



## UvA-DARE (Digital Academic Repository)

### Nuclear voices in the news

*A comparison of source, news agency and newspaper content about nuclear energy over time*

Boumans, J.W.; Vliegthart, R.; Boomgaarden, H.G.

**DOI**

[10.1177/0267323116629879](https://doi.org/10.1177/0267323116629879)

**Publication date**

2016

**Document Version**

Final published version

**Published in**

European Journal of Communication

[Link to publication](#)

**Citation for published version (APA):**

Boumans, J. W., Vliegthart, R., & Boomgaarden, H. G. (2016). Nuclear voices in the news: A comparison of source, news agency and newspaper content about nuclear energy over time. *European Journal of Communication*, 31(3), 260-282.  
<https://doi.org/10.1177/0267323116629879>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

# Nuclear voices in the news: A comparison of source, news agency and newspaper content about nuclear energy over time

European Journal of Communication  
2016, Vol. 31(3) 260–282  
© The Author(s) 2016  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0267323116629879  
ejc.sagepub.com



**Jelle W Boumans and Rens Vliegthart**

University of Amsterdam, The Netherlands

**Hajo G Boomgaarden**

University of Vienna, Austria

## Abstract

While news media are frequently criticized for their alleged increasing reliance on ‘subsidized content’ provided by sources and news agencies, this claim is seldom empirically verified. Based on insights from computer science, this study proposes an approach to quantitatively compare source, news agency and newspaper content over time. Including press releases from two corporate actors and one nongovernmental actor as well as articles of news agencies and newspapers, the approach is applied to the debate about nuclear energy in the Netherlands (2003–2012). Results show no indication of an increased similarity of newspapers’ content with either source content or news agency content, thus providing no justification for the concerns about an increasing dependency of newspapers on subsidized content. Contradicting literature, we found that media content is most similar to the nongovernmental organization’s content, with the exception of one regional newspaper that strongly reflects the local corporation’s content.

## Keywords

Content analysis, journalism, local/regional media, news agencies, sources

The current economic hardship of news media allegedly affects the quality of their performance. Decreasing newsroom capacity, faster news cycles, high levels of competition, declining and unstable readership and falling advertising revenues are just some of the

---

## Corresponding author:

Jelle W Boumans, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands.

Email: [j.w.boumans@uva.nl](mailto:j.w.boumans@uva.nl)

challenges that newspapers face these days (Lewis et al., 2008; Franklin, 2012). Investors and the capital market in general are placing newspapers under further pressure (McChesney, 2008). Concerns with minimizing costs and enhancing profits have led to cuts in investigative journalism and to a greater reliance on official sources, the public relations (PR) industry and news agencies in the production of news (Davies, 2008; McChesney, 2012). Parallel with this weakening of traditional journalistic practices, critical observers have witnessed an impressive growth of the PR industry and diverse forms of 'spin' and news management forms, which deliver this 'subsidized content' to the media (Davis, 2000; Prenger et al., 2011). These dynamics arguably have resulted in a shift of the power balance between sources and journalists in favour of the former (Davis, 2000; Franklin, 2011; Jackson & Moloney, 2015). This article compares organizational press releases with news media content over a period of 10 years to assess whether the growing dependency of journalists on sources has led to an increasing similarity between news and source contents.

Sources are an indispensable part of the news production process and studying these dynamics provides valuable insights into the performance of the news media. News coverage has aptly been described as '*a sampling of sources' portrayals of reality, mediated by news organizations*' (Sigal, 1986: 27–28). The key concern nowadays is that news coverage involves increasingly less mediation and more sampling, even to an extent that source information is literally copied (Lewis et al., 2008). This is believed to compromise media's societal duty to provide an independent flow of information and interpretation. While many journalism studies have shed light on the practices of journalists and the processes involved in the production of news, the origins of news contents – sources and their PR activities – have received less attention (Manning, 2001). Although recent research has demonstrated the pervasive role of the PR industry in the news media (Jackson and Moloney, 2015; Lewis et al., 2008; Reich, 2010; Sissons, 2012), longitudinal empirical research into the relationship between source content and news content remains scarce. A basic assumption in this study is that if indeed sources are becoming more influential over time, news content will increasingly mirror the content of press releases.

Apart from sources, news agencies are another fundamental category in the infrastructure of news (Lewis et al., 2008). Wire content traditionally steers a large part of the daily press and research indicates that its impact is growing (Scholten and Ruigrok, 2009; Rutten & Slot, 2011). At the same time, agencies are increasingly struggling to provide quality and remain profitable at the same time (Boyer, 2011; Rutten & Slot, 2011). There are a number of reasons the role of agencies deserves critical attention. First, agency material is generally taken for granted by journalists: fact-checking of news agency copy is not common practice as journalists assume that the stages of fact-checking, background research and verifications have already been completed (Davies, 2008; Johnston and Forde, 2011). Yet, it has been shown that agencies frequently distribute PR releases with little or no checking or verification of content involved (Johnston and Forde, 2011; Scholten and Ruigrok, 2009). The unchecked reproduction of agency content increases the chances of media disseminating false or distorted information. Second, the fact that newspapers to a large extent rely on agency copy for their coverage of an event is likely to limit the diversity of voices and perspectives in the media landscape.

Although agency content in itself may very well be diverse, without doubt the overall diversity would be greater in a situation where each newspaper produces its own coverage of an event.

Despite their central role in the infrastructure of the news, news agencies have enjoyed remarkably little academic interest. Aside the cited studies above, empirical research is mostly confined to the international agencies (Boyd-Barrett and Rantanen, 2000) and longitudinal research is absent. The absence of empirical data about potential trends makes the often heard claims on increased source dependency hard to maintain. Thus, research on sourcing practices from agencies as well as journalists is in need of a longitudinal perspective.

We rely on insights from the fields of artificial intelligence and information retrieval to empirically address the concerns on media's alleged dependency on subsidized content. The central research questions are (1) *to what degree does the content of the output of sources, agencies and newspapers overlap?* and (2) *how does the degree of content similarity evolve over time?* The Dutch debate on nuclear energy between 2003 and 2012 will function as a case study to address these questions. The dataset includes the contents of three domains: (1) press releases from two business sources and one nongovernmental organization (NGO), (2) articles from the two leading Dutch news agencies and (3) articles from five national and two regional newspapers. The regional newspapers are included because concerns are expressed that their economic hardship would make them particularly reliant on subsidized content (Franklin, 2008; O'Neill and O'Connor, 2008). Considering that regional newspapers are an important factor in the media landscape (Bakker and Scholten, 2011), a better understanding of their relation with sources is vital.

### **Sources: Journalist's grateful partners**

For more than four decades now, studies of the sociology of news production have shed light on the interaction between sources, journalists and news media organizations. Based on observations of journalistic practices, Tuchman (1973) described the task of news organizations as 'routinizing the unexpected' by developing processes that ensure a steady amount of news in a timely fashion. One such routine is the systematic reliance on sources (Cottle, 2003), whose press releases are designed to support the journalist's 'linear, routine writing process' (Van Hout and Jacobs, 2010).

With controversial societal or political issues like nuclear energy, involved parties strive for media coverage that reflects their – often conflicting – perspectives. Through coverage in line with their positions, these stakeholders aim to indirectly influence public opinion (Curtin and Rhodenbaugh, 2001). Such strategic communication directed at generating media attention is coined agenda building (Curtin, 1999) and has attracted wide scholarly attention. Previous agenda building studies have found that not all sources are equally successful in influencing the media agenda. Typically, business and institutional sources tend to be attractive sources for journalists because they are well accessible, well-resourced and provide a regular supply of 'information subsidies' (Davis and Cottle, 2003). The perceived legitimacy of a source is also determined by the 'cultural capital' (differing competences, skills and assets) that sources possess (Anderson, 1997: 9). Being regarded newsworthy in their own right, the dominance of institutional sources

has been widely demonstrated (Hall et al., 1978; Manning, 2001; Rafter, 2014; Shehata, 2010). Similarly, a study that differentiated between different source types showed that corporate PR material is nearly four times as likely to be reflected in general news stories as the press releases of NGOs or pressure groups (Lewis et al., 2008). It must be noted that this study concerned general domestic news. When it comes to environmental issues, the media agenda has been found to be highly dependent on particularly government agencies and NGOs, with corporate sources being less successful in promoting their 'green agendas' (Curtin and Rhodenbaugh, 2001). The specific case of nuclear energy is an environmental as much as an economic and a safety issue. While it is not the article's core interest to assess which source type is the primary definer, the multifaceted nature of nuclear energy makes it an interesting issue to see which type of source is most successful in promoting its agenda.

Empirical agenda building studies usually concentrate on two major independent variables affecting media agendas: real-world conditions and events and the activities of actors. Focusing on the degree of content overlap between actor's information provision and the media, our study falls under the latter category. The importance of these press releases for journalists has been confirmed in various news contexts, including health issues, the environment, and national and foreign policy (Denham, 2010). While the relationship between information subsidies and news media has thus generated wide scholarly attention, longitudinal analyses of this relationship are scarce, if at all existent. By comparing press releases from two industrial sources and one NGO with news output over a period of 10 years, this article sheds light on media source dynamics over time.

When assessing newspapers' reliance on sources and news agencies, a categorization of different types of newspapers is useful. This study differentiates between national and regional newspapers, as previous investigations have shown significant differences between national and local/regional coverage (Hansen, 2011). Regional and local newspapers traditionally fulfil a unique function in democracy, namely to hold local or regional powers to account (Freer, 2007; O'Neill and O'Connor, 2008). The current economic situation could make it harder for these newspapers to perform this task. Operating on narrow markets, the steady decrease in circulation rates and advertising revenues weighs heavy on regional newspapers (Kerrigan and Graham, 2010). The Dutch regional newspaper sector is in decline; the number of journalists working for the Dutch news corporation Wegener, the largest publisher of regional newspapers, has decreased from 2000 in 2001 to 900 in 2012 (Dohmen, 2012). This trend is visible in other countries as well: in the United Kingdom, for instance, the number of regional titles has halved over the last 50 years (Franklin, 2008). Local newsrooms are characterized as 'a pressurized and demoralized working environment' where it is 'all too easy for journalists to become dependent on the pre-fabricated, pre-packaged "news" from resource-rich PR organizations' (O'Neill and O'Connor, 2008). We hypothesize this reliance to be strong in particular with regard to a local (in our case nuclear) organization, because previous research has demonstrated a strong reliance of regional newspapers on local sources (Prenger et al., 2011).

*H1.* Regional newspapers show greater similarity with the content of news agencies and regional sources than national newspapers.

## The central role of news agencies

News agencies are a dominant factor in the world of the news (Boyd-Barrett and Rantanen, 2000; Forde and Johnston, 2013). Traditionally, they occupy a unique position – as no other party, they monitor newsworthy events all over the world and generate news from a wide variety of sources. To illustrate, a journalist at the German department of news agency Associated Press (AP) receives on an average day about 4000 to 5000 press releases that strive for agency coverage (Boyer, 2011). The perceived prominence of a source is a key filtering principle; the more prominent the politician, public figure, organization or expert is, the more likely their statements would be turned into agency releases (Boyer, 2011). This tendency to rely on pre-packaged information provided by credible sources has increased along with the numerous challenges that agencies face (Boyer, 2011; Carlson, 2009; Forde and Johnston, 2013). First, technological developments have made news omnipresent and online accessible in a split second. The agencies have thus lost their traditional monopoly on ‘raw’ news and are now faced with multiple competitors that have entered their market. Second, in the digital age, it is increasingly difficult to secure exclusivity of the content; as soon as a medium publishes content online, other media profit for free by recycling the copy (Phillips, 2010). Third, news organizations have put the screws on the agencies. This is particularly the case in the Netherlands, where the exceptionally high level of concentration of the newspaper market places the agencies in a disadvantageous position at the negotiation table (Rutten & Slot, 2011: 4). The agencies seek to remain profitable by cutting substantively in personnel and the pioneering of new markets (Manning, 2008); The Dutch market leader, *Algemeen Nederlands Persbureau* (ANP), for instance, offer their expertise to organizations in writing successful press releases (Rutten & Slot, 2011).

While the number of employees has been steadily reduced over the past years, the level of output that ANP produces has increased by 40% over the period 2004–2009. These developments have led the president of ANP to express his concern that the agency has reached a ‘critical borderline of providing required quality’ (Rutten & Slot, 2011: 25). The precarious situation of the agencies has not withheld news media to strongly rely on news agencies for their content (Hijmans et al., 2011). On the basis of a content analysis involving nearly 60,000 articles from 9 national newspapers between 2006 and 2008, Otto Scholten and Nel Ruigrok (2009) found an across the board trend of increased reliance on agency copy for news and background information. Within 3 years, the percentage of newspaper articles that were (partly) based on agency content increased from 23.9% to 27.6%. The study furthermore demonstrated that in the largest national newspapers, 20–30% of the articles containing agency content did not attribute the agency as a source.

The sections above motivate our expectation that newspapers increasingly rely on news agency content and that both agencies as well as newspapers increasingly rely on source content. This translates into the following hypothesis:

*H2.* The degree of content similarity between the three domains has increased over time.

Furthermore, if the reliance on pre-fabricated news is indeed related to the economic decline of the newspaper sector, we may expect this trend to be strongest for regional newspapers, which are, as argued above, particularly troubled by the economic developments:

*H3.* The degree of content similarity with sources and agencies over time has increased more strongly for regional newspapers than for national newspapers.

## The case of nuclear energy

The enduring issue of nuclear energy is a suitable topic to assess the overlap of content from different domains, since it has generated public debate for decennia. The extraordinary complexity of the issue, in technical as well as psychological terms (Van Dam, 2003), ensures a wide-ranging spectrum of viewpoints. A further advantage is that with both corporate as well as nongovernmental stakeholders, different source categories are involved. The environmental organization Greenpeace has been a declared opponent of the nuclear industry and actively seeks media attention for its viewpoints. The corporate sources are represented by the two largest nuclear organizations that are active in the Netherlands: global nuclear service provider *Nuclear Research and consultancy Group* (NRG), whose core business is the production of nuclear isotopes for the pharmaceutical market, and *Elektriciteits Produktiemaatschappij Zuid* (EPZ), an energy producer that exploits a nuclear power plant. In light of the literature, the inclusion of an ‘official’ governmental source would have had our preference. Yet due to the government’s highly fragmented communication on the issue, no single governmental source published enough textual material to allow for a reliable comparison.

## Method

The central aim of the article is to compare the content of press releases from sources with the contents of news agencies and newspapers by means of an automated technique. This section describes the technique, which has shown to produce results that are consistent with a traditional, manual content analysis (Boumans & Vliegthart, 2014). For the remainder of the article, the three categories *sources*, *agencies* and *newspapers* will be referred to as ‘domains’. Within these domains, the collection of texts from each specific source, agency or newspaper is called a ‘corpus’. Thus, all articles of news agency ANP together form one corpus, which combined with the corpus of *Novum* forms the domain of the agencies.

## Constructing meaning through word choice

The meaning of a text can be derived from the words that the text consists of. Our research aim is to compare similarity between collections of texts. To assess this, we rely on a ‘bag of words’ representation of the texts, meaning that the order of the words is not taken into consideration. Theoretically, word order can substantially change the nature of

a text and the meaning of a specific word depends on the context in which it appears. In practice, however, for many purposes, the list of words that a text is reduced to is an adequate reflection of the general meaning of that text (Grimmer and Stewart, 2013). We argue that the higher the content similarity between an agency corpus and a newspaper corpus, the more likely it is that the general meaning of their content overlaps. Likewise, the higher the content similarity between the corpus of a source and that of an agency or newspaper, the more successful the source has been in promoting its perspective. In the next section, we shall describe our proposed technique, and illustrate with an example that the similarity score introduced below is a realistic indicator of the degree to which two texts share meaning.

### *Tf-idf and cosine similarity*

The most straightforward technique to assess similarities between (collections of) texts is to compare the presence and frequency of terms (i.e. words) in these texts. Yet research on automated text analysis has demonstrated that raw frequency as a measure is not the best indicator to assess a text's meaning (Robertson, 2004). One problem with this measure is that it considers all terms to be equally important and informative, where in fact many terms tell us little about the meaning of a specific text. This is particularly the case for the most common terms in language. The widely applied *Term Frequency Inverse Document Frequency* (tf-idf) measure corrects for this problem by assigning less weight to terms that occur frequently, and more weight to uncommon terms (Jones, 1972). An example of the latter in our case would be 'isotope', a very specific term that is not frequently encountered. When the term appears across two documents, these documents arguably have something substantive in common. Using tf-idf word scores, we can extract the substantive terms that form the basis of the comparisons between domains. Tf-idf is calculated by multiplying the number of times the word appears in a document by the log of the total number of documents, divided by the number of documents that the word resides in. Consequently, it increases with the frequency of the term  $j$ , but decreases as the term occurs in more documents ( $k$ ) in the set of ( $N$ ) documents (Leydesdorff and Welbers, 2011). The formula for the tf-idf score of a term in document  $j$  reads as follows:

$$w_{i,j} = tf_{i,j} \times \log \left( \frac{N}{df_i} \right) \quad (1)$$

where  $tf_{i,j}$  is the number of occurrences of the term in  $j$ ,  $df_i$  is the number of documents containing the term and  $N$  is the total number of documents.

Tf-idf enables us to make a thorough comparison between texts. Next, a tool to standardize the degree of similarity is required. To this end, we rely on one of the most frequently applied tools from the field of data mining: *Cosine similarity* (Tan et al., 2006). An important explanation for its popularity – it is applied across disciplines varying from medicine to engineering (Tan et al., 2006) – is the fact that the outcome is easy to understand. The similarity score can take any value between 0 and 1, depending on the degree of similar tf-idf terms. If the cosine similarity score is 0, two (sets of) documents do not



share any terms. A score of 1 implies that the two (sets of) documents are identical in their use of words.

The cosine similarity is defined as follows (where  $x$  and  $y$  represent the term frequencies of the two (sets of) documents):

$$1 - \cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|} \quad (2)$$

To facilitate an understanding of the measure and its values, Appendix 1 provides an example of three texts and their joint cosine scores.

A software tool has been developed to calculate the tf-idf scores and create two-dimensional ‘word-document’ matrices that represent the collections of documents per corpus. In a word-document matrix, each row describes a document and each column corresponds to a word. The tool then uses these matrices to calculate the cosine score for a specified time period and specified relation (e.g. Greenpeace and ANP). These data are read into SPSS, after which analyses of correlation and regression are performed to answer our research questions.

### Description dataset

One of the advantages of a computer-assisted approach is that it allows analysing the entire population of the material rather than just a sample. This is particularly valuable because the dynamics of the relationship between sources and media tend to form patterns; sources often dominate in the news discovery phase, while in the ‘news gathering phase’, journalists do (Dimitrova and Stromback, 2009). Because similarity scores between the press releases and media articles over the period 2003–2012<sup>1</sup> are aggregated to a monthly level, these short-term fluctuations in the degree of similarity are accounted for.

All articles of both national wire services, ANP ( $n=753$ ) and *Novum* ( $n=194$ ) are included, as well as a wide spectrum of newspapers, the contents of which are digitally available through the Lexis Nexis database: national quality newspapers (NRC ( $n=374$ ) and *Volkskrant* ( $n=377$ )), a popular newspaper (*Telegraaf* ( $n=170$ )) and two regional papers (*Noordhollands Dagblad* ( $n=491$ ) and *Provinciale Zeeuwse Courant* ( $n=720$ )). The selection of the regional papers is based on the region where the two nuclear reactors in the Netherlands are situated: EPZ is active in Borssele, Zeeland (*Provinciale Zeeuwse Courant*) and NRG in Petten, North Holland (*Noordhollands Dagblad*). Relevant articles are retrieved through a search term that included ‘atomic energy’ OR ‘nuclear!’ OR ‘reactor’, AND ‘Borssele’ OR ‘Borsele’ OR ‘Petten’.<sup>2</sup>

Press releases of the following sources are included: energy producer EPZ ( $n=110$ ), nuclear research and service provider NRG ( $n=100$ ) and nongovernmental environmental organization Greenpeace ( $n=183$ ). The press releases of the sources are retrieved from the organizations’ websites. For EPZ and NRG, all press releases on the websites are included. In the case of Greenpeace, only the releases that were labelled by the organization with the tag ‘nuclear energy’ were selected. The data have been aggregated to a monthly level of analysis.

## Results

The result section is structured on three types of analysis. First, a correlational analysis will deliver insight into the extent to which the publication of information subsidies and media attention is correlated, offering a first understanding of the media–source dynamics. Second, the cosine similarity measure is applied to gain insight into the extent to which the content of the different corpora overlaps. Third, the similarity levels between the domains will be regressed on time to assess possible trends in reliance on source content by agencies and newspapers.

### *Prominence of the debate*

Figure 1a and 1b show the amount of published articles on nuclear energy per source and per media domain per year.

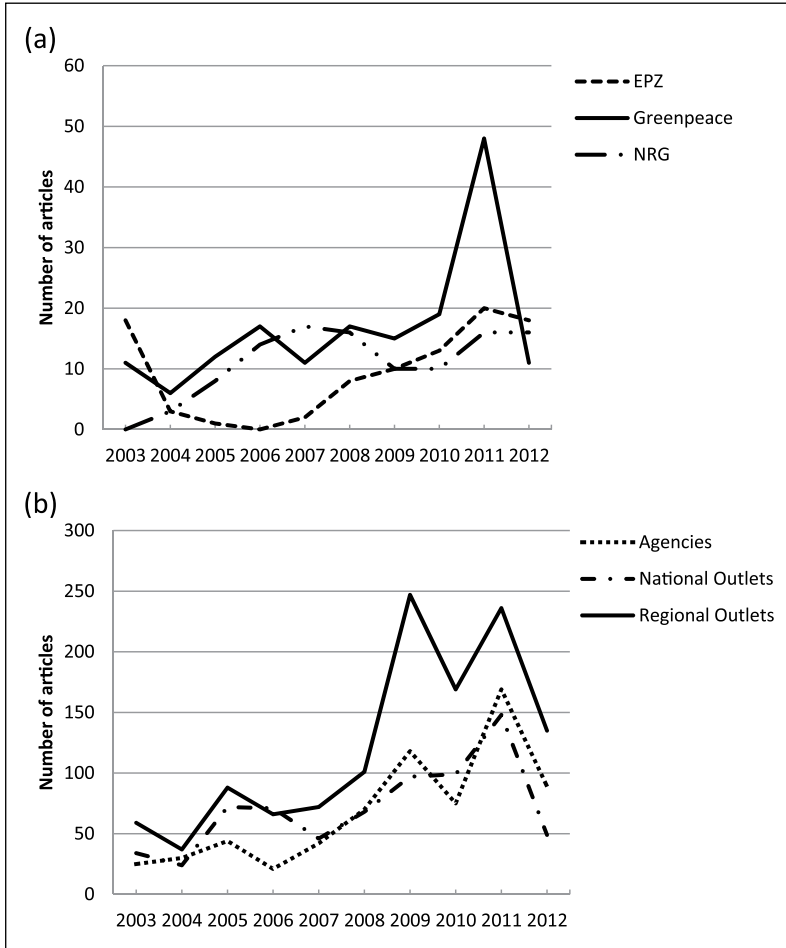
Figure 1a shows that overall, the three sources are publishing more stories over time. In most years, Greenpeace has the highest publication frequency, up to 48 articles in 2011: more than the combined output of the two industrial organizations. From Figure 1b, we can infer that in this same year, media attention was at its peak as well. Table 1 provides insight into the degree of correlation between the specific corpora (individual sources, agencies and newspapers), thus to what degree the visibility of output of these sources overlaps within months.

Table 1 shows that the amounts of press releases from Greenpeace, NRG and EPZ are not correlated. In other words, the sources do not appear to communicate their message in the same time frame. When we assess the correlation of the sources with the agencies and newspapers, it can be noticed that the amount of press releases of Greenpeace correlates substantially higher with the amount of articles of the agencies and newspapers compared to the other two sources (see last column). The strongest correlations between agencies and newspapers can be found between ANP and the national newspapers. It becomes clear that Greenpeace, the agencies and the newspapers often share the same time frame to publish on nuclear energy. Whether there is also a shared understanding in terms of the contents of the articles is addressed in the next section.

### *Cosine similarity*

The cosine similarity is a measure that informs about the degree of similarity between the collections of texts. Table 2 depicts the cosine values of the various relationships.

We start our discussion of Table 2 by looking at the relationship between the sources. While the low similarity between Greenpeace and the industrial sources EPZ and NRG (.22 for both relations) will not come as a surprise, the even lower similarity between NRG and EPZ (.19) is interesting. The finding seems to indicate that the two organizations, apart from choosing separate publication moments, also share little commonalities in their messages concerning the issue of nuclear energy. In other words, in their external communication, the two leading nuclear companies do not promote the nuclear issue from a uniform point of view, but rather from their own unique perspectives.



**Figure 1.** (a) Number of articles per source, 2003–2012 and (b) number of articles per media domain, 2003–2012.

With respect to the content overlap between sources and agencies, results suggest that there is a moderate degree of overlap between the content of the two domains, with considerable differences between the two agencies. The content of *Novum* is overall more in line with the sources than the content of ANP is (.49 versus .35). Of the three sources, Greenpeace’s content is generally most similar to the agencies and their content shows a relatively high similarity with the content of *Novum* (.59).

Regarding the relation between sources and newspapers, the results are consistent with the previous findings: Greenpeace’s content is overall most well reflected by the newspapers. There is one notable exception; however, the content of regional newspaper *Noordhollands Dagblad* scores very low on similarity with Greenpeace (.19), but fairly high with respect to NRG (.53). Examination of the terms with the

Table I. Correlation of amount of articles between corpora.

Sources	Sources		Agencies		National newspapers			Regional newspapers			Mean <sup>a</sup>
	EPZ	Greenpeace	ANP	Novum	NRC	Volkskrant	Telegraaf	Noordhollands Dagblad	Provinciaalse Zeeuwse Courant		
Sources	NRG	.14	.36**	.32**	.19*	.10	.22**	.30**	.38**	.27	
	EPZ	.11	.20*	.07	.17*	.11	.08	.09	.15	.12	
	Greenpeace		.32**	.35**	.33**	.31**	.36**	.30**	.44**	.34	
	Mean <sup>b</sup>		.29	.25	.23	.17	.22	.30	.41		
Agencies	ANP		.60**	.59**	.58**	.59**	.69**	.33**	.46**		
	Novum		.36**	.42**	.44**	.44**	.25**	.55**	.55**		
Newspapers	NRC		.78**	.78**	.55**	.55**	.14	.45**	.45**		
	Volkskrant		.63**	.63**	.63**	.63**	.13	.36**	.36**		
	Telegraaf		.35**	.35**	.35**	.35**	.41**	.41**	.41**		
	Noordhollands Dagblad		.30**	.30**	.30**	.30**	.30**	.30**	.30**		
	Df	141	141	141	141	141	141	141	141	141	

EPZ: *Elektriciteits Produktie maatschappij Zuid*; ANP: *Algemeen Nederlands Persbureau*; NRG: Nuclear Research and consultancy Group.

<sup>a</sup>Mean correlation of sources with news agencies and newspapers. For EPZ and NRG, only the regional newspaper that corresponds with the location of the company is included in the calculation.

<sup>b</sup>Mean correlation of news agencies and newspapers with sources. For the regional newspapers, the correlation with the non-regional corporation is excluded from the calculation.

\* $p < .05$ ; \*\* $p < .01$ .

**Table 2.** Mean cosine similarity between sources, agencies and newspapers 2003–2012.

Sources	Sources			Agencies			National newspapers				Regional newspapers			Mean <sup>a</sup>
	EPZ		Greenpeace	ANP		Novum	NRC	Volkskrant	Telegraaf	Noordhollands Dagblad	Provinciaalse Zeeuwse Courant			
	.19	.22	.22	.33	.42	.3	.39	.29	.53	.21				
NRG				.28	.47	.38	.36	.33	.12	.45		.38		
EPZ				.45	.59	.59	.55	.5	.19	.51		.32		
Greenpeace				.35	.49	.42	.43	.37	.36	.48		.48		
Mean <sup>b</sup>					.53	.71	.72	.66	.5	.65				
ANP						.64	.64	.59	.36	.61				
Novum							.88	.81	.48	.76				
NRC								.83	.58	.76				
Volkskrant									.52	.74				
Telegraaf										.44				
Noordhollands Dagblad														

EPZ: *Elektriciteits Produktie maatschappij Zuidi*; ANP: *Algemeen Nederlands Persbureau*; NRG: *Nuclear Research and consultancy Group*.

<sup>a</sup>Mean cosine similarity of sources in relation to news agencies and newspapers. For EPZ and NRG, only the regional newspaper is included in the calculation.

<sup>b</sup>Mean cosine similarity of news agencies and newspapers in relation to sources. For the regional newspapers, the similarity score with the non-regional corporation is excluded from the calculation.

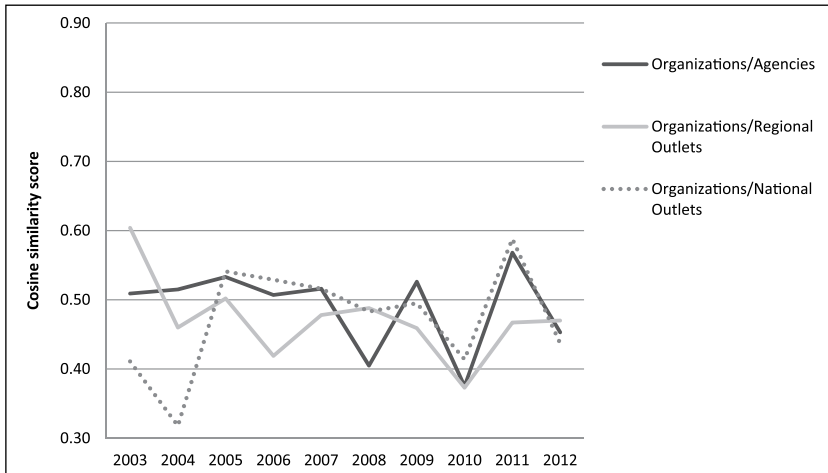
highest tf-idf score confirms that *Noordhollands Dagblad* covers the issue of nuclear energy in largely the same language as NRG. Appendix 2 provides an overview of the most prominent terms per corpus and shows that the abbreviation 'NRG' is on top of *Noordhollands Dagblad*'s ranking. Among the most prominent terms that NRG communicates in their press releases are 'isotopes', 'medical' and 'research', terms that stress the organizations activities in the pharmaceutical world and that are also highly ranked in the coverage of *Noordhollands Dagblad*. These terms are absent in Greenpeace's list, which consists of terms that refer to waste, danger (Chernobyl) and transport. While generally speaking the content of Greenpeace is thus most closely associated with that of the newspapers, there are noticeable differences in the source–newspaper relationships when we differentiate between types of newspaper. The quality newspapers NRC *Handelsblad* and *Volkscrant* hold the highest scores regarding content similarity with the sources (mean of .42 and .43, respectively). Apparently, in their reporting on the issue of nuclear energy, there are commonalities with the content of the sources. This relatively high similarity between the two papers, with .88 being the highest similarity score in the matrix, suggests that the content of NRC *Handelsblad* and *Volkscrant* is very similar regarding nuclear energy.

As is the case with the relationship between NRG and *Noordhollands Dagblad*, so do EPZ and the regional newspaper *Provinciale Zeeuwse Courant* share a moderate degree of similarity (.45). Unlike *Noordhollands Dagblad*, however, the content of *Provinciale Zeeuwse Courant* is even stronger related to the content of Greenpeace (.51). From Table 2, it has become clear that in most instances the content of Greenpeace is considerably more reflected in particularly agency and national newspaper content than the industrial content.

Concerning the relation between agencies and newspapers, the results show that the degree of overlap between newspapers and news agencies is varied, with similarities ranging between .25 (ANP/*Noordhollands Dagblad*) and .63 (*Novum/Volkscrant*). Counter intuitively, of all newspapers the quality newspapers are the ones most similar with the news agencies: *Novum/NRC Handelsblad* scores .61. The regional papers have the least in common with the agencies in terms of contents, which is contradictory to our expectation that regional newspapers would be more similar to agency content than their national counterparts (H1).

### Trends over time

The section above has given insight into the general strengths of the various relationships over the entire research period. As set out in the theoretical section, literature suggests that similarities between corpora may have increased over time. Regression analyses are applied with the independent variable time ( $x$ ) explaining the degree of similarity between two domains ( $y$ ). Data of the three domains have been aggregated to a yearly level, thus every analysis is based on 10 data points. Possible trends in the degree of similarity can be established in this way. Figure 2 presents an overview of the cosine similarity score of the organizations with the agencies and outlets.



**Figure 2.** Cosine scores of organizations with media, 2003–2012.

Figure 2 can partly address our expectation that the degree of content similarity between the three domains has increased over time (H2). As it becomes apparent from the Figure 2, there is no clear trend for any of the three relations of the organizations with the media. While the similarity score between organizations and outlets has increased marginally between 2003 and 2012, regression analyses show that this increase is not statistically significant. Concerns about sources becoming increasingly influential are thus not confirmed by this analysis. From Figure 2, it becomes clear that the degree of similarity differs considerably between years and we will briefly zoom in on one ‘dip’ (2010) and one ‘peak’ (2011). In 2010, the media agenda was dominated by the discussion on a potential new nuclear plant, as an assessment of the media headlines reveals. The headlines of the industry’s press releases indicate that only a fraction of the releases address this issue. And while judged from the NGO’s headlines, Greenpeace does aim to partake in this discussion, the NGO’s relatively low similarity scores with the media indicate the organization is less successful in accessing the media agenda than in most other years. The ‘peak’ year 2011 is marked by the nuclear incident in Fukushima, Japan, resulting in closely aligned agendas that were largely dominated by this disaster – particularly so in the cases of the media and Greenpeace. It appears that the incident provided a ‘window of opportunity’ for the NGO’s oppositional standpoint. To address the second part of our hypothesis, regarding our expectation that newspapers and news agencies are increasingly similar in content, we turn to Figure 3.

Figure 3 gives a visual insight into the inter-media relations. The first thing that can be noted is that the scores are structurally higher than the scores in Figure 2. Regarding the individual relations, one score is significantly increasing over time: between regional and national outlets ( $r=.02$ ,  $p<.05$ ). There is no indication that newspapers – either regional or national – are increasingly relying on news agency content. The results of the regression analyses imply a rejection of our hypotheses on an increased content overlap between sources, agencies and newspapers (H2 and H3).

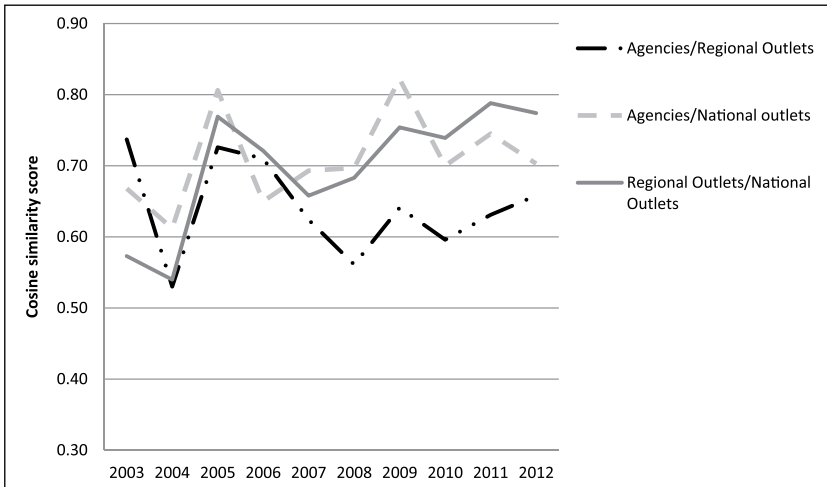


Figure 3. Inter-media cosine scores, 2003–2012.

## Discussion

Studies of journalism and news production have shown that in the relationship between sources and journalists, sources increasingly lead the dance (Carlson, 2009; Prenger et al., 2011). This study demonstrates that news agencies structurally join these ballroom sessions as well. Particularly national newspapers reflect agency content to a substantial degree. This finding is not in line with the rationale that quality newspapers' relatively large resources leads to more unique content (Scholten and Ruigrok, 2009). It may well be that the current financial and time constraints limit the newspaper's aspirations to provide unique coverage. The findings confirm the central role of news agencies in the world of the news and evoke concerns on the plurality of voices that the mainstream media offer. The ultimate consequence of these developments could be a homogenous and uncritical news landscape. The troubling economic situation of the Dutch news agency market notwithstanding, this study found no signs of national news agencies increasingly relying on the content of the selected sources. An interesting finding is that throughout the entire period, there are considerable differences between the two rivaling agencies in terms of content similarity to the sources' press releases: the content of challenger *Novum* is significantly more similar to source content.

Apart from the role of news agencies, the impact of source content was central to the study. The content of Greenpeace was most well reflected in both the agency copy as well as the newspapers' content. This finding contradicts notions that NGOs are generally less successful in securing media access than business sources, yet is in line with earlier findings of Greenpeace being a gatekeeper or primary definer of issues (Anderson, 1997: 8, 167). As David Deacon (2003: 122) remarked, Greenpeace is increasingly 'attuned to the news values of large media organizations'. As such, it can be considered representative for the limited group of globalized, well-resourced NGOs that have tailored their strategies to attract mainstream media's interest. The vast majority of NGOs



and pressure groups, however, do not enjoy these resources and are less successful in gaining media access (Waisbord, 2011).

It is worth restating that written press releases are only one aspect of the complex and multifaceted relationship of sources and journalists (Reich, 2010). Relationships between journalists and PR practitioners are more complex and involve a range of written as well as verbal interactions, many of which are not easy to trace (Lewis et al., 2008). The increasingly sophisticated practice of PR professionals to ‘spin’ news is a prime example of PR influence that is difficult to examine systematically. However, as press releases still account for a substantial part of PR activities (Forde and Johnston, 2013), they provide inevitably conservative, yet verifiable evidence. Furthermore, it is likely that the content of the press releases is in line with the messages that PR professionals promote through other channels like face-to-face contact. Thus, while it should not be seen as an absolute measure for sources’ impact on journalistic content, we expect the similarity score to provide an accurate indication for sources’ success in promoting their interests.

Another point that should be taken into account is the study’s time frame. Since the organizations’ press releases have only been digitally archived since 2004, this study has focused on the last decennium. Finding no significant increase of media reliance on the content of the agencies or included sources during this period, it must be noted that relevant factors like decreasing newsroom capacities and a professionalization of sources’ communication have been at play for some time before the start of our research period. It is well possible that a study covering a longer period will sketch a different picture.

From a methodological viewpoint, the application of cosine similarity to compare texts from different domains has lived up to its expectations and provided useful insights. The approach reduces the complexity of the texts by evaluating and comparing the most significant terms. While this is admittedly a simplification of the texts’ meaning, we posit that in general, the significant terms are indicative of the organizations’ key message. When these terms are also present in media content, there is thus an overlap in agendas.

The next step would be to sophisticate the research design in such way that causality can be established. While analyses like these typically require larger datasets than the one used for this study, time series analysis, for instance, could clarify the extent to which economic, technological and organizational factors are accountable for the degree in content similarity. It could also model the impact of key societal or political events on content overlap.

While in this specific case, it has been shown that the news media do not increasingly rely on the press releases of the selected organizations, this does not imply that there is no growing dependency on sources in general: the number of included sources is too small to make generalizations. Additionally, the hypothesis is tested in a very specific context – nuclear energy. Empirical research on media–source dynamics in different contexts, including a wider array of potential sources, is much needed. While nuclear energy is an important political issue as such and we have no specific reasons to expect patterns for other issues to be very different, it remains an empirical question whether we find the same similarity patterns for cases such as foreign or even celebrity news. Keeping in mind that source checking plays only a ‘modest role in the journalistic process’ (Diekerhof and Bakker, 2012: 252) and that source crediting is not always as transparent as ought to be (Reich, 2010; Scholten and Ruigrok, 2009), critical research on the dynamics between

sources and media remains of vital importance. The approach presented in this article can aid in this crucial task.

Finally, concerning our findings on the regional newspapers, the relatively low similarity with the national news agenda confirms that the regional newspapers add distinctly different perspectives to the media landscape. The content is likely to be more oriented towards local and regional rather than national developments. The relatively high similarity with the corporations' content is in line with Deirdre O'Neill and Catherine O'Connor's (2008) observation that regional journalism may well have become all too dependent on pre-packaged news from resource-rich PR organizations. If anything, this finding legitimizes further research on the relation between sources and journalists on a regional level.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### Notes

1. Not all media data are available for the entire period: content of *Novum* is available since 2004 and content of *Noordhollands Dagblad* since 2007.
2. Dutch equivalents for these terms are used ('kernenergie' for atomic energy, 'nucleair!' for nuclear!). The selection of relevant articles is based on the presence of the combined search terms and is limited to domestic news. Editorials and letters are excluded from the analysis.

### References

- Anderson A (1997) *Media, Culture and the Environment*. London: UCL Press.
- Bakker P and Scholten O (2011) *Communicatiekaart Nederland: Overzicht van Media en Communicatie* [Dutch Communication Map: Overview of Media and Communication]. Deventer: Kluwer.
- Boumans J and Vliegenthart R (2014) Safety first versus op de barricaden. Een inhoudsanalyse van het nucleaire debat in Nederland. [Safety first versus on the barriers. A content analysis of the nuclear debate in the Netherlands] *Tijdschrift voor Communicatiewetenschap* 42(4): 358–377.
- Boyd-Barrett O and Rantanen T (2000) European national news agencies the end of an era or a new beginning? *Journalism* 1(1): 86–105.
- Boyer D (2011) News agency and news mediation in the digital era. *Social Anthropology* 19(1): 6–22.
- Carlson M (2009) Dueling, dancing, or dominating? Journalists and their sources. *Sociology Compass* 3(4): 526–542.
- Cottle S (2003) *News, Public Relations and Power*. London: SAGE.
- Curtin P (1999) Reevaluating Public Relations Information Subsidies. *Journal of Public Relations Research* 11(1): 53–90.
- Curtin P and Rhodenbaugh E (2001) Building the news media agenda on the environment: A comparison of public relations and journalistic sources. *Public Relations Review* 27(2): 179–195.
- Davies N (2008) *Flat Earth News: An Award-Winning Reporter Exposes Falsehood, Distortion and Propaganda in the Global Media*. London: Chatto & Windus.

- Davis A (2000) Public relations, news production and changing patterns of source access in the British National Media. *Media, Culture & Society* 22(1): 39–59.
- Davis A and Cottle S (2003) *News, Public Relations and Power*. London; Thousand Oaks, CA: SAGE.
- Deacon D (2003) Non-governmental organizations and the media. In: Cottle S (ed.) *News, Public Relations and Power*. London: SAGE, pp. 99–115.
- Denham B (2010) Towards conceptual consistency in studies of agenda-building processes: A scholarly review. *Review of Communication* 10(4): 306–323.
- Diekerhof E and Bakker P (2012) To check or not to check: An exploratory study on source checking by Dutch journalists. *Journal of Applied Journalism & Media Studies* 1(2): 241–253.
- Dimitrova D and Stromback J (2009) Look who's talking. *Journalism Practice* 3(1): 75–91.
- Dohmen A (2012, February 29) Nog Meer Banen Schrappen? Onmogelijk [Cutting even more jobs? Impossible]. Online news article. Retrieved from <http://www.nrc.nl/handelsblad/2012/03/05/nog-meer-banen-schrappen-onmogelijk-12268903>
- Forde S and Johnston J (2013) The news triumvirate: Public relations, wire agencies and online copy. *Journalism Studies* 14(1): 113–129.
- Franklin B (2008) *Pulling Newspapers Apart: Analysing Print Journalism*. London: Routledge.
- Franklin B (2011) Sources, credibility and the continuing crisis of UK journalism. In: Franklin B and Carlson M (eds) *Journalists, Sources, and Credibility: New Perspectives*. London: Routledge, pp. 90–107.
- Franklin B (2012) The Future of Journalism. *Journalism Studies* 13(5–6): 663–681.
- Freer J (2007) UK regional and local newspapers. In: Anderson P and Ward G (eds) *The Future of Journalism in the Advanced Democracies*. Burlington, VT: Ashgate Publishing, pp. 89–103.
- Grimmer J and Stewart B (2013) Text as data. The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 97: 1–31.
- Hall S, Critcher C, Jefferson T, et al. (1978) *Policing the Crisis: Mugging the State, Law, and Order*. London: Macmillan.
- Hansen A (2011) Communication, media and environment: Towards reconnecting research on the production, content and social implications of environmental communication. *International Communication Gazette* 73(1): 7–25.
- Hijmans E, Schaafraad P, Buijs K, et al. (2011) Wie schrijft ons nieuws? een analyse van voorverpakt nieuws en brontransparantie in binnenlandse nieuwsonderwerpen in vier Nederlandse en Vlaamse kwaliteitskranten. [Who writes our news? An analysis of pre-produced news and source transparency in domestic news items in four Dutch and Flemish quality newspapers] *Tijdschrift Voor Communicatiewetenschap* 39(2): 77–91.
- Jackson D and Moloney K (2015). Inside Churnalism: PR, journalism and power relationships in flux. *Journalism Studies*. Epub ahead of print 10 March. DOI: 10.1080/1461670X.2015.1017597.
- Johnston J and Forde S (2011) The silent partner: News agencies and 21st century news. *International Journal of Communication* 5: 115–129.
- Jones K (1972) A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28(1): 11–21.
- Kerrigan F and Graham G (2010) Interaction of regional news-media production and consumption through the social space. *Journal of Marketing Management* 26(3–4): 302–320.
- Lewis J, Williams A and Franklin B (2008) A compromised fourth estate? UK news journalism, public relations and news sources. *Journalism Studies* 9(1): 1–20.
- Leydesdorff L and Welbers K (2011) The semantic mapping of words and co-words in contexts. *Journal of Informetrics* 5: 469–475.
- McChesney R (2008) *The Political Economy of Media: Enduring Issues, Emerging Dilemmas*. New York: Monthly Review Press.

- McChesney R (2012) Farewell to journalism? Time for a rethinking. *Journalism Practice* 6(5–6): 614–626.
- Manning P (2001) *News and News Sources: A Critical Introduction*. London: SAGE.
- Manning P (2008) The press association and news agency sources. In: Franklin B (ed.) *Pulling Newspapers Apart: Analysing Print Journalism*. London: Routledge, pp. 247–255.
- O’Neill D and O’Connor C (2008) The passive journalist: How sources dominate local news. *Journalism Practice* 2(3): 487–500.
- Phillips A (2010) Old sources: New bottles. In: Fenton (ed) *New Media, Old News*. London: SAGE, pp. 87–101.
- Prenger M, van der Valk L, Van Vree F, et al. (2011) *Gevaarlijk Spel: De Verhouding tussen PR & Voorlichting en Journalistiek* [A Dangerous Game. The Relation between PR & Information and Journalism]. Diemen: AMB.
- Rafter K (2014) Voices in the crisis: The role of media elites in interpreting Ireland’s banking collapse. *European Journal of Communication* 29(5): 598–607.
- Reich Z (2010) Measuring the impact of PR on published news in increasingly fragmented news environments. *Journalism Studies* 11(6): 799–816.
- Robertson S (2004) Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation* 60(5): 503–520.
- Rutten P and Slot M (2011) Zijn de persbureaus te verslaan. De positie van Nederlandse persbureaus in de nieuwsketen. [Can the press agencies be defeated? The position of Dutch press agencies in the news Chain]. *Research Report TNO*.
- Scholten O and Ruigrok N (2009) *Bronnen in het Nieuws: Een Onderzoek naar Brongebruik in Nederlandse Dagbladen 2006–2008* [Sources in the News: A Study of Source Usage in Dutch Newspapers 2006–2008]. Amsterdam: Nederlandse Nieuwsmonitor.
- Shehata A (2010) Marking Journalistic Independence. Official dominance and the rule of product substitution in Swedish press coverage. *European Journal of Communication* 25(2): 123–137.
- Sigal L (1986) Who? Sources make the news. In: Manoff R and Schudson M (eds) *Reading the News: A Pantheon Guide to Popular Culture*. New York: Pantheon Books, pp. 24–43.
- Sissons H (2012) Journalism and public relations: A tale of two discourses. *Discourse & Communication* 6(3): 273–294.
- Tan P-N, Steinbach M and Kumar V (2006) *Introduction to Data Mining*. Boston, MA: Pearson Education.
- Tuchman G (1973) Making news by doing work: Routinizing the unexpected. *American Journal of Sociology* 79: 110–131.
- Van Dam H (2003) *Kernthema’s (valedictory speech)* [‘Core Themes’]. Delft: Technical University Delft.
- Van Hout T and Jacobs G (2010) News production theory and practice: Fieldwork notes on power, interaction and agency. *Pragmatics* 18(1): 59–85.
- Waisbord S (2011) Can NGOs change the news. *International Journal of Communication* 5: 142–165.

**Appendix I.** Cosine similarity score example.

Press release Greenpeace,  
July 2008

**Yet another French nuclear plant leak**

On July 7, 30,000 liters of radio-active water leaked from another nuclear plant.

Again, a leak has been discovered at a nuclear plant in France on Friday July 18. A crack in a pipeline of the plant in Romans-sur-Isere, south of Lyon, caused the leakage of uranium. It is the second leak within two weeks at a nuclear plant. The new leak once again demonstrates that nuclear energy remains dangerous and can lead to severe contamination.

Press release NRG, July 2008

**NRG prepared for development new nuclear plants**

Nuclear expertise center NRG is ready to actively contribute to expanding the share of nuclear energy in the Dutch energy mix. The recently published annual report makes clear that NRG is planning to build a new nuclear plant in Finland, a fusion reactor in France and prolonging the lifespan of nuclear plants in England. This is the concrete result of the new strategy to shift NRG's activities from maintaining expertise to the development of new nuclear plants. The company's new position neatly fits the government's standpoint on the role of nuclear energy in the Netherlands, as expressed in the Energy Report of June 2008. The government works out different scenarios for nuclear energy in order for the next cabinet to make a decision on the development of one or more new nuclear plants in our country. NRG views this as a promising step towards the sustainability of our national energy supply and to keep up with the latest developments around nuclear energy in Europe. The current share of nuclear energy in Europe amounts up to 30 percent and will need to be maintained to secure a safe, reliable and affordable future energy supply. This implies that in the next twenty years, 60 to 70 third generation nuclear plants will need to be build. This generation produces a smaller volume and shorter active waste and uses around 20 percent less uranium. NRG focuses her expertise on supporting the government and the international market. The realization of new plants in the coming years involves an investment of tens of billions of euros. The time has come for the Dutch business world to position itself into a share in this.

NoordHollands Dagblad  
article, July 2008

**NRG shifts strategy towards new nuclear plants**

Petten, July 15 2008. The Netherlands will build new nuclear plants in the near future. Nuclear Research & consultancy Group (NRG, Petten) is convinced of this. To avoid missing the boat, the organization has drastically changed course. In the past years, without prospects of more nuclear energy, NRG among others maintained nuclear expertise in the Netherlands. Currently, NRG focuses more prominently on developing new plants. The daughter of Energy Research Centre Netherlands (ECN) has plans for a new to build her own new plant well. NRG, owner of the high flux reactor (hfr) in the dunes, wants to build a new reactor in Petten after 2015, the so-called Pallas project.

According to NRG, it provides nuclear expertise for a large new plant in Finland, and it is involved in prolonging the lifespan of existing nuclear plants in England. In the shape of a material study, NRG also contributes to an experimental fusion reactor in Cadarache, in the South of France.

---

**Appendix I. (Continued)**

Press release Greenpeace,  
July 2008

The French nuclear watchdog ASN decided on the 11<sup>th</sup> of July that the Tricastin-plant temporarily needs to be shut down. Because of a leak, 30,000 liters radio-active water spilled into the soil and the rivers Gaffiere and Lauzon. All activities in the plant have been cancelled and new safety measures are introduced to avoid future incidents.

As a consequence of the accident, all recreational areas are closed, the agriculture can no longer use water from the rivers and fishing is forbidden. The two leaks are yet another example of the risks that come along with nuclear energy. While the nuclear industry attempts to convince everyone that nuclear energy is safe, accidents like these continue to prove them wrong. Rather than wasting time on nuclear energy, we ought to invest in a real energy revolution. Greenpeace has been campaigning internationally against nuclear plants for years. Earlier this year, Greenpeace-activists protested against the realization of two nuclear plants in Finland and France

Press release NRG, July 2008

NoordHollands Dagblad  
article, July 2008

The optimism about the future is great in Petten. The atom scientists have calculated that in the period til 2030, around 70 new nuclear plants will be built in Europe. NRG is growing, although it is hard to find enough qualified personnel. President R. Stol has faith however. At the University of Delft for instance, there is an increase of students applying for nuclear studies.

---

NRG: Nuclear Research and consultancy Group; NHD: *NoordHollands Dagblad*.

Mutual terms are underlined. Cosine scores: Greenpeace–NRG .260; Greenpeace–NHD .122; NRG–NHD .394.

Appendix 1 presents three articles: one from Greenpeace, one from NRG and one from NHD. Upon reading the texts, we notice that the content of NRG and NHD appears similar; they both discuss the organization's future strategy of building new plants. The contents of the two texts are far from identical, but the texts do show a fair degree of overlap both in terms of vocabulary (see mutual terms, underlined) as well as with regard to their

meaning. This is reflected in the cosine similarity score of .394 for the two texts. The Greenpeace text, on the other hand, has less in common with the other two texts. This becomes clear from the relatively low number of mutual terms and is reflected in low similarity scores of .122 (Greenpeace–NHD) and .260 (Greenpeace–NRG). The fact that this latter score is higher than the previous is explained by the higher amount of terms that Greenpeace and NRG share: both mention ‘nuclear energy’ multiple times and refer to ‘uranium’, the ‘future’ and ‘safety’. The news article does not include these latter terms and refers less often explicitly to energy. Because the terms ‘nuclear’, ‘energy’ and ‘plant’ occur frequently across all three texts, these terms have a relatively lower individual tf-idf score than for terms that occur less often.

Appendix 2. Most prominent tf-idf terms per corpus.

Sources	Agencies			National newspapers			Regional newspapers		
	NRG	Greenpeace	ANP	Novum	Volkskrant	NRC	De Telegraaf	Noordhollands Dagblad	Provinciaalse Zeeuwse Courant
1 Nuclear fuel	Flux	Nuclear waste	Greenpeace	Delta	ANP	Waste	Delta	NRG	Zeeland
2 Transport	Authorities	Waste	Isotopes	Isotopes	Europe	Europe	Essent	Pallas	Delta
3 Used	Isotopes	Radioactive	Delta	Greenpeace	Minister	Cabinet	RWE	Isotopes	EPZ
4 Container	Hoge	Clean	NRG	NRG	Waste	Minister	Cabinet	Zijpe	Zeeuws
5 France	Medical	Activists	Minister	Ministry	Cabinet	ANP	Minister	Million	Viaanderen
6 Company	Research	Russia	Waste	Research	Energy supply	Environment	Million	Bibu	Nuclear
7 COVRA	HFR	Delta	Transport	Construction	Environment	Delta	Construction	Medical	CDA
8 Recycling	Waste	Verhagen	Cabinet	EPZ	State	Greenpeace	Secretary of state	Dunes	Bevelanden
9 Minister	Production	Uranium	Haag	Minister	Economy	State	NRG	Research reactor	Waste
10 Measurement	Events	Contamination	Verhagen	Cda	NRG	Energy supply	Environment	Flux	Construction
11 Unit	Pallas	France	EPZ	Spokesman	Greenpeace	Economy	Waste	ECN	Walcheren
12 Waste	Company	Sustainable	Nuclear waste	RWE	Closure	Construction	CDA	High	Covra
13 Hague	Radioactive	Chernobyl	Spokesman	Safety	Isotopes	Research	European	Construction	Duiveland
14 Recycling factory	European	Construction	CDA	Cabinet	Ministry	EPZ	Energy company	Research	Schouwen
15 Inspection	Performed	Teuling	Medical	Zeeuwse	European	CDA	EPZ	Production	Tholen

EPZ: Elektriciteits Produktie maatschappij Zuid; NRG: Nuclear Research and consultancy Group; ANP: Algemeen Nederlands Persbureau.