways. First, noun labels, compared to adjectival descriptions, are more strongly associated with stereotypic inferences. For instance, when a person is described as 'a Jew', this is more defining and recipients tend to more strongly expect stereotypical Jewish habits. Adjectival references in contrast (i.e., 'Jewish'), are seen as just one characteristic that is less profound and immutable (Carnaghi et al., 2008). Second, labels can be expressed in more generic or more specific form. Here we distinguish generic labels (*Muslims are..*) from subsets (*some Muslims are...*), subtypes (*fundamental Muslims are..*) and individuals (*my Muslim neighbor is..*). Statements that make use of generic labels play an important role in transferring category knowledge (Cimpian and Markman, 2008; Gelman, 1988) and can therefore be particularly influential in spreading stereotypes.

2.1.2. Bias in describing characteristics and behavior

In describing characteristics and behavior, bias is mainly observed in the selection of information. Characteristics and behaviors are more likely mentioned when they fit the existing stereotype (Klein et al., 2008, among others). This tendency has a reinforcing effect. When people are more frequently exposed to stereotype-congruent information, this leads to a continuous confirmation of existing (negative) stereotypic associations, as has for instance been shown for negative news concerning immigrants (Schemer, 2012). At the same time, there can be opposing movements where counter-stereotypical information is provided, for instance because it is socially unacceptable to express prejudiced stereotypic beliefs (Ruscher et al., 2005). However, this does not necessarily disconfirm the stereotype. It depends on the form on which this information is delivered.

Research on linguistic form shows that stereotype incongruent information is often formulated differently than information that is congruent with an existing stereotype (Beukeboom, 2014). Research on the Linguistic Intergroup Bias (Maass et al., 1995; Wigboldus et al., 2000) shows that stereotype-congruent behavior is more likely to be described in terms of enduring personality traits (e.g., *the woman is emotional*), but more behavioral and concrete when it is stereotype-incongruent (e.g., *the man is crying*). Similarly, the Stereotypic Explanatory Bias shows that stereotype-incongruent behaviors are more often explained (Hammer and Ruscher, 1997). And the Negation Bias shows that stereotype-incongruent information is more likely described with a negation that simultaneously introduces a stereotype congruent term (e.g., using 'not oppressed' to describe female Muslims rather than 'independent' or 'free') (Beukeboom et al., 2010). With these biases stereotype incongruent behaviors are framed as unexpected one-time events.

We are not aware of any NLP approach that aimed to specifically study the above described biases, yet there are previous approaches that addressed identifying stereotyping. The following subsection will outline this related work in NLP and explain how microportrait extraction provides a next step.

2.2. Stereotyping in NLP

Using NLP to identify issues of stereotyping is still a relatively unexplored field. Nevertheless, recent studies have addressed various aspects of this topic. We distinguish three related themes: 1) studies that investigate how biases and stereotypes are reflected in language models, among others searching for approaches to avoid that such connotations are reflected by algorithms applying these language models, 2) studies that aim to identify offensive or stereotypical language use and 3) studies that develop software to study how stereotyping is reflected in language.

Most studies fall in the first category. Howard and Borenstein (2017) provide an overview of ways in which our biases are reflected in machine learning algorithms. Notably, Caliskan et al. (2017) show that human biases are reflected in distributional models for both 'morally neutral' concepts such as flowers and insects as well as concepts where this is problematic such as gender and race (Caliskan et al., 2017). The same insights with regards to gender bias had previously been made by Bolukbasi et al. (2016), who also propose a method for 'fixing' this bias.

Binns et al. (2017)'s title refers to the observation of bots taking over bad language. They build a classifier aiming to detect when language is offensive. A similar approach is found in the research of Tulkens et al. (2016) which aims to detect racist discourse. Both studies reveal the challenge involved in creating a gold standard for such studies due to the subjective nature of determining when a statement is crossing the line.

Related work that studies bias in linguistic expressions directly, and is therefore closest related to this research, includes van Miltenburg (2016) and van Miltenburg et al. (2016). Their studies show that image descriptions include cultural biases (e.g., people will describe a child as "black" or "Asian", but white children are simply "a child") and a tendency to use negation to indicate that something is different from what is expected (e.g., a man without shirt). Joseph et al. (2017) investigate to what extent we can observe stereotyping in Twitter, distinguishing between *affective stereotyping* and *semantic stereotyping*, where the former refers to the feeling we associate with a label and the later refers to other associations (such as the activities associated with a specific profession).

The main difference between the studies mentioned above and the approach proposed in this paper is that microportraits aim to reflect the story that is being told of a single entity or group with a single article. By extracting these stories on a large scale, we can then look for patterns and see e.g., how a specific group is portrayed in the media. From this point of view, the study that is probably closest related to this work is Card et al. (2016) who study frame detection through persona description, showing that extracting small stories around persona provides useful information in detecting frames annotated in the Media Frames Corpus (Card, Dallas and Boydstun, Amber E and Gross, Justin H and Resnik, Philip and Smith, Noah A, 2015), (Card et al., 2015). Their approach combines syntactic relations and labels applying a Dirichlet process and using the outcome in a Bayes model for identifying frames. Our study mainly differs from this approach in that it offers the patterns of co-occurring descriptions directly to social scientists rather than offering potentially identified frames. As such, our approach stimulates bottom-up investigation of expressed stereotypes and can be seen as complementary to the work by Card et al. (2016). The outcome provides the possibility of going beyond simple choices between labels and properties, but also allows researchers to investigate how labels, properties and roles relate to each other. We will explain the idea behind the model in more detail in Section 3.1.

2.3. Muslim Stereotyping in the media

The process of 'media logic' influences the way in which the public debate as found in the media is held. With more and more people turning to commercial news outlets and social media for their information (Sonck and de Haan, 2015), the debate is increasingly dominated by entertainment and simplification, focusing on conflict and persons (Welbers et al., 2015). Such news coverage leads to a higher level of (political) cynicism among the general public (Cappella and Jamieson, 1996; Wolfsfeld, 2011). Populist parties use these feelings of unease among the public with one-liners that fit well into the media logic of the media. As a consequence, negative stereotypes and prejudice can take hold of the public which may very well harm relations between communities and provoke societal conflict (Rehman, 2007).

In the United States, discrimination toward Arab Muslims increased after September 11th, 2001 (Sheridan, 2006). Stereotypes, prejudice, and discrimination toward Arab Muslims increased even further in the wake of international terrorism by extremist groups who claimed to have ties to Islam (e.g., ISIS). This is also true for the Netherlands (Adriaansen et al., 2010; Ruigrok et al., 2017b).

Two common components of the Arab Muslim male stereotype are (i) that Arab Muslims are part of the out-group (Saeed, 2007), and (ii) that they are angry, violent, and often terrorists, personified by images of Osama bin Laden (Jackson, 2010). Muslims are presented as terrorists 81% of the time on television (Dixon and Williams, 2015). Many individuals report being afraid of Muslim and/or Arab men, often because they are perceived as violent and a threat to America (Gottschalk and Greenberg, 2008; Sides and Gross, 2013). Muslim women, on the other hand, are stereotyped as oppressed (Stadlbauer, 2012). Specifically, many inhabitants of non-Muslim countries believe that the veil of Muslim women (e.g., hijab, niqab, and burqa) is a representation of oppression (Wagner et al., 2012).

3. Introducing Microportrait Extraction

The goal of this section is to introduce microportrait extraction as an NLP task. After defining what microportaits are, we explain how they can be used to study stereotyping in Section 31). We then describe the first implementation of microprotrait extraction (Section 32).

3.1. Microportraits and Stereotypes

Microportraits are designed to study framing and stereotyping. The idea is that by exploring how specific people are described on a large scale, researchers can identify common patterns in descriptions of people who share certain characteristics. For the use case in this paper, for instance, we explored how people that are explicitly labeled as "Dutch" or "Muslim" are described in Dutch media. The basic units of a microportraits are **descriptions**. A description can be a **label** assigned to an entity, a **property** assigned to them or a **role** they play in a specific event. For instance, the expression *the pious Muslim smiled* contains the following three descriptions: the label *Muslim*, the property *pious* and the agent or arg0 role in *smiling*. The **microportrait** of a person is the collection of all descriptions of this person within a single article. Table 1 illustrates the microportraits for the referents of *Muslim* and *John* in the snippet *the pious Muslim smiled when John waved at him*. All descriptions related to the same referent share the same identifier (*docIdt2* for the Muslim and *docId3* for John in Table 1).

When applied to large volumes of text, researchers can use microportraits to identify which labels, properties and activities tend to co-occur and what choices writers make when describing a person. For instance, do they choose to indicate that someone belongs to a religious group by using a property (*a Muslim man*) or a label (*a Muslim*)? What other properties and labels are used when talking to individuals from this group? Do certain sources talk in terms of "us" and "them"? When do writers feel inclined to make a specific origin, religious background, hobby or some other feature (such as looks or achievements) explicit?

identifier	relation	label
docIdt2	label	*Muslim*
docIdt2	property	*pious*
docIdt2	arg0	*smile*
docIdt2	arg2	*wave*
docIdt2	label	*him*
docIdt3	label	*John*
docIdt3	arg0	*wave*

Table 1: Illustration of microportrait

More subtle forms of stereotyping can be identified by investigating how descriptions of groups differ when talking about a specific theme. Do people from certain backgrounds easily receive labels such as *thief*, *criminal* and *perpetrator* (asserting involvement and making it part of their being) whereas others are *suspect* of, e.g. *stealing* (leaving the option of innocence open and highlighting (possibly incidental) behavior)? This can be investigated by searching for descriptions related to a specific topic and investigating how various groups are described in relation to them. A mixed approach would start by a bottom-up approach to identify typical labels, properties and roles and then using the outcome in a top-down study for examining details in choices of representation.

Microportraits provide profiles of people, groups and other entities within a document. Without prior knowledge of the stereotypical traits associated with a group, a microportrait in isolation cannot provide insight in stereotyping. It can also not be used to identify the emergence of new biases that may influence the stereotypical views in society. By investigating patterns in microportraits extracted from a large corpus, we can however identify biases in the kind of information provided when specific groups are described (stereotypical content) and whether there are differences in how this group is presented (stereotype reflection in form). Microportraits can be used to investigate several dimensions of stereotypes depending of whether one takes a bottom-up approach, a top-down approach or a mixture of the two.

The bottom-up approach is the most straight-forward way of studying stereotyping through microportraits. It can be applied by simply calculating the Pointwise-Mutual Information score of descriptions that co-occur in the same microportrait. We can use the outcome of these calculations to see, for instance, which descriptions are typically used for Muslims and how they compare to descriptions typically used for other groups.

3.2. Microportrait extraction

We have implemented a first version for extracting microportraits from Dutch news and applied this to study identify how Dutch media talk about Muslims and immigrants from Muslim countries compared to how they talk about Dutch people. This subsection describes the current implementation of our system for Dutch microportrait extraction, which combines syntactic patterns with entity coreference resolution. Because it does not provide the deeper semantic interpretation provided by semantic roles yet, we refer to the microportraits coming out of the current implementation as *vanilla microportraits*.

We use a small pipeline that forms a subpart of the pipeline for event extraction developed as part of the larger NewsReader (Vossen et al., 2016) and BiographyNet (Fokkens et al., 2017) pipelines for event extraction. The pipline includes the Alpino parser for dependency parsing (Bouma et al., 2001) and the ixa-pipe named entity recognizer (Agerri and Rigau, 2016). Both modules are run using a wrapper that provides output in the NLP Annotation Format (Fokkens et al., 2014a, NAF).[1] We implemented a new system for Dutch coreference resolution, based on the Stanford multisieve-entity coreference resolution approach (Lee et al., 2013) that applies coreference resolution on top of this small pipeline.[2]

In the first step, we extract descriptions at a sentence level. We start by taking nouns as labels and then identify their properties by extracting their modifiers and attributes through copula constructions using the dependency structure. We also use syntactic dependencies to identify the 'roles' an entity plays. Alpino outputs 'deep' dependencies that indicate that the subject of a passive sentence has an object relation with the main verb. We take the simplified assumption that (deep) subjects are agents, (deep) objects are patients and (deep) indirect objects recipients. Prepositional complements receive the label *prep-role*, where *prep* corresponds to the surface form of the preposition. In the ideal form, the roles in microportraits would consist of semantic roles, but the only open source semantic role labeler for Dutch we are aware of yielded worse results than our current patterns on a development set.

The outcome of this step yields so-called 'nano-portraits', which are descriptions related to an entity within a clause. In the second step, we combine the nano-portraits related to the same entity using the output of the coreference resolution module. This results in collections of descriptions related to the same referent within a document: the so-called 'vanilla microportrait'. The next section will explain how we used automatically extracted vanilla microportraits in order to investigate whether Muslims are stereotyped in a discriminatory manner in Dutch media.

4. Muslims in Dutch News

In this section, we describe the outcome of a pilot study using microportrait extraction for identifying stereotyping of Muslims. The outcome of this study was used to shed light on how Dutch media talked about Muslims in political news during election time. From a computational linguistics point of view, this study serves the purpose of exploring what microportraits have to offer for this type of investigation. In particular, we explore whether they allow communication scientists to go beyond methods that are commonly used, such as word clouds based on Tf-idf scores.

[1] This Alpino wrapper is available at: `https://github.com/cltl/morphosyntactic_parser_nl` and the IXA-named-entity recognizer at `https://github.com/ixa-ehu/ixa-pipe-nerc`.

[2] The implementation is available under the Apache license under: `https://github.com/antske/coref_draft`

	Dutch	Muslims
labels and properties	famous, average, Dutch origin, fast, beautiful, free	radical, moderate, conservative, Sunni extremist, pious
roles (agent)	take, miss, win, break, drive, make, score	insult, convert, adhere, rape, murder, extinct

Table 2: English translation of most typical descriptions

4.1. Method

We collected news articles from national daily newspapers, online newsites and news blogs from the period from September 5th up until March 15th 2017. This resulted in a total of 622,480 articles. The use case focuses on articles that address politics in election time. We therefore extracted articles about politics that appeared in the period from January first to March 15th 2017, covering the election day for the Dutch parliament and 2.5 months leading up to it, from this set. We consider an article to be about politics when it explicitly mentions one of the political parties or one of its prominent members or a member of the parliament or government. This resulted in a selection of 15,573 articles.

In this set of political articles, we compare which words, labels, properties or roles occur typically with people labeled as *Dutch* and which are typical for people labeled as *Muslim*. We investigate both the full set of news articles as well as the subset mentioning politicians or political parties.[3]

We use two approaches: first, we use Latent Dirichlet Analysis (LDA) in order to identify which words typically occur in articles referring to Dutch people, Muslims or both. In the next step, we extracted microportraits from these texts in order to identify what was said about Muslims and Dutch people, respectively, on a more fine-grained level. We investigated the most typical labels and properties as well as the most typical roles.

4.2. Analysis

The outcome of the exploratory study using LDA reveals that articles that only refer to Dutch people are mainly about (winning) sports, the most typical words in news articles about Muslims are *aanslag* "assault", *president* (idem) and *amerikaanse* "American", due to the discussion about Trump's proposal to refuse Muslims from certain countries entrance to the United States. Articles that mentioned both "Muslims" and "Dutch" seem mainly to be directly related to the elections, society and integration (with typical words such as *stemmen* "vote", *integratie* "integration" and *democratie* "democracy").

The results obtained through microportrait extration are presented in Table 2. The table provides English translations of the labels, properties and roles most typically mentioned when Dutch articles explicitly talk about Dutch people or Muslims.

The terms in Table 2 indicate a shocking difference in the way media portray Muslims. However, this list is still anecdotal. In order to gain further insight into whether Dutch people are indeed generally portrayed positively whereas Muslims are portrayed negatively, we let four student assistants annotate descriptions indicating whether they provided a positive or negative picture of the person being described. The most frequent descriptions from our corpus were presented out of context (so that annotators did not know whether they applied to Muslims, Dutch people or neither) and annotated independently by at least two annotators each. We then assigned negative scores (-1) to descriptions annotated as negative and positive scores (+1) to descriptions labeled as positive in our microportraits. The results are presented in Table 3. For all three categories, results indicate that people labeled as 'Muslim' are described more negatively than people called 'Dutch'. Where the former are on average portrayed (slightly) negatively, descriptions associated with Dutch people are neutral to positive. Results are significant for all three categories.

	Dutch	Muslim
label	0.0000	-0.020
property	0.050	-0.073
roles	0.008	-0.140

Table 3: Positive/Negative reporting

5. Initial Validation and Annotations

In this section we present our validation steps and the first steps towards creating a gold standard for evaluation.

5.1. Validation

We performed an initial validation of the method by checking the precision 1,058 descriptions from randomly selected articles. Manual inspection revealed that 98.1% of the descriptions were correctly extracted from the text. Furthermore, 87.2% of the descriptions was placed into the correct microportrait. These results seems suspiciously high, but this is mainly due to the the far majority of descriptions being expressed by basic substructures in a sentence that the Alpino parser analyzes correctly despite possibly making errors in more complex parts of the sentence. The high result of the validation of the placement in the correct microportrait is high, because assignments are partially due to local structures (such as the adjective modifying a noun or a predicative structure).

Most mistakes we found in detecting descriptions occurred in long sentences. There is no reason to assume that journalists use longer sentences when talking about Dutch people or Muslims. We can therefore expect errors to be equally distributed over the two classes and there are currently no indications of our approach leading to a bias in the overall outcome.

The validation looks promising, but a more detailed and solid evaluation of the method is necessary. We are currently in the process of creating gold standard data. The current status of this gold standard is described in the coming subsection.

5.2. Annotation Instructions

In order to provide a better evaluation, a gold standard evaluation set is currently under development. As is common

[3]This research is also reported (in Dutch) in Ruigrok et al. (2017a). More details can be found at `https://www.microportretten.nl`.

for new complex annotation tasks, annotations are carried out in multiple rounds leading to updates in the annotation guidelines. Annotators used the following procedure:

- mark all labels used to refer to an entity
- mark all modifiers of each label as 'property'
- connect each property to the appropriate label
- mark all activities and connect them to the appropriate label, indicating whether the label plays the role of 'agent','patient' or some other role
- connect all labels referring to the same entity to the same external identifier

During the first annotation cycle, annotators were instructed to annotate all microportraits occurring in the text. This resulted in low inter agreement scores. The main reason for these low scores were that annotators found it difficult to annotate the full text. Even after an additional revision round, annotations were incomplete and which annotations were missing differed from one annotator to another. We therefore launched a second cycle in which annotators only annotated the full microportrait if it contained a label or property from a predefined set (including the Dutch words for Muslim, Christian, Jewish, Belgian, German, Moroccan, Turkish, Dutch and derived terms). The annotations following these new guidelines are currently ongoing.

6. Discussion

The outcome of our use case reveals that microportraits can provide a different perspective than more basic approaches such as LDA. In particular, our method detected severe forms of stereotyping that remained unnoticed when looking a word co-occurrence alone. It must be taken into consideration, however, that the final evaluation of the technology is ongoing. Solid evaluation in interdisciplinary studies involves intrinsic evaluation (performance of the coreference resolution tool and microportrait extraction) and extrinsic evaluation (do the mistakes made by the tools introduce a bias that influences the research questions) (Fokkens et al., 2014b). At this point, more elaborate evaluation of the accuracy of our tools is necessary.
The high precision scores in our validation are encouraging. As mentioned in Section 51, we did not find any indication of a bias or errors that would systematically miss (positive) descriptions of Muslims or (negative) descriptions of Dutchmen. It is therefore unlikely that the outcome of our pilot study is the result of a bias in the tool. An additional indication that the results are likely to be accurate is that the topics covered in the news during the investigated period provide a plausible explanation. The observation that the label 'Dutch' is mainly used when stressing sport achievements is in line with the outcome of the LDA investigation. The corpus contained several articles addressing the events of New Year's Eve in Cologne as well as crimes committed by ISIS which are both topics likely to contain the negative stereotypical descriptions we identified for Muslims.
Nevertheless, the indications outlined above cannot replace a proper evaluation: when results are checked rather than evaluated on an independently created datasets, borderline cases will often be considered 'correct' leading to higher results. Moreover, these verifications do not give insight into the recall of the system. The correspondence between microportraits and covered topics supports our outcome, but does not provide a guarantee. The completion of a gold standard dataset is therefore essential for continuing this line of research. This is ongoing work and we plan to report the outcome of this evaluation in future work.

7. Conclusion

This paper introduced microportraits: the collection of all descriptions of a single entity (a person, group, object or event) found in a single document. We summarized insights from communication science on how stereotypes can be reflected in language use and showed how microportraits can be used to study stereotyping in text.
We implemented a basic pipeline for microportrait extraction for Dutch.[4] The pipeline extracts so-called 'vanilla microportraits' which consist of descriptions based on syntactic patterns. All tools implemented for this research are freely available under the Apache license.
We applied this to investigate stereotyping of Muslims in the Dutch media comparing how people or groups labeled explicitly as 'Dutch' or 'Muslim' are described. Whereas a basic bottom-up study using LDA mainly indicates themes that are discussed (sports for Dutch, terrorism and political crisis for Muslims), the microportraits provide a more detailed and specific insight into how groups are portrayed.
Evaluation of the tool and methodology are currently ongoing. We carried out validation checks that indicate solid performance and the patterns that emerge from the microportraits can be explained by observations from the data. Though this is promising, validations checks are merely indicative and solid evaluation is needed. The creation of a gold-standard evaluation set is ongoing. Nevertheless, we believe that the preliminary outcome of our use case clearly illustrates the potential use of microportraits. As such, the main contribution of this paper is the introduction of this new NLP task that is of high interest for researchers in the social sciences wanting to investigate stereotyping.

8. Acknowledgements

The work presented in this paper was funded by the Netherlands Organization for Scientific Research (NWO) via VENI grant 275-89-029 awarded to Antske Fokkens. The pilot study was funded by the Stichting Democratie and Media (Democracy & Media Foundation).

9. Bibliographical References

Adriaansen, M., van Praag, P., et al. (2010). Nieuwe scheidslijnen en de turbulente relatie tussen politiek, media en burgers.

Agerri, R. and Rigau, G. (2016). Robust multilingual named entity recognition with shallow semi-supervised features. *Artificial Intelligence*, 238:63–82.

[4] https://github.com/cltl/micro-portraits

Beukeboom, C. and Burgers, C. (2017). How stereotypes become shared knowledge: Biased language use in communication about categorized individuals.

Beukeboom, C. J., Finkenauer, C., and Wigboldus, D. H. (2010). The negation bias: when negations signal stereotypic expectancies. *Journal of personality and social psychology*, 99(6):978.

Beukeboom, C. J. (2014). Mechanisms of linguistic bias: How words reflect and maintain stereotypic expectancies. *Social cognition and communication*, 31:313–330.

Binns, R., Veale, M., Van Kleek, M., and Shadbolt, N. (2017). Like trainer, like bot? inheritance of bias in algorithmic content moderation. In *International Conference on Social Informatics*, pages 405–415. Springer.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems*, pages 4349–4357.

Bouma, G., Van Noord, G., and Malouf, R. (2001). Alpino: Wide-coverage computational analysis of dutch. *Language and Computers*, 37(1):45–59.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Cappella, J. N. and Jamieson, K. H. (1996). News frames, political cynicism, and media cynicism. *The Annals of the American Academy of Political and Social Science*, 546(1):71–84.

Card, D., Boydstun, A. E., Gross, J. H., Resnik, P., and 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)*, volume 2, pages 438–444.

Card, D., Gross, J., Boydstun, A., and Smith, N. A. (2016). Analyzing framing through the casts of characters in the news. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1410–1420, Austin, Texas. Association for Computational Linguistics.

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., and Arcuri, L. (2008). Nomina sunt omina: on the inductive potential of nouns and adjectives in person perception. *Journal of personality and social psychology*, 94(5):839.

Cimpian, A. and Markman, E. M. (2008). Preschool children's use of cues to generic meaning. *Cognition*, 107(1):19–53.

Dixon, T. L. and Williams, C. L. (2015). The changing misrepresentation of race and crime on network and cable news. *Journal of Communication*, 65(1):24–39.

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

Fokkens, A., Soroa, A., Beloki, Z., Ockeloen, N., Rigau, G., van Hage, W. R., and Vossen, P. (2014a). Naf and gaf: Linking linguistic annotations. In *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, pages 9–16.

Fokkens, A., Ter Braake, S., Ockeloen, N., Vossen, P., Legêne, S., and Schreiber, G. (2014b). Biographynet: Methodological issues when nlp supports historical research. In *LREC*, pages 3728–3735.

Fokkens, A., ter Braake, S., Ockeloen, N., Vossen, P., Legêne, S., Schreiber, G., and de Boer, V. (2017). Biographynet: Extracting relations between people and events. In *Europa baut auf Biographien*.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive psychology*, 20(1):65–95.

Gottschalk, P. and Greenberg, G. (2008). *Islamophobia: making Muslims the enemy*. Rowman & Littlefield.

Hammer, E. D. and Ruscher, J. B. (1997). Conversing dyads explain the unexpected: Narrative and situational explanations for unexpected outcomes. *British journal of social psychology*, 36(3):347–359.

Howard, A. and Borenstein, J. (2017). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics*, pages 1–16.

Jackson, L. (2010). Images of islam in us media and their educational implications. *Educational Studies: Journal of the American Educational Studies Association*, 46(1):3–24.

Joseph, K., Wei, W., and Carley, K. M. (2017). Girls rule, boys drool: Extracting semantic and affective stereotypes from twitter. In *CSCW*, pages 1362–1374.

Klein, O., Tindale, S., and Brauer, M. (2008). The consensualization of stereotypes in small groups. *Stereotype dynamics: Language-based approaches to the formation, maintenance, and transformation of stereotypes*, pages 263–292.

Lee, H., Chang, A., Peirsman, Y., Chambers, N., Surdeanu, M., and Jurafsky, D. (2013). Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, 39(4):885–916.

Maass, A., Milesi, A., Zabbini, S., and Stahlberg, D. (1995). Linguistic intergroup bias: differential expectancies or in-group protection? *Journal of Personality and Social Psychology*, 1(68):116–126.

Rehman, J. (2007). 9/11 and the war on terrorism: The clash of 'words', 'cultures' and 'civilisations': Myth or reality. In M.N. Craith, editor, *Language, Power, and Identity Politics*, pages 198–215, New York. Palgrave Macmillan.

Ruigrok, N., Fokkens, A., Gagenstein, S., and van Atteveldt, W. (2017a). Stereotyperende microportretten van moslims in het (politieke) nieuws. Technical report.

Ruigrok, N., van Atteveldt, W., Gagestein, S., and Jacobi, C. (2017b). Media and juvenile delinquency: A study into the relationship between journalists, politics, and public. *Journalism*, page 1464884916636143.

Ruscher, J. B., Cralley, E. L., and O'Farrell, K. J. (2005). How newly acquainted dyads develop shared stereotypic impressions through conversation. *Group Processes & Intergroup Relations*, 8(3):259–270.

Saeed, A. (2007). Media, racism and islamophobia: The representation of islam and muslims in the media. *Sociology Compass*, 1(2):443–462.

Schemer, C. (2012). The influence of news media on stereotypic attitudes toward immigrants in a political campaign. *Journal of Communication*, 62(5):739–757.

Sheridan, L. P. (2006). Islamophobia pre–and post–september 11th, 2001. *Journal of interpersonal violence*, 21(3):317–336.

Sides, J. and Gross, K. (2013). Stereotypes of muslims and support for the war on terror. *The Journal of Politics*, 75(3):583–598.

Sonck, N. and de Haan, J. (2015). Media: Tijd in beeld.

Stadlbauer, S. (2012). A journey to a "pure islam": Time, space, and the resignification of ritual in post 9/11 faith testimonies of muslim women. *Narrative Inquiry*, 22(2):348–365.

Tulkens, S., Hilte, L., Lodewyckx, E., Verhoeven, B., and Daelemans, W. (2016). The automated detection of racist discourse in dutch social media. *Computational Linguistics in the Netherlands Journal*, 6(1):3–20.

van Miltenburg, E., Morante, R., and Elliott, D. (2016). Pragmatic factors in image description: The case of negations. In *Proceedings of the 5th Workshop on Vision and Language*, pages 54–59. ACL.

van Miltenburg, E. (2016). Stereotyping and bias in the flickr30k dataset. In Jens Edlund, et al., editors, *Proceedings of Multimodal Corpora: Computer vision and language processing (MMC 2016)*, pages 1–4.

Vossen, P., Agerri, R., Aldabe, I., Cybulska, A., van Erp, M., Fokkens, A., Laparra, E., Minard, A.-L., Aprosio, A. P., Rigau, G., et al. (2016). Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Knowledge-Based Systems*, 110:60–85.

Wagner, W., Sen, R., Permanadeli, R., and Howarth, C. S. (2012). The veil and muslim women's identity: Cultural pressures and resistance to stereotyping. *Culture & Psychology*, 18(4):521–541.

Welbers, K., van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., and Schaper, J. (2015). News selection criteria in the digital age: Professional norms versus online audience metrics. *Journalism*, 17(8):1037–1053.

Wigboldus, D. H., Semin, G. R., and Spears, R. (2000). How do we communicate stereotypes? linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, 1(78):5–18.

Wolfsfeld, G. (2011). *Making sense of media and politics: Five principles in political communication*. Taylor and Francis.

10. Language Resource References

Card, Dallas and Boydstun, Amber E and Gross, Justin H and Resnik, Philip and Smith, Noah A. (2015). *The media frames corpus*. School of Computer Science, Carnegie Mellon University, 2.0.