Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Lingue Moderne per la Comunicazione e
la Cooperazione Internazionale
Classe LM-38

Tesi di Laurea

Multi-Dimensional Analysis and Public Communication of Science and Technology: A Corpus-based Study on the Controversy over Fracking

Relatore
Prof.ssa Maria Teresa Musacchio

Correlatore: Laureanda
Prof. Federico Neresini   Virginia Zorzi
                                                                                      n° matr.1060662 / LMLCC

Anno Accademico 2014 / 2015

Table of contents

Chapter 1

Delineating a corpus-based approach to the controversy over fracking

1.1 Introduction

This study analyses the media coverage of a controversy which has received widespread international attention in recent years: that over the application of hydraulic fracturing – also known as fracking – technologies to extract shale gas from underground reserves. Having long developed an interest concerning the challenges posed by environmental problems to the dominant models of human growth and development, I first became interested in the controversy during a brief correspondence kept with a scientific committee member of the Italian section of ASPO (Association Study of Peak Oil and Gas), who sent me a detailed report about these controversial technologies[1] (Rossi, 2014). As a matter of fact, the enormous opportunities that this new form of oil and gas extraction seems to offer are faced with heavy environmental effects frequently ascribed to the application of these technologies. This has ignited a lively debate which has spread throughout different media and now involves a number of different actors (including citizens, oil and gas companies, geologists, engineers and other specialized researchers, economists, political institutions and environmental organizations), each claiming their own arguments in the dispute. After a first approach to the subject through the ASPO report, further research into the potential advantages and drawbacks attributed to fracking practices revealed the extremely conflictual and unstable condition of the controversy, stretching from local protests to official reports, from national and international policy to environmental campaigns. One of the aspects that most attracted my attention was the considerable space devoted to this dispute by the media; hence my interest in analysing the way in which it has been presented to the

---

[1] Hydraulic fracturing is in fact part of a process consisting of the combination of vertical drilling, horizontal drilling and the actual fracking; hydraulic fracturing, or fracking, is often used to refer to the whole extraction process. Therefore, fracking can be regarded both as a single technology if all stages are considered together, and as a combination of technologies. In this study, both singular and plural forms will be used.

public through mass communication. Therefore, the analysis will be focused on linguistic and communicative aspects characterizing newspaper and magazine reports on the controversy, with the aim of finding some detectable trends reflecting linguistic, social and cultural patterns underlying the representation of fracking provided by the media.

1.2. A brief insight into the fracking technology and its implications

Resulting from industrial exploration, testing, research and trials carried out in the U.S. and dating back to the 1980s (although basic versions of it have been attested as far back as World War II and earlier) , hydraulic fracturing is based on drilling vertically into the ground down to the shale layers to reach and extract gas and oil. Shale is defined as a type of sedimentary rock which, due to geological factors, is particularly rich in hydrocarbons, which it holds in its pores. Once shale has been reached, the drilling is turned parallel to the surface (horizontal drilling). To make the rock release the hydrocarbons, large quantities of water are pumped at high pressure, causing the surrounding rock to break in several small cracks. To ease this operation, the water is mixed with sand, to prop open the cracks after the fluid retrieves, and with chemicals – whose identification is made difficult by industrial secrecy policies – to reduce friction and thicken the water. Thus, due to the change in pressure, hydrocarbons held in the rock pores stream from the cracks into the well and then towards the surface, where they are gathered, stored and picked up by trucks to be hauled. Dozens of fractures can be practiced along the horizontal leg of a well. Most of the gas or oil is yielded within months from the initial fracturing; however, always owing to enormous pressure differences, hydrocarbon molecules within more or less one meter from the cracks can travel from inside the shale towards them, continuing to produce hydrocarbons in slowly declining amounts over the following years. The relatively short lifecycle of a well, and the fact that it can drain fuel from no more than some 70 meters around the hole imply that new wells have to be constantly drilled, and they have to be within relatively reduced distance from each other, to avoid leaving any "valuable

hydrocarbons behind" (Gold, 2014:28-30). Fracking has been made practicable at economically acceptable rates and has therefore been massively implemented on several areas of the United States (particularly in its central and eastern areas) since around 2005-2006. This has considerably increased the U.S. gas and oil production, with relevant consequences on national economy and a relevant reduction of energy imports from overseas. The surge in domestic energy production has been regarded as a revolution, and has generated expectations of long-term domestic fossil fuel abundance, thus transforming the vision of the nation's energy future, to the point of envisaging the U.S. as a possible natural gas exporter. Furthermore, natural gas is considered a low-carbon energy source with respect to conventional oil and especially coal, thus allowing fracking supporters to define it as a 'cleaner' and 'greener' form of energy, with regards to global warming preoccupations. Finally, the idea of readily available, nationally-produced cheap energy, together with the undeniable energy abundance occurring in the U.S. since the elaboration and application of current fracking techniques, have had a wide international appeal. Governments and energy companies have thus started to imagine and plan the development of fracking industries also outside the U.S.

However, concerns and environmental drawbacks related to fracking started to be felt already in its wake. These include:

- The disposal of large quantities of wastewater. The mixture of water and chemicals (some of which are suspected to be toxic and carcinogenic) used for the actual fracturing of the shale and then retrieved in the final stages of the process need in fact to be treated with adequate technologies in order not to cause severe environmental pollution.

- Aquifers contamination. If the borehole is not well isolated from the surrounding rocks, or if the local geology makes it possible, residual waters which remain, even in small quantities, inside the well and within the cracks might come into contact with aquifers employed either in agriculture or for human usage.

- Earthquakes. It has been attested that the crack-opening operations cause micro-seismic activity, seldom perceived by humans. However, the frequency of larger earthquakes has increased in the U.S. areas most intensely interested by hydraulic fracking activities. Moreover, it is suspected that even more serious

consequences might take place if fracking happens to interact with geological faults, whether active or not.

- Natural gas leakages. Gas (especially methane) can leak out of wells into the surrounding environment, causing air pollution – and explosions in some extreme case – and increasing the greenhouse effect: if it leaks into the atmosphere, methane has a more powerful greenhouse effect than carbon dioxide.

These are just a few examples (Rossi, 2014: 3,4) of the numerous doubts cast on the reliability and usefulness of this newly applied technology: it seems to pose serious risks to both the environment and people's health, although there is no scientific agreement about the extent to which fracking can be held responsible for any of those actual or potential damages. Even more relevantly, fracking has spread fast since 2006, permanently affecting the American landscape, the life of numerous local communities and the U.S. energy policy.

1.3 Core aspects of the study

The kind of linguistic study here proposed takes an interest in the main trends expressed by the media in communicating the controversy. Therefore, it seeks to detect these trends by means of a macro-analytical approach, applied to a relatively large amount of data: in other words, by means of corpus analysis methods. The mass media can easily offer great quantities of available information in this sense, especially for what concerns retrieving materials from large databases available online. To avoid data dispersion and the mixing of exceedingly diverging trends, then, newspapers and magazines (both in paper and online edition) were selected as the primary sources of the analysed texts. Only texts produced in English were included in the corpus, for several reasons: firstly, the methodological procedures adopted only existed for the English language (adapting it to other languages would have required a completely different project and completely different competences); secondly, a contrastive perspective would have limited in-depth analysis of the data, and has been therefore put aside; thirdly, the controversy was originated in an English-speaking country, and the English-

speaking world is probably where the debate over fracking was (and still is) most widespread, thus offering more complete sets of observable data.

Dealing with major objects of scientific and technological controversy, however, quite easily involves extending one's interests beyond merely lexico-grammatical or semantic observations, precisely as corpus analysis cannot be considered separately from the circumstances which generated the corpus (Stubbs, 1996: 83). There is a number of issues at stake in the controversy, which could be involved in linguistic analyses: political and economic agendas, citizens' and activists' claims, technical and scientific debate, the role of the mass media, climate change, relationships among science, technology and society, public perceptions of innovation and risk, and so on. Although they are treated as different concepts, for obvious expository and conceptualizing necessities, they cannot be completely separated from each other, nor from the corpus itself, because they all represent social practices. As such, they do not happen in a vacuum, but shape and are shaped reciprocally among each other, in complex and contingent interaction processes (Bijker, 1992: 1-13). Thus, among the possible perspectives that could be adopted in order to give a more comprehensive interpretation of the trends detected in the corpus, a sociological approach was suggested to me, and proved to be a valuable interpretive tool. The sociological fields of Science and Technology Studies (STS) and Public Communication of Science and Technology (PCST) studies address some aspects which are pre-eminent in the controversy, such as the relationship between science, technology, innovation and society, and the way scientific and technological knowledge is produced and communicated. These approaches were therefore used as an integration to the results and interpretation of the corpus analysis section of the study.

1.4 A brief state-of-the-art overview

Corpus analysis is based on the centrality of naturally occurring language as the object of study of linguistics, in contrast to the elaboration of theories exclusively based on author-designed examples. This approach acknowledges language as a social practice, which actively reproduces society while being produced within it. In this

corpus-based theoretical frameworks, the saussurian dualist concept of language (strictly dividing *langue* – the abstract rules underlying the use of language – and *parole*, the actual occurrences of language in use) is abandoned in favour of a more unified and continuous idea of language (Firth, Halliday, Sinclair in Stubbs, 1996: 44-46), where deep underlying grammatical choices and contingent occurrences are regarded as two different perspectives from which the very same phenomenon can be observed (Webster and Halliday in Halliday, 2005: 193-198). From this point of view, language emerges as an inherently probabilistic system(Halliday, 2005: 238), where frequency in use instantiates probability in the system, and register variation is defined as the skewing of some of the overall probabilities. Hence the importance of corpus-attested language and the use of reference corpora to evaluate data against a body of language use representing the whole system. Corpus analysis in these studies was however limited to small corpora (much smaller than the one I intended to collect) and focused on isolated aspects of the language, or offering interpretations often accompanied by manual checks and individual text analysis. Also Bednarek's account (Bednarek, 2006) addressed extremely relevant aspects to the subject matter of this study, such as evaluation and the author's stance, analysing the discourse of a corpus of 100 newspaper articles. However, her research method could only be applied in a combination of micro and macro-analysis. Another concept connected to my research purposes is that of metadiscourse (Hyland, 2005), which refers to the linguistic tools used by a writer to convey and negotiate meaning with the reader. However, this is such a wide and context-specific concept that it makes large-scale automated analysis quite difficult. Therefore, Hyland selected some classes of words which quite systematically revealed certain metadiscourse functions, and applied them to different corpora, including a set of scientific versus popular science articles. The comparison of different genres could also offer useful frameworks in characterizing the corpus as pertaining to certain media sub-genres: those of newspapers and magazines, with special attention to science popularization. In this regard, sociolinguistics analysed genre awareness on the part of writers (Berkenkotter, Huckin, 1995) and, more relevantly for this study, genre comparison through corpus analysis has been thoroughly employed in the writing of the *Longman Grammar of Spoken and Written English* (Biber et al., 1999), although no account is provided there about popular science and science in the news. Moreover, a

suitable set of detectable and quantifiable items had to be identified to be used in this analysis. The idea of researching lexico-grammatical (and therefore also syntactical) patterns of co-occurrence to detect communicative functions differentiating texts of various genres was developed by Biber (1988) in the form of Multi-Dimensional Analysis, which was adopted to assess written versus spoken genre variation through six different "Dimensions of Variation" for the English language. Biber's analysis was applied to several commonly acknowledged English language genres and sub-genres, including press reportage (relevant to the current analysis), and was then used to develop a typology of texts (Biber, 1989). It did not, however, touch the characterization of popular science texts, nor were later applications of Multi-Dimensional Analysis (Biber, Conrad, 1998) concerned about these aspects. Further research conducted by Biber with Connor and Upton (2007) addressed science articles, but was based on the micro-analysis of rhetorical moves (consequently, not applicable to a relatively large corpus). A Discourse Studies approach to the discourse of science popularization was adopted by Myers (2003) in questioning widespread traditional assumptions about public understanding of science and the status attributed to scientific experts, without however considering merely linguistic aspects in detail. In another paper by the same author (Myers, 1998), rhetorical strategies related to the discourse of environmental sustainability were analysed and discussed through a small corpus of leaflets, but the analysis was focused on the perception of the corpus on the part of readers, and involved experiments based on focus groups with actual readers, falling outside the kind of research applied to this study.

The analysis of different views of science and technology in relation to society is among the focal points of STS and PCST studies. The majority of the considered works come from an Anglophone context (which, in any case, largely applies to the analysed texts), while some other studies are at least partially concerned with the Italian context. These studies focus on the dynamics inside and outside the scientific community, with a special interest towards critical situations and debated issues. Other related subject matters include: the role of scientific journalism; the nature of institutional activities and campaigns dealing with science and technology and the trust they generate among lay people; public understanding of science; the role of experts and their media representation; the communication of risk; the elaboration of new proposals to consider

and shape scientific and technical knowledge production and build the interaction between the scientific community, politics and all the other parts of society (Trench, Bucchi, 2008). STS and PCST seem to share some elements: for example, the concepts of 'boundary work' and 'boundary object'(Bucchi, 2000; Neresini, 2011)[2], and a strong interest in debates and controversies. As for the STS works I have examined, another core issue they approach is the way technological innovation occurs, is shaped and perceived within society, especially in the light of the growing pervasiveness and influence of technological applications in everyday life (Bijker, 1992; Neresini, 2011; Lorenzet, 2013). Particular attention is placed in challenging current stereotypical distinctions between science, technology and their outcome in society (Neresini, 2011). Media coverage and its principles have also been addressed within STS (Hilgartner, Bosk, 1998), while controversies, in all their complexity and comprehensiveness are the main object of study of controversy mapping, a sociological approach related to Actor-Network Theory (Venturini, 2010).

1.5 Research question

The purpose of this study is to analyse the way in which the highly controversial technology of hydraulic fracturing and all the debate surrounding it have been communicated, in recent years, by English language newspapers and magazines. The first step towards an evaluation of the discursive practices employed in the analysed media consists in a corpus-based analysis capable of identifying the most frequent linguistic structures and the most frequent lexical items chosen to represent the controversy in the texts. This analytical stage is aimed both at the corpus on the whole and at some of its sub-sections. In particular, newspaper articles, magazine articles (especially from popular science magazines) and letters to the editor are going to be compared and analysed in detail. The following step consists in observing the obtained results and their interpretations through sociological perspectives, in order to detect:

- The way in which the technical and scientific knowledge over hydraulic fracturing is produced, managed, received and reacted to by the various actors

---

[2] However, these two notions are not going to be adopted in this study.

involved in the controversy (all seen through the representation provided by the corpus);

* The way in which the relationships among different actors are described;

* The way in which the highly controversial nature of the debate affects the previous two points;

* Which values, assumptions and paradigms underlie the media coverage of the controversy, and which (if any) of these values, assumptions and paradigms are being negotiated and challenged within the controversy.

Therefore, this study intends to integrate corpus-based linguistics methods with sociological approaches, in order for the former to be enriched by possible implications of the obtained results, and for the latter to be complemented by a specifically linguistic basis.

1.6 Methods of analysis

1.6.1 The corpus

The corpus consists of 928 articles mainly taken from newspapers and magazines, all written in English and published between 2010 and 2014. Most articles were published in the U.S. and British press, but also Canadian, Irish, South African and a couple of Indian newspapers were included.

The majority of these texts were selected and downloaded through Lexisnexis , a corporation providing legal services which include computer-assisted legal research thanks to the accessibility to legal and journalistic documents. Through the Lexisnexis Academic search engine (www.lexisnexis.com/hottopics/lnacademic/), articles from both paper and online editions are extracted from archives and made available for downloading. The part of the corpus thus collected had already been divided by Lexisnexis into several categories: 'newspapers', 'newsletters', 'industry trade press', 'news', 'blogs', 'web-based publications', 'magazines and journals', 'newswires and press releases', 'company directories and profiles'. Every category was collected in a single document. At the beginning of each text were specified: byline, text length and the section of the newspaper (e.g. 'local/national news', 'editorial', 'opinion', 'letters')

from which the article came from. To these texts, I added a group of articles specifically taken from *New Scientist*, again using the Lexisnexis search engine, and several articles collected separately, from *Scientific American* and *The Economist*. The Lexisnexis classification, however, proved inconsistent and not very useful for the purpose of my analysis: it filed the same article in more than one category, some categories contained just one text, and it did not provide distinctions between local and national papers, nor between broadsheets and tabloids. These distinctions were impossible to introduce manually in a corpus of nearly 1000 texts; however, during the process of separation that had to be carried out manually to create individual articles from the larger documents, double occurrences were eliminated as accurately as possible and the classification was adjusted.

Therefore, after the separation, cleaning and adjustments, the corpus resulted as consisting of the following sections:

- Newspapers (698 texts); this includes texts from both paper and online resources, as they had been mixed in the Lexisnexis output files. Of these, 29 text files are specified as being web-based.

- Magazines and Journals (93 texts); this includes a very small group (6 texts) from magazines covering the fields of book reviews, legal news and commerce , plus the articles from *The Economist* (22 texts) and those from popular science magazines: *Scientific American* (30 texts) and *New Scientist* (35 texts).

- Letters (109 texts); despite the impossibility (and in certain cases the pointlessness) of dividing all articles according to the different sections they came from (as specified by Lexisnexis), it was chosen to create this one distinction, since it appears to be relevant in this context. Being generally written to the newspaper from people external to it, letters to the editor tend to have a different structure from other kinds of articles; moreover, their purpose is more oriented towards expressing the author's opinion and claims, often dealing with persuasion of the reader into agreeing with the message. Thus, all files explicitly marked as belonging to the 'Letters' section of their newspapers were separated from ordinary articles. This does not mean that the rest of the corpus does not contain articles written in order to express the author's perspective: the texts marked as 'Editorial' and 'Opinion' were left in the 'newspapers' section, and it

cannot be excluded that some of these (especially the 'Opinion' ones) were actually letters written to the editorial offices. The main reason for keeping this classification was that a careful scrutiny of each text would have been too time-consuming; furthermore, the expression of the author's viewpoint can be present at different levels in a text, and a clear boundary between the purposes of persuasion and conveying information cannot be drawn. Separating letters to the editor from the rest of the 'newspapers' material seemed however a quite 'safe' operation, as they distinguish themselves from the great majority of ordinary 'newspapers' texts, and generally show a different authorship.

- Company directories and profiles (3 texts); this includes a very short list of descriptions of a company, a magazine and a grass roots group all dealing with fracking activities or issues.

- Newswires and Press Releases (25 texts); these texts all came from the same source, what appears to be as a publisher of local monthly magazines, and were signed by institutions or citizens group dealing with the issue of fracking; some of these releases had the form of a letter, others of an argumentative text, others were more informative in style; they however did not seem to belong to the typical newspapers article category and therefore were excluded from it but kept while analysing the whole corpus.

1.6.2 Methods for corpus analysis and data interpretation: a first outline

I individuated two different phases in my analysis of the corpus: one focused on lexico-grammar, with particular attention to grammatical patterns and syntax; the other focused on lexical and semantic aspects. For what concerns the first part, the methodological approach which appeared to best respond to my research purpose was Biber's Multi-Dimensional Analysis (MD Analysis), providing that I could find a software capable of reproducing it. Not only does MD Analysis allow to analyse a large corpus, but it also encompasses a considerable amount of linguistic features to be detected and counted in the corpus, and organizes them into functional groups ('factors'), providing interpretive frameworks for each of them. Moreover, MD Analysis has been used to assess linguistic variation within the English language: therefore, the textual genres identified and analysed by Biber served as reference for my

own results. Further information on the differences internal to the corpus could then be derived from applying MD Analysis to the corpus sections. The starting point of the lexico-semantic part of the study were instead word frequency counts; these were calculated for the corpus and its sections, and then elaborated to extract several lexical information, including for instance the weight of single words or groups of words on the corpus, keywords, lexical variability. Some instances of metadiscourse as devised by Hyland were also included in the frequency counts. All methods adopted are thoroughly explained in their specific chapters.

Finally, the corpus, together with the analysis results and their interpretation, were contextualized in a sociological perspective; thus further – and hopefully more comprehensive – interpretations were proposed regarding the main trends and frames characterizing the representation of the controversy in the corpus.

1.7 Note on the limitations of this study

This is by no means a comprehensive and complete account of the controversy over hydraulic fracturing; rather, it analyses fracking through its media representation. This is in turn limited to a certain kind of mass media, and only refers to the Anglophone context. Another serious limitation, mainly due to time constraints, is the lack of diachronic and diatopic perspectives: thus the corpus is considered as a whole, regarding all involved countries and relative cultures together, and treating a five-year period as a moment, a sort of framed stage in the development of the controversy. Moreover, the literature research might be deemed poor, supported by few authors and fewer theories. I defend this work by considering it a first attempt at approaching a rather undocumented sector of corpus linguistics: the relatively innovative element of this research (the study of a public technoscientific[3] controversy in both linguistic and sociological terms) partly implied an autonomous process of analytic tool development. On the one hand, this went beyond what could be found in the studied literature; on the other hand it sometimes became central with respect to the research of analogous or

---

[3] In Science and Technology Studies, technoscience refers to the growing proximity of scientific research and technological application contexts (Bucchi and Lorenzet in Lorenzet, 2013). Furthermore, this term allows for an overcoming of stereotypical distinctions between science and technology (Neresini, 2011), which are closely connected and develop together.

related existing work. Finally, the large-scale, computer based approach here adopted clearly excludes in-depth verification of the communicative functions emerging from the analysis. Yet it allows to take into consideration a statistically relevant set of data, thus making it possible to discuss results whose validity can be much more extensive than those of individual text analyses.

Chapter 2

Multi-Dimensional Analysis

2.1 Choosing a suitable method of analysis

To analyse the collected articles according to my research purposes, I needed a method which could be applied to a relatively large corpus, and which provided me with reasonably complete lexico-grammatical information. I found that Douglas Biber's Multi-Dimensional (MD) Analysis could be a useful tool to study lexico-grammatical patterns present in the corpus and try to interpret them, also by comparing the corpus to some text genres and typologies already studied, labelled or commonly acknowledged in the English language.

In the study on which I based my work (Biber, 1988), MD Analysis was elaborated and used to assess linguistic variation among different genres, and particularly between spoken and written registers. The corpus used in Biber's study mostly contained texts in British English, but there were also some instances of texts by American and Canadian authors. The corpus was expressly collected to include texts produced in a wide range of contexts and with a wide range of purposes, in order to represent a sort of 'general sample' of spoken and written language. All texts were categorized according to different, commonly acknowledged 'genres' and sub-genres, which reflected to some extent the different contexts and purposes considered. Meanwhile, after a review of previous research, a set of linguistic features with relevant communicative functions was gathered (67 in total), ranging from lexical to grammatical and syntactical items; subsequently, frequency counts of each feature were carried out on each text of the corpus thanks to a proper tagging program, and, after a normalization of each text to a length of 1000 words, descriptive statistics for each feature (mean frequency, maximum and minimum occurrences in any text, range between maximum and minimum occurrence, and standard deviation) were computed on the whole corpus.

Yet, MD approach is based on the assumption that no single linguistic feature can account for complex communicative functions underlying linguistic variation; rather, this could be achieved by analysing the co-occurrence patterns among linguistic

features. Therefore, using a factor analysis statistical method, the relatively high number of variables counted (namely, the 67 linguistic features previously collected) were clustered into a reduced number of 'factors', groups of variables sharing strong co-occurrence patterns; after several statistical tests (Biber, 1988: 82-85) and a rotation routine, a 7-factor solution was selected; later in the study, however, only the first five factors appeared as both fully relevant and offering adequate chances of interpretation. For this reason, and since factor 6 seems to have scarce relevance for the topic on which I am researching, I will only use factor 1 to 5 in my analysis.

To understand the structure and nature of these factors, it will be useful to briefly describe how the patterns of co-occurrence, or the correlation among variables works in MD Analysis. A correlation among linguistic features can be either positive or negative. A large positive correlation means that the features in question systematically occur together, while a large negative correlation means that those features systematically co-vary in a complementary way. The same kind of pattern is indicated by factor weights or loadings, although they have no one-to-one correspondence with correlation coefficients: every linguistic feature receives a certain factor weight or loading for the factor(s) it is assigned to; roughly speaking, this means the degree to which a feature represents a factor. Factor loadings can be either positive or negative – meaning that features with a positive loading tend to systematically co-vary in a complementary fashion with features with a negative loading – but the sign does not influence the importance of a loading, which is instead expressed by its absolute value. Therefore, each factor is made to include all the features whose loading on that factor – positive or negative as it can be – is considered important enough (for the final MD Analysis, the cut-off was fixed at .35).

The communicative functions that each factor, as combination of linguistic features, is observed to convey are then assessed and used for an interpretation. Every factor is thus treated and defined as a Dimension of linguistic variation; it also can be said that different Dimensions of variation underlie different factors. Biber's Dimensions have therefore a continuous nature: the clusters of variables they represent can be present ad different degrees in texts, from the more negative-loading variables to the positive-loading ones. In fact, the name assigned to each Dimension reflects both its extremes of characterization. In Biber's study, after the extraction of the 7 factors from the

frequency counts of linguistic features for the whole corpus, the presence and influence of Dimensions on individual texts was measured through 'factor (or Dimension) scores', computed for each text separately to "characterize the text with respect to each factor" (Biber, 1988: 93). A factor score in a text is computed by summing the number of occurrences of the features having salient loadings on that factor. To calculate factor scores, all frequencies have to be standardized to a mean of 0.0 and a standard deviation of 1.0: this allows to "retain the range of variation for each linguistic feature while standardizing the absolute magnitudes of those frequencies to a single scale" (Biber, 1988: 94), in order to prevent very frequent features to have an inordinate influence on the computed score. Once each text had been assigned its factor score, average factor scores were computed for the texts within each genre section, and the different genres could be compared according to their Dimension scores, which allowed further interpretation of the textual Dimensions. Some genre categories had texts whose Dimension scores could differ considerably, while some texts from different genres were noticed to have similar scores. Finally, the existence of eight text types (comprising the same texts, written and spoken, used for MD Analysis) was suggested on the basis of a computerized clustering of Dimension scores of all texts(Biber, 1989). Therefore, while genre categories can be considerably heterogeneous with regards to Dimension scores, text-types can gather texts categorized in different genres, with very similar Dimension scores.

2.2 Summary of Biber's Dimensions of variation

I here try to summarize the main characteristics of the first five Dimensions of variation as described in Biber's *Variation across Speech and Writing*. For a brief description of genres, see Biber (1988: 65-71); for a description of text-types, see Biber (1989).

- Dimension 1 (D1) represents, at its positive extreme, an involved and non-informational focus, which is typical of a discourse whose primary purposes are interactive or affective and constrained by production circumstances, for instance a casual conversation. Features with positive loadings on factor 1 are, for

example: present tense, 'private verbs' (e.g. *think*, *feel*), first and second person pronouns, *that*-deletions (e.g. *I think [that] I'll go*), contractions, pro-verb *do*. These features can also be said to express a fragmented, reduced and generalized kind of content. On the contrary, the negative side of D1 is characterized by features such as nouns, word length, prepositional phrases, type/token ratio and attributive adjectives, all of which can be interpreted as sharing a high informational focus and providing the opportunity to integrate information and choose words precisely and accurately, as might be in academic prose. D1 has been given the interpretive label 'Informational versus Involved production'.

- At the 'positive' end of its continuum, Dimension 2 (D2) has features like past tense and perfect aspect verbs, third person personal pronouns, 'public verbs' (e.g. *report, say*), present participial subordinate clauses (e.g. <u>*Stuffing his mouth*</u> <u>*with cookies, Joe ran out the door*</u>) (Biber, 1988: 233) which can be interpreted as marking a narrative purpose. The two features having salient negative loadings on factor 2 are present tense and attributive adjectives. The former of the two quite intuitively indicates an opposition between reporting past events (narrative attitude) with past tense and perfect aspect verbs, and dealing with more immediate matters with present tense. The latter seems to reflect the use of elaborated nominal referents, which could here prove to be more frequent in non-narrative discourse. Thus, texts with a high positive score on this Dimension (for example, a typical novel) generally have a narrative focus, while texts with a low D2 score (for example, most press reviews) tend to be non-narrative. D2 has been given the label 'Narrative versus Non-narrative Concerns'.

- Dimension 3 (D3) marks the distinction between two different kinds of reference. Texts with a high D3 score tend to present a high frequency of three relative constructs: namely, WH relative clauses on object and subject position and pied-piping constructions (e.g. *the manner <u>in which</u> he was told*) (Biber, 1988: 235). Despite having functional differences among them, all these constructs can build an explicit and elaborated identification of referents in a text. In D3, they co-occur with phrasal coordination and nominalizations, which reflect an integrated and informational kind of communication, both aspects which are compatible with an explicit reference (as is seen, for example, in academic prose).

Conversely, texts with a low D3 score tend to have very low frequencies of the above mentioned features while presenting many adverbials, especially time and place adverbials, which are more frequently used for deixis than they are for internal text reference (e.g. *mentioned above*). This kind of reference (which can be found, for example in sport broadcast) is non- specific and situation dependent, so D3 was assigned the label 'Explicit versus Situation-Dependent Reference'.

- Dimension 4 (D4) has only features with salient positive loadings; some examples are prediction modals, suasive verbs (e.g. *insist, suggest, agree*), conditional subordination, necessity and possibility modals. All these features occurring together are suggested to convey an overtly persuasive and argumentative kind of discourse, where some point of view (be it the author's or not) is made explicit; a kind of text where this might occur is professional letters. Conversely, if a text has a low D4 score, it either does not have a persuasive focus, or it expresses persuasion in some different, more covert ways, not accounted for in this MD Analysis study. D4 has been labelled 'Overt Expression of Persuasion'.

- Dimension 5 (D5)appears to measure the level of abstractness of a text: its 'positive features' include adverbial subordinators, agentless passives and *by*-passives. The frequencies of these passive structures are made excluding the counts of another two 'positive features' for D5: adverbial past participial clauses (e.g. *Built in a single week, the house would stand for fifty years*) and past participial WHIZ deletions (or participial clauses functioning as reduced relatives, e.g. *the solution produced by this process*; both last examples taken by Biber, 1988: 233). All these passive constructions usually have abstract and/or technical content and a formal style; subordinators and other conjuncts apparently mark the complex logical relations typical of an abstract/technical/formal discourse, as might be academic prose. The feature with the greatest negative loading (slightly smaller than the cut-off limit) on this Dimension is type/token ratio, which might unexpectedly point to the relatively low lexical variety of highly abstract (and therefore probably technical or specialized) discourse, which repeatedly uses a small set of technical vocabulary

for the sake of precision and to avoid ambiguity. As can be imagined, two genres with low average D5 scores are fiction and face-to-face conversation. This Dimension has been thus labelled 'Abstract versus non-Abstract information'.

2.3 Applying Multi-Dimensional Analysis to a new corpus

My purpose is therefore to try to apply MD Analysis on the corpus of articles about fracking, to see how and how much each textual Dimension is present there. This of course requires using Biber's 1988 whole-corpus means and standard deviations as a reference at the stage of standardization, in order to make the 'fracking corpus' comparable to Biber's corpus, which virtually represents the general language. MD analysis as described in Biber's *Variation across speech and writing* was developed in the 1980s. It basically worked with: a tagging program – called the 'Biber tagger' – able to identify all the linguistic features used for this study within the computerized corpus of selected texts; a program which operated the frequency counts on the tagged texts; multivariate statistics programs (or packages) which allowed to carry out all the statistical descriptions, to extract the factors and calculate loadings and scores, to analyse the obtained results and to test and check the outcome of each of these routines. Apart from the Biber tagger and a reference to the SAS standard statistical package, no explicit reference is made as to the names of the programs used in the study.

Some thirty years later, this kind of devices has of course evolved significantly, but MD analysis of a particular corpus always requires some software for automatic tagging, frequency counts, and statistics (in this case, the tools for the extraction of factors and testing of statistical routines are not necessary, since Biber's factorial structure will be used). The first solution I thought about was to get the Biber tagger and use Professor Laurence Anthony's freeware concordance program AntConc 3.4.3, to then calculate loadings and other figures with Microsoft Windows Excel, which provides also some statistical functions. Unfortunately, the Biber tagger, which is owned by Northern Arizona University proved to be totally unavailable.

Therefore, I considered using a freeware simple POS (Part Of Speech) tagger and try to detect and count Biber's 67 Linguistic features through AntConc. The first tagging

attempt was done with Laurence Anthony's TagAnt, built on the TreeTagger, a tool developed by Helmut Schmid at the University of Stuttgart, whose accuracy is around 96%. TagAnt allowed me to tag each text file separately, which would generate individual-text results as in Biber, but would require filling a table with figures of all 67 linguistic features for all the 928 articles manually. The program can generate horizontally tagged texts, which seemed to be the most suitable form to use in the concordance program later. However, although TagAnt functioned correctly, there were problems for feature identification, once the tagged files were input in AntConc. This was due to the lack of options for the detection of a sequence of items decided by the user: for example, to look for perfect aspect verbs, the program should recognize a tag/tags indicating any voice of the verb *have*, followed by up to two optional tags indicating adverbs, in turn followed by the tag/tags indicating past participle verbs. Instead, AntConc only provides the possibility of setting one or more words as 'target' and one or more words as 'context word(s)'. Moreover, even trying using the 'target' and 'context' options did not yield appreciable results, since the program probably did not receive the right search input (probably my mistake, but still unresolved) or did not work well with tags. This made much more difficult the identification and counting of Biber's linguistic features, as they encompass and combine simple grammatical items, syntactic structures, lexical classes and punctuation. The same operation was attempted with another tagger, namely the CLAWS POS tagger for English (in its free online version, Free CLAWS WWW Tagger), developed at Lancaster University. Its accuracy is of 96-97% and it uses a tagset (here C7 was chosen) which is more complex and detailed than the TreeTagger. Free CLAWS WWW Tagger can tag up to 100000 words at a time. This meant that either the 928 texts had to be entered and tagged one at a time, or the whole corpus had to be unified and divided in parts of maximum 100000 words to be tagged more quickly, which would later imply either a new separation into individual articles or a MD analysis carried out on a unified corpus. Furthermore, even with this tagset, AntConc did not produce appreciable results.

A better solution was found in a program called MAT (standing for Multi-Dimensional Analysis Tagger) developed by Nini (2015), which is freely downloadable and is defined in its webpage and manual as "a replication of Douglas Biber's *Variation Across speech and Writing* tagger". MAT contains a tagging module and an analysing

module which operate on the text or corpus input by the user: the tagging module includes a first preliminary POS tagging based on the Stanford Tagger and a second stage which detects and tags the linguistic features used by Biber in MD Analysis; the analysing module calculates the Dimension scores of the text or corpus and compares it to the mean scores of some genres and text types as proposed by Biber (1988, 1989). Using the input file(s), the program thus generates: a Stanford-tagged preliminary version of the text or corpus; a MAT-tagged version with a tagset including Biber's 1988 linguistic features; three tab delimited files reporting the statistics needed to perform the Dimensions Analysis; graphs which plot the text or corpus with respect to each textual Dimension and compare it to the genres and text-types respectively studied and proposed by Biber.

MAT was tested for reliability on the LOB and the Brown corpus, both used in Biber's 1988 study; the figures it generated are reported on the MAT manual (Nini, 2014: 10-12,14-16) against those generated by the Biber tagger. In general, the results obtained are quite similar to Biber's, showing that MAT is reliable in replicating Biber's study. The most noticeable differences lie in the scores assigned to the Brown corpus (which however are said not to invalidate the general outcome of the test), and in D3 scores, which are inflated in MAT. This difference is due to statistical figures regarding adverbs, but no cause has been individuated so far. Consequently, D3 scores have to be treated carefully when using MAT. For what concerns the present study, after some informal tests on another corpus, the program appeared to be reliable enough in genre and text-type comparison and was therefore used to analyse the corpus of articles about fracking.

2.4 MAT MD-Analysis of the corpus

It was decided to carry out a MD analysis with MAT , first on the whole corpus, and then on three sections of it: 'Newspapers', 'Magazines and Journals', 'Letters'. In commenting the results obtained, several elements are taken into account: prior expectations, the comparison made by the program with Biber's text-types and genres, the distribution (expressed by standard deviations) of the scores of single texts around

the mean values for the corpus or section in question. As for the comparison with genres, MAT builds its parallel with a selected group of genres from Biber's study: 'Conversations', 'Broadcasts', 'Prepared speeches', 'Personal letters', 'General fiction', 'Press reportage', 'Academic prose', 'Official documents'. In this analysis, however, comparisons are mainly drawn with genres considered as close to or dealing with the corpus about fracking. Consequently, particular attention has been given to 'Press Reportage', as the corpus is made of press articles, and 'Academic Prose', since the debate over the use of fracking technologies involves public communication of science and technology; this, especially in cases of controversial issues, implies a complex interaction of the popular and popular science media with more specialized sectors of information and communication, such as those produced by the scientific community.

It will be useful to remember that Dimension scores express the continuous nature of Dimensions and can be roughly said to measure the difference between a particular text and an 'average language' represented by a general corpus of English texts : this means that scores which are far from a value of 0 are close to one of the continuum ends, and make the text more marked for one of the extremes of that Dimension; on the contrary, values closer to 0 make the text unmarked for that Dimension, that is the text is close to the average values of the sample representing the language in general. At the same time, if some text or corpus has a Dimension score which is similar to that of one of Biber's genres or sub-genres, this does not imply that the text or corpus in question is altogether similar and comparable to that genre; rather, the similarity is very much limited to the linguistic features involved in the Dimension considered.

2.5 MD Analysis results: whole corpus

Dimension 1: 'Involved versus Informational production'

D1	mean	minimum value	maximum value	range	standard deviation
CORPUS	-12,28	-28,68	15,88	44,56	6,28

The mean score of the corpus is -12, 28. This mirrors the markedly informational, non-involved focus of these texts, as was expected of a journalistic corpus. A brief

check of the texts with the highest and lowest D1 scores (both belonging to the 'Newspapers' section) confirmed that the tagger has been quite reliable and that the statistics reflect the use of linguistic features gathered in this Dimension. The 'lowest' article was scored -28, 68, while the highest was scored 15,88, thus the 44,56 range appears quite wide. Nevertheless, in comparing it to the genre ranges obtained by Biber, the difference in size between the corpora being analysed have to be taken into account. This corpus consists of 928 texts, and its range is indeed wider than most genre ranges in Biber (1988: 122-125), but these genre categories consist of a maximum of 80 texts(Biber, 1988: 67). Furthermore, the statistical measure of standard deviation for D1 is 6,28, meaning that 68% of the texts have a D1 score between -6 and -18,46. It could therefore be argued that the corpus is relatively homogeneous for D1.

Text-type comparison

The average D1 score of -12,28 is used by MAT for a comparison with Biber's (1989) text types; the type whose D1 characterization is closest to that of this corpus is 'General narrative exposition' (no exact score can be found in Biber's work, but according to his graphs it should be around -10). The distinctive characteristics of this text type are a marked informational focus (this aspect is in common with the corpus) and a moderate narrative concern (unlike the corpus, as will be explained later in the chapter) , where narration is mainly used as a part of the expository information being conveyed (Biber, 1989: 31). The cluster of texts representing this type is the most general and largest of all eight clusters in Biber's typology; it includes texts from many different genres, of which general fiction, biographies, academic prose, texts on religious subjects are just some examples; besides, 73% of the 'Press reportage' genre is included here, which is consistent with MAT results, for this corpus, for D1.

Genre comparison (see graph 2.1 in the appendix)

As for genre comparison, 'Academic prose' is the genre with the closest average D1 score (-14,9) to the average score of the corpus; 'Press reportage' is also quite close (-15,1). Therefore, of the three, 'Press reportage' results as having the most markedly informational and least involved focus with respect to academic prose; the corpus is

almost three points less informative, or more involved than 'Academic prose'. Although all three scores are markedly low and there is not an excessively wide gap between their average scores, the somehow less informative, more involved features of the corpus might be interpreted, in particular by considering the articles with a positive D1 score. Some of the linguistic features which, from the data provided by MAT, are classified as 'overused' and seem to increase individual D1 scores for the three most 'involved' articles are WH-clauses (similar to *that*-clauses and complementing verbs, e.g. *I believed what he told me*) , possibility modals, causative subordination, pro-verb *do*, direct WH- questions, pronoun *it*, amplifiers (e.g. *very*). The first three of them can be interpreted as indicating the development of a line of reasoning, as if some viewpoint were being explained. The other features can give some information about the style adopted in the explanation: it seems to consist in an interactive focus, almost conversational in that it uses some generalized features typical of time-constrained productions, as if an argument were taking place. Other linguistic features observed while skimming through these articles included present tense, first person pronouns, second person pronouns, *be* as main verb. The corpus indeed consists of articles about a highly controversial technology, with obvious reliance on topical events and situations (which might increase the use of present tense). Moreover, the controversy involves different stances being frequently expressed and/or reported, which might lead, in some cases, to a somehow more conversational and personal attitude in writing, in the cases where really neat positions are being explained in a more assertive and 'involved' way.

In spite of the 'involved' elements present, the corpus is mainly informational in focus: writing about a controversial technology whose impacts on the environment and on people's health is an object of debate, involves technological, scientific, legal, economic issues among others. This implies conveying (at least at a surface level) a high amount of information to describe the debate around gas extraction and energy production with this technique.

Dimension 2: 'Narrative versus Non-Narrative Concerns'

D2	mean	min	max	range	standard deviation
CORPUS	0,32	-7,38	16,88	24,26	3,61

The mean D2 score of the corpus is 0,32, making the corpus generally unmarked for this dimension. The articles with highest and lowest D2 scores were checked to see if the tagger and statistics had yielded reliable results. Of the texts which resulted as more narrative, some reflected a typical narrative style, with episodes being reported using past tense, perfect aspect verbs, public verbs. Some others do not show such a recognizable structure: there is a combination of descriptions of situations through present tense and past events told using perfect aspect and few past tenses. In my opinion (and according to the figures provided by MAT), what gave them high D2 scores were mainly present perfect verbs, public verbs (for reported speech) and synthetic negation, also appearing in phrases such as *NO fracking*, *No drilling* and similar slogans, employed in anti-fracking campaigns. As for the checked articles with the lowest D2 scores, their structure is undoubtedly non-narrative, mainly describing situations and processes with present tense. Overall, the corpus can be considered unmarked for D2; After MAT-generated results have been tested, narration might at most be slightly less present than what is detected by MAT scores. The range of the corpus for D2 is 24,26, the more 'narrative' text being scored 16,88 (higher than any text from Biber's corpus) and the less 'narrative' being marked -7,38 (lower than any text from Biber's corpus), with a standard deviation of 3,61. It can therefore be said that, in spite of few texts with extremely marked scores, more than half of them remains approximately between 4 and -3,3.

Text-type comparison

With respect to Biber's text types, two of them have a similarly unmarked, although negative, D2 score, namely 'Informational Interaction' and 'Intimate Interpersonal Interaction' , which broadly represent two kinds of conversations: the former aimed at conveying information, the second at maintaining interpersonal relationships. The corpus therefore seems to have an amount of narrative content more similar to that of these kind of exchanges than, for example, to that of 'General Narrative Exposition' type, which has a slightly more narrative score (around 2), or of 'Scientific exposition' and 'Learned Exposition' (which represents mainly humanities and social sciences

Academic prose, but also press reportage and other genres), which are both situated further below the zero line for their markedly non-narrative style.

Genre comparison (see graph 2.2 in the appendix)

Among the genres used for comparison by MAT, the ones with the average D2 scores closest to the corpus are 'Personal letters' (0,3) and, more relevantly, 'Press reportage' (0,4) which both (predictably, as explained above) have narrower ranges and standard deviations than the corpus does. Academic prose is less narrative, with a mean D2 score of -2,6. Consequently, the corpus is in line with Biber's values for the Press; it shows a quite balanced situation between the description of situations and processes (which might be thought of as the provision of more technical, legal or scientific information linked to hydraulic fracturing), and the narration of episodes (which might be thought of as the development and consequences upon the land and communities of the application of this technology). Both these ways of writing can be used in a controversy to support the validity or harmfulness of some technique: for example, descriptions can either advocate the damaging nature of fracking or its safety; episodes of communities suffering environmental damage after the installation of a well in their area can be used to protest against fracking, while events concerning the advantages brought by the production of low price energy thanks to fracking can be used to promote it. An analysis of the most frequent words and expressions in the corpus might help in the interpretation of such modalities. The values around the minimum and maximum D2 scores for the corpus show a greater variety in the amount of narrative content among the articles, with respect to the texts collected by Biber; however, this might be due to the higher number of texts involved.

Dimension 3: 'Explicit versus Situation-Dependent Reference'

D3	mean	min	max	range	standard deviation
CORPUS	4,97	-9,97	19,07	29.04	3,9

The corpus has a mean score of 4,97 for D3, that is, markedly explicit in reference. However, as tests to assess the reliability of MAT already indicate, the scores assigned by the program are higher than those assigned by the Biber tagger, so the results for the corpus should not be considered as completely comparable to Biber's D3 scores. The difference is generally between one and four points (Nini, 2014:10-16), thus, even so, the corpus score can supposedly be considered positive, although less marked. However, the brief preliminary examination of the articles distributed near the extreme values (-9,97 and 19,07) revealed, in my opinion, little connection between the D3 scores and the actual presence of either explicit or context-dependent reference in the texts. The articles with the highest scores are indeed likely to have been scored so because of their use of phrasal coordination (which, according to Biber's description (1988:245) has the function of integrating and expanding information), and because of a moderate frequency of nominalizations, but these do not seem to be sufficient, by themselves, to constitute explicit reference in a text. *Wh*-relative clauses, those with the clearest impact in creating explicit reference, are not present. The same applies for pied-piping constructions. Moreover, in spite of the absence of adverbs referring to an implicit context, items with explicit reference such as *the government* or *the country*, or such as proper nouns of people or places not identified by any integration, may in fact refer to local realities linked to the context in which the article is published, in the same way as the reference of adverbs can be situation-dependent. The texts which were assigned the lowest D3 scores (and should therefore have shown a substantially more situation-dependent reference) have a higher frequency of place and time adverbials; most of them actually refer to situations not specified in the articles; in some cases, however, words such as *beneath* or *underneath* in prepositional functions were tagged as adverbs. Finally, the adverb *underground*, appearing on average once in more or less 300 texts out of 928, should not be considered as an indicator of non-explicit reference, as it generally does not have a reference which changes according to the situation, at least when used in reference to oil and gas extraction.

Overall, what this part of MD analysis seems to better indicate is the presence of place and time adverbs in texts with low scores. However, the observations made above about texts with extremely different D3 scores does not reveal, in my opinion, relevant differences in explicitness or implicitness of reference. Besides, since the standard

deviation is of 3,90, the majority of texts is distributed within a much narrower range than the actual corpus range of 29,04. Considering also the difference between MAT scores and the scores published in Biber, it seemed wiser to avoid using D3 to evaluate reference in this corpus.

D4: 'Overt Expression of Persuasion'

D4	mean	min	max	range	standard deviation
CORPUS	1,27	-9,27	22,15	31,42	4,24

The mean score of the corpus is 1,27; this is not a marked result, but the fact that a journalistic corpus is averagely placed on the positive side of the continuum is telling of the use that can be done of overt persuasion when dealing with a controversial issue. The corpus does not appears to be extremely heterogeneous, with a range of 31,42 – the maximum score of 22,5 and minimum of -9,27 are respectively higher and lower than any of the values from Biber's analysis – and a standard deviation of 4,24; it is however less homogeneous than for D1. Some of the articles with the highest and lowest scores were checked to assess the presence and function of D4 linguistic features; the scores assigned appear to be reliable, as, in most cases, high D4 scores correspond to texts with markedly argumentative structures, mostly oriented to persuasion. It is important to note that these structures do not necessarily express the author's voice trying to influence the audience; overtly persuasive structures might also be included in a piece of reported speech where some stance is being communicated. Conversely, the articles with very low D4 scores completely lack the argumentative kind of discourse represented by the cluster of linguistic features in D4; as a result, their style is more assertive and factual. However, this by no means excludes a persuasive purpose or the expression of subjectivity; when present, they are simply communicated using different tools, which can make persuasion either covert, hidden by factual style, or more similar to an assertion of the author's voice as true.

Text-type comparison

For what concerns the comparison with Biber's text types, the positive D4 score of the corpus takes it very close to the two types called 'Imaginative narrative' (mostly fiction texts, but also prepared speeches, with a "mixing of narrative focus and involved presentation" (Biber,1989: 31)) and 'Intimate interpersonal interaction' (briefly described in the discussion of D2 scores). The 'General narrative exposition' type is situated below the 0 line, with a D4 score of approximately -1, marking a certain distance between the average amount of overt persuasion expressed in this text-type, largely consisting of press texts, and that present in the corpus. Predictably, the average score of the corpus is also more persuasive than 'Scientific exposition' and 'Learned exposition' (both described in the D2 discussion, and both situated below 'General narrative exposition' for D4), despite the technical and scientific issues involved in the representation of the controversy about hydraulic fracturing. This plainly marks a different approach to science and technology on the part of the press, if compared to specialized texts– especially during a controversy.

Genre comparison (see graph 2.3 in the appendix)

In the MAT graph plotting the position of the corpus with respect to Biber's genres, there is only one genre with a mean D4 score higher than the corpus , and it is 'Personal letters' (D4 score of 1,5). Yet, if all genres categorized in Biber's 1988 study have to be considered, there are several which show a more overtly persuasive style; among them 'Press Editorials' relevantly stands out with a score of 3,1. This does not diminish, however, the fact that the corpus has an averagely positive score, even with respect to the 'Press reportage' genre, which has a slightly negative D4 score (the mean is of-0,7). Academic prose also has a negative mean D4 score (-0,5) but shows a high internal variability, with a range of 24,6 and a standard deviation of 4,7 (and it is important to remember that 'Academic prose' consists of a group of only 80 texts, which marks its variability even more).

What emerges from the analysis of D4 is therefore a use of overtly persuasive linguistic devices which varies throughout the corpus, but is relatively frequent if compared to the press and academic genres studied by Biber. On the one hand, this is

was not expected of a corpus mainly composed of newspaper articles, where persuasion is generally thought not to show, if present. On the other hand, this corpus is focused on a controversial subject; the higher frequency of overt persuasion can be explained with the high density of stances overtly expressed and reported by the media. This is one of the features that distinguishes mass media from specialized texts, in the representation of debated subjects: stances and different perspectives sometimes are made more explicit, whereas in specialized texts overt subjectivity is less frequently admitted; moreover, in popular media, viewpoints are not always explained in technical or scientific terms, but other aspects often take part in the debate (among which the social, economic, moral consequences of the decisions which may be taken) (Bucchi, 2000, 2008). However, this overt and argumentative kind of persuasion is not the only way in which authors can express opinions; stance has clearly been detected even in texts with a low D4 score. It can be therefore argued that , in participating in the construction of public debate about fracking, newspapers provided mainly informational contents, but this information was complemented by a relatively frequent use of persuasion in its different forms, both reported and directly expressed by the author.

Dimension 5: 'Abstract versus Non-Abstract Information'

D5	mean	min	max	range	standard deviation
CORPUS	0,55	-3,92	11,04	14,96	2,71

The mean of the corpus is 0,55, meaning that the corpus is unmarked for D5. While checking the scores of the single articles, I noticed that 16 texts shared the same score (-3,92, which is also the lowest): since, however, it was not possible to know whether this was a program error or some coincidence, I decided to continue my analysis as if those texts were correctly assigned that score. The preliminary inspection of the articles found around the 'extreme' scores (-3,92 and 11,04) showed a quite clear distinction between articles with high and low D5 scores, especially concerning the use of conjuncts and adverbial subordinators; also passive structures, together with past participial clauses and WHIZ deletions only appear in texts with high scores. Overall, articles with largely positive scores show a moderate tendency towards passivization,

sometimes including agent deletion, but even more than that, they appear rather elaborate and formal in approaching their subject, generally showing a detached attitude. On the contrary, articles which received very low scores resulted in more direct styles and a simpler use of syntax and sentence construction in general. This led me to regard D5 in the first place as a good measure of formality and elaborateness and, in the second place, as a reasonably reliable measure of abstractness by means of passivization in this corpus. These features seem to be present in an average and unmarked way throughout the corpus: the range is the smallest of all Dimensions, and so is the standard deviation, resulting in a relatively homogeneous characterization for formality and elaborateness.

Text-type comparison

The text type closest to the corpus is 'General narrative exposition', but it has a negative D5 score (more or less -0,4); the only types on the positive part of the continuum, together with the corpus, are 'Scientific exposition' and 'Learned exposition'. As a consequence, even if unmarked for D5, the corpus appears to be slightly more formal in language – and therefore appears to get slightly closer to specialized prose – than a general corpus of informational and narrative texts would be.

Genre comparison (see graph 2.4 in the appendix)

In the genre comparison the corpus is very close to 'Press reportage' which has a mean score of 0,6, so that it seems to be in line with press articles in general. The corpus, however, with a maximum D5 value of 11,04, contains instances of texts much more formal than those contained in 'Press reportage' (whose most abstract/formal text reaches a score of 5,5). The mean score of 'Academic prose' lies well above that of the corpus and of 'Press reportage', at 5,5 – which seems somehow to reflect an average measure between those of 'Learned exposition' and 'Scientific Exposition' in the text-types graph. If the internal variation of the genres being compared to the corpus are considered, 'Press reportage' appears more homogeneous, with a range of 9,9 and a standard deviation of 2,4, while 'Academic prose' shows a much greater variation, with a range of 19,2 and a standard deviation of 4,8. Taking into account the different sizes

of the groups considered, 'Academic prose' (80 texts) results as the most heterogeneous, followed by 'Press reportage' (44 texts), while the corpus (928 texts) results as the most homogeneous.

The corpus thus appears to be unmarked for D5, which means that, in the great majority of texts, the register used is in an intermediate position between a formal and elaborate style and a plainer and more straightforward one. Moreover, the kind of abstractness obtained through passivization does not seem to be pervasive throughout the corpus. The average score of the corpus confirms the expectations in terms of genre comparison, since it is very similar to Biber's values regarding the Press. Therefore, although the subject of fracking is quite complex and has many technical and scientific implications, it can be said that the language used by these newspaper articles to represent it does not follow the approaches that specialized writing might adopt in such cases (see the levels of abstractness reached by the 'Scientific exposition' text type), but remains within the criteria used by the Press in general. It can be intuitively derived that technical elements have been considerably simplified in favour of other aspects of great social, political, economic, moral, legal interest.

2.6 MD Analysis results: sections of the corpus

In being analysed, the three sections 'Newspapers', 'Magazines and Journals' and 'Letters' will be taken together and considered with respect to each Dimension in order to make comparisons among them easier. The sections of the corpus will then be compared to Biber's genres. These sections have different sizes, both among them and with respect to Biber's genre-groups. Therefore, it will be useful to remember that: the section with predominant influence on the whole corpus is that of 'Newspapers', with 698 texts; 'Letters' , with 109 texts, make up for something more than one tenth of the corpus; 'Magazines and Journals', with 93 articles, are also approximately one tenth. For a greater clarity, from now on, the corpus sections will be written in lower case, while Biber's categories will be in upper case. Given the more specific nature of the groups of texts being analysed, and following criteria of closeness to the corpus in terms of genre, circumstances of production and purposes of production, it was chosen to

consider one more genre, 'PRESS EDITORIALS', and some sub-genres which formed part of 'PRESS REPORTAGE', 'ACADEMIC PROSE' and 'PRESS EDITORIALS: 'POLITICAL PRESS REPORTAGE', 'FINANCIAL PRESS REPORTAGE', 'SOCIETY PRESS REPORTAGE', 'PERSONAL EDITORIALS', 'LETTERS TO THE   EDITOR',   'NATURAL   SCIENCE   ACADEMIC   PROSE', 'POLITICS/EDUCATION ACADEMIC PROSE', 'TECHNOLOGY/ENGINEERING ACADEMIC PROSE'. Of course, their Dimension scores and statistical descriptions were all taken by Biber's study.

Dimension 1: 'Involved versus Informational production' – corpus sections

D1	mean	min	max	range	standard deviation
Newspapers	-12,41	-28,68	11,21	39,89	5,93
Mag&Jour	-13,8	-26,9	4,83	31,73	5,33
Letters	-8,77	-20,91	15,88	36,79	7,48
CORPUS	-12,28	-28,68	15,88	44,56	6,28

Predictably, letters to the editor proved to be, on average, more involved than the rest of the corpus − but always on the 'informational' side − while the articles from magazines (most of which are from popular science magazines) resulted as the most informative, although the difference with 'Newspapers' is small (just 1,39 points). The absence of a wide gap between the two (and the fact that both their mean D1 scores are higher than Biber's 'PRESS REPORTAGE' and 'ACADEMIC PROSE') might be explained by the controversial nature of the topic they deal with, as this may have(if slightly) increased the use of some 'involved' features, for example when a strong claim is made or reported, or when the interest of a whole community (be it local, national or cultural) is recalled through the use of inclusive first person pronouns. In the MAT-generated table reporting the standardized scores (z-scores) of all the relevant linguistic features counted, for instance, the standardized average frequency of first and second person pronouns for the corpus (which have a positive, that is 'involved' weight for D1) is slightly higher than that of third person pronouns (which have a negative D1 loading).This means that, in this corpus, first and second person pronouns − more

'involved' – tend to appear more frequently with respect to the average of the general corpus than the 'less involved' third person pronouns, which might have influenced the D1 scores of both corpus and sections. However, it is not possible to make a complete comparison between the different groups analysed, for every linguistic feature involved, because no standardized frequency is available for Biber's texts, so it is impossible to calculate which individual linguistic features determined the difference in D1 scores between them. In 'Letters', which is markedly more involved instead, the informational content is likely to have been partly omitted, or taken for granted, in favour of a more involved participation to the debate. For what concerns internal variability, these sections appear slightly more homogeneous than the general corpus, with the exception of 'Letters', with its high standard deviation (7,48); this might be caused by the fact that, presumably, this section is the least constrained in terms of writing rules and standards.

Genre and sub-genre comparison

Dimension 1: 'Involved vs Informational Production' sections and genres

max
min
▲ D1 average

Letters PRESS EDITORIALS Newspap Mag&Jour ACADEMIC PROSE PRESS REPORTAGE

Dimension 1: 'Involved vs Informational Production' sections and sub-genres

A chart showing vertical lines with blue triangle markers for various genres along the x-axis, with y-axis values ranging from -35 to 20.

X-axis categories (left to right):
- Letters
- LETTERS TO ED.
- PERSONAL EDITORIALS
- Newspap
- Mag&Jour
- TECH./ENGIN. ACADEMIC PROSE
- POLITICS/EDU. ACADEMIC PROSE
- SOCIETY PRESS REPORTAGE
- POLITICAL PRESS REPORTAGE
- FINANCIAL PRESS REPORTAGE
- NATURAL SCIENCE ACAD. PROSE

Apart from the differences between the more informational parts of the corpus and its closest genres, the average scores of the three sections do not seem to contradict the scores calculated for genres and sub-genres by Biber. 'PRESS EDITORIALS' has a

mean of -10 (1,23 points below 'Letters'); two of its sub-genres, 'LETTERS TO THE EDITOR' and 'PERSONAL EDITORIALS', have a mean score of respectively -9,9 and -11. All these groups involving a higher load of personal viewpoint expression are situated above the newspapers, magazines and academic articles. However, 'Letters' and Biber's 'LETTERS TO THE EDITOR' are situated above -10, while 'PERSONAL EDITORIALS' is on average less involved. For what concerns the more informational groups, 'Newspapers', 'Magazines and Journals' and Biber's 'ACADEMIC PROSE' have average D1 scores in the -10 /-15 bracket, and 'PRESS REPORTAGE' has a score of -15,1. Yet if Biber's genres are split into subgenres, the only group remaining in the -10 /-15 bracket together with 'Newspapers' and 'Magazines and Journals' is 'TECHNOLOGY/ENGINEERING ACADEMIC PROSE', while all the others have lower scores. This closeness in terms of informational density between the two sections and 'TECHNOLOGY/ENGINEERING ACADEMIC PROSE' might point to a connection these groups can have because of the kind of subject they cover – fracking is primarily an oil and gas extraction technology – even though, for many aspects, what can be a shared core subject is treated quite differently in popular and specialized media. The remaining sub-genres have mean D1 scores within the -15/-20 bracket, with 'NATURAL SCIENCE ACADEMIC PROSE' as the least 'involved'. The groups with higher internal variability (considering both range and standard deviation) are the corpus sections and 'ACADEMIC PROSE'. Finally, it is important to remember that, whatever the degree of involvement of the analysed texts, average D1 scores are all markedly below the 0 line, therefore all groups are more informational than 'involved'; however, the sections of the corpus are the only groups containing texts with positive D1 scores, which is telling of a higher level of involvement which can be reached in writing about a controversial issue.

Dimension 2: 'Narrative versus Non-Narrative Concerns' – corpus sections

D2	mean	min	max	range	standard deviation
Newspap	0,88	-7,38	16,88	24,26	3,56
Mag&Jour	-0,38	-5,75	9,76	15,51	2,81
Letters	-2,11	-7,19	11,27	18,46	3,34
CORPUS	0,32	-7,38	16,88	24,26	3,61

'Newspapers' and 'Magazines and Journals' maintain D2 scores near 0, which means they are unmarked for Dimension 2 (as the whole corpus is). There is, however, a small difference between the two: 'Newspapers' have a positive D2 score of 0,88, while 'Magazines and Journals' have a negative score of -0,38. It can thus be said that, in newspaper articles from this corpus, related episodes are (as expected) slightly more frequent than in magazine articles, which probably focus less on stories and give approximately the same space to narration and non-narration (in this case, the latter might be the description of fracking and its effects). 'Letters' have a mean score of -2,11 , the only section with a markedly negative score, which might be explained by the prominence given to comments in this kind of text. 'Newspapers' appear as the less homogeneous section, with the widest range (it contains both the article with the lowest and the highest D2 scores in the corpus) and highest standard deviation, but it is also the section with the highest number of texts.

Genre comparison

Dimension 2: 'Narrative vs Non-Narrative Concerns': sections and genres
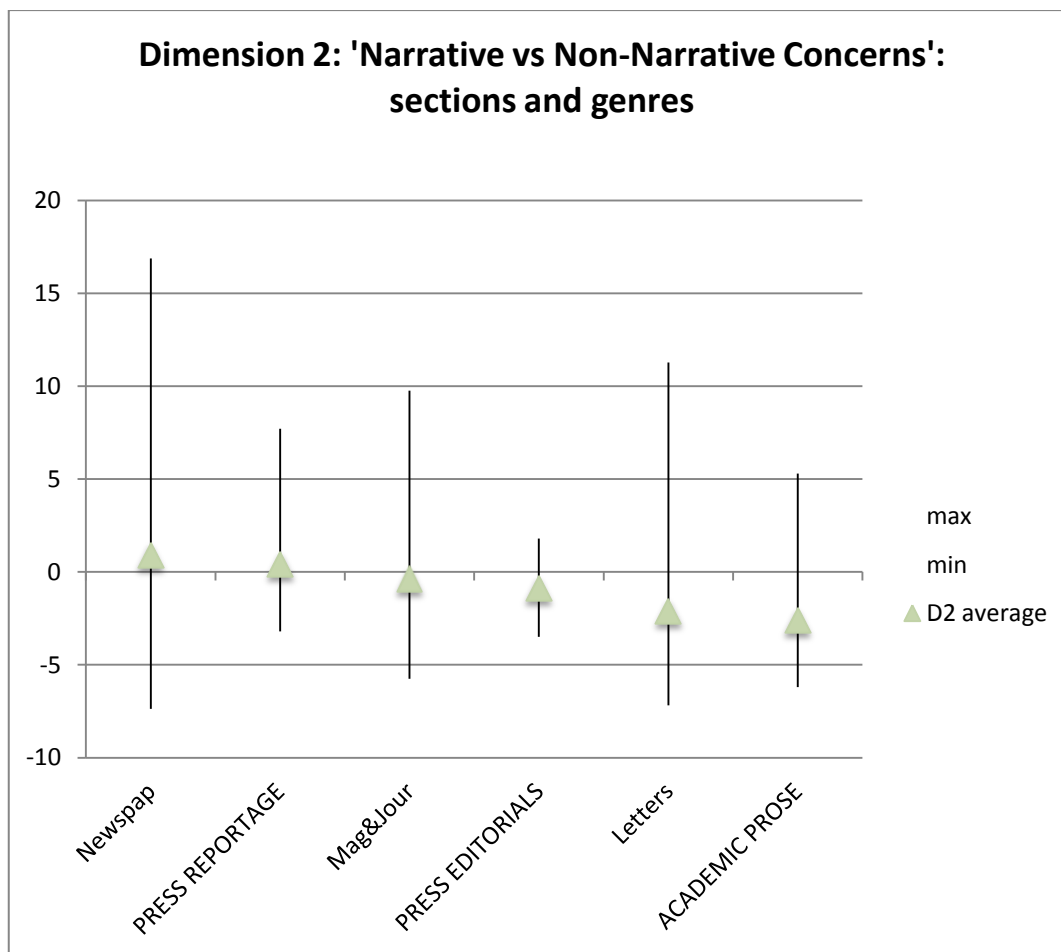
With respect to Biber's genres, 'PRESS REPORTAGE' (mean D2 score of 0,4) seems, on average, consistent with 'Newspapers', even though the former has got much more internal variability than the latter, at least partly explicable by their difference in size. 'ACADEMIC PROSE' , whose narrative load could be expected to be close to that of 'Magazines and Journals', has a mean D2 score of -2,6, much lower than 'Magazines and Journals'. This can lead to conclude that articles from popular science, or covering legal and economic issues in dedicated magazines did not approach the controversial subject of fracking in a purely technical way, comparable to what is generally done in specialized texts , but rather were more similar, in their use of narration, to the general trends in the popular press. In its non-narrative connotation, 'ACADEMIC PROSE' is closer to 'Letters', although the two have plausibly been assigned negative scores for different reasons: most likely, letters are non-narrative because their main focus is

commenting on episodes, facts and ideas, rather than reporting what happened (episodes are just mentioned or hinted to). Conversely, the genre that in Biber would correspond most to 'Letters', namely 'PRESS EDITORIALS', is much closer to being unmarked, but shows a negative average score of -0,9. This is probably because the texts collected in this 'PRESS EDITORIALS' groups referred to episodes in a slightly more extended way (this genre includes both letters and editorials, with the latter usually being fully developed articles, where relating episodes might have a different function).

Sub-genre comparison
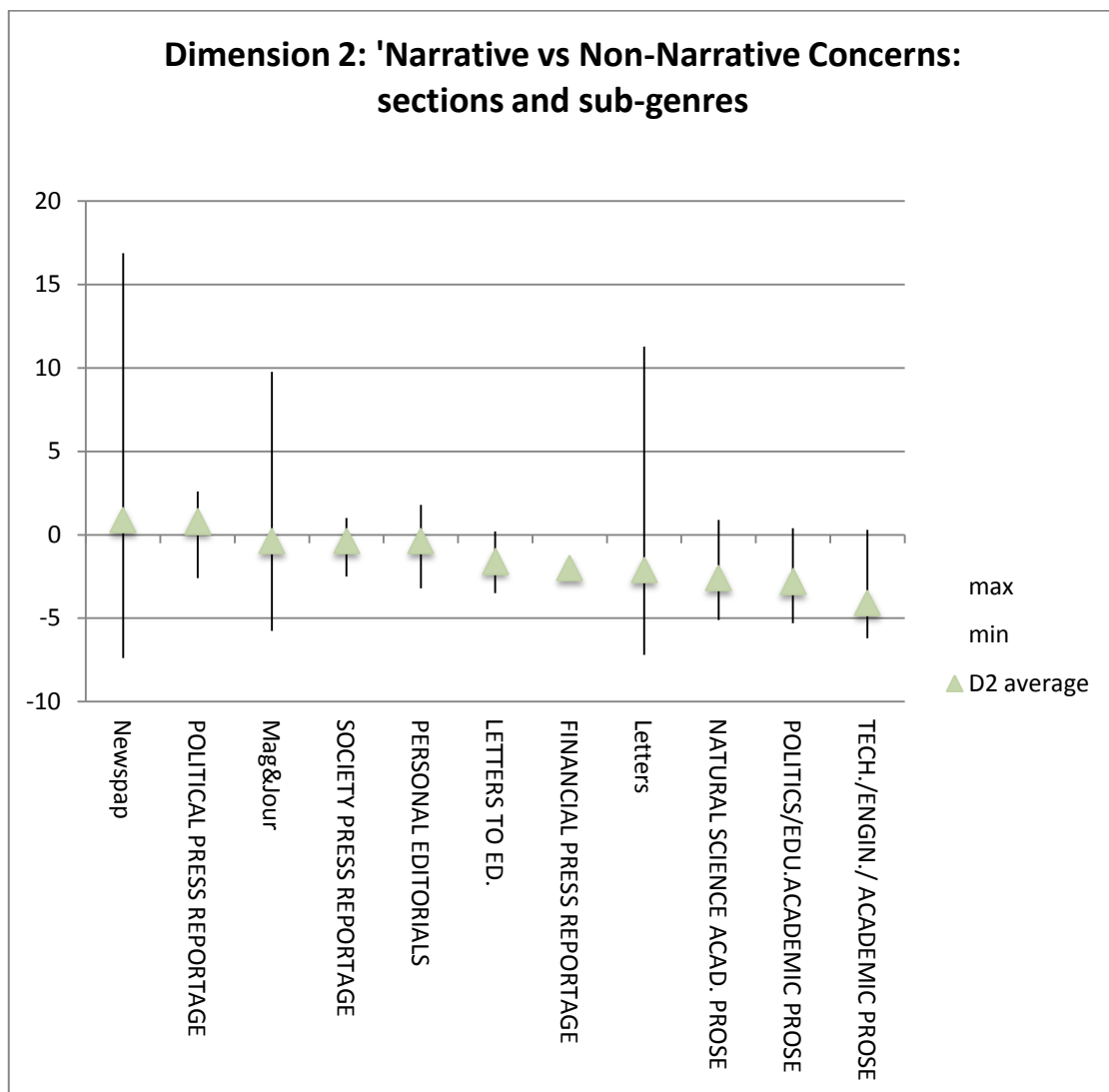
Sub-genres highlight the distinctions that can be drawn within genres with respect to D2. The only groups with positive average scores are 'Newspapers' and 'POLITICAL PRESS REPORTAGE', both near 1. 'Magazines and Journals', 'SOCIETY PRESS

REPORTAGE' and 'PERSONAL EDITORIALS' all have very similar scores (between -0,38 and -0,4). 'LETTERS TO THE EDITOR' is somewhat more marked (D2 score of -1,6) while the lowest scores were assigned to a heterogeneous group of categories covering the fields of financial economics ('FINANCIAL PRESS REPORTAGE') ,of comments (the already mentioned 'Letters') and of academic research ('NATURAL SCIENCE ACADEMIC PROSE', 'POLITICS/EDUCATION ACADEMIC PROSE', TECHNOLOGY/ENGINEERING ACADEMIC PROSE', which has the most markedly non-narrative score: -4,1). Thus the frequency of narrative devices, although not markedly high, is higher in the articles from 'Newspapers' and in those reporting about politics. One reason for this might be that both groups make a similar use of narration in representing debates and their development. The texts from magazines, the editorials collected in Biber's study and those texts which Biber categorizes as coming from the Press and dealing with society are unmarked for D2: narration appears balanced with what could be a description and characterization of the present (probably more frequent in magazines and 'society articles') or a comment (probably more frequent in 'PRESS EDITORIALS'). Instead, a predominance of comments and personal opinions with respect to narration can be found in the two groups containing letters to editors ('LETTERS TO THE EDITOR' and 'Letters'). What instead lowers the scores of the academic sub-genres is presumably a specialized and mainly expository kind of language, mostly written using present tense.

As mentioned above, D3 is not used to evaluate reference in this corpus, and therefore the scores elaborated by MAT are not used here to draw comparisons between corpus sections and Biber's genres and sub-genres.

D4: 'Overt Expression of Persuasion' – corpus sections

D4	mean	min	max	range	standard deviation
Newspap	1,38	-9,27	22,15	31,42	4,1
Mag&Jour	0,07	-6,59	10,53	17,12	3,43
Letters	2,05	-9,27	18,11	27,38	5,2
CORPUS	1,27	-9,27	22,15	31,42	4,24

The section of the corpus with the highest average frequency of overtly persuasive features is 'Letters' (D4 score of 2,05); the section with the lowest average D4 score is 'Magazines and Journals' (0,07), while 'Newspapers' (1,38) are close to the score of the whole corpus (it is also the most influential group on the corpus). These values could be expected. Firstly, 'Letters' contain texts whose authors' main purpose is to make their own opinion explicit and to give convincing reasons for it. Secondly, the articles from 'Magazines and Journals' are expected to be informative and factual in focus, and not to show overt persuasion (although it might always be present in other forms). Thirdly, in 'Newspapers', articles are expected to be mainly informative and focused on reporting episodes and current affairs; they can however include opinion articles or argumentative passages, or they can report overtly biased claims and statements. Concerning the homogeneity within the sections, those with higher internal variability are 'Letters' and 'Newspapers': it is interesting to note that the texts with the highest, but also the lowest D4 scores in the corpus are included in these two sections, in spite of their average values.

Genre comparison

Dimension 4: 'Overt Expression of Persuasion' sections and genres
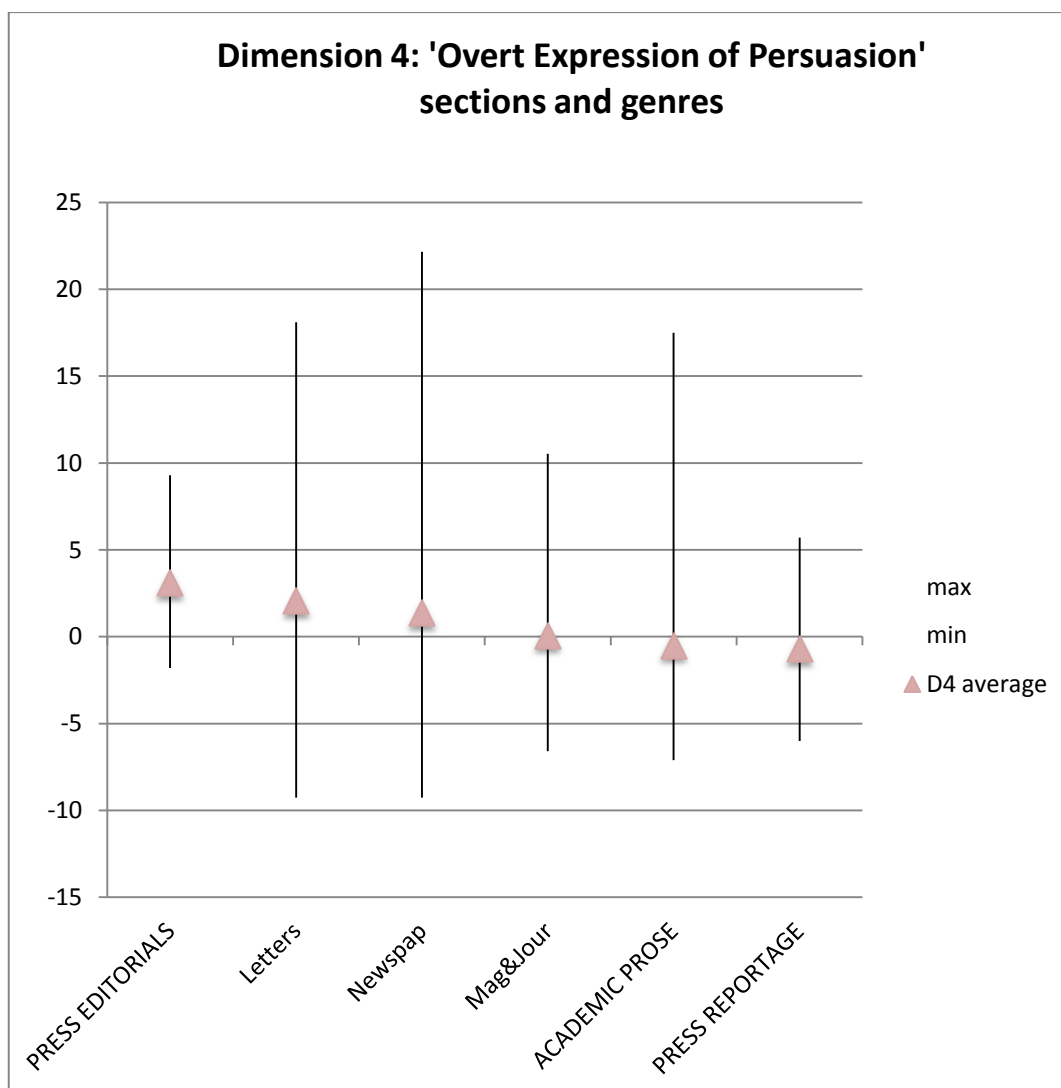
It is also interesting to plot the average scores of these sections with respect to Biber'
genres: while all the corpus sections (as analysed above) have positive scores, and so
does 'PRESS EDITORIALS', which is more overtly persuasive than 'Letters', Biber's
'ACADEMIC PROSE' and 'PRESS REPORTAGE' both have negative scores, albeit
very small in absolute value. As a result, 'Magazines and Journals', 'ACADEMIC
PROSE' and 'PRESS REPORTAGE' are on average unmarked for D4. Consequently, it
can be supposed that , while it is frequent in editorials and letters, overt argumentation
and stance explanation can also be found in magazines and newspaper articles (either
present or reported) and in academic texts (used in some academic fields to support a
thesis), but in these three groups its presence is not predominant. On the other hand,

newspapers from the corpus, with fracking as their subject, seem to show a more overtly persuasive attitude than their Biber counterpart (that is 'PRESS REPORTAGE') maybe because of the presence of opinion articles and editorials in the section, but probably also due to the controversial nature of the covered topic, which fosters a wider use of persuasion through overt linguistic tools in the public construction of the debate. In terms of internal variability, 'ACADEMIC PROSE' is close to 'Letters' and 'Newspapers'. The lack of uniformity of the 'ACADEMIC PROSE' genre will be more visible as its sub-genres, which have been assigned quite distant D4 scores from one another, are compared to the corpus.

Sub-genre comparison

Dimension 4: 'Overt Expression of Persuasion' sections and sub-genres
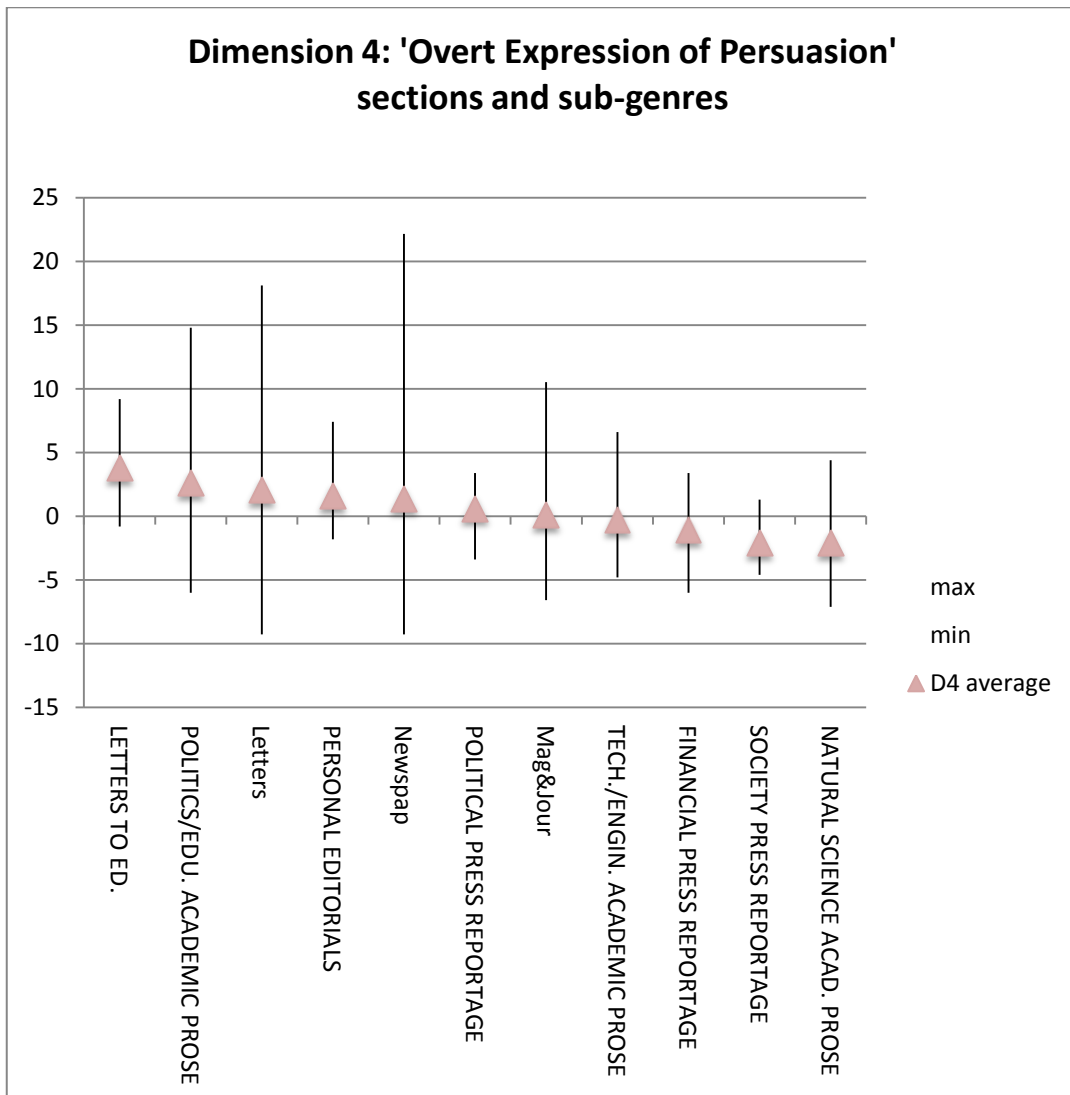
If Biber's sub-genres are considered, LETTERS TO THE EDITOR' has the highest mean score among all considered groups (3,8) and is surprisingly followed by 'POLITICS/EDUCATION ACADEMIC PROSE' (2,6), with 'Letters' (2,05) and 'PERSONAL EDITORIALS' (1,6) right below. This 0-4 bracket is completed by 'Newspapers' (1,38) and finally by 'POLITICAL PRESS REPORTAGE' and 'Magazines and Journals', which both have unmarked scores (respectively 0,6 and 0,07). As could be expected, the groups dealing with letters or editorials are situated on the positive side of D4 the continuum with relatively marked scores. 'Letters' are between 'LETTERS TO THE EDITOR' and 'PERSONAL EDITORIALS'; I would explain this location with the relatively more heterogeneous value of 'Letters' with respect to Biber's group, which makes the section a sort of intermediate group: not as persuasive as 'LETTERS TO THE EDITOR', nor as close to 0 as 'PERSONAL EDITORIALS'. 'POLITICS/EDUCATION ACADEMIC PROSE' is the academic sub-genre with the highest mean D4 score; its core subject might have influenced the presence of argumentation, even though it is actually impossible here to verify if such a high D4 score for an academic genre is due to argumentation as an expository tool to support some thesis or to the explicit representation of different stances (either directly expressed by the authors or reported by them). Thus, apart from groups dealing with letters or editorials, whose markedly persuasive nature has already been accounted for, a kind of parallel can be drawn between the other two sections and Biber's politics-centred sub-genres; their closeness in this context might be explained by the frequency of debate and/or argumentation shared by these groups. This could lead to think that the media have represented the issue of fracking as a controversy with a structure quite similar to that in which political debates are represented, at least for what concerns the presence of different opinions and claims.

On the negative size of the D4 continuum is located the rather unmarked 'TECHNOLOGY/ENGINEERING ACADEMIC PROSE' (-0,3) . The factor or factors preventing texts on highly specialized technical subjects from being markedly not persuasive (as would be expected) cannot be verified here: it is not clear whether these texts contain argumentative devices with an expository purpose or if overt persuasion has any particular function for them. 'TECHNOLOGY/ENGINEERING ACADEMIC PROSE' is followed by 'FINANCIAL PRESS REPORTAGE', with a mostly non-

overtly persuasive style (its score is -1,1), and 'SOCIETY PRESS REPORTAGE' with 'NATURAL SCIENCE ACADEMIC PROSE' (both -2,1). Therefore, in these low-score sub-genres overt persuasion tends not to be used, in favour of a more detached and unbiased style, probably with a more informative and descriptive focus and at least no evident persuasive intention on the part of the author. In comparing sections and sub-genres, it can finally be argued that, in terms of overt persuasion, the issue of hydraulic fracturing has been generally treated by popular media with a kind of political-debate approach, rather than with a more scientific, financial/economic or social approach.

Dimension 5: 'Abstract versus Non-Abstract Information'– corpus sections

D5	mean	min	max	range	standard deviation
Newspap	0,49	-3,92	11,04	14,96	2,7
Mag&Jour	1,1	-3,92	8,3	12,22	2,48
Letters	0,68	-3,41	9,55	12,96	2,95
CORPUS	0,55	-3,92	11,04	14,96	2,71

The corpus has a mean score of 0,55, which indicates that linguistic features revealing formality, elaborateness and passivization are used with an average frequency and balanced with features typical of a direct and syntactically simple style. Of the three sections, 'Magazines and Journals' is the one with the most marked score (1,1), resulting, in line with expectations, as the group of texts showing greater formality and complexity in form. The remaining sections have scores lower than 1; 'Newspapers' is the section with the lowest score (0,49), which mirrors the general tendency of the whole corpus and probably indicates that, with respect to magazine articles, newspaper articles follow rules of greater conciseness and directness in the representation of the controversy about hydraulic fracturing. 'Letters' are situated slightly above 'Newspapers', with a D5 score of 0,68. The minimal difference between the two might be caused by the greater freedom from genre restrictions which the authors of letters might have enjoyed; this could, in some cases, have allowed them to adopt a more complex style in dealing with this controversial issue. For what concerns internal variability, all of the three sections have a range approximately between 12 and 15 –

quite low with respect to the other Dimensions – with 'Newspapers' as the most similar to the general corpus. Yet it must not be forgotten that 'Letters' and 'Magazines and Journals' have a much smaller size than 'Newspapers' and their degree of internal variability has therefore to be considered higher. In this context, 'Letters' appear as particularly heterogeneous, with a range of 12,96 and the highest standard deviation of all groups including the corpus (2,95); this might be consistent with the fact that this is the less constrained section of the corpus with respect to genre rules and restrictions.

Genre comparison
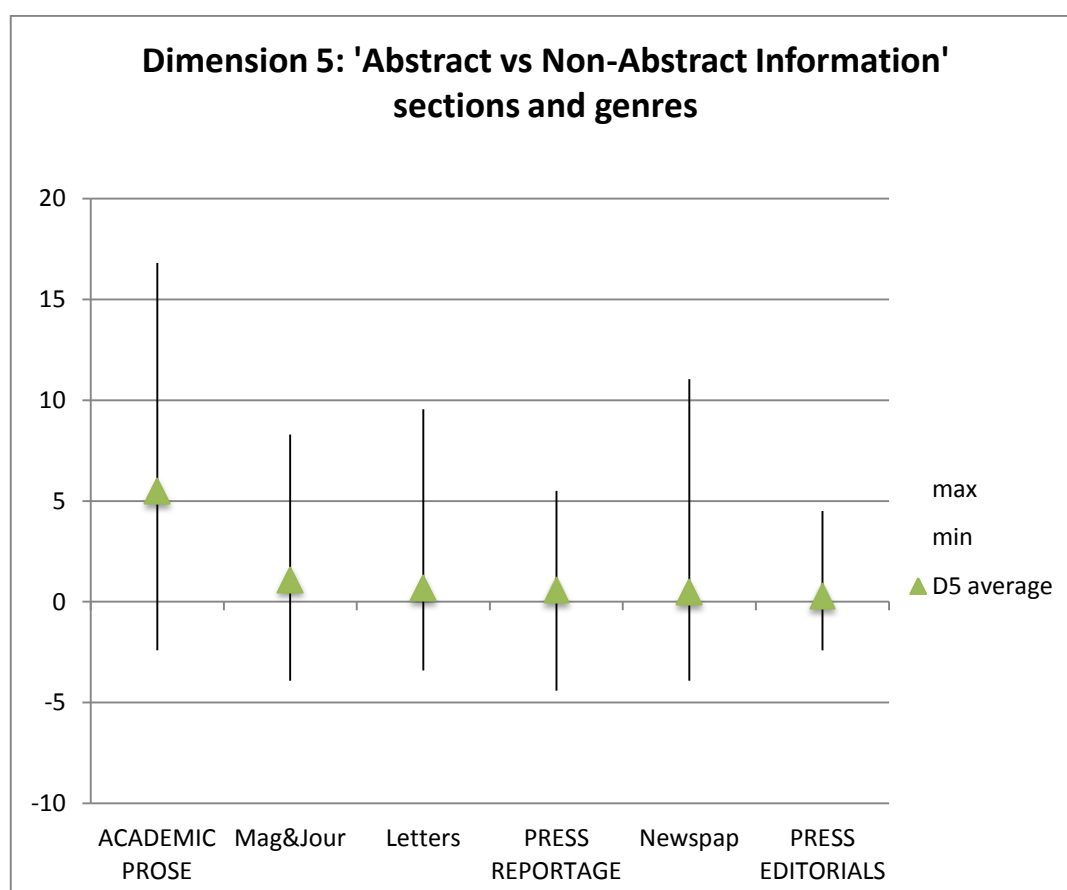
What emerges from the comparison with Biber's genres is the substantial difference between 'ACADEMIC PROSE', which is located in the markedly abstract/elaborate part of the D5 continuum (D5 score of 5,5) and the sections of the corpus together with the other genres considered: apart from 'Magazines and Journals' (1,1), all the other groups have a score between 0,68 and 0,3: all positive scores, but rather unmarked.

Thus, 'Magazines and Journals' use formality and elaborateness in a way which is more similar to that of press articles than to that of specialized studies, even though it has a D5 score higher than that of all the other groups. It can also be noticed that 'PRESS EDITORIALS' has the least elaborate/abstract score (0,3), unlike 'Letters'(0,69), which are situated above both 'Newspapers' (0,49) and 'PRESS REPORTAGE' (0,6). This might contradict the interpretation of the relatively higher-than-expected D5 score of 'Letters' in the previous paragraph. Yet it can also be caused by the fact that 'PRESS EDITORIALS' consists of different sub-genres, not all of which are comparable to 'Letters'. However, the analysis of subgenres (see next paragraph) continues to show a difference between the more specific sub-genre of 'LETTERS TO THE EDITOR' and 'Letters', thus revealing that this slight difference in formality/elaborateness might be caused by the topic of the texts, combined with the greater freedom from genre restrictions: fracking raises environmental, health and legal problems among others, and this might have led some authors to use a more complex structure in their letters.

Sub-genre comparison

Dimension 5: 'Abstract vs Non-Abstract Information' sections and sub-genres
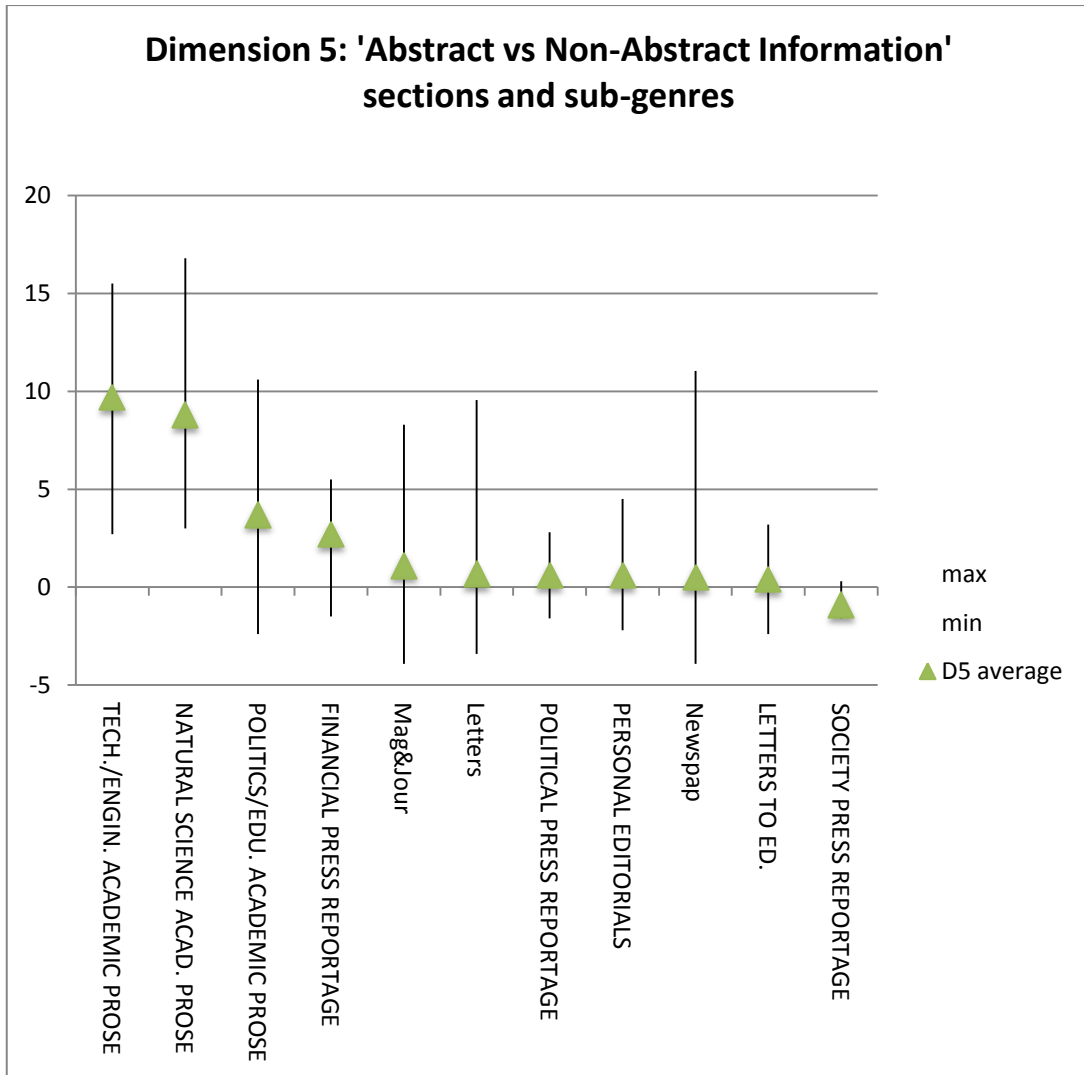
Subgenre comparison mirrors the plotting of genres, but at the same time reveals a more complex situation. The three sub-genres chosen from 'ACADEMIC PROSE' have the three highest D5 scores, but there is a wide gap between 'TECHNOLOGY/ENGINEERING ACADEMIC PROSE' together with 'NATURAL SCIENCE ACADEMIC PROSE', whose mean scores are near 10 and whose texts all have positive scores, and 'POLITICS/EDUCATION ACADEMIC PROSE', which contains also texts classified as non-abstract, and has an average score of 3,7 – nearer to 'Magazines and Journals' than to the other academic sub-genres. Although this sub-genre remains markedly abstract and elaborate (there is an important difference here between specialized and non-specialized writing), once again texts associated to politics

54

appear closer to the corpus than texts with a technological or scientific focus do. Right below the academic sub-genres comes 'FINANCIAL PRESS REPORTAGE'(2,7), the only Press-reportage sub-genre with a score higher than 1 in this group; its relatively elaborate and abstract language probably comes from the highly technical aspect of the field it covers. All the other sub-genres are located in the unmarked bracket between approximately 1 and -1.The sub-genre with the lowest score (-0,9) is 'SOCIETY PRESS REPORTAGE'; this is a small score in absolute terms, but this group of press articles centred around society is the only one on the negative side of the D5 continuum. Therefore, probably because of the field this sub-genre deals with, 'SOCIETY PRESS REPORTAGE' articles appear to be relatively less abstract/formal/elaborate than articles about politics and, in this case, articles about the controversy of fracking.

2.7 Conclusion

Applying Biber's MD analysis to the corpus helped to define some of its characteristics and to understand how the controversy of hydraulic fracturing was approached and represented in this sample of newspaper and magazine articles. In most cases, the corpus and its sections showed high levels of heterogeneity if compared to Biber's genres, with higher ranges, and with texts whose scores exceeded Biber's measurements, thus resulting anomalous with respect to their genre and to the rest of the corpus or section. This might indicate that the corpus has a more diversified nature, or might be due to the greater size of the corpus with respect to the corpus used by Biber in elaborating MD analysis.

As is revealed by its average D1 score, the corpus places its focus on conveying information. It should be remembered that the quality or quantity of such information cannot be assessed by MD analysis: rather, the analysis can reveal how frequently information is provided in general. On average, letters to the editor are the most involved, while the articles from 'Magazines and Journals' (mainly representing popular science) are the least involved. However, features indicating involvedness of language are on average more present in the corpus than in those texts collected by Biber to represent press reportage. Thus the controversial aspect linked to the hydraulic

fracturing issue might have affected the construction of facts, events and figures on the part of the media, in that the informational part continues to be predominant, but involved features are used slightly more often than usual to convey a message of lack of certainties and conflict between different voices and different attempts at defining the boundaries of the debate.

These representational choices seem to be supported by a mixture of narration and non-narration (mainly description or comment) which makes D2 scores quite unmarked, and is in line with Biber's values for the Press. Both narrative and non-narrative texts can in fact contribute to represent a controversy, and both aspects can coexist in the same text. On average, the least narrative section in the corpus is 'Letters' (probably because of a predominance of comment over narration), while 'Magazines and Journals' are unmarked – although on the non-narrative side of the continuum – and newspaper articles are the most narrative section. The corpus uses narration with a high frequency if compared to academic or specialized texts. This supports the assumption according to which narration is a communicative device normally employed by popular media in describing subjects with complex scientific and technical implications; the same subjects are plausibly treated in mainly non-narrative ways by specialized media.

Another important characteristic of the corpus is the high frequency of overt persuasion in relation to Biber's samples of both press reportage and academic articles. In my opinion, this might work alongside the already mentioned use of involved features in conveying the idea of conflict and controversy. Numerous actors have taken part in the debate about fracking, and so different competing stances are reported in the media, with each actor trying to support and strengthen their own message. One of the ways in which readers can be persuaded into agreeing with some stance is that of argumentation and overt persuasion. Thus the connection between this feature and the controversy being represented (either reported or with the author directly engaging in it) becomes quite clear. On average, letters to the editor predictably resulted as the most overtly persuasive section, while popular science articles resulted as the least overtly persuasive and newspaper articles was situated between them.

As could be expected of a corpus of press articles, the level of abstractness, complexity and formality of the language used is not high, on average. This does not imply that such a complex subject as the development and consequences of a gas and

oil extraction technology is always dealt with in extremely plain and simple terms: rather, the use of abstract or elaborate linguistic structures is not particularly high, and is probably balanced by less complex and elaborate structures. That is consistent with Biber's study, as both the corpus and his 'PRESS REPORTAGE' texts are averagely unmarked for D5. Academic texts resulted as much more elaborate/formal/abstract in language than the corpus, although, since no scientific articles about fracking have been analysed here, a direct comparison between different texts on the same subject is not possible. The popular science and magazine articles included in 'Magazines and Journals' proved to be slightly more elaborate/formal/abstract than the rest of the corpus, yet remaining near the unmarked area for D5. Popular science media are generally seen as channels of public communication of science and technology working as intermediaries between the scientific community and the lay public, which seems to be confirmed by these D5 values, not as low as those of newspapers, nor as high as those of specialized texts. If, however, this result is contextualized with the relatively high values of involvedness and overt persuasion of the whole corpus, including popular science texts, such intermediation appears somehow atypical, since it does not seem to present scientific and technological information as it is generally regarded, that is unbiased and unequivocal. Debate and uncertainty regarding the effects of fracking on people and the environment thus seems to affect the role of popular science, as well as the average degrees of involvedness and overt persuasion in newspapers.

Another interesting observation can be made about several similarities between the corpus (or its sections) and groups of texts dealing with politics: 'POLITICAL PRESS REPORTAGE' articles have a level of narration which is higher compared to other 'PRESS REPORTAGE' sub-genres, and thus similar to the 'Newspapers' section; 'POLITICS/EDUCATION ACADEMIC PROSE' and 'POLITICAL PRESS REPORTAGE' have levels of overt persuasion which are higher compared to other 'PRESS REPORTAGE' and 'ACADEMIC PROSE' sub-genres, and thus more similar to the corpus; in terms of abstractness and formality of language, 'POLITICS/EDUCATION ACADEMIC PROSE' proved to be less abstract/formal than the other academic subgenres, and thus closer (albeit still higher in D5 score) to the corpus, while 'POLITICAL PRESS REPORTAGE' showed an unmarked use of formality/elaborateness/abstractness similar to that of the corpus, differently from

'SOCIETY PRESS REPORTAGE', which resulted as less elaborate/ abstract. This possible connection between the representation of politics and that of a controversial technology is not analysed in detail here – it would require more specific research – but can be regarded as a suggestion of a certain approach on the part of popular media to such technical and complex issues; an approach which resembles the representation of a political arena rather than a technical, scientific, legal or economic treatment of a newly applied technology.

Chapter 3

Lexico-semantic analysis

3.1 Introduction

The second part of this analysis deals with the lexico-semantic aspect of the texts, differently from MD Analysis, which researches a combination of grammatical and lexical choices in meaning production. Here, instead, the choice of words in the corpus was analysed according to their frequency, both for single words, and for whole fields of related words.

In broad terms, the purpose of this lexico-semantic analysis is to understand, through a research into the contents expressed throughout the corpus:

- the concepts which have received particular attention;
- the associations that might have been created among them;
- the presence of different opinions about the debate over fracking;
- some clues to the level of directness and lexical variability of the corpus;
- some clues to the construction of any 'actor' taking part in the development of the controversy.

Also here, the distinction between various sections of the corpus was taken into consideration, to see if any differences in the choice and use of words could be detected and interpreted as reflecting different ways of approaching the same topic. In this part of the analysis, a classification prior to that used in the MD Analysis was employed, where the 'Newspapers' section was still separated from the 29 articles classified as 'Web-based publications'. Therefore, of the sections taken into consideration to compare the three main sub-genres present in the corpus, that of 'Newspapers' is slightly smaller than that used in the MD Analysis – 669 articles instead of 698 – but the difference, which accounts for around 4% of the 698-text section, should not have caused any sizeable alteration to the results.

3.2 Methods of analysis

Use of software

This part of corpus analysis was carried out using two concordance programs, namely Wordsmith Tools and AntConc. Their basic functions are the same – concordance, wordlists and keywords – but are performed in different ways, offering different additional figures and options to be set. Wordsmith is more sophisticated and allows a greater deal of figures to be considered. AntConc was used as a complement to integrate Wordsmith results, especially in the keyword computation, where the corpus was compared to two different reference corpora in the two programs.

What has been analysed

The first step towards a lexico-semantic analysis was creating wordlists where words are ranked according to their frequency; it was decided to do so for the whole corpus and for its three sections. Then a keyword lists were generated to see which words were characteristic of the corpus with respect to a reference corpus representing a general sample of the language. Afterwards, statistic measures of the lexical variability and level of directness were taken into consideration both for the corpus and for its sections. Finally, some classes of metadiscourse devices were counted in the corpus and sections.

Wordlists

To obtain results as relevant and accurate as possible, some preliminary processes had to be carried out:

- A general English stoplist was applied (Castello, 2008:58) to eliminate words which are generally extremely frequent but not relevant in terms of content; this list was edited by adding some words, specific to the corpus, whose extremely high number of occurrences was due to their usage by Lexisnexis in the initial description of every collected file; among these words are *length* (used to describe the length of each article in words), *words* (after the cardinal number in the length description), *byline* (of course present at the moment of giving the

60

article writer's name), *pg* (standing for *page*), *edition/editions* (often present next to the name of the source newspaper or magazines in phrases such as *online edition*), *news* (often used in describing the section of the newspaper from which the texts came from). After having taken as many as possible of these "deceptive" items off the wordlist, the actual trends in terms of word choice appeared clearer. In addition, attention was to be paid to some words whose frequency was altered because part of their occurrences did not have to be counted. One example was *new*, which at first sight appeared as one of the most frequent adjectives, but this figure was instead inflated because *new* forms part of some toponyms such as *New York*, *New Zealand*, *New Jersey*, *New Scientist*.

- A lemmatization list (Someya, 1998) was also used, in order to group the different inflected forms of a word and consider them as a single item; this gave priority to the content of words with respect to their different grammatical forms, as deemed adequate for a lexico-semantic analysis. The only item manually added to the lemmatization list was *fracking*, which included *frack, fracks* and *fracked* – neologisms deriving from the abbreviation of the expression *hydraulic fracturing*. I preferred to set the –ing form as the lemma, appearing in the main list, as this was the most important and the most used one, so much as to be treated as the exemplification of the whole class.

After the wordlists had been created, lists of the 100 most frequent words were extracted for corpus and sections. Each list contained, among other figures, the number of occurrences of each word and their relative frequency in terms of percentage. Thanks to this percentage frequencies, it was possible to make comparisons between corpus and sections and, above all, between different sections: this would yield results as for words appearing in different sections (some of which would probably not appear in all of them) and their respective frequency.

Every word belonging to these lists potentially reveals some piece of information regarding the way in which the controversy has been framed and presented. However, their role in the framing of the topic cannot be analysed individually, since words and their uses form a complex system of relationships, which constitutes the discourse characterizing the corpus; besides, they can be used with different meanings and functions on different occasions. Therefore, I tried to describe this complexity by

grouping words into some sort of categories or 'fields', each representing a different aspect of the controversy over fracking. To understand how present each of these fields was in every section, I had to find a way to compute their frequency with a method comparable among the sections. Therefore, prior to any analysis, the sum of occurrences of all the items in each 100 words list was computed list by list. Afterwards, these sums were turned into percentage numbers with respect to the total of running words of their respective sections. Since all three percentages were between 17% and 19% of the total of running words of their respective section, they were considered roughly comparable, because they represented similar portions of their sections. The weight and main characteristics of the individuated fields were analysed in each section list, by taking into account only those tokens included in the 100 most frequent words of each wordlist. More precisely, the percentage of the total occurrences of the words belonging to the same field was calculated against the sum of occurrences of all the items in each 100 words list. Then, other words considered relevant but not classified in any field were also analysed and compared among the sections. Possible relationships, associations and frequent collocations among words were searched for and reported.

Keywords and Weirdness

To complete the research of words and expressions peculiar to the corpus and/or to its sections, two Keyword lists were generated, to assign a degree of keyness (that is, of how characteristic a word is of the corpus it belongs to) to the words occurring in the corpus: one using Wordsmith, with the British National Corpus (BNC) wordlist as a reference; the second using AntConc, with the Corpus of Contemporary America (COCA) wordlist as a reference. Another perspective from which particularly frequent words in a corpus can be looked at is that of 'weirdness', the ratio between the percentage frequency of a word in the analysed corpus and the percentage frequency of the same word in a reference corpus. Consequently, it was decided to calculate weirdness on a reduced list of extremely frequent words in the corpus, as a further element to assess those words, called 'signature words' which are typical of the corpus because of their extraordinary frequency with respect to a reference corpus (Ahmad in Palumbo, Musacchio, 2010: 70).

Lexical variety and directness

Among the data automatically generated by Wordsmith, together with the frequency counts, is type/token ratio (TTR), a measure of lexical variability of which a standardized version (STTR) is also shown; to extract the STTR, the program considers a certain number of words from the beginning of each text (called the basis, which can be set by the user) and then computes an average among all their TTR. As the texts I am working with were on average quite short, and the program did not include texts smaller than the basis in the operation, I decided to set the STTR basis at 300 words. Within the same set of data, measures of average word and sentence length were also provided by the program and included in the analysis, as indicators of the level of directness vs elaborateness of the texts – to be combined with the data of MD Analysis about abstract and elaborate style.

Metadiscourse

The last part of the lexico-semantic analysis is about those self-reflective words and phrases which function beyond the propositional content of texts (without necessarily ruling it out) as tools in the hands of the writer to negotiate meanings with the reader, thus establish some textual relationship between the two as members of a particular community. Among others, Ken Hyland (2005: 37) defined as metadiscourse any linguistic device used to perform these functions. As metadiscourse is too complex and vast as a category for automated research, and metadiscourse devices cannot be included in a fixed set of clearly defined items, in his study Hyland proposes a classification of some words which very often work as pieces of metadiscourse, and describes each category with its own purposes. Researching the importance and frequency of use of these categories in the corpus can contribute to an understanding of several operations more or less subtly carried out by the authors in textually interacting with the public. Consequently, in my analysis, those categories with a particularly relevant relation to the corpus and its topic were transformed into lists and cleared of any item which could have caused ambiguities (mainly question marks, not accepted by AntConc, and imperative verbs, whose forms are identical to base forms and to nouns which are spelled the same as their corresponding verbs) and therefore was not suitable

for automatic research. Afterwards, the presence of each category in the corpus and its sections was assessed through frequency counts with the edited lists as search terms.

3.3 Results and Interpretation

3.3.1 Results and Interpretation: wordlists

Having created 100-words comparable frequency lists for the corpus and its sections (see tables 3.1.1-3.1.4 and 3.2 in the appendix), I identified three major groups which I have called 'fields' (in a broader meaning than that of semantic field), each describing a different aspect of the controversy over fracking:

- The first field includes words used to write about what happens when fracking technologies are applied: in other words, how fracking works. The analysis of this group of words can help to understand how the process of fracking is described and explained to an audience whose technical background concerning chemistry and oil or gas extraction is mostly poor. These are therefore the words which have become common when writing or talking about fracking, the expressions that the lay public can remember more easily – those most easily recalled when fracking is mentioned. I have called this field 'Fracking procedures'.

- The second field concerns the aspects which make hydraulic fracturing controversial, that is those expressions which refer to the possible negative impact of this technologies on the environment and on the communities affected by the proximity of shale gas or shale oil extraction facilities. I have called this field 'Controversial side'.

- In the third field, words referring to any person, group of people, institution or organization mentioned in the texts have been gathered, to outline the identification of actors involved in the controversy. This field was called 'Actors'.

As specified above, however, the discourse (or discourses) employed in the corpus consist of complex systems of choices and word use: this means that the fields I have created do not have to be considered any definite, absolute and totally objective entities;

moreover, their boundaries are blurred, and some words were assigned to more than one field, with their occurrences counted in both.

Results for 'Fracking procedures'

Three quite typical descriptions or definitions of fracking among those found in the corpus are reported below, to give an idea of the predominant lexical choices:

> Since August last year, the council has processed 13 resource consents for fracking. The practice, in which a mixture of chemicals and water is injected into the ground at high pressure to help release gas, is a major driver of a renewed oil and gas boom in Taranaki. (Taranaki Daily News - New Zealand, October 15, 2012)

> Fracking is where water and chemicals are injected into rocks at high pressure to extract gas from the cracks. (Lancaster Guardian December 20, 2012)

> Fracking, or hydraulic fracturing, is a method gas companies are using to release natural gas from the vast Marcellus Shale. A mixture of water, chemicals and sand is injected into the shale to fracture it. (Fracking, or hydraulic fracturing, is a method gas companies are using to release natural gas from the vast Marcellus Shale. A mixture of water, chemicals and sand is injected into the shale to fracture it. (Vindicator - Distributed by McClatchy-Tribune Business News October 20, 2011)

Words included in the 'Fracking procedures' field

The first word to be included in this field was quite intuitively *fracking*, together with its 'extended' counterpart *fracturing*, lemmatized by the software into the lemma '*fracture*', also included in the list because it always indicates the breaking of underground layers which characterizes fracking procedures. As a consequence, also *hydraulic*, always used in the expression *hydraulic fracturing* in the corpus, was counted. Of each item mentioned, all the singular and plural, or the different verbal forms possible are implied in the same place, due to the lemmatization process.

Other important components of this field are: *gas* and *oil*, sources of energy and object of fracking practices (it is possible that some of the occurrences of the plural *gases* was referred to greenhouse gases rather than natural gas; this number is however so low as not to alter the counts); *water*, used in the process, together with *chemicals* (appearing in the wordlist as *chemical*, together with its singular form); *use* in all its forms, a generic verb whose object are often the components employed in fracking; *rock*, a general term; the more specific *shale*, either pre-modifying the words *gas* and *oil*

or in head position; *drill*; *process*, generic term referring to the application of fracking technologies, as well as *operation*; *wells*; the occurrences of the adjective *natural* when modifying the noun *gas*; *methane*, more specific than *gas* or *natural gas*; *underground*; *deep*, mainly used in reference to what happens underground during or after fracking; *surface*, used to locate things and processes especially with reference to the last part of fracking practices; *release*, said mainly of fossil fuels getting out of the cracks opened in the shale; *pressure* and the occurrences of *high* modifying *pressure*; *extract*; *practice*, almost exclusively referred to fracking in this corpus; *involve*, which is in turn almost always used in the expression *fracking involves*, followed by a sort of explanation of the process; *exploration*, a preliminary procedure to fracking, included in this field as strongly associated to prospective hydraulic fracturing operations.

Newspapers-fracking procedures		
Word	**Freq.**	% in the section
FRACKING	6518	2,516
GAS	2244	0,866
WATER	1424	0,550
SHALE	850	0,328
OIL	849	0,328
USE	792	0,306
DRILL	785	0,303
FRACTURE	581	0,224
PROCESS	565	0,218
NATURAL (gas)	467	0,18
CHEMICAL	547	0,211
HYDRAULIC	466	0,180
ROCK	387	0,149
WELLS	381	0,147
RELEASE	327	0,126
UNDERGROUND	240	0,093
HIGH (pressure)	168	0,06
PRESSURE	239	0,092
EXTRACT	217	0,084
PRACTICE	216	0,083
INVOLVE	241	0,093
EXPLORATION	215	0,083
total occurrences	18719	7,2
% tot. occurrences of 100 words lst	36,9	

Table 3.3.1.'Fracking procedures' field in Newspapers

Magazines and Journals-fracking procedures		
Word	**Freq.**	% in the section
GAS	972	1,296
FRACKING	795	1,060
SHALE	405	0,540
OIL	348	0,464
WATER	329	0,464
DRILL	264	0,352
NATURAL (gas)	166	0,221
WELLS	191	0,255
USE	185	0,247
CHEMICAL	177	0,236
METHANE	160	0,213
FRACTURE	131	0,175
ROCK	123	0,164
HYDRAULIC	91	0,121
RELEASE	84	0,112
DEEP	83	0,111
OPERATION	77	0,103
FLUID	72	0,096
PROCESS	65	0,087
SURFACE	65	0,087
total occurrences	4783	6,4
% tot. occurrences of 100 words lst	36,4	

Table 3.3.2. 'Fracking procedures' field in Magazines and Journals

66

Letters-fracking procedures		
Word	**Freq.**	% in the section
FRACKING	922	2,730
WATER	238	0,705
GAS	230	0,681
USE	89	0,264
CHEMICAL	76	0,225
OIL	68	0,201
SHALE	62	0,184
DRILL	59	0,175
PROCESS	56	0,166
WELLS	50	0,148
NATURAL (gas)	36	0,107
FRACTURE	29	0,086
total occurrences	1915	5,7
% tot. occurrences of 100 words lst	30,9	

Table 3.3.3.'Fracking procedures' field in Letters

Word distribution

As shown in table 3.2 in the appendix, The section where *fracking* has the highest relative frequency is 'Letters' (2,73%), followed by 'Newspapers' (2,52%) and, at a relatively considerable distance, by 'Magazines and Journals' (1,1%). *Fracking* is the most frequent word (excluding those blocked by the stoplist) in the corpus, in 'Newspapers' and in 'Letters'. As for 'Magazines', the most frequent word is *gas*, which in turn has much lower relative frequencies in the other sections. Further difference can be noticed between *fracking* and *hydraulic fracturing*, respectively the synthetic, easy-to-remember word also used in slogans, versus the more precise and explicative term. Overall, and in all sections, *fracking* is far more frequent than *hydraulic fracturing*; however, the ratios between these two synonyms vary according to the sections: for the whole corpus, the occurrences of *hydraulic fracturing* are approximately 6% of those of *fracking*; the percentage rises by one point for 'Newspapers' and reaches 12% for 'Magazines and Journals', while Letters' is the section where *hydraulic fracturing* is the least frequent with respect to *fracking* (the percentage is 2%). This is also confirmed by the fact that *fracture* has a markedly

reduced presence in letters compared to other sections, while *hydraulic* does not even appear among the 100 most frequent words in 'Letters'.

Water is most frequently used in 'Letters', where its association with people's concerns for contamination deriving from fracking emerges with strength; its frequency remains high in 'Newspapers', while 'Magazines and Journals' insist less on this element. Due to this frequent proximity to words such as *contamination* and *pollution*, besides its function as a component of hydraulic fracturing operations, *water* was also included in the 'Controversial side' field.

The use of *shale*, a word coming from geology, in very frequent in all sections. It is however far more frequent in 'Magazines and Journals' than in any other section; the greater specificity of this term, and probably the fact that it is not as self-evident as *rock* nor as popular as *fracking* might have had a role in its distribution. Moreover, this is often a pre-modifier in expressions such as *shale gas* or *shale deposits* or *shale reserves*, besides being part of proper names like in *Marcellus Shale*; this lowers its frequency as an individual word, which indicates a particular type of stone. After a manual check of more or less a hundred concordances, it seems that explanations or definitions of *shale* are rare (the only one was found in an article from *New Scientist*), either because its meaning is (more or less justifiably) taken for granted, or because it is not considered relevant enough in the debate.

Overall, 'Magazines and Journals' as a section uses different words in a more even distribution than the other sections do, which can be thought to imply a slightly more 'pluralistic' use of words in describing fracking procedures. This can be observed for example in the top part of table 3.2, where the difference in frequency among the first words is more homogeneous than in other sections. Also exemplificative can be the fact that this is the section whose 100-word list contains the highest number of different words referring to the three main resources derived from fossil fuel (*oil*, *gas*, and *coal*) and to energy resources in general: of these *methane*, *coal* and *power* (in the corpus mostly used in its 'energy' meaning, also as a verb) do not appear in any other list; *fuel* only appears in 'Magazines and Journals' and 'Letters'. Another example is the generic term *process*: it is much less used in 'Magazines and Journals' than in the other sections; however, this seems to be compensated by *operation*, only present in the 'Magazines and Journals' 100-word list. The other word with a similar meaning,

practice, appears only in the 100-word list of 'Newspapers', but with a much lower frequency if compared both to *process* in all sections and *operation* in magazines.

In analysing the set of words used in the 'Letters' section for this field, on the contrary, a more basic vocabulary emerges, with respect to the other sections: from the absence of *hydraulic* in the 100-word list (reflecting the marginal role of the extended expression *hydraulic fracturing*) to that of *rock*, *release*, *pressure*, *underground*, *methane*, *operation*, *extract*, *involve*, *practice*, *exploration*, *deep*, *surface*, *fluid*. It is important to note that the absence of these words from the analysed 100-word piece of the general wordlist does not imply their absence from the general wordlists altogether: rather, it has a very low frequency, in any case lower than that of the words included in the analysed 100-word lists.

Overall, as clarified in tables 3.3.1, 3.3.2 and 3.3.3, the 'Fracking procedures' field accounts for 36,9 % of the total occurrences of the 100 most frequent words in 'Newspapers', for 36,4% in 'Magazines and Journals', and for 30,9% in 'Letters'. Consequently, it can be said that both newspapers and magazine articles definitely give considerable attention in terms of space to the explanation and description of fracking, but while magazines allows for a wider range of words and concepts around the central topic of drilling for shale gas, newspapers seem to insist on a narrower set of sub-topics, of which one example is the continuous repetition of *fracking*, resulting in my opinion much more popular, mysterious, new and impressive than the longer, technical-sounding *hydraulic fracturing*. In letters to the editor, on the other hand, the explanation and description of fracking has a smaller weight, although still relevant, on the whole set of analysed words. This may be due to the scarce interest, on the part of the writers, in informing the audience; rather, procedures and definitions are probably taken for granted, while the main purpose is to express and argue for one's own opinions and claim one's own position in the debate; this aspect is also confirmed by the low informative focus revealed for these texts by the MD Analysis.

Results for 'Controversial side'

There are several reasons for people's concerns for and opposition to fracking. One is the contamination which affects underground water resources after water, sand and

chemicals (also called *fracking fluids* in the corpus) are pumped into the drilled shale and can come into contact with aquifers used for human consumption (see chapter 1); it is possible that these fluids stream through leaks to the surface after being used, thus contaminating the whole surrounding environment. That regarding the chemical substances employed together with water and sand in the fracking process is a kind of 'debate- within -the –debate'. Firstly, it is not clear which chemicals are used for fracturing purposes, and drilling companies are reluctant to disclose information about what is exactly included in fracking fluids. Secondly, according to the opponents of fracking, these fluids are toxic and highly dangerous. Another controversial aspect are gas leakages, both in the air (therefore potentially causing explosions) and in the water coming from aquifers (therefore potentially poisoning water). Moreover, the minor explosions happening when high pressure liquids are pumped into the shale are suspected to cause tremors and earthquakes. A further argument put forward by environmental groups is the role of fracking with regards to climate change, it being a natural gas extraction technique, alleged to make available huge quantities of fossil fuels and thus boosting carbon emissions and dramatically lowering the interest in non-polluting or renewable forms of energy together with the financial support from governments.

Words included in the 'Controversial side' field

Therefore, the following words were included in the 'Controversial side' field: *fracking*, as it was used both in describing how the technology works and in writing about its consequences; *water* and *groundwater*, whose contamination is very frequently mentioned; *air*, also felt as an element which could be contaminated by gas leakage from fracking; *chemical*, both a component of the 'Fracking procedures' and the 'Controversial side' fields for the reasons stated above; *environmental*, when not part of proper nouns such as *Environmental Protection* (*Agency* or *Department*), as it is otherwise mostly associated with *damage, effect, concern* (also included in the list), and the like; *risk*; *impact*; *toxic*; *ban* and *moratorium*, often advocated and sometimes imposed by governments on drilling companies to prevent fracking operations considered dangerous; *cause*, almost always associated to negative consequences of

fracking, whose most frequent objects when used in its verbal form are *tremor*, *damage* and *earthquake* (also part of this field); *health*, heavily affected by fracking according to its opponents; *anti*, always used as a prefix in *anti-fracking* when writing about activist, groups and campaigns acting within the controversy; *waste*, here indicating the waste deriving from shale gas extraction – a factor of environmental contamination and something which should be disposed of; the noun phrase *climate change*; *underground*, as what happens underground is the object of both explanations and concern; *controversial*; *contamination* and *contaminate*; *protect*, which mostly has *health*, *environment*, *people*, *kids* and *natural beauty* as its objects and is used in supporting the anti-fracking cause.

Newspapers-controversial side		
Word	**Freq.**	% in the section
FRACKING	6518	2,516
WATER	1424	0,550
ENVIRONMENTAL	444	0,231
CHEMICAL	547	0,211
CONCERN	467	0,180
BAN	451	0,174
ENVIRONMENT	392	0,151
ANTI	345	0,133
CAUSE	327	0,126
IMPACT	299	0,115
RISK	285	0,110
MORATORIUM	282	0,109
HEALTH	256	0,099
PUBLIC (health)	45	0,017
EARTHQUAKE	254	0,098
UNDERGROUND	240	0,093
POTENTIAL (impact, risks, contamination, dangers)	74	0,028
CONTROVERSIAL	237	0,091
total occurrences	12887	4,9
% tot. occurrences of 100 words lst	25,4	

Table3.3.4.'Controversial side' field in Newspapers

Magazines and Journals-controversial side		
Word	**Freq.**	% in the section
FRACKING	795	1,060
WATER	329	0,464
CHEMICAL	177	0,236
METHANE	160	0,213
ENVIRONMENTAL	111	0,148
HEALTH	90	0,120
CONTAMINATION	80	0,107
CAUSE	78	0,104
(climate) CHANGE	39	0,052
GROUNDWATER	76	0,101
RISK	76	0,101
IMPACT	72	0,096
DRINK (associated to water, polluted water)	63	0,084
PUBLIC (health)	20	0,026
total occurrences	2166	2,9
% tot. occurrences of 100 words lst	16,2	

Table 3.3.5.'Controversial side' field in Newspapers

Letters-controversial side		
Word	**Freq.**	% in the section
FRACKING	922	2,730
WATER	238	0,705
CHEMICAL	76	0,225
BAN	59	0,175
CAUSE	59	0,175
ENVIRONMENTAL	57	0,169
CONCERN	53	0,157
HEALTH	53	0,157
ENVIRONMENT	49	0,145
WASTE	46	0,136
GROUNDWATER	43	0,127
RISK	41	0,121
CONTAMINATION	40	0,118
TOXIC	39	0,115
PROTECT	35	0,104
AIR	32	0,095
CONTAMINATE	32	0,095
(climate) CHANGE	13	0,038
total occurrences	1887	5,6
% tot. occurrences of 100 words lst	30,4	

Table3.3.6.'Controversial side' field in Letters

Apart from the differences in the use of *fracking*, already discussed above since it was considered also part of the 'Fracking Procedures' field, there are other differences among the sections of the corpus, of which those which appeared more relevant are here reported.

Word distribution-Water

Water, already briefly analysed, is given great importance in this context. Its presence as a word is particularly relevant in 'Letters' (0,7% of the total running words of the section), slightly lesser in 'Newspapers' (0,5%) and in 'Magazines and Journals'(0,43%). Furthermore, more than half of the occurrences of the word *supply*, in 'Letters' is part of *water supply*. My interpretation for this frequency distribution is influenced by the strong connection emerging in these texts between water and human activities and life; here, water usage represents the most immediate concern, from a

human point of view, because any negative consequence of fracking proved to first be felt through water. The higher relevance of water in the letters section seems to point to the centrality of a feeling of concern towards the potential dangers brought by fracking; these texts can be thought of as a more direct expression of people's reaction to the newly announced technology, compared to news stories created in newspapers or popular science articles. As those who write letters to newspapers are mostly those somehow affected or concerned in some way by the application of fracking, it is plausible to believe that most of these letters mention the possible effects of fracking on people (independently from the author's positive or negative attitude towards it, which cannot be assessed text by text in this kind of analysis).

Word distribution-Other words

Eighteen different words, among the 100 most frequent in 'Letters', are part of this field; 'Newspaper' have sixteen and 'Magazines and Journals' have twelve. Among the 100 most frequent words in 'Magazines and Journals' , words like *concern*, *ban*, *moratorium*, *anti*, *toxic*, *protect,* reflecting the immediate reactions to a potentially harmful innovation, cannot be found. Rather, the 100-word list of 'Magazines and Journals' seems to deal in a detached way with water contamination, the chemicals contained in fracking fluids, the causal link between fracking and the risks its negative effects pose, also on human health; it is also the category where *climate change* has its highest relative frequency, which is telling of a more generalized approach to the topic overall.

It is interesting to notice, here and elsewhere in the corpus, how *climate change* and *global warming*, which commonly refer to the same phenomenon, have asymmetrical uses in the corpus, in that *climate change* is much more frequent than *global warming*. In the whole corpus, the former is more than four times more frequent than the latter (see table 3.3.6b). Research conducted about the different perception of these two expressions and their use in political campaigns[4] revealed that while *global warming* was found capable of triggering stronger emotional reactions and evoke catastrophic outcomes perceived as threatening, *climate change* was found to be less engaging, due

---

[4] Reported in *The Guardian*, May 27 2014, online edition; retrieved May 2014 from www.theguardian.com/environment/2014/may/27/americans-climate-change-global-warming-yale-report

to its more neutral connotation which weakened risk perceptions. 'Newspapers' is the section where the largest difference can be noticed; the other two sections show a more balanced situation, with 'Magazines and Journal' as the section with the smallest difference (still though, with *climate change* more than twice more frequent than *global warming*). It cannot be said whether this asymmetrical use results from precise strategic choices, or *climate change* simply prevails for its authoritative origins (it is the formulation used by specialists). The final effect, though, might, according to that research, be that of minimizing the extent of potential damages brought by fracking on the long run.

	Corpus		Newspapers		Magazines and Journals		Letters	
	Raw frequency	Relative % frequency	Raw frequency	Relative % frequency	Raw frequency	Relative % frequency	Raw frequency	Relative % frequency
climate change	163	0,04	102	0,03	39	0,05	13	0,03
global warming	35	0,008	13	0,005	17	0,022	4	0,011
climate change rawfrequency/*global warming* raw frequency	4,65		7,84		2,29		3,25	

Table 3.3.6b Distribution of climate change vs global warming in the corpus and sections

The 'Newspaper' articles 100-word list covers a relatively high number of the controversial themes – one example of this is the frequency of *earthquake*, appearing only in this list –, with the exception of the concept of contamination and the climate change and fracking waste issues. Neither does 'Letters' have the climate change issue among its main topics (although *change* appears in the 100-word list), but it does otherwise cover a relatively wide range of 'Controversial sides', from the waste disposal, to the toxicity of chemicals used, to air contamination. This seems to support the above formulated suggestion that 'Letters' gives a greater attention to the immediate consequences of fracking on people than other sections do. Conversely, the section with the least space given to this part of the controversy is 'Magazines and Journals', whose articles seem focused on describing what fracking is and what it implies, and seem to approach its controversial side in a more detached way, rather than calling attention to the direct reporting of local communities' difficulties, an aim which instead seems to be more present in newspaper articles. In all sections, finally, the frequency of the verb

cause might be telling of a quite high level of consciousness/awareness regarding the correlation (either affirmed or negated) of fracking with the environmental damage occurring where fracking is applied. As showed in tables 3.3.4, 3.3.5 and 3.3.6, the percentage weight of this field on the total occurrences of each 100-word list is 25,4% for 'Newspapers', 16,2% for 'Magazines and Journals' and 30,4% for 'Letters', again confirming the hypothesis characterizing 'Letters' as the section which most emphasizes this aspect of fracking with a focus on immediate particular concerns, followed by 'Newspapers' which can be said to approach these aspects in a less rich and involved way, and characterizing instead 'Magazines and Journals' as the least lexically and grammatically involved section, which treats the controversial aspects of fracking in a more generalized way.

Results for 'Actors'

Words included in the 'Actors' field

This field was made to include more or less any mention of people's presence and actions within the controversy. Its components are: institutions like *state*, *government*, *country* (of which *UK*, *America* and *California* are the only proper nouns appearing in these lists), *university*, *agency* and *department* – often part of their proper nouns, as in *Environmental Protection Agency*, whose acronym *EPA* is part of the lists as well; *company* and *industry*; *people*, *population*, *resident* and the noun *public* (whose frequency as a noun has been counted with the help of the TagAnt tagging program; *council*, *community*, *county* (plus *Butte*, name of a Californian county where fracking is practiced), and *local* when used as pre-modifier of *community*, *council*, *government*; *group*, often referred to groups of anti-fracking activists; *green*, when used to mention *Green Parties* (also called the *Greens*).

Newspapers-actors		
Word	**Freq.**	% in the section
STATE	802	0,310
COMPANY	660	0,255
GOVERNMENT	529	0,204
COUNCIL	505	0,195
COUNTY	486	0,188
PEOPLE	465	0,180
GROUP	450	0,174
INDUSTRY	422	0,163
PUBLIC (as a noun)	109	0,420
COMMUNITY	322	0,124
COUNTRY	263	0,102
RESIDENT	232	0,090
GREEN Party, Greens	126	0,048
LOCAL	72	0,028
UK	223	0,086
total occurrences	5666	2,2
% tot. occurrences of 100 words lst	11,2	

Table 3.3.7.'Actors' field in Newspapers

Magazines and Journals-actors		
Word	**Freq.**	% in the section
STATE	185	0,247
INDUSTRY	177	0,236
COMPANY	146	0,195
PUBLIC (as a noun)	17	0,022
AMERICA	73	0,097
COUNTRY	71	0,095
UNIVERSITY	70	0,093
AGENCY	67	0,089
YORK	69	0,092
EPA	64	0,085
GOVERNMENT	64	0,085
DEPARTMENT	63	0,081
total occurrences	1066	1,4
% tot. occurrences of 100 words lst	8,0	

Table 3.3.8.'Actors' field in Magazines and Journals

Letters- actors		
Word	**Freq.**	% in the section
COUNTY	94	0,278
INDUSTRY	93	0,275
PEOPLE	72	0,213
STATE	72	0,213
COMPANY	62	0,184
COMMUNITY	56	0,166
PUBLIC (as a noun)	19	0,056
BUTTE	51	0,151
CALIFORNIA	50	0,148
GOVERNMENT	50	0,148
LOCAL	16	0,047
COUNCIL	43	0,127
total occurrences	678	2,0
% tot. occurrences of 100 words lst	10,9	

Table 3.3.9.'Actors' field in Letters

Word distribution

A considerable difference between 'Magazines and Journals' and the other sections results from this analysis, and again strengthens the hypotheses assigning to 'Magazines and Journals' an approach which is less focused on people's direct reactions and comments to the debate, broader in perspectives and less insisting on the same terms than other sections. For what concerns this field, 'Magazines and Journals' shows in fact a smaller number of actors, which are mainly public institutions; it also includes – differently from the other sections – *university* among its 100 most frequent words, although with a relatively low percentage frequency; a brief manual check of the concordances of *university* in 'Magazines and Journals' confirms its function of providing a and authoritative basis to the contents of the articles, by citing either academic reports and studies on the subject, or the opinion of experts working in academic departments relevant to the subject. Although 'Magazines and Journals' is the first section mentioning universities among its most frequent actors, the importance of backing up what is being written by means of official or academic documents is present throughout the whole corpus. It is expressed by the words *report* and *study*, evenly distributed throughout the sections and whose relative percentage frequency is between 0,1% and 0,2%. With regards to actors, however, newspaper articles and letters to the editor maintain a general perspective in words such as *state* and *government*, but at times narrow their framing to local communities, or direct their attention to the general public more often than magazines do. Compared to 'Newspapers', 'Letters' make a greater use of *people* and a lesser use of *government*, rather referring to counties as political organizations. The frequency of *UK*, *California* and *Butte* are probably due to the newspapers from which the texts are retrieved (it cannot be excluded that they are also used in the Lexisnexis headings, or part of the name of the source newspapers), while *America* more plausibly shows a particularly high frequency because of the fact that fracking began to be practiced in the United States, and American states have the territories which are most affected by this technology. Another important category of actors which stands out in all sections is that concerning the more strictly economic aspect: *industry* and *company*, extremely frequently indicating oil and gas companies, are at the top of these field lists in all sections. Overall, the weight of this field on the total occurrences of the 100 most frequent words accounts for 11,2% in 'Newspapers',

8% in 'Magazines and Journals' and 10,9% in 'Letters'. The distance between different sections is quite small, but in line with the proposed interpretation; moreover, the differences might show more from a qualitative point of view than from a quantitative one in this field.

Other groups of words

Apart from the three fields I have identified, several other observations about the lexical choices detected among the corpus sections appeared particularly relevant.

Reporting verb *say*

The very frequent use of *say*, in particular its simple past form, clearly tells about the journalistic nature of the corpus, rich in direct and reported speech. The percentage frequency of say is higher than that of reference corpora both in the whole corpus (see tables 3.4.1-3.4.4)and throughout its sections. Yet the distribution of this verb is not homogeneous across sections: it has a very high percentage frequency in 'Newspapers' (slightly more than 1% of the whole section), while it is considerably less frequent in 'Magazines and Journals' (almost 0,5%); 'Letters' is the section where its relative frequency is lowest (0, 14%). This could be justified by the fact that while letters are mainly written to argue the writer's opinion and try to persuade readers into agreeing with them, the function of reporting other people's speech might be limited to add strength to the author's arguments, whereas in newspapers, reporting has a more pervasive function, that of informing the readers about events through people's declarations and communicating messages which contribute to the development of the controversy, always making a selection on the base of relevance in the context of news-making (see Hilgartner and Bosk, 1988). 'Magazines and Journals' is probably less concerned with reporting and more centred around expository writing, so that the frequency of quotation or reported speech is lower in this section.

Modal verbs

Will is the most frequent modal verb both in the corpus considered as a whole and in its sections. Being a prediction modal, it reflect s a high density of forecasts, and its

level of keyness in the corpus (see tables 3.4.2, 3.4.3, 3.4.4) shows that its usage is typical of the corpus. The act of forecasting implies that the propositional content of what is being communicated is perceived with a relatively high level of certainty; this happens also when opinions about the future are expressed, and this is probably what made the use of *will* so frequent. This aspect is most evident in 'Letters', where *will* has a percentage frequency of 0,46%, and in 'Newspapers' (0,43%); it is less stressed in 'Magazines and Journals' (0,33%).

Would is also among the most frequent words; it does not form part of the keyword lists, which means that its use is not particularly typical of the analysed corpus; apart from the less frequent function of expressing future in the past, *would* is mainly associated to conditionals, and specifically to describing situations which are not certain, expressing conjectures or different degrees of probability. A similar function also pertains to the other possibility modals *could* and *may*, all appearing among the 100 most frequent words in all sections; *may*, however, had to be disambiguated from the name of the month, so its occurrences showed to be less frequent than they seemed. As for their distributions, *could* is less frequent in 'Letters', while *may* only had a relative frequency above 0,1% in 'Magazines an Journals'. *Can* is also very frequent; as well as *may*, it was disambiguated from the noun *can* (referring to the metal container) and other forms not tagged as modal verb: the resulted frequencies were only slightly lowered. The high frequency of possibility modals was also revealed by the MD Analysis results for D1 (see paragraph 2.5 in chapter 2).There are not outstandingly frequent lexical verbs which modals accompany in this corpus, so rather than the complete lexical meaning, it is the kind of modalization which is being taken into consideration here. Modals seem therefore to reflect a framing of the controversy through hypotheses and forecasts, in a continuous discussion about what is possible and what is not, what is likely or certain to happen in the future and what is not. That is why the adjective *potential*, whose frequency is above 0,1% in 'Newspapers' and of 0,086% in 'Letters' (while it does not appear in the 100-word list of 'Magazines and Journals') can be added to complete this part of the analysis: it is indeed very often referred to the consequences of fracking, both negative and positive. This kind of depiction of the controversy through possibility and prediction is probably adopted in mass

communication to characterize the debate as pluralistic and lively (and therefore more entertaining).

Locations of fracking

Another semantic area which can be found in the analysed lists is that of the places where fracking happens or is planned; several nouns are used for this purpose; the most frequent are *area*, *site* and *land*; this category sometimes also includes nouns like *county* or *country*, whose meaning includes both the political organization and the territory controlled by it. 'Magazines and Journals' is the section where these nouns have the smallest relative frequencies; it is probable that most articles have a more general consideration of the events related to fracking, rather than focusing on particular and individual cases of fracking application, which instead happens in the other sections. 'Letters' make a quite large use of *area* and *land*, avoiding *site*; these texts seem therefore to prefer a general term (*area*) and the polysemic *land* (either as property to be sold and bought, or a more evocative concept appealing to the emotions of readers). 'Newspapers' seem instead to prefer *site*, more directed towards the functional characteristics of a territory.

Political and legal action

Among the other words still not considered, the theme of legal measures to regulate fracking on a local basis; this theme appears to be frequent only in 'Newspapers', as most of the words pertaining to it are found only in the 100-word list of this section. The first aspect which stands out about regulations is the prohibition to practice fracking: *ban* and *moratorium* are in fact the most frequents words, and their role as a reaction to the possible damage caused by fracking made me categorize them in the 'Controversial side' field. Other words appearing in the lists are *rule*, *regulation* and *bill* (mostly used in its meaning of written suggestion for a new law, and occasionally accompanying *regulation* and *moratorium*). Also here, a word of a different grammatical category can be added, because it contributes to the weight of this theme of laws and regulations: it is the verb *allow*, mostly used to write about permissions to drill for shale gas. Overall, 'Newspapers' appears here as a section which includes the

political and legal actions concerning fracking as a fundamental part of the controversy, which needs to be communicated to the public.

Positive concepts

Words expressing potentially positive concepts about the controversy are not very frequent nor numerous in this corpus, possibly because the media coverage of the controversy is based upon the negative effects of fracking rather than the advantages it could imply. The idea of environment contamination and legislation battles probably give an effect of drama (see Hilgartner and Bosk, 1988)which results more appealing than potential business expansion and financial advantages brought by investments on fracking companies, maybe also because sometimes these are perceived as only affecting gas and oil industry; in any way, the references to possibilities of boosting the economy and gaining a greater energy independence do not seem to be given much space in the corpus, at least form the lexico-semantic point of view and in comparison with controversial aspects (affirmed or negated as they may be). The most frequent 'positive' words in the corpus are *safe*, *clean*, *benefit* and *create* (this last one always has *jobs* as its object, since one of the advantages that the growing shale gas industry has been said to bring is the creation of numerous job opportunities). Quite surprisingly in my opinion, these words can only be found among the 100 most frequent words in the 'Letters' section, except *job*, which also appears in that of 'Newspapers'. Apart from the fact that it cannot be assessed whether these words are affirmed or negated in the texts, the ideas of safety, clean environment, benefits and job creation are only frequent (although not with very high percentage frequencies) among letters to the editor. A daring interpretation might suggest that the drama effect is less important here with respect to 'Newspapers', thus allowing for a more frequent reference to hopes and positive intentions; all themes with which 'Magazines and Journals' are not concerned.

Other words

Several words from the100-word lists generated for this analysis have not been taken into consideration, nor grouped in any field or category of those analysed above. Most of them did not show particularly relevant information, because they were used in

generic ways and their frequency did not highlight any keyness in the keyword list, so they were left out of the explanation and interpretation of the results. What follows is a list of words which showed some relevance or raised interest. After each word, the section(s) in whose lists they were found is specified, and followed by a brief description of its characteristics in the corpus.

- *Energy* (all sections): its relative frequency is approximately between 0,3 and 0,4 sections. It is of course among the most frequent words overall, pointing to the main issue underlying the debate: the need to guarantee sufficient energy supplies to sustain human activities.

- *Call* ('Newspapers'): this verb may be interpreted at first sight as marking the presence of definitions, specifically the kind of definition typically followed by the passive form of the verb *call* and the specific term to be defined. Nevertheless, it does not seem to be the case here, where *call*, whose most frequent collocates were the prepositions *for* or *on* (in the phrasal verbs *call on* and *call for*) refers to claims, especially when describing protests and the relative requests for political measures. Therefore, this verb is very frequently employed in representing the aspect of requests regarding action, laws or measures regarding fracking.

- *Support* ('Newspapers') is one of the words which contribute to representing the controversy; it is the various opinions and factions taking part to the controversy which receive support. It is interesting that *support*, which has a positive meaning, is not counterbalanced, in terms of relative frequency, by any negative item, such as *oppose*, *fight*, *protest*, in any of the sections.

- *Emission*, *world*, *global* ('Magazines and Journals'): *emission* is used when mentioning *carbon* or greenhouse gases, and appears to indicate an attention to one of the most important factors of climate change. This might in turn point to a broader approach, which connects the controversy over fracking to a vast perspective of global energy issues. This is consistent with the above explained hypothesis of a generalized perspective on the debate on the part of 'Magazines and Journals'. A further confirmation of such assumptions is the presence, only on the 100-word list of 'Magazines and Journals', of the words *world* and *global*.

- *Fact* ('Letters'): using the concept of factuality to make a statement more powerful and true to the eyes of the reader is not clearly detectable through automated corpus analysis .Here *fact* is most commonly used in expressions such as *the fact that* or *in fact*, and it would require a detailed manual analysis to try to understand what is given the value of a fact, on which basis and for which purpose.

- *Know* (all sections): this verb is not among the Keywords for the corpus; yet its presence seems important in my opinion, as the theme of public knowledge of science and technology is recurrent and relevant throughout the corpus; it cannot be said what is considered to be known and what is not according to the corpus, but this remains a relevant issue, reflected by the presence of this verb.

- Note: *business*('Newspapers'), *opinion* and *letter*('Letters'), *national* ('Newspapers' and 'Letters') *time*(all sections) were left out of the analysis as their high occurrences were due to their appearance on the headings used by Lexisnexis to categorize texts.

3.3.2 Results and interpretation: Keyword lists and weirdness

Keywords

Keyword lists are used to detect the most typical words of the corpus as a whole. The four keyword lists which created (tables 3.4.1-3.4.4) were not calculated on exactly the same basis, so their results are different from one to another, especially in their calculation of keyness; however, they seem to have some elements in common, as the words with the highest keyness values are more or less similar throughout the lists.

The first wordlist (table3.4.1) was created with Wordsmith Tools. The log-likelihood statistic (the default option in both Wordsmith Tools and AntConc) was used to calculate keyness. The whole BNC frequency list was then used as a reference. In this sense, this list can be considered more accurate than the others, as the reference list includes many more words. In creating the list, however, the program used a wordlist it had previously produced, similar to the one which was used to analyse word frequencies above: this wordlist had been generated using both a stoplist (not including the

supplements mentioned at the beginning of the chapter) and a lemmatization list (not including the grouping of the forms *frack*, *fracked*, *fracks* under the lemma *fracking*), therefore the frequencies used by Wordsmith to elaborate word keyness were the result of a lemmatized corpus with filtered stopwords. Lemmatization has probably inflated the keyness of lemmatized items by increasing the frequencies of the lemma . Nevertheless, the list should help to have an approximate idea of the differences in frequency between corpus and reference corpus.

The second wordlist was created with AntConc, always keeping the log-likelihood statistic. It was created trying to replicate the settings of the Wordsmith keyword list as much as possible, which means that the same stoplist and roughly the same lemmatization list (without hyphens, which created problems with AntConc) were applied to the initial wordlist. A frequency list including the first 5000 most frequent words from the COCA corpus was used as a reference.

As using the whole BNC wordlist created quite different results from using a 5000-word list like that of the COCA corpus (the only freely downloadable list), a third list was created with AntConc, where all settings remained the same with respect to the previous one, except for the reference list, which consisted of the first 5000 words extracted from the BNC frequency list.

Finally, another keyword list was created with AntConc, without lemmatizing the initial wordlist, to see which and how relevant the differences were. In this list, the COCA reference list was applied. In tables 3.4.1-3.4.4, the first fifty entries of each keyword list have been considered.

Despite the evident numerical difference in keyness values between the list realized with the whole BNC wordlist and those realized with 5000-word lists as reference, and between the lemmatized and non-lemmatized lists, all of them showed high levels of keyness for words which also appeared among the 100 words of the three section wordlists. In terms of content, the main differences were noticed between the Wordsmith-generated lists and the AntConc-generated ones. Obvious differences were also observed between the lemmatized AntConc lists and the non-lemmatized one, but the content remained quite similar between the three AntConc lists.

The Wordsmith-generated list includes a greater number of words which did not appear in the previous analysis: they refer to the waste products of fracking

(*wastewater*), to names of region or states (*Karoo*, *Pennsylvania*) and to an individual fracking company, mostly mentioned in reference to the controversy in the UK (*Cuadrilla*). Another main difference between the Wordsmith-generated list and the AntConc ones is the presence of the verb *be* (or its forms in the non-lemmatized one) *have* (or its forms) in the latter, and not in the former; besides, modals are much more present in the AntConc lists than they are in the Wordsmith one (where only *will* appears, and is not among the fifty words with highest keyness). This is probably due to the way in which the wordlists were elaborated by the program, although no detailed reason has been found. Words such as modals, generic verbs like *make*, generic nouns like *year* and *report* do not form part of this reduced keyword lists. *Say* shows instead a high keyness, confirming its representativeness for the corpus as a whole and for at least two of its sections.

A large part of these fifty-keyword lists can be categorized into the field of 'Fracking procedures': *fracking* is always at the top of the list (even without its own lemmatization), and is followed by *gas* and the other elements and actions involved in the procedures. As for the other fields, 'Controversial side' is also present in words such as *water*, *chemical*, *anti* (counted alone by Wordsmith and together with *fracking* in *anti-fracking* by AntConc), *moratorium*, *concern*. There is however some difference between the Wordsmith-generated list, which places these words at relatively higher ranks, and the AntConc-generated lists, where these words can be found more frequently towards the last ranks. Moreover, the Wordsmith-generated list is the only one including the verb *contaminate*, while *contamination* is in common with the non-lemmatized list. The field of 'Actors' appears as that with the least keywords, among these lists of fifty. Except for *Pennsylvania*, *Cuadrilla* (both already mentioned before) and *California* (present in the last AntConc list, there are not any other proper nouns; the common noun with the highest keyness is everywhere *state*; *industry* and *company*, with lower levels of keyness, also have similar ranks among the lists. *Government* and *community* can be found respectively only in the two AntConc lemmatized lists and in the non-lemmatized one. Nevertheless, all these differences do not seem to reveal any really relevant information, as they seem determined either by the program or by the considerably different ways in which the lists were created (lemmatization, size of the reference list), rather than by the reference corpora themselves.

Apart from the differences among the lists, what emerges from all of them is a confirmation of the importance of the presence of explicative or descriptive elements regarding the mode of operation of hydraulic fracturing, in other words, of the informational focus also revealed by the Multi-dimensional analysis. The Wordsmith-generated keyword list also shows a particular relevance of the words indicating the possible damage brought about by fracking. These are the lexical aspects which distinguish the corpus from the general language.

Weirdness

Weirdness was calculated for the first thirty words in the Wordsmith-generated list, also here with frequencies of lemmatized words, again due to the previous elaboration of the wordlist by Wordsmith.

N	Word	weirdness	Keyness
Weirdness-full corpus			
1	FRACKING		93791,9
2	FRACK		2850,91
3	KAROO	3446,184	1983,82
3	SHALE	2129,424	14402,6
4	CONTAMINATE	948,7559	2386,19
5	HYDRAULIC	599,777	5771,2
6	FRACTURE	502,3818	7194,67
7	MORATORIUM	431,4918	3083,99
8	DRILL	289,6084	9921,48
9	ANTI	274,9615	3411,25
10	GROUNDWATER	266,5011	2488,89
11	METHANE	245,0359	2619,78
12	EARTHQUAKE	182,0038	2655,51
13	GAS	123,6205	25918,9
14	CONTAMINATION	121,2651	2166,74
15	WELLS	115,2397	4703,95
16	FLUID	55,00352	2047,07
17	CHEMICAL	50,35816	4975,59
18	BAN	49,96422	3458,03
19	OIL	34,0998	6785,28
20	ENERGY	29,60222	6653,83
21	ENVIRONMENTAL	25,69053	3765,86
22	ROCK	21,06555	2251,72
23	WATER	15,97329	7728,3
24	CONCERN	15,32598	2176,89
25	NATURAL	14,75503	2825,69
26	COUNTY	14,2379	2084,73
27	SAY	12,34754	10262,8
28	INDUSTRY	9,869677	2047,76
29	STATE	7,757347	2689,42
30	COMPANY	6,667809	1950,87

Table 3.5. Weirdness values calculated for the words with the thirty highest keyness values in the Wordsmith-generated keyword list (excluding the words with altered frequencies linked to Lexisnexis). From left to right, the table shows: rank, word, weirdness, keyness. The words are ordered according to decreasing weirdness values, colored in blue as they were manually calculated.

Here, the order of the words is slightly different form that shown in the keyword list which served as a basis: after *Karoo* (82 of the articles of the corpus come from South African newspapers), the first signature word is *shale*, a word used both to refer to a certain kind of rock and, above all, to indicate the source of energy (shale gas, shale oil, shale) at the origin of the controversy. The weirdness of *fracking* and *frack* could not be computed because the denominator in the ratio to be calculated (the relative frequency in the reference corpus) was 0. There is quite a great difference between *Karoo* and *shale* and the following word on the list, *contaminate*, whose weirdness is less than half of that of *shale*. Weirdness decreases gradually down to *wells*, touching alternatively concepts indicating stages and components of fracking with elements felt as potentially dangerous. There is a gap in weirdness between *wells* and *fluid*, and from there downwards weirdness decreases regularly, reaching levels which remain high, but very far from those words at the top. In general, the signature words seem much more balanced between 'Fracking procedures' and 'Controversial side'.

3.3.3 Results and Interpretation: other statistics

Lexical variability

text file	file size	tokens (running words) in text	tokens used for word list	types (distinct words)	type/token ratio (TTR)	standardi sed TTR	STTR std.dev.	STTR basis
Corpus	2430478	388986	381362	17604	4,62	58,95	40,29	300
Newspapers	1623579	259041	254119	13392	5,27	58,41	41,26	300
Mag&Jour	461505	74988	73545	8394	11,41	60,49	39,64	300
Letters	209493	33771	33009	5058	15,32	59,95	38,78	300

Table 3.6.1. Lexical variability statistics for corpus and sections; the STTR, relevant as it compares corpus and sections, is in red.

The standardized type/token ratios calculated for the different sections are quite similar among each other, and standard deviations seem to indicate great internal differentiation. However, some minor differences can be observed among the average measures of different sections: the lowest figure is that of 'Newspapers' (58,41), while 'Letters' show greater lexical variability (59,95) and 'Magazines and Journals' have the

highest value (59,95). One possible interpretation of the difference between newspaper articles and letters to the editor might be a higher degree of norms and restrictions characterizing news-article writing, which may also have affected the vocabulary used; if a certain set of words has been considered suitable to making news stories and be 'sold' to the public, then it is likely that most of the vocabulary about that story will be taken from there. On the contrary, for their own nature, origin an purpose, 'Letters' does not have to respect as many norms as conventional articles do, but for brevity and any minor adjustments made by the editorial staff before the publication. For what concerns 'Magazines and Journals', its STTR appears to confirm the previous observations about a more pluralistic approach of this group of texts to the issue of fracking, although the distance separating it from the other two sections is quite limited: differences in lexical variability among these sections is not extreme, although perceivable.

Average word and sentence length

text file	mean word length (in characters)	word length std.dev.	mean sentence length (in words)	sentence length std.dev.
Corpus	5,02	2,70	21,43	11,98
Newspapers	5,03	2,70	21,36	11,15
Mag&Jour	4,96	2,71	21,73	13,43
Letters	4,95	2,65	20,14	13,05

Table 3.6.2. Mean word and sentence length for corpus and sections. The average values are in red.

The results for average word length are also quite close to each other, with a standard deviation which does not exceed 3, but seem to contradict those for STTR: high lexical variability might be thought to co-occur with longer and therefore complex words in more elaborate styles. Here, however, the more lexically variable sections seem to have on average shorter words. This might be determined by the fact that many of the most common and key words (for example *fracking*, *government*, *environment*, *environmental*) are quite long, and have their highest percentage frequency in 'Newspapers', which might have increased the average word length figures.

The results for average sentence length appear very similar from section to section, with standard deviations showing a highly variable situation (which seems quite normal if we think of journalistic texts). The only section with an average sentence length below 21 is 'Letters'; this section might be slightly different from the others because in

its texts opinions are probably expressed in a slightly briefer and less elaborated way than in elsewhere: in 'Letters', argumentation tends to be immediate, both for space availability and persuasive efficiency.

3.3.4 Results and interpretation: metadiscourse

Hyland's study (2005) divided metadiscourse into two dimensions of interaction: the interactive one, which helps to guide the readers through the text, and the interactional one, which helps the writer to express comments and views engaging with the reader in different ways (these comments and views do not necessarily reflect the author's personal beliefs).

From the interactive dimension, two categories were chosen for this analysis:

- Evidentials, which refer to information from other texts. Here, they might have been used to refer to declarations, experts' opinions, reports or studies, in order to back up any assumption about fracking. This is the list of the evidentials used in this analysis: *to cite, cite, to quote, quote, ref., according to, cited, quoted.*

- Code glosses, which elaborate the propositional meaning and can be found, for example, in definitions. As this corpus deals with the communication of a technological issue, combined with scientific elements (chemicals, pollution and gas leakages, climate change), to a supposedly lay public (at least for the majority) code glosses could point to the purpose to explain these technical issues to the readers. This is the list of the code glosses used in this analysis: *( ), as a matter of fact, called, defined as, e.g., for example, for instance, I mean, i.e., in fact, in other words, specifically, such as, that is, that is to say, that means, this means, viz., which means.*

From the interactional dimension, all the five categories were chosen for this analysis:

- Hedges, which allow the author to withhold complete commission to a proposition; as a result, hedges make the proposition open to negotiation, by asserting that what is being conveyed is not an absolute truth, but rather one of the possible point of views. This is the list of the hedges used in this analysis: *About, almost, apparent, apparently, appear, appeared, appears, approximately,*

argue, argued, argues, around, assume, assumed, broadly, certain amount, certain extent, certain level, claim, claimed, claims, could, couldn't, doubt, doubtful, essentially, estimate, estimated, fairly, feel, feels, felt, frequently, from my perspective, from our perspective, from this perspective, generally, guess, indicate, indicated, indicates, in general, in most cases, in most instances, in my opinion, in my view, in this view, in our opinion, in our view, largely, likely, mainly, may, maybe, might, mostly, often, on the whole, ought, perhaps, plausible, plausibly, possible, possibly, postulate, postulated, postulates, presumable, presumably, probable, probably, quite, rather, relatively, roughly, seems, should, sometimes, somewhat, suggest, suggested, suggests, suppose, supposed, supposes, suspect, suspects, tend to, tended to, tends to, to my knowledge, typical, typically, uncertain, uncertainly, unclear, unclearly, unlikely, usually, would, wouldn't.

- Boosters, which emphasize certainty about the proposition, maximizing the author's commitment to what is written and closing down alternative views. As this corpus is a sort of arena where suspects, suppositions and vague opinions mix with absolute certainties, precise forecasts and reference to authoritative sources, also boosters might have had an important function in stating different perspectives as scientific truths or self –evident statements. This is the list of the boosters used in this analysis: *Actually, always, believe, believed, believes, beyond doubt, certain, certainly, clear, clearly, conclusively, decidedly, definite, definitely, demonstrate, demonstrated, demonstrates, doubtless, establish, established, evident, evidently, in fact, incontestable, incontestably, incontrovertible, incontrovertibly, indeed, indisputable, indisputably, must, never, no doubt, obvious, obviously, of course, prove, proved, proves, really, show, showed, shown, sure, surely, think, thinks, thought, truly, true, undeniable, undeniably, undisputedly, undoubtedly, without doubt.*

- Attitude markers, which signal the writer's affective attitude (for example surprise, agreement, importance, etc. …) to propositions. As here, differently from scientific texts, involvement, empathy and drama have play an important role in 'Letters' and 'Newspapers', attitude markers might have been used to communicate in an (at least) seemingly open and unmediated way the feelings of

the author or of these whose words were reported, towards what was being said. This is the list of the attitude markers used in this analysis:*!, Admittedly, agree, agrees, agreed, amazed, amazing, amazingly, appropriate, appropriately, astonished, astonishing, astonishingly, correctly, curious, curiously, desirable, desirably, disappointed, disappointing, disappointingly, disagree, disagreed, disagrees, dramatic, dramatically, essential, essentially, even, expected, expectedly, fortunate, fortunately, hopeful, hopefully, important, importantly, inappropriate, inappropriately, interesting, interestingly, prefer, preferable, preferably, preferred, remarkable, remarkably, shocked, shocking, shockingly, striking, strikingly, surprised, surprising, surprisingly, unbelievable, unbelievably, understandable, understandably, unexpected, unexpectedly, unfortunate, unfortunately, unusual, unusually, usual.*

- Self-mentions, which make explicit reference to the author, and in this corpus might have had some functions (mostly in 'Letters') in affirming the author' s position about the controversy and in placing the author within a broader group, thanks to the use of inclusive devices (such as first person plural pronouns and possessives). This is the list of the self-mentions used in this analysis: *I, we, me, my, our, mine, us, the author, the author's, the writer, the writer's.*

- Engagement markers, which overtly build a relationship with the readers by explicitly addressing them, thus highlighting their presence. This is the list of the engagement markers used in this analysis: *(, ?, the reader's, by the way, do not, incidentally, let us, let's, one's, our, take a look, take as example, us* (excluding instances of US)*, we, you, your.*

Graph 3.7.1. Metadiscourse markers in the corpus

Graph 3.7.2.Metadiscourse markers and corpus sections.
N.B. the difference among categories have not been deemed reliable, so
only the differences among sections have been considered
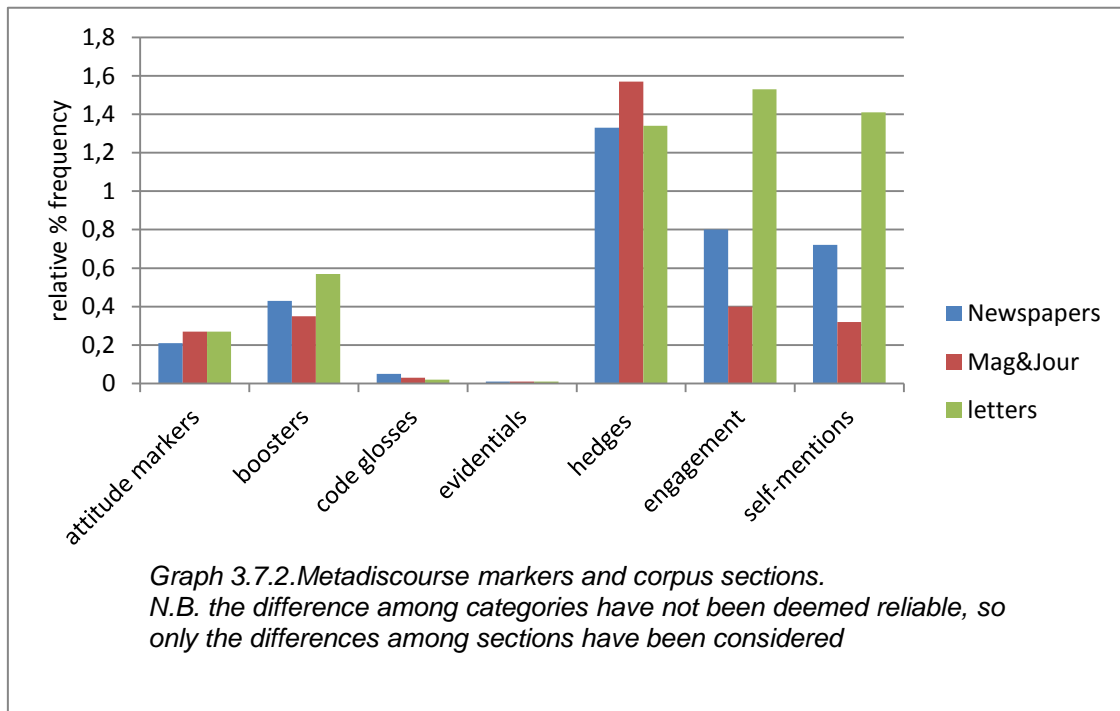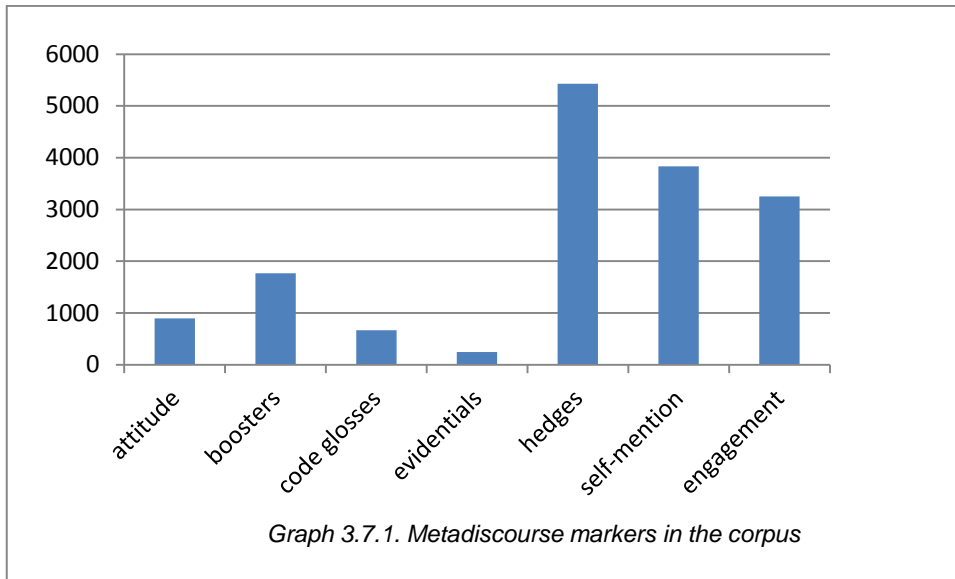
This part of the analysis is considerably limited: first, the counts on the categories did not allow to see which were the frequencies of individual items, so that only the weight of whole categories can be evaluated. Secondly, the categories were heavily modified in order to avoid ambiguities in the computation of frequency counts, so the results cannot be compared to Hyland's. Thirdly, categories are not statistically comparable, as every category consists of a different number of items, so that the larger categories

might have more possibilities of being frequent (even it always depends on the words in question); moreover, the results are considered bearing in mind that words do not always have the same functions and that these metadiscourse categories are not definite objects to test against the corpus. Having said that, these categories maintain their value in that they represent sets of words with similar functions, and the possibility of counting them in the corpus and sections gives at least the opportunity to compare the presence of the same category among different sections (which in fact yielded some interesting results, as can be seen in graphs2.6.1 and 2.6.2). This is therefore the aspect here analysed, after having calculated the relative percentage frequencies of the occurrences categorizing each category on the total running words of each section.

Code glosses and evidentials

If compared to the other categories (even though this does not have any statistical value), code glosses have really low frequencies. This could mean that rephrasing, definitions, clarifications and explanations are not very frequent in any of the sections (whereas this aspect was expected to be frequent in popular science). In the corpus, apart from the brief descriptions of fracking reported in the 'Wordlist results' sections, few definitions of any of the technical details, or of those notions of chemistry and geology scattered throughout the corpus are provided. They probably have not been considered necessary to the representation of the controversy. Evidentials appear even less frequent. This could be interpreted as absence of references; despite that, it has been observed previously that words such as *report* and *study* (and *university* for 'Magazines and Journals') occur quite frequently; therefore, these words could be used in the corpus with different formulations than those present among evidentials (among which the only item which can plausibly be used in an article is *according to*).

Hedges

Hedges reaches its highest point (almost 1,6%) with 'Magazines and Journals', in line with the observations made above, which suggest that this section has a relatively more pluralistic, generalized and open approach than the other sections do. This goes hand in hand with the use of hedges as described above. The other sections show a

relative frequency of around 1,35%. This sort of combines with the figures from boosters in a somehow complementary situation, where 'Newspapers' and 'Letters' show a narrower perspective in viewpoint expression (and also in general lexical choices, as noticed above).

Boosters

Boosters show meaningful differences across sections: 'Letters' appear as the section where these means of closing down alternative opinions and downplaying the role of the reader are most frequent. This is indeed a kind of text where opinions are more or less firmly expressed; the controversial nature of the topic and the importance it has for the life of many communities might have sharpened the debate, and consequently, also opinions and beliefs. 'Newspapers' show a lower frequency for this category (it is around 4%, while 'Letters' are around 0,6%); the same controversial aspects might have intensified the debate here (both expressed in opinion articles and in reported speech), but the informative part might have limited the use of boosters. 'Magazines and Journals' is where boosters are least frequent (below 4%); this was quite predictable as popular science articles are usually more concerned with explaining often than they are with boosting any viewpoint.

Attitude markers

Attitude markers have more or less the same relative frequency in all sections, only a bit less frequent in 'Newspapers' (around 0,2%) and near 0,3% in the other two sections. 'Letters' seem justified in that in this kind of text, the affective attitude of the author is more likely to emerge. 'Magazines and Journals' instead probably used these items for the purpose of reporting opinions and attitude or express the author's point of view in a slightly more affective way than expected. Some of the items of this category, however, could also be employed in a formal context, where the author expresses their point of view on some aspect of the controversy. Overall, a similar level of attitude markers among the sections seems to indicate that the controversy has affected all these journalistic styles in the same way regarding affective attitude expression.

Self-mentions

As could be expected, 'Letters' is the section which makes the largest use of first person pronouns and possessives (probably much more than phrases such as *the author* and *the author's)* where the involvement of the writer (see also D1 in MD Analysis) shows itself most. As can be seen in graph 2.6.2, 'Newspapers' follow at a considerable distance: also here there is a certain level of self-expression probably most of its occurrences are in the forms of inclusiveness. Finally, 'Magazines and Journals' is predictably the section where the presence of the author is more often hidden. Popular science (especially the most authoritative magazines) is seldom concerned with first person interventions; as in scientific texts, an impartial stile is pursued most of the times. Instances of self- mention are found also in 'Magazines and Journals', but they are probably inclusive instances or part of reported speech.

It was interesting, having analysed first person pronouns and possessives. to check the most frequent collocates of *our* in the corpus: it came out that the most frequent collocates are all linked to natural resources (water, environment, energy). Starting from the meaning of *our* , 'of or relating to us or ourselves especially as possessors' , this result is telling of the way people often perceive the surrounding environment and its elements, that is, almost as a granted property.

Engagement markers

Engagement markers have a quite similar distribution to that of self-mentions. It is possible that the two categories work together and give a measure of the level of interaction present in the texts. While 'Letters' are probably more oriented towards a direct author-reader interaction, the other sections probably show this interaction more in reported speech.

3.4 Conclusion

The following points summarize the main findings resulting from the complete corpus analysis, with special attention to the different sections of the corpus:

- 'Magazines and Journals' emerge as the section whose articles approach the controversy in the most detached way; its levels of involvedness and overt persuasion are averagely lower than those of the other two sections, although they remain slightly higher than the average score of Biber's 'PRESS REPORTAGE'. Moreover, this section devotes more space to the 'Fracking procedures', compared to the 'Controversial side' and the 'Actors', among which only those with a national, global or general and institutional dimension are frequent. The perspective adopted on the whole issue is on average broader, often characterized by references to a higher number of elements, and less concerned with immediate risks. Moreover, the detached attitude is confirmed by a lower use of engagement markers, compared to the other sections. Considering all these aspects together, it appears to be the section which most aims at reproducing the standards of scientific communication. The fact that it is the section which makes the largest use of hedges might also refer to the their frequency in scientific articles, although Hyland argues that this class of words has a different function in scientific and popularization articles (Hyland, 2005:98-99).

- 'Letters' are characterized as the most internally variable section, according to the ranges and standard deviations obtained with MAT, which could be justified by a minor presence of writing norms with respect to professional articles. On average, it results as the most involved and overtly persuasive section. This is further substantiated by the frequency of self-mentions, engagement markers and boosters. 'Letters' also appears as the least concerned with technical details, referring much more frequently to the possible consequences of hydraulic fracturing, in particular the short-term ones. It is the only section among whose 100 most frequent words, items with positive connotation were noticed. The stakeholders appearing most frequently here are both general and local, with a slight preference for the local ones.

- As expected, 'Newspapers' appears as a section whose features are, in some measure, in an intermediate position between the other two sections. This is true of the average levels of involvedness and overt persuasion, and of the items appearing in the different fields. The 'Fracking procedures' field includes some words which are more generic than those in 'Magazines and Journals', and this

section devotes more space to this field than letters do. In referring to the 'Controversial side' of fracking, it seems to include a larger and more detailed set of concerns than 'Magazines and Journals', but its 'Controversial side' field has a smaller weight on the whole section than that of 'Letters. As for the 'Actors' section, it includes more items with respect to the other sections (among these items, words referring to political actors generally have higher frequencies than in the other sections).

- 'Newspapers' and 'Magazines and Journals' are characterized by a frequent use of the reporting verb *say*: this was expected of a journalistic corpus. The frequency of *say* is higher in 'Newspapers', around twice that of 'Magazines and Journals' and ten times that of 'Letters'. 'Newspapers' is probably the section which is most concerned with reporting different opinions about the controversy. The presence of conflicting opinions about fracking and the consequent uncertainty pervading the whole corpus is emphasized by the frequency of prediction and probability modals (respectively *will* and *might*, *can*, *could*, *may*).

- All sections have economic actors (*company*, *industry*) among the most frequent items in the 'Actors' field.

- Apart from 'Letters', whose internal variability has been mentioned in the second point of this list, the other sections generally show quite wide ranges (except for D5) and quite small standard deviations in the MD Analysis results. This points to a relatively homogeneous situation, with few instances which are extremely distant from the average in the corpus. For D1, the most distant exceptions from the mean score are mainly represented by extremely informative and not involved texts; for D2, they are represented by highly narrative texts; for D4, by extremely persuasive texts; for D5, by texts extremely elaborated in language. These articles can therefore be said to show which positions along the continuum of each dimension are the most divergent in the corpus.

Chapter 4

Re-elaboration and interpretation of the results: a sociological approach to corpus analysis

4.1 Introduction

The media coverage of the controversy over fracking can be studied to explore complex relationships between science and society, and among different parts of society. This analysis is limited to a sample of newspaper and magazine articles, and cannot account for other media such as television, websites or popular science books; it can however yield some information about how the controversy has been framed, characterized and presented. The relationship between science and society has been object of sociological studies for a couples of decades, for example in Science and Technology Studies (STS) and Public Communication of Science and Technology (PCST) studies. Furthermore, the contextualization of social dynamics into a controversial environment allows them to be analysed in particular circumstances, regarded by some sociologists as moments of potential change and re-discussion of otherwise fixed social and epistemological paradigms (Bucchi, 2000; Lorenzet, 2013; Venturini, 2010).

The focal points of the controversy which sees hydraulic fracturing, or fracking, as the object of a lively debate, may be (perhaps simplistically) summarized as follows: on the one hand, the proponents of fracking claim that it can provide enormous quantities of cheap energy from fossil fuels, with all the consequences this might have on everyday life, national economies, international political economy, and so on. On the other hand, its opponents maintain that it has serious impacts on the environment and on people. Therefore, the main question that could be asked, once the linguistic analyses have been carried out, is: what information can this corpus reveal about its representation of the controversy, from a sociological perspective? As a first attempt to

answer it, the corpus, and consequently the controversy, have been connected with theories developed within STS and PCTS, to start outlining possible interpretations.

4.2 Reviewing the corpus through sociological interpretations

4.2.1 The continuum model of science communication, the phenomenon of deviation and a possible classification of the controversy

One possible approach is that elaborated by Bucchi (2000, 2008). Analysing the process of scientific knowledge communication, he refers to a model elaborated by Cloître and Shinn (Bucchi, 2008) and consisting of a continuum joining four main levels of communication of science, each with its particular characteristics:

- The *intraspecialist level*, taking place among specialists researching in the same discipline and typically appearing in specialized scientific journals;
- The *interspecialist* level, touching different disciplinary areas and intended to reach specialists from different fields;
- The *pedagogic* level, which characterizes science text books, and where science starts to be presented in the form of consolidated theories;
- The *popular* level, typical of science articles in the daily press, or of television documentaries.

Texts distributed along this continuum gradually change the way in which scientific content is presented to the respective audiences: from specialized contexts, where the exposition is provisional and marked by tentative conclusions, to the popular press, where science is found to have undergone a process of factualization, with no trace of doubts or disclaimers (Fleck in Bucchi, 2008: 61). This led Bucchi to imagine a graphic representation of the continuum as a kind of funnel, where the meaning of scientific knowledge is created in an open context made of hypotheses and different possible interpretations (the large edge of the funnel) and is then transmitted as less and less ambiguous, down to the opposite end (the narrow one), where all uncertainties have been abandoned. In a normal, routine situation, scientific knowledge usually flows from within the scientific community (interspecialist and intraspecialist levels) to the lay

public [5] (pedagogic and popular levels), undergoing the above described transformation. The newly consolidated scientific facts are in turn generally accepted and reproduced by the scientific community as an ideal of knowledge (Bucchi, 2008: 62). There are, nevertheless, cases in which scientific facts (and the possible related technological applications) are not so straightforwardly accepted, and become object of debate not only in the stages of their formulation within the specialist groups, but also among the lay public. On these occasions, scientific knowledge follows atypical trajectories with respect to that described in the continuum model, and doubt pervades the pedagogic and popular levels too. This kind of phenomena have been studied and described by Bucchi through the concept of 'deviation' of the conventional flow of scientific communication towards the public (Bucchi, 2008: 63). In cases of deviation, technical and scientific discourse directly reaches the public without being previously agreed within the specialized levels. This deviated passage to the public can give new and unexpected directions to the controversy, and these can in turn influence specialized circles. In this sense, a case of deviation can represent a sort of subversive moment in the production and communication of technoscientific knowledge.

If the corpus was to be classified according to the continuum model, it would pertain to the popular level, with magazine articles leaning slightly towards the pedagogic level. The whole corpus, including its most technical and science-oriented contents, is in fact presented in a popularized way, which is not considered purely scientific. MD Analysis seems to confirm that: as shown by the average score of the corpus for D5, the kind of language used is not as elaborate and abstract as in academic prose texts, nor extremely plain and simple (just in line with Biber's reference 'PRESS REPORTAGE' texts). Only 'Magazines and Journals have a higher score (1,1), which is more markedly located on the elaborate style, and might point to the fact that the position of magazines in the continuum of science communication is slightly closer to the pedagogic level than newspaper texts are. The popularising aspect of the corpus is also substantiated by the MD Analysis results for D2, where, differently from academic texts and similarly to press reportage texts, the corpus (especially 'Newspapers' and 'Magazines and

---

[5] Although here the notion of 'public' is kept for analytical purposes, it is essential to remember that the public is not a fixed entity. On the contrary, it is complex and heterogeneous, and the forms and internal distinctions characterizing it are shifting as a result of different possible contexts (Michaels, Irwin and Einsiedel in Bucchi, Trench, 2008:173-4).

Journals') emerges as a mixture of narrative and non- narrative style: as usually happens with the popularization of science and technology, narration has been included among the linguistic devices used to represent scientific and technical issues.

Differently from the description given of these levels in Cloître and Shinn's model, however, what dominates in the descriptions about hydraulic fracturing are not established scientific facts, nor the celebration of technological innovation. It is rather uncertainty which pervades these texts, together with a multitude of different accounts and opinions regarding the way fracking is performed, the economic advantages it brings about, the environmental and human costs it implies. This aspect is shown by the MD Analysis results, in that the corpus was found to be mostly informational in style, but slightly more involved than the average values indicated by Biber for the press. Another element I think indicates the presence of different ideas is the use of overt persuasion (expressed by D4 in MD Analysis); although not the only way to express persuasion, the overt devices gathered in D4 had, once again, slightly higher values than the average for their corresponding genres in Biber's study, especially in the 'Newspapers' and 'Letters' sections, the nearest ones to the popular end of the continuum. Besides, the high frequency of modals (some of which appearing among AntConc-generated keyword lists) affecting propositions in different degrees of probability, points either to the density of forecasts (in the case of *will*) or to the widespread uncertainty (in the case of *might*, *may*, *can*, *could*). The metadiscourse category of hedges certainly adds strength to the tentativeness surrounding most assessments of fracking.

Moreover, in the representation of the controversy made by the media sampled in this study, knowledge seems to spring from many different sources, and then to spread among different audiences, generating various feedbacks: geologists, engineers, oil and gas companies, research institutions, politicians, the public, activists and environmentalists[6] are involved in the debate at the same time. Research results are made to interact with surveys by environmental organizations, both are made to interact

---

[6] As has been specified about the notion of 'public', also that of 'environmentalist groups' or 'organizations', here employed, should be regarded as including a diversified range of different subjects, whose actions and purposes can diverge, combine and shift according to the context, in a continuous challenge to position themselves in relation to changing scientific orthodoxies, public opinion, and institutional stances (for a full account on the role of environmental action groups and other NGOs in PCST, see Yearley in Bucchi, Trench, 2008: 159-172).

with citizens' claims, and so on. This kind of communication has been defined as *cross-talk* (Bucchi in Lorenzet, 2013: 46), and Lorenzet (ibid.) identifies it as the predominant communicative model in the controversies he defines as 'global', which spread internationally and involve several actors and media. Global controversies are deemed capable of generating far-reaching debates, touching issues such as the relationship between humans, technology and the environment. It is also the case of fracking, even though the debate might have been initially based on a more local dimension, that of people living near the wells and suffering from damages due to fracking. This might lead to categorize the controversy as also 'local'; in the same study, 'local controversies' are indeed described as starting with bottom-up campaigns from local communities wanting to be involved in the decision-making process regarding the construction of particular facilities, often perceived as potentially harmful, in the area where they live. However, the predominant ways and contexts in which local controversies are communicated (local media and microblogging social media, with citizens as the main source of communication about the issue) do not reflect the width reached by the controversy over fracking in more or less all means of mass communication.

Thus, as a consequence of the widespread dispute, topics and opinion exchanges which would otherwise happen within a community of experts take place also in the rest of society. This of course implies a debate which acquires many more elements than it would do in a specialized research context: visibility is given to bodies of knowledge not only originated by university departments and company directives, but rather coming from direct experience of the drilling sites, awareness of environmental causes and so on. All these features could be regarded as fitting Bucchi's concept of deviation, although, due to further elements argued later in the chapter, this debate does not seem to be fully comparable to the cases described by Bucchi.

4.2.2 Framing the controversy as a social problem

The controversy over fracking received extensive coverage by the media in the period considered by this study and still does today. The way in which certain issues gain public attention thanks to media coverage is addressed in an STS paper by Hilgarnter and Bosk (1988) through the concept of 'social problem', defined as "a

putative condition or situation that is labelled as a problem in the arenas of public discourse and action" (Hilgartner, Bosk, 1988: 55). Social problems are hosted by public arenas, which have different 'carrying capacities' (determined by the number of topics which can be covered at a certain time by the same public arena). At the basis of this analysis lies the assumption that the identification and importance of a social problem are arbitrarily attributed to it, regardless of its objective, intrinsically harmful conditions. Moreover, according to this theory, "public attention is a scarce resource, allocated through competition in a system of public arenas" (ibid.). Consequently, competition happens among a high number of potential social problems , or issues claiming public attention, but only some of them receive significant attention and become successful. From this point of view, despite being formulated simultaneously as a problem and as an enormous economic opportunity (which may differentiate it from full social problems), fracking might have followed the same discursive path of a successful social problem, and of one that stays relevant for several years. To be successful and become visible, however, potential social problems must have a marketable design which suits at least one public arena (although this success easily spreads across different arenas, such as newspapers, television and internet. For this reason, not only does competition take place among different issues, but also among different designs (that is, different constructions and interpretations) of the same issue. As a consequence, it can be reasonably presumed that the coverage of the controversy over fracking is not determined by its inherent characteristics alone, but also by a process of "collective definition" (Hilgartner, Bosk, 1988: 70). In this sense, each public arena has 'principles of selection', criteria upon which the selection of problems is based: these vary in quality and importance in different arenas, but the authors of the study identify some which seem to apply to all of them. Three of these principles are considered to be particularly relevant to the controversy here analysed:

- Drama: casting social problems in dramatic ways and making frequent reference to hard facts through authoritative sources makes them more successful, and creates an appealing combination of factuality, urgency and emotional rhetoric;
- Culture: a potential social problem which is related to broad cultural preoccupations is more likely to receive a good coverage;

- Politics: Social problems which fall within the range of discourse of political forces exerting influence on a given arena are more likely to be successful in that arena.

Moreover, due to space and time limitations typical of the mass media as public arenas, the more succinct a message becomes, the more successful it is. This is also valid for scientific accounts (see Dunwoody in Bucchi, Trench, 2008:19), as will be explained about the relationship between science and society outlined below.

Following Hilgartner and Bosk's line of reasoning, it is therefore legitimate to concede that the corpus shows a framing of the controversy over fracking which has been at least partially shaped according to these selection principles. Moreover, the role played by the predominant interpretation (or interpretations) of the controversy stands out also outside the media themselves, if space in public arenas is considered as being connected to space in the political agenda and to the public understanding of the topic. However, these influence patterns are not to be taken for granted nor intended as linear. The framing provided by the media does only partially affect public opinion. Indeed, the former only partially reflects the latter. In other words, the media suggest interpretations, they do not impose them. This leaves the audience with wide margins for further appropriation and elaboration (Neresini, 2011: 132-137).

4.3 Re-contextualization of the linguistic results within sociological frameworks: possible interpretations

Having suggested possible connections between some sociological theories and the controversy, together with its related discursive practices, the interpretation of my linguistic analysis results can now be pushed further, aiming at a more comprehensive description. This is an attempt to develop a deeper understanding of underlying discursive and cultural patterns present in the corpus, borrowing some sociological notions and concepts I found relevant and meaningful in this context. It is important, then, to bear in mind that what follows is my own interpretation of the media coverage of fracking. Besides, I am not mapping the whole controversy, nor am I arguing that this is also what happens outside the media: as clarified above, they can partially reflect the

dominant mind-set of the culture within which they are produced, and can partially influence public opinion and agendas, but there is no direct link between the two, nor universal rules of interaction. Finally, in reproducing the interpretive trajectories leading the reasoning, the paragraph headings are formulated as questions, with the paragraphs working as tentative answers. The integrations that had to be made to the lexico-semantic analysis are all reported in tables 4.1 and 4.2.

	Corpus		Newspapers		Magazines and Journals		Letters	
	Raw frequency	Relative % frequency	Raw frequency	Relative % frequency	Raw frequency	Relative % frequency	Raw frequency	Relative % frequency
renewable energy/energies/ source(s), renewables	135	0,03	81	0,03	16	0,02	27	0,07
non-renewable energy/energies/ source(s), non-renewables	1	0,0002	0,0003	0	0	0	0	0
scientific*, science, sciences (*excluding Scientific American)	223	0,05	133	0,05	63	0,08	10	0,02
activist(s), protest(s), protester(s), campaign(s), campaigner(s), environmental group(s), environmentalist(s)	784	0,2	607	0,2	78	0,1	27	0,07
bridging fuel(s)	20	0,005	3	0,001	9	0,01	0	0
bonanza, abundance, abundant	48	0,01	23	0,008	19	0,02	5	0,01

Table 4.1. Additional frequency counts

	COCA	COCA-newspapers and magazines	Corpus
renewable(s)	4042	2174	135
non-renewable(s)	66	21	1
renewable(s) raw frequency/*non-renewable(s)* raw frequency ratio	61,24	103,52	135,00

Table 4.2. Raw occurrence of "renewable" versus "non renewable" in the corpus and COCA

4.3.1 Does the corpus show any bias as a function of the space devoted to different aspects of the controversy?

Fracking opponents could claim that the media do not devote enough space to the risks posed by fracking, in favour of the inviting perspective of a new abundance of fossil fuel energy. However, this is not what my lexico-semantic analysis showed, in none of the three sections. Among the 100 most frequent lemmatized words in the

corpus and sections (extracted excluding the words listed in the stoplist) none of the fields or word groups identified seemed to refer to the advantages of fracking technologies, or to the potential benefits their application could bring about (see 'positive concepts', paragraph 3.3.1 in chapter 3). The positive concepts identified were mostly found in letters, and they may refer to possibilities of making fracking procedures safer (*safe*), of respecting the environment (*clean*), or of creating new jobs (*create*, *job*). These words seem to be used with a view to environmental causes, or generally more immediate, everyday-life concerns (even if their individual uses and connotations cannot be ascertained here). Frequency counts of the words *bonanza*, *abundance*, *abundant* (those whose occurrences were almost always connected to positive economic consequences of fracking) were computed and added together, to verify the extent of their presence in the corpus and sections (see table 4.1): there were 48 instances in the whole corpus, 0,01% of the total of running words; 23 of them were found in 'Newspapers' (0,008% of the total of running words of the section), 19 in 'Magazines and Journals' (0,025%, most of which occurred in texts from *The Economist*), 5 in 'Letters' (0,014%). These three words certainly cannot account for any instances where the profitability of fracking is mentioned, envisaged or celebrated. Yet their relatively low frequencies, compared with the lexico-semantic analysis of the most frequent 100 word listed in chapter 3 seem to reveal a set of semantic choices clearly addressing concerns, risks and damages, rather than abundant and cheap energy and all that goes with it. On the contrary, an additional group of words referring to anti-fracking activities (*activist(s), protest(s), protester(s), campaign(s), campaigner(s), environmental group(s), environmentalist(s)*) showed relatively high percentage frequencies (see table 4.1) – 0,2% in the whole corpus, 0,07 % in 'Letters', 0,1% in 'Magazines and Journals', 0,2% in 'Newspapers' –, which are not entirely justifiable by the higher number of items in the list with respect to that formed by *abundance*, *abundant* and *bonanza*.

This is not to imply that the media are biased towards anti-fracking campaigners: in the first place, my analysis cannot identify the connotation given to the concepts, actions and people mentioned. In the second place, such a high frequency of the 'Controversial aspect' field with respect to a hypothetical 'Advantages' field could find an explanation in Hilgartner and Bosk' s 'principles of selection'. Drama might

definitely have played a role in the lexico-semantic choices of the corpus: contaminated water, earthquakes, air filled with explosive methane; families and communities whose life have been heavily affected by the arrival of drilling companies; all these elements offer a highly dramatizing potential to be employed in the news, probably more marketable than that offered by optimistic economic forecasts and accounts of drilling companies' increasing business (strongly interconnected factors which might be perceived as distant from common people's life). Moreover the slightly higher-than-average involvedness of language, together with all the elements indicating an intense and lively debate (for example overt persuasion and modals of prediction and possibility) contribute to create an emotional kind of rhetoric. The selection principle of culture is also relevant here for two main reasons: first, the environmental impacts that fracking is suspected to have can easily evoke the broad themes of environmental protection and sustainable development, so frequent in the media and by now so popular at least in Western culture. A second reason might be represented by the utter dependence of most human civilizations on fossil fuel energy supplies, and the subtle and deeply felt urgency with which every possible energy source needs to be tracked down and fully exploited; together with it come the preoccupations about the future of fossil fuel-based economies and, again in an environmentalist perspective, the awareness of greenhouse gas emissions generated through fossil fuels, which contribute to the almost globally acknowledged issue of climate change. The political aspect of the media coverage of fracking is recognized as relevant, but perceived as somewhat awkward to address without adequate competence regarding both international and local political scenes, and with the same words (for example, *green* and *conservative*) indicating different parties in different countries. As this aspect would require a study for itself, it is going to be left out of this interpretation.

4.3.2 What kind of relationship does the corpus represent between science and society?

The traditional conception of public communication of science, which dates back to the beginning of the 20th century, gives scientists the status of knowledge producers and possessors, and gives the public the status of ignorant and passive receivers of scientific knowledge, which needs to be transferred to them through linguistic

adaptation (hence the metaphor of translation) in order to be understood. These widespread assumptions imply and legitimate professions (mainly popular science journalists and writers) in charge of mediating between scientists and the public by translating scientific knowledge. This in turn allows the scientific community to proclaim itself and be perceived by the public as extraneous to the communication of scientific knowledge. The continuum model rejects the metaphor of translation, arguing that scientific knowledge, as it passes through the various levels described, does not involve any kind of translation, but rather a change in the notions themselves (Bucchi, 2008: 58-62). It however appears to maintain a certain distance between the scientific community and the general public (though it must be admitted that the specialized levels of communication are influenced by the other levels). Moreover, according to widespread beliefs, science is ideally separated from technology (intended as the application of innovative discoveries in the scientific field) and the two are ideally separated from the consequences of technical applications, which often tend to be polarized into positive or negative, depending on the use that is made of them (this aspect is analysed in further detail in chapter 5). If combined with the traditional conception of science communication, the whole process – from scientific discoveries, through innovative technologies, to their applications – sees the lay public in a condition of constant cultural delay with respect to technological innovation. Although all these assumptions, together with the, should not be taken for granted and have in fact been challenged (Neresini, 2011, Bucchi:2008), they continue to be extremely influential in the way science and technology are understood and communicated today. Furthermore, they contribute to justify the idea of a passive public, and of specialized groups as extraneous not only to science popularization, but also to potentially negative consequences of technical applications.

Controversies, however, as has been explained in paragraph 4.2.1, can be characterized by particularly frequent instances of interaction among the specialized community, political institutions and the public (both inside and outside the media) in a way which is more complex and less linear than the traditional account assumes. A debated issue, for example, could favour initiatives, projects and activities on the part of public and research institutions to foster communication between public and researchers and promote a positive attitude towards science and technology. On the other hand,

citizens' or activists' organizations can ask for and reach channels of communication with political institutions or the scientific communities, in order to analyse in depth any technical concern and protect their interests or defend their cause.

Also the debate about fracking appears to have generated a wide and complex range of dynamics linking the many actors taking part in the controversy. However, I argue that the corpus reflects an idea of distance between the scientific community and the lay public. Quite relevantly, the frequency lists do not seem to reveal any real opening up to the public of the contents and procedures of researches aimed at assessing the effects of fracking. There is a marked focus on the process of fracking itself, but its definitions and explanations seem highly standardized (characterized by the same, very frequent words), and not very detailed. Only 'Magazines and Journal's show greater lexical variability than the other sections do with respect to the 'Fracking procedures' field, which has been interpreted as revealing a slightly less standardized and more context-dependent approach. Nevertheless, the attitude towards fracking and other technical procedures seems overall characterized by approximation and conciseness: the ideas, for instance, of mixing water with chemicals, pumping this mixture into the wells, opening cracks in underground rock layers to fetch natural gas from there seem sufficient to reason and discuss profusely and comprehensively about fracking. The extremely low frequency of code glosses, although not completely reliable statistically, may indicate the lack of discursive practices aimed at disclosing scientific and technical details to the public. Another example is the absence, in the wordlist analysis results, of the importance of the many different geological characteristics specific to each different area, which affect the feasibility and the effects of fracking. Specific terms can appear, but do not necessarily point to detailed accounts of scientific and technical methods. An example is s*hale*, an extremely common word in the corpus, whose specifically geological meaning seems almost never to be explained throughout the corpus (see chapter 3), and which has been very often used as a premodifier of *gas* and sometimes attributed the meaning of 'energy obtained from shale gas and oil' as a consequence of metonymical use. The simplification and conciseness of technical and scientific accounts surrounding fracking in the corpus might be explained as part of a marketability strategy which privileges succinct messages (see paragraph 4.2.2), and is

mostly justified as responding to the audience's modes of information consumption (Bucchi , 2008: 59).

I suggest that this way of displaying technical issues contributes to preserve the status of scientists, researchers and engineers as authoritative drivers of uncertainty reduction (Dunwoody in Bucchi, 2008: 20). The words *science*, *sciences*, *scientific* are mostly used as a decisive element in trying to settle debated aspects and suppress irrational behaviours. Even if added together, they have however a relative frequency which is not enough to appear in the considered 100-word lists, which might indicate that scientific concerns faced stiff competition with other aspects (such as the political ones). The percentage frequency of these three items added together is highest in 'Magazines and Journals' (0,08%), lowest in 'Letters' (0,02%) and halfway in 'Newspapers' (0,05%), reflecting the importance that the aspect of scientific confirmations might have had in each section (see table 4.1). The value of scientific facts remains in any case high in the corpus, as shown by other elements in the lexico-semantic analysis. For example, the reporting verb s*ay* is extremely frequent; *study* and *report* are also among the most frequent words. It is therefore plausible to think that specialists are often involved in the debate: their opinion and summaries of their researches are probably reported in a hierarchically high position with respect to other sources, and are probably also used to back up other people's opinions. Concerning the way in which scientists' words are reported, it has also been argued that, in debated cases, when credible experts make contradictory claims, journalists try to maintain a stance of objectivity and balance. Objectivity is created by avoiding focusing on the validity of a claim, stressing instead the accuracy of the report; balance is reached by comprehensively representing as many truth claims as possible (ibid.), regardless of their validity. It is not being implied that specialists' opinions within the debate should not be regarded as valuable or valid. Rather, it is argued that the specialized groups' activities are only partially exposed in the corpus, and their published results are generally regarded as utterly consolidated and authoritative, while the tentativeness and uncertainty of research, its possible inconsistencies and unorthodox methods and instruments, the social and personal components of scientific research are hardly mentioned. Moreover, the lexico-semantic analysis did not reveal reference to initiatives, projects and activities promoting the interaction between common people

and members of the scientific community. In this sense, the media representation of this controversy does not fit into the definition of deviation proposed by Bucchi. In his study he also suggested that deviation might be initially induced on purpose by some members of the scientific community, who are involved in a dispute happening inside their specialized technoscientific sector, and act to deviate the flow of communication towards the public to allow their debate to reach wider audiences (the general scientific community, political authorities, the lay public). Deviation could thus reveal itself as a sort of tool which scientist taking part in the debate try to control to obtain visibility and support from other social groups, be it to make their opinions prevail or to legitimize their position. In this case, more than usual, the popular media become a way not only for scientists to communicate to the public, but also to communicate among themselves to negotiate the terms and outcomes of the dispute. In the case of fracking, however, the controversy was not aroused owing to a crisis within the scientific community, which lead to disclose technical and scientific notions to the public, as described in the case-studies exposed by Bucchi (2000). Rather, it was the heavy impacts reported from several families and communities after having their backyard drilled which caused the protests. Moreover, the way scientific processes are represented could be defined as a mainly 'frontstage' discursive practice: results (however uncertain and in conflict among different sources) are directly presented and commented, without revealing any preliminary stages of experimentation research and of debate among experts (what would be defined as 'backstage' for the public) (Bucchi, 2000: 127-130) which precede the popularization of technical and scientific knowledge. It cannot be excluded, however, that the members of the scientific sectors involved in the controversy are using the media to communicate amongst themselves. On the whole, it can be said that the predominant interpretation that the corpus seems to give of the relationship between science and society is largely referred to traditional conceptions of public communication of technical and scientific knowledge.

4.3.3 What kind of relationship does the corpus represent between politics and the public?

As can be seen from the analysis of the 'Actors' field in chapter 2, politicians and political institutions are present in all the corpus sections:

- 'Newspapers' seem to place a focus on national and local governance organizations: *state* is the most frequent actor; *government* – which can refer both to national and local governments – and *council* – usually referred to local forms of government – have the highest relative frequencies of all sections; *county* can indicate both an area and the political institutions ruling in that area – in its political interpretation, it is an expression of local government ; *local* itself, when modifying *community*, *council, government*, *group* expresses the importance of regional political forces in the decisions to be taken regarding the potential application of fracking technologies; finally, *green* is also among the 100 most frequent words in 'Newspapers'; its frequency is partly due to its political use in referring to Green parties, traditionally concerned with the protection of the environment.

- 'Magazines and Journals' seem less focused on local political organizations, with words such as *state*, *America*, *country* and *government* (with a much lower relative frequency than in newspaper articles) in its list. Moreover, *agency*, *EPA* and *department* indicate, to a certain extent (their relative frequencies are not extremely high), the importance of the role of such institutions in the articles of this section.

- *County* is the most frequent actor in 'Letters', indicating the probable predominance, in these texts, of local political institutions. This is supported by the use (although not frequent after disambiguation) of the adjective *local* and by the presence of *council* among the 100 most frequent actors in 'Letters'. However, also *state* and *government* appear in the 100-word list of 'Letters'.

Of all the sections, 'Newspapers' is the section where words related to political institutions have the heaviest weight (1,07% of the total running words – these figures have been calculated on the basis of the field analysis in chapter 3, and are not therefore reported in tables 4.1). 'Magazines and Journals' and 'Letters' (respectively 0,7% and 0,8%) seem to devote slightly less space to politics in terms of frequency, maybe in favour of other aspects: in the former section, these aspects might be technical and scientific or economic issues; in 'Letters', they might be peoples' opinions, claims and concerns. The percentages however point to a high relevance of the field of politics in all sections. Besides, it could be worth remembering that the MD Analysis revealed, in

some cases, more resemblance between the corpus and textual genres related to politics than between the corpus and texts related to science and technology (or even engineering): this could be an indicator of the great political relevance that this issue has.

In terms of associations revealed from the lexico-semantic analysis, the political world appears entitled to either impose bans and moratoria on fracking, or to allow it. On the other hand, while governments decide about permissions to drill, common people or activist groups appear to call for measures and regulations, protest against fracking or support it, be concerned for potential risks and eager to know more about it from governments and expert commissions. The relationship which emerges from the corpus between politics and the public is limited to this strictly separated role distribution. Unfortunately, the methods used in this study do not allow to ascertain what kind of connection takes place in the corpus between public claims and political action. What is noticeable, however, is the absence from the wordlists considered of any items referring to instances of exchange, collaboration or direct interaction between the two.

4.3.4. Does the corpus show any paradigm (cultural, cognitive, social) which is being re-discussed and negotiated in the context of the controversy over fracking?

As explained at the beginning of the chapter, in the works I have adopted as my sociological guide into the interpretation of the corpus, controversies are sometimes described as moments of potential change in social, cultural and cognitive paradigms; these can regard for example the relationship between science and society, the status and boundaries of some disciplinary fields and the communication of and public attitude towards science and technology. As has been argued in paragraph 4.3.2, the predominant notion of public communication of science and technology in the corpus seems quite close to the traditional conceptions explained in Bucchi (2008), so it is highly improbable that any paradigm concerning this field is being challenged. An attempt at researching traces of other possible cultural and social transformations happening within the corpus can then start from the assessment of the main terms of negotiation of the controversy. For this purpose, the debate could be ideally comprised

within two propositional extremes: whether to continue to implement fracking technologies and further extend their use, or to quit them for good.

The extreme of the debate which opposes fracking, and in particular environmental groups, emerges, more than other actors, as trying to take advantage of the doubts cast on the validity of the hydraulic fracturing technological innovation in order to promote new priorities in the elaboration and evaluation of energy production research and technologies. In other words, they seem to be striving to change the predominant paradigm of energy resources exploitation. Among their priorities are the enhancement of energy sources alternative to fossil fuels and the consequent decrease of greenhouse gases emissions. The shift to alternative forms of energy production is also associated to lesser immediate impact and risks linked to new technology implementation. It could moreover be concluded that environmentalists acting in this controversy are trying hard to increase the scientific legitimization of their cause, in order to give scientific relevance to the negative impacts of fracking on the environment. The presence of these actors can be revealed by the group of words referring to anti-fracking activities created and researched earlier in the chapter (consisting of the items *activist(s), protest(s), protester(s), campaign(s), campaigner(s), environmental group(s), environmentalist(s)*), and by many of the occurrences of the items in the 'Controversial side' field, almost all of which deal with environmental concerns.

The challenge posed by anti-fracking movements and environmental groups to the currently predominant model of energy production policies can be contextualised into a broader dispute: that about climate change (or global warming) and environmental protection. The debate over global warming has become widespread and powerful over the last decades, causing growing attention to environmental concerns in many different social contexts. As a consequence, these have become important in gaining political support, managing industrial policies, selling products to the public, and also appealing to the audience in mass communication, which might explain the frequency of *environment* and *environmental* in the corpus, and of *climate change* especially in 'Magazines and Journals'. The role of science and technology in finding ways of supplying the energy we need is therefore also connected to the issues of global warming and environmental protection, and it is strongly present in the corpus,

especially for what concerns the assessment of the effects that shale gas and oil extraction technologies can have on people, animals and the whole environment.

Given its discursive relevance, the climate change issue cannot be ignored, not even by fracking advocates. This has generated a sort of 'intermediate' argument between the perceptions of enormous economic opportunity and environmentally dangerous practice associated to fracking: it is the idea of fracking as a 'bridging fuel' to a lower carbon future, mainly concerning the United States. The 'bridge' metaphor exemplifies a process in which the rising production of natural gas, made available in large quantities by hydraulic fracturing, widely replaces coal in the United States electricity sector, thus reducing carbon dioxide emissions (which is regarded as a step towards the avoidance of significant climate change). At the same time, this should give the nation the time and opportunity to develop those forms of energy production (from renewable and carbon-free sources) which will finally and decisively drive the United States (and desirably the rest of the world) out of climate change concerns. Thus, in a seemingly paradoxical attempt at mediating between the profitability of fracking and the needs for environmental reassurance, drilling for natural gas becomes a way to favour renewable energy and reduce the risk of significant climate change. The idea of fracking as a 'bridging fuel' has been looked for in the corpus, by counting the occurrences of *bridging fuel* and checking the occurrences of *bridge* connected to this concept (table 4.1): the frequency of such expressions is very low in the corpus, and completely absent from 'Letters'. It probably has not been deemed extremely relevant in the development of the controversy. Besides, its intermediate position might not have been functional in the journalistic construction of the debate. Instead, the interest towards renewable energies is present in the corpus, although not part of the 100-word lists analysed (see table 4.2): the sum of the occurrences of *renewable energy/energies*, *renewable source(s)*, *renewables* is 0,03% of the total running words in the whole corpus, 0,03% in 'Newspapers', 0,02% in 'Magazines and Journals', 0,07% in 'Letters'. Curiously enough, there is instead only one occurrence of *non-renewable* in the whole corpus (appearing in 'Newspapers'), despite the fact that the dispute is centred around a non-renewable source of energy. As indicated by the frequencies of these words in the newspaper and magazine sections of the Corpus of Contemporary American (COCA), *renewable* and *renewables* are generally much more frequent than *non- renewable* and

non-renewables, partly justifying their strongly asymmetrical use in the corpus. On the other hand, the ratios between the frequencies of the two lemmas (occurrences of *renewable* plus *renewables* on occurrences of *non-renewable* plus *non-renewables*) calculated for both COCA and the corpus reveal that the difference in use is larger in the corpus (where the ratio is 135 – that is *renewable(s)* is 135 times more frequent than *non-renewable(s)*) than in COCA (where the ratio is approximately 103). Moreover, this difference could be expected to be reduced in a corpus mainly concerned with non-renewable sources than renewable ones. The notion of bridging fuel, together with the interestingly high frequency of reference to renewable energy sources might indicate a tendency to admit the importance of climate change awareness in relation to the controversy over fracking. Nevertheless, this does not necessarily result in a genuine intention to change the predominant energy production and consumption paradigm.

The corpus contains, to sum up, some elements suggesting the strong influence of the climate change issue on the media coverage of this controversy. Acting within this context, anti-fracking and environmental groups appear to be striving for a change in energy production paradigms. This should indicate that, also thanks to the anti-fracking cause, something is moving towards a reformulation of the basic principles regarding the relationship between man and environment. However, I argue that, despite the frequent claims for bans and moratoria, and the anti-fracking protests, both the extreme propositional terms of the controversy, as presented in the corpus, are ultimately expression of the very same cultural paradigm. This assumption is grounded on the hypothesis that, according to the results of the linguistic analysis of the corpus, the anti-fracking statement, at the negative extreme of the controversy, has an ambiguous and vulnerable basis.

Indeed, the pervasiveness of concerns about fracking is counterbalanced by hesitation about its actual effects (as shown, for example, in the large use of possibility modals), and the fact that hydraulic fracturing is seen as maybe dangerous is not enough to exclude its feasibility. It could be said that even the environmentalist cause, together with the concerns and damages suffered by local communities, become a quite weak argument if confronted with the possibility of elaborating a safer way to do it. The actual extent of damages suffered by the families and communities affected by oil and gas companies is not being discussed here. Many other issues connected to the impact

of fracking on people living above exploitable shale sites are not touched in the analysis of the corpus, as they do not show from frequency counts. Some examples are: the distinction between owning the surface rights and the *mineral rights* (appearing 12 times throughout the whole corpus)of an area; the ways in which companies appropriate the land; the partition of whole inhabited areas in drilling pads; the continuous need to drill new holes to sustain the business; the noise, smell and look of a drilling site (in a house backyard or inside a farm). However, none of these seems to have particular relevance in the corpus. On the contrary, the stress is on less everyday-life, less bureaucratically complex and more potentially dramatic aspects: water contamination, earthquakes, air pollution and dangerous methane leakages. Experts still have to find a shared answer as for whether these aspects they are unavoidable or can be somehow fixed or eliminated by more advanced technologies (one more reason to designate the scientific community as the chief context where a resolution might be found), which keeps the debate open. Furthermore, even though protests, concerns and the awareness of serious risks posed by fracking technologies appear widespread and deeply felt, there are some elements that are never ever called into question throughout the corpus:

- The legitimacy of human energy demands – a quite relevant issue in prospecting possible actions directed at reducing greenhouse gas emissions. The growth rates of world energy consumption (tripled in the last 45 years – Hughes, 2013:3) do not seem in fact to receive much attention. Rather, what seems important is to grant growing supplies to meet the demand. As renewable sources cannot provide for these supplies, this inevitably becomes an element that implicitly supports the shale gas industry.

- The nature and extent of the benefits granted by shale gas in the American economy. While the great profitability of fracking seems to be taken for granted, and plans for the future are made relying on the fracking industry, this same profitability is questioned (Rossi, 2014; Hughes, 2013, 2014); among the main arguments supporting this thesis, is the relatively short life of shale gas wells, which are said to peak after some years after being drilled to then relentlessly decline. Another aspect is related to the effectiveness of shale gas and oil in terms of net energy yield (the difference between the energy input required to extract

these hydrocarbons and the energy contained in the final product). It is argued (ibid.) that this net yield is generally much lower than for conventional resources, such as the crude oil produced by the OPEC countries. Consequently, according to these accounts, fracking could finally result in an extremely costly operation whose financial success is induced exclusively by generous investments and barely kept afloat by the continuous drilling of new wells.

- Our persistent reliance on finite energy resources such as fossil fuels – something that should be taken into consideration, especially if energy consumption is expected to grow, while the sources from which more than 80% of the world energy is obtained (ibid.) are going to peak (if they have not yet) and then decline (see the Association for the study of Peak Oil & gas at www.peakoil.net/). In the corpus, the lemma *peak* with reference to the peaking and declining oil production is only used in three articles out of 928.

These are just a few hints at a large body of knowledge strongly connected to this controversy, but almost excluded from its mass media coverage, at least for what concerns this corpus analysis. This definitely affects the characterization and interpretation of hydraulic fracturing technologies. Here fracking appears as a dangerous answer to a legitimate question, that of high energy production maintenance. The question, that is the most important condition allowing for fracking practises, is legitimized both by the media analysed in the corpus and by what most research institutions appear to communicate through newspapers and magazines in the corpus. I argue that this generalized attitude seems to be based upon a deeply rooted paradigm, which underlies the way man-environment relationships are conceived in western cultures, and frames a large part of discursive practices about climate change and energy. It defines the way in which surrounding resources are perceived and used, and gives humans a universal right to appropriate any resource felt as useful, in the quantity and time desired, regardless of consequences. In spite of all anti-fracking claims, no challenge is posed to current energy consumption and to the way natural resources are perceived by most of the world population. The collocates found for *our* in the lexico-semantic analysis (mainly referred to environmental resources) partly confirm this kind of attitude towards the environment and its resources (see paragraph 3.3.4 in chapter 3). And it is the constant reconfirmation of this paradigm which will keep fracking alive in

the debate, independently from its environmental impact, until any other factor will concretely contradict its value as a source of cheap energy. Paradoxically, the fact it contaminates water appears far more worrying than the perspective of years of research and investment into a field which will sooner or later lead to a dead end.

4.4 Conclusion

The sociological implications of this corpus analysis have helped to further elaborate and interpret the results of the linguistic analyses in the previous chapters. The main findings of this interdisciplinary approach can therefore be outlined:

- Traits of technical and scientific knowledge popularization have been identified and contextualized in the continuum model of science communication. However, the controversial circumstances in which the texts have been produced generate some 'atypical' elements with respect to the continuum model; the most visible are the mixing of scientific and popular claims in the same arena and the predominance of uncertainty regarding scientific and technical knowledge about hydraulic fracturing.

- Differently from what could be thought from the point of view of fracking opponents, the corpus does provide media coverage to anti-fracking claims, more than it does for the economic advantages brought by fracking technologies where they are applied. This might be explained with the greater marketability of the potential risks posed by fracking in the context of news making.

- Nevertheless, traditional conceptions about scientific knowledge communication and the relationship between the scientific community and the rest of society appear to exert a strong and widespread influence upon the discourse of the corpus. Scientific and technical procedures do not seem to be revealed to the public (apart from the process of fracking itself, whose descriptions remain, however, vague), thus keeping a certain distance between the status of technical and scientific groups working within the fields related to fracking and common people or activists. Moreover, there is no relevantly frequent mention of interaction or positive collaboration between them.

- The relationship between specialists and political institutions appears in mentions of studies and reports which are commissioned to research groups by public institutions and in mentions of public departments or agencies (the clearest example is that of the U.S. Environmental Protection Agency – EPA) which regularly collaborate with these experts. At the same time, the relationship between the public and political institutions (worrying, asking for regulations or bans, protesting for the former, versus granting, banning or postponing fracking authorizations for the latter) seems to preserve a certain (at times conflictual) distance between the two.

- Controversies can be said to represent a moment for potential change and re-negotiation of cultural boundaries and paradigms. The corpus represents the struggle of anti-fracking environmentalist groups to make their way towards greater scientific credibility and more widespread consideration. Moreover, some anti-fracking groups acting in the context of the broader field of the climate change debate are trying to take advantage of the serious preoccupations caused by fracking to enhance a shift from high-emission and non-renewable fossil fuels to carbon-free renewable forms of energy production. However, it is argued that this is not a sufficiently deep modification to actually induce any real change in the way western cultures frame the relationship between humans and environment resources –a relationship chiefly based on reckless exploitation. The concerns over fracking are often connected to short-term damages, still to be fully and officially acknowledged. It is instead argued that some characteristics of shale gas, mostly neglected in the corpus are to be found in more long-term features. First comes its finite nature, something that cannot be confuted by any authoritative report, and should affect any projections, forecasts and expectations concerning its production and consumption. Second, the environmental impact of shale gas production, as any energy production system (and ultimately any human activity) should be calculated taking into account all the energy input and the environmental impact involved in the whole process (for instance, from the drilling of the well to the storage and burning of the fuel).

This account is certainly incomplete and affected by a lack of competence in most of the disciplinary areas touched in the text. However, one important concept should at

least have been showed: the controversy over fracking was grammatically, lexically and semantically (to sum up, linguistically) framed in one of the many possible interpretations that could be attributed to it. It is highly probable that this framing depends on both norms of news selection and construction and on deeply rooted cultural paradigms.

Chapter 5

Conclusion

5.1 A brief overview of results and interpretations

A corpus of nearly 1000 newspaper and magazine texts about the controversy over fracking has been collected and then examined using MD Analysis and elaborating on word frequency counts. The MD Analysis revealed that the corpus uses a slightly more involved and more overtly persuasive language than the average values for the press, which can be ascribed to the controversial nature of the subject. This controversial aspect was supported, in the lexico-semantic analysis, by the fields of words identified (those regarding the 'Fracking procedures', the 'Controversial side' of fracking and the 'Actors' taking part in the controversy) and by the high frequency of both prediction and possibility modals. Possibility modals, together with the many opinions expressed through predictions and overt (or covert, not detectable) persuasion, contributed to give an idea of widespread uncertainty, partly subverting the public construction of popularized technoscience, which is usually 'packaged' and presented as an established theory or technology. The representation of fracking, however, meets the expectations for what concerns the language of popularization, in that, in approaching a theme with technical and scientific implications, it is unmarked both in the use of narration versus description and in the degree of complexity of the language. Among the various sections, 'Magazines and Journals' emerges as the most detached, and less lexically and conceptually standardized section. 'Newspapers' shows a lesser lexical variability with respect to the main fields, is more concerned with local issues than magazines are and is the section which seems to leave more scope to politics. It is also the section where the reporting verb *say* is most frequently used, a typical aspect of this genre, perhaps reinforced by the need to place a focus on accuracy and comprehensiveness, reporting a number of conflicting opinions whose validity is not fully assessed. This aspect is less stressed in 'Magazines and Journals', whereas 'Letters' is the section least concerned

with it. 'Letters' is the most involved, less informative and most persuasive section. As expected, it mainly focuses on commenting. Moreover, it is the section which shows the largest internal variability overall, probably owing to the lack of news-making constraints. The role of economic actors is present with considerable frequency in all sections.

Following the integration of the results with data deriving from a sociological approach, several aspects emerged:

- The space devoted to the possible damages caused by fracking appears larger than that devoted to its possible advantages; this has been explained through the news-selection criterion which gives more media relevance to news represented in a dramatic way;

- The relationship between the scientific community and the lay public is certainly affected by the controversy (which allows a greater interaction) but remains a distant one. Scientific and technical procedures are not disclosed to the public, nor is the public involved in technoscientific or political decisions, except for those protests (enacted by citizens or anti-fracking activists) which succeed in obtaining a ban.

- Although the environmentalist cause (an umbrella term for numerous and diversified claims, expressed by many different groups, variously connected among themselves) seems to be striving to change the current paradigms of energy production and the related course of action regarding climate change, there seems to be a deeper cultural paradigm that remains unchanged. It is the western exploitation and consumption model, which also ultimately justifies hydraulic fracturing.

5.2 Other observations

It is important to bear in mind that analysing a controversy means accepting the inevitable complexity of collective life. This implies that any attempt at building and maintaining simplification, whether in acting within or in representing or analysing a controversy (or in all of these together)requires a lot of work. This does not mean that

simplification should be deserted. In fact, it is often necessary to understand and consequently participate in collective life. Nevertheless, it must be performed carefully, especially with controversies, where some elements of society are constructed and consolidated while others are simultaneously deconstructed and loosened. This continuous and dynamic process has been compared to the changes of state typical of flowing magma, where the solid rocks affected by the heat of the flow melt and join the flow, while the lava at the margins of the stream cools down and solidifies (Venturini, 2010). Thus, controversies might be described as 'magmatic'. Equally, it is important to bear in mind that the conclusions drawn in this study only refer to the period covered by the corpus (besides some sporadic references to earlier events, or to later events happening at the time of writing), therefore limited to a five-year long stage of that 'magmatic flow'. It follows that these conclusions do not intend to delineate events still to happen. Studying a controversy does not mean being able to predict which course it will take in the long-term future (let alone by only considering its media coverage). Envisaging what will happen, for example by imagining possible future outcomes regarding the development of fracking and its consequences in the next decades, can be certainly seen as part of the processes we usually carry out to make sense of our existence. The importance of these processes is revealed in the corpus by the unusually high frequency and the keyness values of the modal *will*. However, as powerful and influential as they may be, long-term forecasts, such as that envisaging "a supply of natural gas that can last America nearly 100 years"[7] are often communicated and received without taking into account a rather important element. Long-term predictions, like technologies, are indeed shaped according to the current social context, that in which they are produced, whereas societies change, as change their underlying paradigms, and, as a consequence, the way technoscience is developed and understood. If it is true that conceptions about the future of a certain technology can be determinant for its future outcomes, it is also true that no one today can be certain about those outcomes, because the future involves different contexts, different paradigms, different societies (and at least partly different people) from those of today (Neresini, 2011).

---

[7] Words pronounced by the U.S. President Obama in the *Remarks by the President in State of the Union Address* on January 24, 2012 (transcript retrieved in June 2015 at www.whitehouse.gov/the-press-office/2012/01/24/remarks-president-state-union-address).

Another element worth taking into account, albeit often overlooked, concerns the way in which technoscientific innovation and the attendant risks are commonly rationalized. It is a widespread assumption that technologies are, in themselves, nor bad nor good – in other words, neutral – while it is their application which can have positive or negative consequences. This assumption has been challenged, by arguing that technologies (and the technical and scientific research at their basis) do not exist separately from their applications; they are all framed within social practices, which disproves their neutrality. Likewise, a number of other elements contribute to the shaping of technologies, from conception to final application: among these are politics, economics, the underlying system of technical knowledge, the research instruments used and theories about the behaviour of the natural environment (Neresini, 2011; Bijker, 1992). If technoscientific innovation is regarded as resulting from the combination of many different factors, this implies that compromises have to be reached (in more or less conflictual ways) and that positive together with negative consequences are connected to these technologies all along their lifecycle, from its very beginning. Despite this ambivalence and complexity are normally perceived as part of our existence, they sometimes tend to be conceptualized differently: for what concerns technologies, this results in the ideal separation of 'pure' technologies from their good or bad applications. Such distinction leads to a dualist perception of innovation, according to which technologies can become either good and therefore useful, or bad, due to the risks they pose (Lupton in Neresini, 2011: 117-121). In turn, the meaning of *risk*, originally comprising both a positive and a negative feature, has been reduced to its negative self (so that the meaning of *risky technology* is almost synonymous with *dangerous technology)*, while a useful technology only produces positive effects[8].

In the case here studied, where the controversial element marks this uncertainty, fracking technologies are still debated, and their ambivalence stands out in the plurality of viewpoints expressed in the corpus. Opposite tendencies can be found, attempting at providing a definitive characterization of fracking as either risky or useful. In order to

---

[8]According to some sociologists, this semantic restriction is due to the growing feeling of unpredictability concerning the consequences of human actions, especially those dealing with the technoscientific enhancement. At least with regards to the cultural systems considered in this study (mainly linked to western models), scientific knowledge, traditionally attributed absolute truth values, has proved mostly incapable of guiding mankind through its choices and the implied consequences (Bucchi and Neresini in Neresini, 2011: 117-121).

observe these tendencies, I found it useful to distinguish between short-term and long-term perspectives:

- Now (short-term): in some cases, fracking has proved not only risky, but concretely dangerous. Interestingly enough, as argued in chapter 3, the verb *cause* (whose relative percentage frequency in the corpus is 0,12), either affirmed or negated, might indicate the tendency to link fracking to the environmental and health consequences described, in spite of the different attitudes adopted. However, it is also argued by some stakeholders that negative consequences are being caused by a wrong, or not fully developed application of these technologies. It seems that they are thus trying to re-separate the technology from its applications, in order to drag the perception of fracking towards the useful side. The media, as operatives in Hilgartner and Bosk's theory[9] (1988: 57), attempt to " "surf" on the shifting waves of social problems" (Nolan , Hilgartner and Bosk in Hilgartner and Bosk,1988: 67) and represent this critical moment, exploiting its appeal to the public (as both the involved and persuasive features in MD Analysis and the lexico-semantic analysis seem to show).

- Then (long-term): according to some of the groups who oppose fracking, characterizing it as dangerous, this practice does not only have inherent immediate environmental effects, but it will also have a bearing on global warming, because, as all fossil fuels, shale gas and oil burning produces carbon dioxide. Conversely, the advocates of fracking argue that shale gas is a 'cleaner' energy than coal and oil, and that it will assure the U.S. sufficient energy supplies to sustain its economy and to develop even 'greener' energy solutions to finally solve the problem of climate change. It is as if fracking advocates had appropriated long-term concerns to maintain that this is in any case a useful and therefore good technology. This particular frame allows for and is corroborated by declarations like that made by the U.S. president (reported above), or by the U.S. Energy Information Administration (EIA), whose reports state that the national gas production will continue to grow significantly at least until 2040 (Hughes, 2014: 301; EIA, 2015: 20). Similar claims have recently been made

---

[9] In their paper, Hilgartner and Bosk define operatives as the groups or individuals who publicly present social problems.

about the UK. For example, Britain's leading shale gas explorer company affirmed that developing a UK shale industry would provide the country with "an alternative to imports from Russia and the Middle East, at a time when North Sea oil and gas output is in decline, cutting costs for industry and creating jobs"[10] . Some consider these projections as incomplete, and far too optimistic (Hughes, 2013, 2014; Rossi, 2014), thus arguing that the shale revolution will lead to a disastrous failure in the next years, with serious environmental damages, energy shortage and no alternatives to contrast global warming.

While the short-term perspectives appeared to be widely covered by the corpus, only some of the long-term aspects received attention. The climate change issue, related to carbon dioxide emissions, is present especially in magazine articles, and, less frequently, in letters to the editor and newspaper articles; the idea of 'bridging fuel' is absent from letters, and not frequent in the rest of the corpus; the questioning of official estimates on future energy production in the U.S. or in other countries does not emerge from the lexico-semantic analysis (even though there is no evidence that it has been completely ignored). Thus, in the conflict between useful and risky or dangerous characterizations of hydraulic fracturing, it seems to me that the arguments in favour of the 'useful' option have revealed a certain versatility in keeping open the possibility of fracking as an ultimately good technology. This versatility is strengthened by the little relevance given by the analysed media to viewpoints confuting official projections regarding natural gas production. Moreover, as observed in chapter 4, the corpus appears to reflect a cultural system deeply rooted in a western capitalist model, whose first concern is short-termed and whose priority is self- perpetuation by sustaining human activities through adequate supplies. The characterization of hydraulic fracturing (either polarized or not) is likely to be strongly influenced by such cultural paradigms, which are at the basis of its initial ambivalence. Finally, the economic aspect (mainly expressed in the form of the actors *industry* and *company* in the corpus) has probably had an extremely relevant role in this characterization, especially in relation to the context of energy production, which is key to the survival of western lifestyles and models as they are today.

---

[10] *Financial Times Weekend*, 13 June/14 June 2015, National edition, p.3.

5.3 Further considerations on the role of the public

The controversy over fracking was first started because of the consequences it had on some of the communities where it was first applied; so it can be said that the public played a fundamental role in igniting the controversy. But the fact that residents (and then groups of activists) had to call for regulations and raise protests in order to have a say about fracking indicates that the public was involved only late in this process of technical and scientific knowledge production and application. Yet this is not the only way in which common people can interact with science and technoscientific innovation.

PCST studies have highlighted a number of possibilities (some already experimented, some only theorized) directed towards a new kind of public engagement in scientific knowledge production, communication and application. In an analysis of PCST in several EU countries, Irwin (in Bucchi, Trench, 2008: 199-212) identified different modes of PCST, whose application and usefulness is context-specific. One is the traditional way (which he calls 'first-order thinking'), based on positivistic assumptions and thus regarding science as the exclusive embodiment of truth. First-order thinking practices tend to exclude the public, deemed essentially ignorant and irrational, from both knowledge production and decisional processes. It also, consequently, contributes to protect the current economic stability and interests, and the traditional status of the scientific community. A second mode, defined as 'second-order thinking', started to appear in institutional PCST initiatives at the beginning of the 21st century: notions of openness, trust and transparency were introduced in institutional discourse, according to principles of deliberative democracy (explained in Einsiedel and Irwin, in Bucchi, Trench, 2008: 178-9, 200). The public did not appear anymore as a passive and irrational crowd, but was regarded as capable of rationality and as a potential resource in risk situations. According to Irwin, these two approaches operate in a complex and sometimes ambiguous coexistence, also due to the fact that policy-makers try to use them according to their own interests. Also Einsiedel (in Bucchi , Trench:2008: 173-184) documents the rise, in recent years, of public participation in technoscientific processes. Standing at the innovative end of a continuum which links it to more traditional approaches, these PCST practices are based on three main points: access to information, participation in decision-making and judicial redress when

necessary. Furthermore, the role of social movements as knowledge producers becomes more and more important in this context. According to these new perspectives (which appear close to Irwin's second-order thinking), the lay public, in all its different and shifting components, is considered to bring a "broader range of values" (ibid.) to bear on technologic and scientific questions. Einsiedel also argues that, while the traditional forms of PCST take place relatively late – that is when technology is applied and commercialized – the newer public engagement-inspired forms take place at a much earlier stage (which is referred to as 'upstream engagement'). In line with the deliberative democracy values underlying the new approaches, the way technology is shaped and assessed could also be affected: the public could be involved in technology shaping and design through activities of dialogue and feedback (which is in defined as 'Constructive Technology Assessment'). A further, even more ground-breaking step towards the inclusion of lay public in technoscientific practices is the notion of social shaping of technology, according to which such practices should be completely disclosed to the public. Furthermore, it has been theorized that a complete and pervasive dialogue between the scientific community and the lay public could be implemented: hence the elaboration of concepts such as the 'deconstruction of expertise' and the 'social distribution of science' (ibid.).

Such collective practices do however pose some issues: firstly, public engagement might be manipulated, or used by particularly powerful groups to legitimize their actions, while they should be open to the possibility of those actions being questioned by public engagement (Einsiedel in Bucchi, Trench, 2008: 189); secondly, the more groups are engaged, the more contrasts are likely to arise, the more difficult it is to obtain consensus in decision-making; thirdly, the public (and all the groups taking part in it) have their own interests, and do not necessarily take their decisions by disengaging themselves from these interests (which, argues Einsiedel, would be unrealistic and unfair). That is why it is suggested that forms of public engagement be mediated and combined with other models (what Irwin calls 'third-order thinking') according to the strengths and limits of each of them, and according to the context in which PCST needs to be practiced.

My analysis of the corpus could only partially try to figure out which degree of public engagement was taking place (or was represented as taking place) during the

controversy over fracking. As explained in chapter 4, however, what could be observed is a sort of mixture between Irwin's first and second-order thinking, slightly leaning towards the first-order. As a matter of fact, the public does not appear as a completely passive, ignorant and irrational crowd: lay opinions are given space, together with more expert ones, and it cannot be ascertained, but neither could it be excluded that part of the technical reports and opinions mentioned were produced by environmental organizations. Besides, words such as *anti-*, *protest*, *campaign* point to an active involvement of sections of the public in the controversy, and words like *ban* and *moratorium* seem to indicate that their requests are considered and sometimes accepted. Nevertheless, in terms of word frequency, there is no opening to the public of technoscientific procedures other than approximate descriptions of how a well is drilled and fracked. Thus lay actors appear completely separated from the technical and scientific world of expertise where such technologies are created. Moreover, there does not seem to be any relevantly frequent reference to public engagement initiatives, such as the examples mentioned by Einsiedel (in Bucchi, Trench, 2008: 176): consensus conferences, citizen juries, scenario workshops, deliberative mapping, online discussions, and so on.[11] The role of economic actors, finally, must not be forgotten, in that their involvement in the controversy and powerful influence are likely to have heavily affected both political decisions and the effectiveness of collective actions.

Surely, enough, newspapers and magazines are not the only media which represent this controversy. Consequently, it is possible that other media reflect a different state of affairs. I am referring, in particular, to the internet (in a broader sense than that of online articles), which more than other media has blurred the boundaries between specialized and lay communication (Trench, 2008: 185-198). This has been made possible thanks to the great variety of sources retrievable, and to the publication of many scientific research documents online, free for anyone to read. Moreover, thanks to the internet, the public has the chance to directly interact with members of the scientific community or with scientific journalists through the social media. As a consequence, if my analysis had included internet-based documents, the outcome in terms of public engagement might have been different. This does not automatically imply, however, that deliberative

---

[11] Einsiendel's study focused on the EU context. Given the fact that the corpus texts were mainly published in the U.S. and in the U.K., the partial shift from a European to a mixed context may be reflected by the corpus.

democracy has been realized, and that public engagement and PCST have been harmoniously integrated thanks to the internet: collective life is extremely intricate and unstable even online (see paragraph 5.1 at the beginning of this chapter). As Trench (ibid.) argues, the World Wide Web favours both phenomena of boundary-breaking and of fragmentation; it favours collaboration, but also competition. It certainly allows the scientific community to practice a 'backstage' discourse (see paragraph 4.3.2 in chapter 4). Yet, backstage practices could become an instrument in the hands of the scientific community to exert a certain control over popularization and deviation cases in order to maintain their traditionally powerful status. Moreover, lay people can find serious difficulties in understanding specialized languages and, given the plurality of sources and the lack of peer- reviewing, those documents cannot always be validated. This is even more so when online arenas host groups with various collocations with respect to the scientific community (outside it, or close to its borders), which results in rather confusing and difficult-to-manage situations, especially from the point of view of lay audiences.

5.4 Some ideas for further research

Suggestions for further development of this research quite obviously start from its limitations. Therefore, using the same methods of analysis (both linguistic and sociological), a diachronic perspective could be introduced by dividing the texts according to the date of publication, in order to observe how the representation of the controversy might have changed over the years. A diatopic perspective could also be adopted (maybe by previously enriching the corpus in order for it to include sections of comparable sizes), so that the various national contexts could be considered and characterized. Along the same lines, other languages could be included in the same kind of corpus-based analysis, maybe in a translation studies perspective, so as to contextualize the results in the respective source cultures. This would however require elaborating MD Analysis methods and software compatible with languages other than English. Another possibility would be that of extending the analysis to other texts genres: in particular, I am thinking about texts produced in websites or social media,

which might help understand what is the role of the internet and which aspects of the above described (end of paragraph 5.2) are predominant in the public representation of this controversy (if any of them emerges as predominant). Other text genres which I think would yield relevant information are: documents produced by public institutions; documents produced by NGOs and other environmental groups; communications made by oil and gas companies; works released by organizations such as the Post Carbon Institute, the Energy Watch Group, and ASPO, which establish themselves as (and in fact represent) scientifically competent alternatives to the official and mainstream body of knowledge, challenging at the same time some of the tenets of western development models.

For what concerns the MD Analysis part, much more in-depth analysis can be carried out about Biber's Dimensions of variation in relation to the texts, by examining which features most influenced the obtained scores, in order to work out much more accurate interpretations. An integration with detailed individual text analysis might also contribute to accuracy, and possibly to the detection of some instances of covert persuasion – which would in turn allow a comparison between overt and covert persuasion in the analysed corpus. Finally, variation among scientific texts and different degrees of technical and scientific popularization could be further examined, in order to extract more detailed patterns regarding scientific discourse.

5.5 Final remarks

This is a glimpse into the complex relationship between linguistic choices and the way reality is shaped and understood, and the corpus-based perspective enables researchers to see it in a considerable amount of texts. Language has been argued to express representational trends which in turn can be said to reproduce the way the controversy has evolved and is evolving (with respect to the time when the study is written). Researching representational trends means assuming that there were other representational possibilities. In fact, there were, and there are always alternatives: the controversy might have been shaped otherwise (see Bijker, 1992); the media might have represented it in a completely different way and I might have conducted a completely

different study. But choices have been made, and every choice had its own reason. Here, linguistic choices have been analysed and matched against political, economic and technological choices, marking those matches and connections which seemed to appear most clearly.

This work has indeed been conceived as part of a complex and wider context, as a (however personal and necessarily limited) attempt at greater awareness of the predominant way in which western culture perceives, understands and represents the relationship humans have with the rest of the Earth. To raise this awareness also means to relativize it, starting to understand that, despite needing a way to systematize and codify reality, societies can also evolve and change their representational and conceptual systems. Raising awareness means understanding that there were, there are, and there will be alternatives; and that every choice has its own reasons, but also has consequences. Moreover, that consequences are not only seen in a single instance, nor only in the short run, but also by looking at the whole system (as broadly as it can be considered, through a corpus or through more extensive sociological analyses), and at long-term developments. In shaping and managing the controversy over fracking, thus, both individuals (politicians, scientists, engineers, geologists, activists, company CEOs, land owners and common people) and whole communities and societies are fully responsible for their own choices, while being at the same time under the – more or less powerful – influence of the other actors[12] involved. It has been argued that our ever-evolving social practices give meaning and legitimize the ways we understand and manipulate natural phenomena (and therefore the way technoscience is present in society), making them plausible and useful. In other words, the way we know and represent the world cannot be separated from the way in which we choose to be in the world (Latour, Jasanoff in Neresini, 2011: 97-99). As social and technical change are so strongly connected, then, if we want to understand either, we have to try to understand both (Bijker 1992: 11). In line with these assumption, this work tries to include the use of language, a social practice, as an important part of that social and therefore also technical process whose components are continuously and reciprocally shaped. The

---

12 I would like here to give actors the meaning Actor-Network Theory-based 'controversy mapping' approach indicates for this term: "not only human beings or human groups, but also natural and biological elements, industrial and artistic products, institutional and economic institutions, scientific and technical artifacts and so on and so forth" (Venturini, 2010).

journalist Russell Gold (2014: 36) summarized the current state of the controversy with these words:

> "The Earth is warming, and once the source rock [the shale, my clarification] is depleted, the era of fossil fuel will end whether we are ready or not. Fracking has changed the energy industry and is changing the world around us. It is here to stay."

This might be interpreted as showing a stark contradiction: in what sense is fracking here to stay, when even shale resources are definitely going to be depleted? Otherwise, this could sound as a perfectly consistent acknowledgment of the fact that hydraulic fracturing is already there. This technological innovation has already taken place, carrying all its consequences with it, and only future choices and future societies as they develop will determine its role in the outcome of energy and climate change issues. After all, to put it in Bijker's words (1992: 3), the technology we get is the technology we deserve.

References

Anthony, L. (2014). *AntConc* (Version 3.4.1) [Computer Software]. Tokyo, Japan: Waseda University Available from http://www.antlab.sci.waseda.ac.jp/ .

Anthony, L. (2014). *TagAnt* (Version 1.1.2) [Computer Software]. Tokyo, Japan: Waseda University Available from http://www.antlab.sci.waseda.ac.jp/ .

Bednarek, M. (2006). *Evaluation in Media Discourse: Analysis of a Newspaper Corpus.* London ; New York : Continuum

Berkenkotter, C., Huckin, T. N. (1995). *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power.* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Biber, D. (1988). *Variation across Speech and Writing.* Cambridge: Cambridge University Press.

Biber, D. (1989). "A typology of English texts". *Linguistics*, *27*(1), 3–43.

Biber, D. et al. (1999). *Longman Grammar of Spoken and Written English.* London: Longman.

Biber, D., Connor, U., Upton, T.A. (2007) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Philadelphia: John Benjamins Publishing Company.

Bijker, W., Law, J. (eds.) (1992). *Shaping Technology/Building Society : Studies in Sociotechnical Change*. London: MIT Press.

Bucchi, M. (2000). *La scienza in pubblico*: *percorsi nella comunicazione scientifica.* Milano: McGraw-Hill.

Bucchi, M., Trench, B. (eds.) (2008). *Handbook of Public Communication of Science and Technology.* London; New York: Routledge.

Castello, E. (2008). *Text Complexity and Reading Comprehension Tests*. Bern: Peter Lang.

Conrad, S., Biber, D. (eds.) (2001) *Variation in English: Multi-Dimensional Studies.* Harlow: Pearson Education.

FREE CLAWS WWW tagger. UCREL, Lancaster University. Available from http://ucrel.lancs.ac.uk/claws/trial.html .

Halliday, M. A. K., (edited by Webster, J.J.) (2005) *Computational and Quantitative Studies.* London; New York: Continuum.

Hilgartner, S., Bosk, C. L. (1988). "The Rise and Fall of Social Problems: A Public Arenas Model" *American Journal of Sociology* Vol. 94, No. 1 (Jul., 1988), pp. 53-78.

Hughes, J. D. (2013) *Drill, Baby, Drill. Can Unconventional Fuels Usher in a New Era of Energy Abundance?* Post Carbon Institute. Retrieved May 2015 from http://shalebubble.org/drill-baby-drill/

Hughes, J. D. (2014). *Drilling Deeper. A Reality Check on U.S. Government Forecasts for a Lasting Tight Oil and Shale Gas Boom.* Post Carbon Institute. Retrieved May 2015 from http://www.postcarbon.org/drilling-deeper/ .

Hyland, K. (2005) *Metadiscourse: Exploring Interaction in Writing*. New YorK: Continuum

Myers G, Macnaghten P. (1998), "Rhetorics of Environmental Sustainability: Commonplaces and Places", in *Environment and Planning A* 30(2) 333 – 353.

Myers G,(2003). "Discourse Studies of Scientific Popularisation: Questioning the Boundaries", in *Discourse Studies*, Vol. 5, No. 2, 05., p. 265-279.

Neresini (2011). *Il nano-mondo che verrà: verso la società nanotecnologica.* Bologna: Il Mulino

Nini, A. 2015. *Multidimensional Analysis Tagger* (Version 1.3) [Computer Software]. Available (February 2015) from: http://sites.google.com/site/multidimensionaltagger .

Observa-Science in society [Website] http://www.observa.it/?lang=it .

Palumbo, G., Musacchio, M. T. (2010). "When a Clue is not a Clue: a Corpus-Driven Study of Explicit vs. Implicit Signalling of Sentence Links in Popular Economics Translation", in: *Rivista internazionale di tecnica della traduzione = International Journal of Translation*, 12 pp. 63-76.

Palumbo, G., Musacchio, M. T. (2010) "Following Norms, Taking Risks: A Study of the Use of Connectives in a Corpus of Translated Economics Articles in Italian*,* in Heine, Carmen/Egberg, Jan (eds.) *Reconceptualizing LSP*. Online proceedings of the XVII European LSP Symposium 2009. Aarhus.

Pozzan, R. *Shale Chaos*, episode from the Italian television programme "Report" available (May 2014) from http://www.report.rai.it/dl/Report/puntata/ContentItem-dbe5e600-c50c-4673-89d6-451aeb09c011.html .

Rossi, M. (2014). "Fracking: la lusinga dei derrick- Miti e realtà dei nuovi idrocarburi non convenzionali" . Retrieved May 2014 from *http://www.aspoitalia.it/index.php/articoli/comunicati- ufficiali/336-fracking-il-fascino-indiscreto-dei-derrick* .

Schmid, H. (1994). "Probabilistic Part-of-Speech Tagging Using Decision Trees". *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK . Retrieved September 2014 from http://www.cis.uni-muenchen.de/~schmid/ .

Scott, M., (2012). *WordSmith Tools* (version 6) [Computer Software]. Liverpool: Lexical Analysis Software.

Someya, Y., (1998). English Lemma List. Retrieved February 2015 from http://www.laurenceanthony.net/software/antconc/ .

Stubbs, M. (1996). *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture.* Oxford: Blackwell.

U.S. Energy Information Administration (2015). *Annual Energy Outlook 2015. With*

Projections to 2040. Retrieved June 2015 at
http://www.eia.gov/forecasts/aeo/pdf/0383%282015%29.pdf .

Venturini, T. (2010). "Diving in magma: How to explore controversies with actor-network theory", in "Public understanding of science", vol. 19(3), pp. 258-273.

Zittel, W. et al. (2013). Fossil and Nuclear Fuels-the Supply Outlook. Energy Watch Group. Retrieved May 2015 from http://energywatchgroup.org/press-release-18-march-2013-fossil-nuclear-fuels-supply-outlook/ .

Appendix

Graphs and tables

Graph 2.1. Dimension 1-corpus (situated at the bottom/right end) vs Biber's genres.
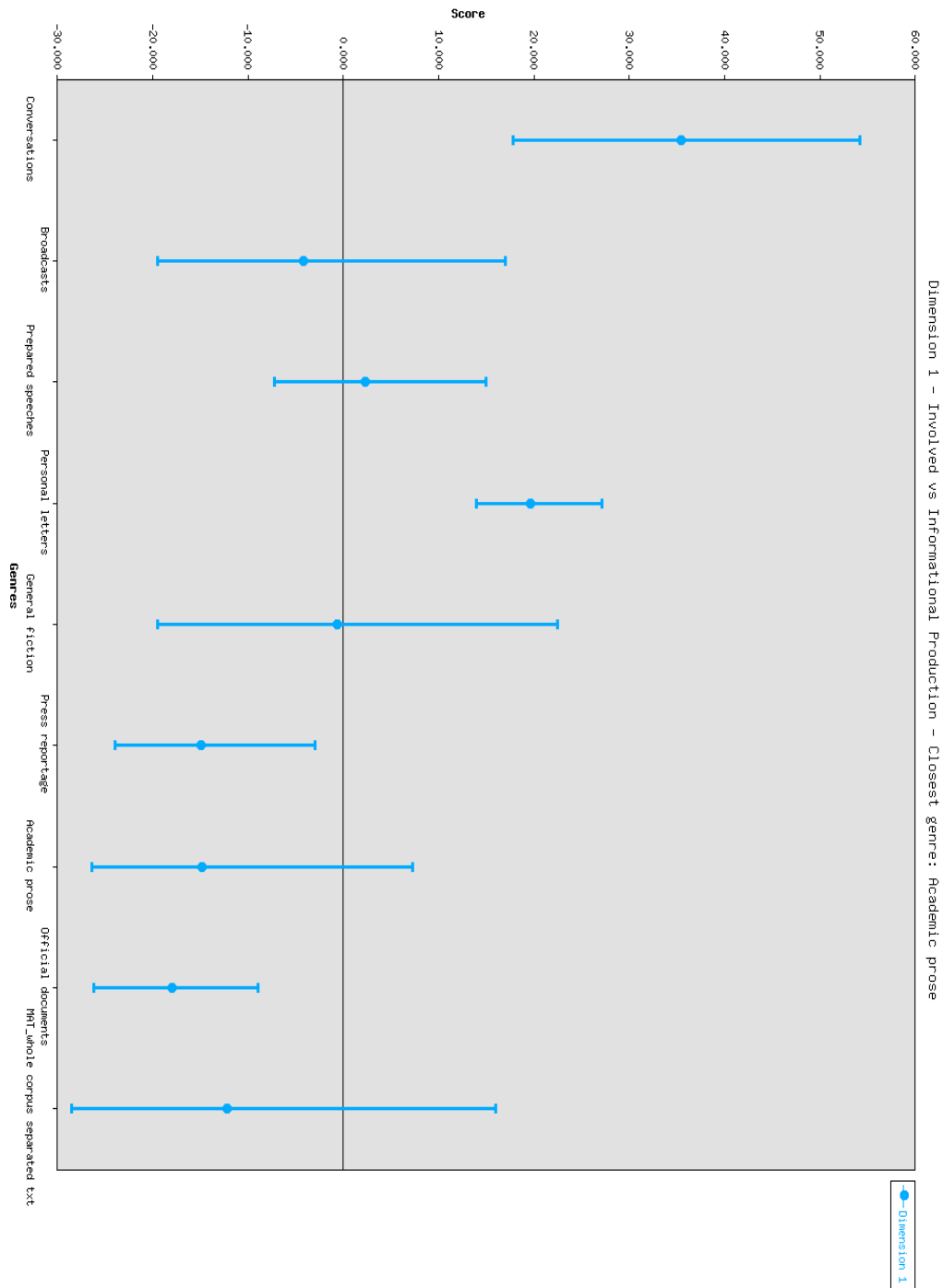
Graph 2.2. Dimension 2-corpus (situated at the bottom/right end) vs Biber's genres.
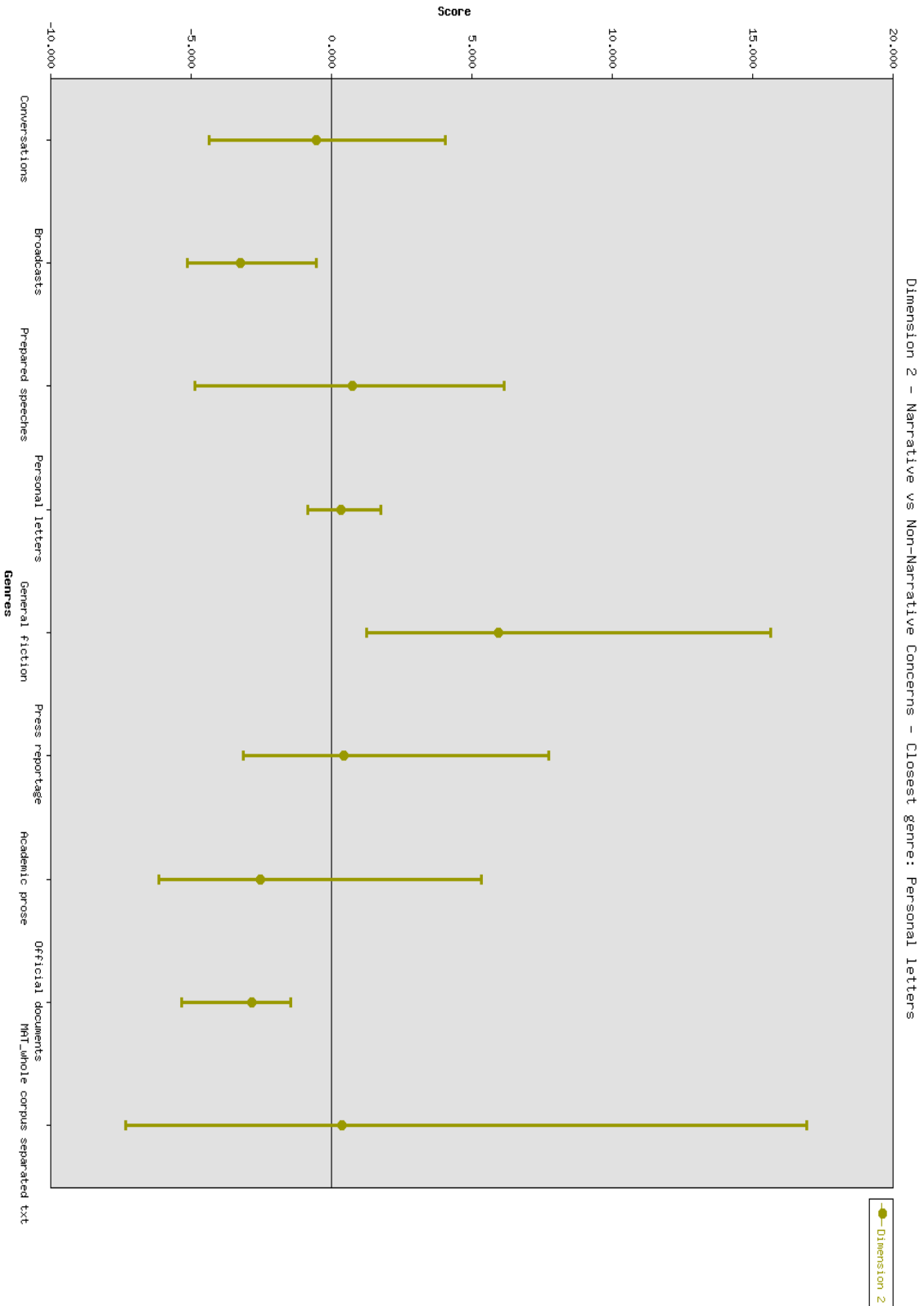
Graph 2.3. Dimension 4-corpus (situated at the bottom/right end) vs Biber's genres.
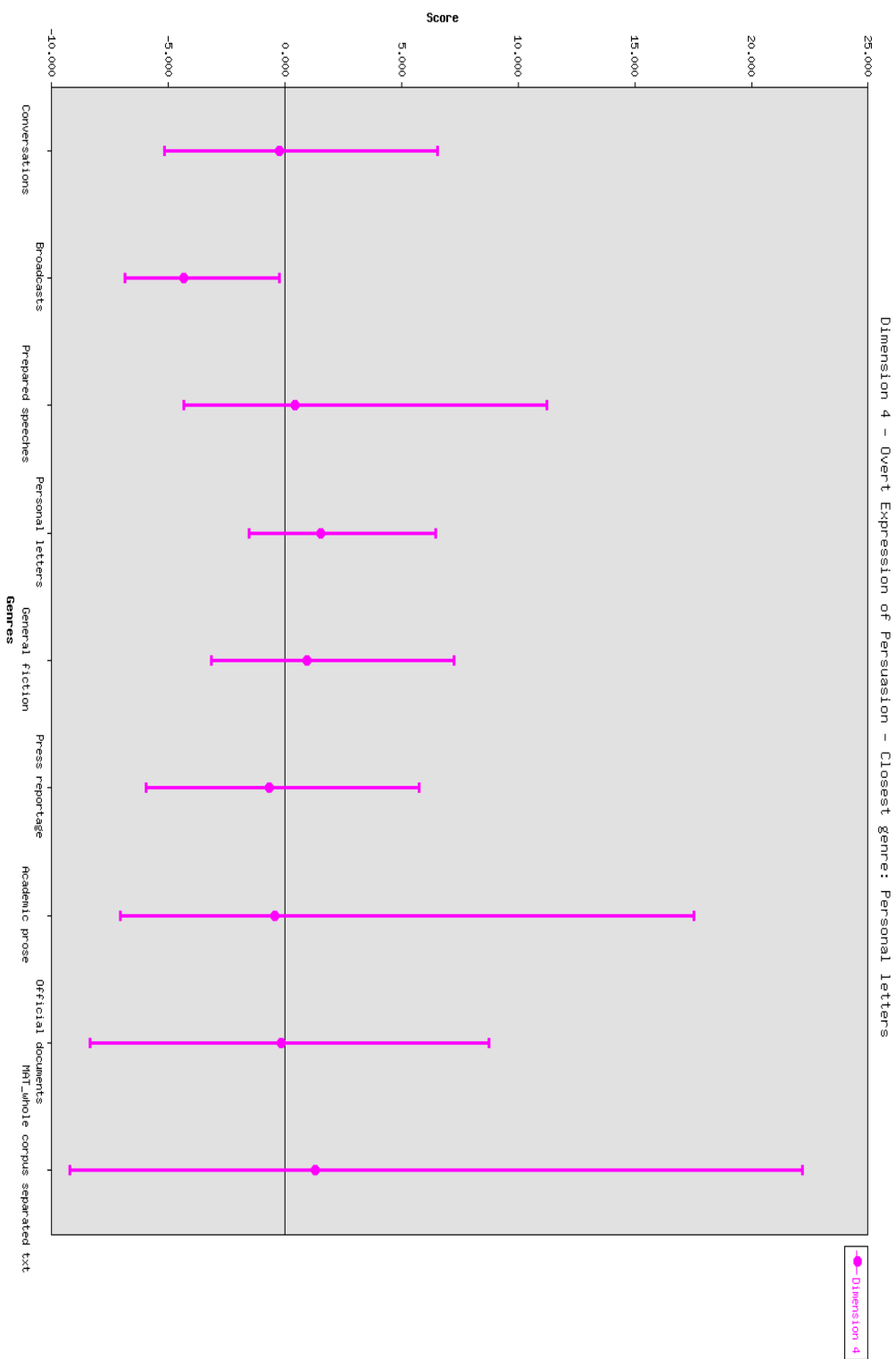
Graph 2.4. Dimension 5-corpus (situated at the bottom/right end) vs Biber's genres.
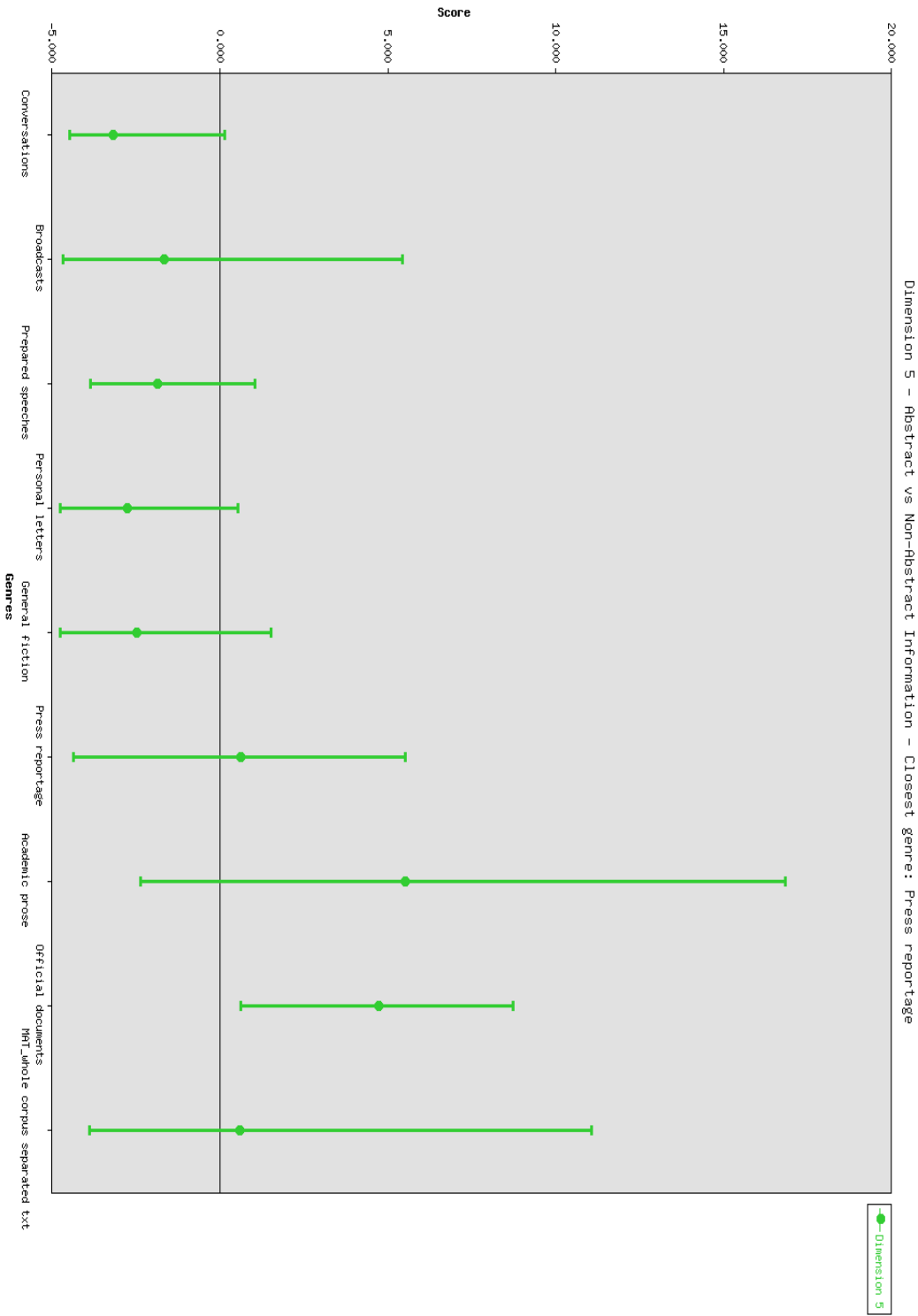
| Wordlist-Full Corpus |
N	Word	Freq.	%	Lemmas
1	FRACKING	8793	2,26	fracking[8439] frack[257] fracked[94] fracks[3]
2	GAS	3581	0,92	gas[3537] gases[44]
3	SAY	3252	0,84	say[315] said[2385] saying[109] says[443]
4	WATER	2136	0,55	water[2091] watered[1] watering[1] waters[43]
5	WILL	1617	0,42	will[1614] ll[3]
6	ENERGY	1403	0,36	energy[1396] energies[7]
7	SHALE	1394	0,36	
8	OIL	1357	0,35	
9	WOULD	1266	0,33	
10	DRILL	1168	0,30	drill[178] drilled[88] drilling[889] drills[13]
11	STATE	1152	0,30	state[823] stated[40] states[279] stating[10]
12	USE	1130	0,29	use[349] used[544] uses[88] using[149]
13	NEW	1094	0,28	new[1087] newer[5] newest[2]
14	COMPANY	939	0,24	company[354] companies[585]
15	MORE	923	0,24	
16	CHEMICAL	874	0,22	chemical[122] chemicals[752]
17	COULD	871	0,22	
18	ENVIRONMENTAL	842	0,22	
19	CAN	829	0,21	
20	NATURAL	814	0,21	
21	FRACTURE	785	0,20	fracture[99] fractured[22] fractures[47] fracturing[617]
22	YEAR	755	0,19	year[409] years[346]
23	INDUSTRY	744	0,19	industry[709] industries[35]
24	PROCESS	722	0,19	process[670] processed[10] processes[28] processing[14]
25	ALSO	681	0,18	
26	GOVERNMENT	674	0,17	government[628] governments[46]
27	WELLS	661	0,17	
28	PEOPLE	626	0,16	people[623] peoples[3]
29	REPORT	619	0,16	report[463] reported[63] reporting[17] reports[76]
30	HYDRAULIC	617	0,16	
31	CONCERN	615	0,16	concern[113] concerned[157] concerning[8] concerns[337]
32	COUNTY	612	0,16	county[546] counties[66]
33	BAN	609	0,16	ban[437] banned[99] banning[38] bans[35]
34	MAKE	606	0,16	make[242] made[222] makes[60] making[82]
35	PUBLIC	587	0,15	
36	NEED	578	0,15	need[305] needed[119] needs[154]
37	COUNCIL	577	0,15	council[491] councils[86]
38	AREA	550	0,14	area[334] areas[216]
39	ROCK	548	0,14	rock[432] rocking[1] rocks[115]
40	TAKE	543	0,14	take[241] taken[98] takes[52] taking[93] took[59]
41	GROUP	542	0,14	group[366] groups[176]
42	ISSUE	535	0,14	issue[300] issued[72] issues[147] issuing[16]
43	ENVIRONMENT	514	0,13	environment[507] environments[7]
44	PLAN	504	0,13	plan[111] planned[46] planning[167] plans[180]
45	TIME	500	0,13	time[264] times[231] timing[5]
46	STUDY	497	0,13	study[318] studied[16] studies[142] studying[21]
47	LOCAL	488	0,13	local[465] locals[23]
48	MAY	486	0,12	
49	RELEASE	479	0,12	release[269] released[137] releases[40] releasing[33]
50	CAUSE	474	0,12	cause[200] caused[161] causes[46] causing[67]

Table 3.1.1. 100 Most frequent words in the whole corpus. A stoplist and a lemmatization list were applied

Wordlist-Full Corpus				
N	Word	Freq.	%	Lemmas
51	HIGH	458	0,12	high[372] higher[60] highest[23] highs[3]
52	RISK	442	0,11	risk[222] risked[2] risking[2] risks[216]
53	HEALTH	436	0,11	
54	KNOW	434	0,11	know [187] knew [11] know ing[6] know n[203] know s[27]
55	SITE	434	0,11	site[216] sited[1] sites[216] siting[1]
56	COMMUNITY	431	0,11	community[192] communities[239]
57	IMPACT	427	0,11	impact[281] impacted[6] impacting[4] impacts[136]
58	INCLUDE	409	0,11	include[74] included[50] includes[43] including[242]
59	LAST	407	0,10	last[399] lasting[6] lasts[2]
60	COME	406	0,10	come[178] came[50] comes[101] coming[77]
61	ANTI	405	0,10	
62	SUCH	402	0,10	
63	FIND	401	0,10	find[101] finding[31] finds[19] found[250]
64	PLACE	390	0,10	place[316] placed[29] places[44] placing[1]
65	ALLOW	388	0,10	allow [205] allow ed[80] allow ing[65] allow s[38]
66	POTENTIAL	388	0,10	
67	COAL	386	0,10	coal[385] coals[1]
68	CALL	384	0,10	call[113] called[179] calling[57] calls[35]
69	SEE	381	0,10	see[248] saw [23] seeing[15] seen[82] sees[13]
70	RESOURCE	378	0,10	resource[81] resources[297]
71	GET	374	0,10	get[233] gets[36] getting[44] got[59] gotten[2]
72	COUNTRY	369	0,09	country[278] countries[91]
73	WANT	368	0,09	w ant[232] w anted[40] w anting[3] w ants[93]
74	NORTH	361	0,09	
75	SUPPORT	358	0,09	support[259] supported[40] supporting[30] supports[29]
76	NOW	355	0,09	
77	EVEN	350	0,09	even[286] evening[64]
78	FLUID	350	0,09	fluid[184] fluids[166]
79	UK	349	0,09	
80	BILL	346	0,09	bill[253] bills[93]
81	MORATORIUM	343	0,09	moratorium[335] moratoriums[8]
82	MANY	341	0,09	
83	EARTHQUAKE	339	0,09	earthquake[57] earthquakes[282]
84	PRESSURE	338	0,09	pressure[310] pressured[3] pressures[24] pressuring[1]
85	IT'S	336	0,09	
86	WASTE	336	0,09	w aste[301] w asted[1] w astes[32] w asting[2]
87	CITY	333	0,09	city[301] cities[32]
88	SUPPLY	333	0,09	supply[145] supplied[8] supplies[175] supplying[5]
89	CHANGE	331	0,09	change[241] changed[32] changes[43] changing[15]
90	GREEN	330	0,08	green[246] greener[12] greenest[4] greens[68]
91	UNDERGROUND	330	0,08	
92	REGULATION	328	0,08	regulation[106] regulations[222]
93	GIVE	326	0,08	give[106] gave[27] given[137] gives[19] giving[37]
94	WAY	322	0,08	w ay[277] w ays[45]
95	MUCH	318	0,08	
96	METHANE	317	0,08	
97	OPERATION	317	0,08	operation[86] operations[231]
98	JOB	316	0,08	job[67] jobs[249]
99	EXTRACT	315	0,08	extract[155] extracted[61] extracting[91] extracts[8]
100	INVOLVE	314	0,08	involve[33] involved[88] involves[178] involving[15]

Table 3.1.1. 100 Most frequent words in the whole corpus. A stoplist and a lemmatization list were applied

				Wordlist-Newspapers
N	**Word**	**Freq.**	**%**	**Lemmas**
1	FRACKING	6518	2,52	fracking[6281] frack[185] fracked[51] fracks[1]
2	SAY	2696	1,04	say[233] said[2153] saying[98] says[212]
3	GAS	2244	0,87	gas[2228] gases[16]
4	WATER	1424	0,55	water[1401] watered[1] watering[1] waters[21]
5	WILL	1125	0,43	will[1124] ll[1]
6	WOULD	927	0,36	
7	ENERGY	880	0,34	energy[875] energies[5]
8	SHALE	850	0,33	
9	OIL	849	0,33	
10	STATE	802	0,31	state[588] stated[35] states[172] stating[7]
11	USE	792	0,31	use[236] used[395] uses[57] using[104]
12	DRILL	785	0,30	drill[106] drilled[48] drilling[623] drills[8]
13	WORD	706	0,27	word[20] worded[1] wording[2] words[683]
14	NEW	687	0,27	new[685] newer[1] newest[1]
15	COMPANY	660	0,25	company[259] companies[401]
16	ENVIRONMENTAL	599	0,23	
17	COULD	582	0,22	
18	FRACTURE	581	0,22	fracture[74] fractured[9] fractures[33] fracturing[465]
19	PROCESS	565	0,22	process[524] processed[9] processes[21] processing[11]
20	NATURAL	552	0,21	
21	CHEMICAL	547	0,21	chemical[61] chemicals[486]
22	MORE	542	0,21	
23	GOVERNMENT	529	0,20	government[499] governments[30]
24	COUNCIL	505	0,19	council[441] councils[64]
25	CAN	494	0,19	
26	COUNTY	486	0,19	county[436] counties[50]
27	CONCERN	467	0,18	concern[79] concerned[125] concerning[6] concerns[257]
28	HYDRAULIC	466	0,18	
29	PEOPLE	465	0,18	people[463] peoples[2]
30	ALSO	464	0,18	
31	REPORT	461	0,18	report[350] reported[43] reporting[6] reports[62]
32	YEAR	457	0,18	year[263] years[194]
33	BAN	451	0,17	ban[316] banned[76] banning[29] bans[30]
34	GROUP	450	0,17	group[309] groups[141]
35	INDUSTRY	422	0,16	industry[408] industries[14]
36	ISSUE	412	0,16	issue[246] issued[39] issues[113] issuing[14]
37	PUBLIC	407	0,16	
38	PLAN	401	0,15	plan[88] planned[38] planning[132] plans[143]
39	AREA	400	0,15	area[247] areas[153]
40	TAKE	400	0,15	take[178] taken[71] takes[38] taking[73] took[40]
41	NEED	395	0,15	need[214] needed[77] needs[104]
42	ENVIRONMENT	392	0,15	environment[387] environments[5]
43	ROCK	387	0,15	rock[308] rocks[79]
44	MAKE	383	0,15	make[154] made[145] makes[25] making[59]
45	WELLS	381	0,15	
46	LOCAL	379	0,15	local[365] locals[14]
47	ANTI	345	0,13	
48	TIME	333	0,13	time[168] times[161] timing[4]
49	CAUSE	327	0,13	cause[139] caused[113] causes[29] causing[46]
50	RELEASE	327	0,13	release[196] released[88] releases[28] releasing[15]

Table 3.1.2.100 most frequent words in 'Newspapers'. A stoplist and a lemmatization list were appl

				Wordlist-Newspapers
N	**Word**	**Freq.**	**%**	**Lemmas**
51	STUDY	325	0,13	study[215] studied[10] studies[88] studying[12]
52	COMMUNITY	322	0,12	community[145] communities[177]
53	SITE	310	0,12	site[161] sites[149]
54	CALL	304	0,12	call[95] called[140] calling[40] calls[29]
55	HIGH	304	0,12	high[263] higher[28] highest[13]
56	PLACE	304	0,12	place[253] placed[21] places[30]
57	IMPACT	299	0,12	impact[200] impacted[4] impacting[1] impacts[94]
58	ALLOW	297	0,11	allow[162] allowed[61] allowing[43] allows[31]
59	KNOW	297	0,11	know[125] knew[6] knowing[4] known[148] knows[14]
60	LAST	296	0,11	last[291] lasting[5]
61	MAY	293	0,11	
62	POTENTIAL	290	0,11	
63	RESOURCE	288	0,11	resource[65] resources[223]
64	CITY	285	0,11	city[259] cities[26]
65	RISK	285	0,11	risk[138] risked[2] risking[2] risks[143]
66	INCLUDE	283	0,11	include[52] included[37] includes[28] including[166]
67	MORATORIUM	282	0,11	moratorium[275] moratoriums[7]
68	GREEN	270	0,10	green[207] greener[7] greenest[4] greens[52]
69	WANT	269	0,10	want[165] wanted[32] wanting[2] wants[70]
70	SUPPORT	268	0,10	support[198] supported[30] supporting[21] supports[19]
71	IT'S	265	0,10	
72	COUNTRY	263	0,10	country[208] countries[55]
73	GET	261	0,10	get[156] gets[30] getting[26] got[48] gotten[1]
74	NORTH	257	0,10	
75	HEALTH	256	0,10	
76	BILL	254	0,10	bill[183] bills[71]
77	EARTHQUAKE	254	0,10	earthquake[39] earthquakes[215]
78	BUSINESS	250	0,10	business[232] businesses[18]
79	GIVE	250	0,10	give[82] gave[23] given[108] gives[12] giving[25]
80	COME	249	0,10	come[103] came[33] comes[63] coming[50]
81	SOUTH	244	0,09	
82	INVOLVE	241	0,09	involve[29] involved[63] involves[143] involving[6]
83	UNDERGROUND	240	0,09	
84	PRESSURE	239	0,09	pressure[229] pressured[2] pressures[8]
85	CONTROVERSIAL	237	0,09	
86	SEE	235	0,09	see[148] saw[16] seeing[9] seen[56] sees[6]
87	RESIDENT	232	0,09	resident[24] residents[208]
88	NOW	231	0,09	
89	SUCH	229	0,09	
90	NATIONAL	226	0,09	
91	UK	223	0,09	
92	REGULATION	222	0,09	regulation[68] regulations[154]
93	LAND	219	0,08	land[194] lands[25]
94	JOB	218	0,08	job[40] jobs[178]
95	EXTRACT	217	0,08	extract[110] extracted[33] extracting[69] extracts[5]
96	RULE	217	0,08	rule[34] ruled[16] rules[153] ruling[14]
97	PRACTICE	216	0,08	practice[186] practiced[1] practices[29]
98	EXPLORATION	215	0,08	
99	SUPPLY	214	0,08	supply[92] supplied[1] supplies[120] supplying[1]
100	FIND	212	0,08	find[57] finding[20] finds[6] found[129]

Table 3.1.2.100 most frequent words in 'Newspapers'. A stoplist and a lemmatization list were a

N	Word	Freq.	%	Lemmas
	Wordlist-Magazines and Journals			
1	GAS	972	1,30	gas[948] gases[24]
2	FRACKING	795	1,06	fracking[735] frack[35] fracked[23] fracks[2]
3	SHALE	405	0,54	
4	SAY	373	0,50	say[47] said[104] saying[4] says[218]
5	OIL	348	0,46	
6	ENERGY	329	0,44	energy[328] energies[1]
7	WATER	329	0,44	water[327] waters[2]
8	DRILL	264	0,35	drill[46] drilled[33] drilling[181] drills[4]
9	MORE	255	0,34	
10	NEW	252	0,34	new[248] newer[4]
11	WILL	249	0,33	will[247] ll[2]
12	YEAR	229	0,31	year[114] years[115]
13	CAN	212	0,28	
14	COULD	210	0,28	
15	NATURAL	194	0,26	
16	WELLS	191	0,25	
17	STATE	185	0,25	state[125] stated[1] states[59]
18	USE	185	0,25	use[59] used[88] uses[13] using[25]
19	CHEMICAL	177	0,24	chemical[39] chemicals[138]
20	INDUSTRY	177	0,24	industry[167] industries[10]
21	COAL	173	0,23	
22	WOULD	173	0,23	
23	€	162	0,22	
24	METHANE	160	0,21	
25	COMPANY	146	0,19	company[42] companies[104]
26	MAKE	137	0,18	make[49] made[49] makes[27] making[12]
27	MAY	137	0,18	
28	ENVIRONMENTAL	136	0,18	
29	FRACTURE	131	0,17	fracture[17] fractured[7] fractures[14] fracturing[93]
30	EMISSION	130	0,17	emission[1] emissions[129]
31	FIND	124	0,17	find[27] finding[9] finds[8] found[80]
32	ROCK	123	0,16	rock[92] rocking[1] rocks[30]
33	ALSO	122	0,16	
34	MUCH	121	0,16	
35	SUCH	116	0,15	
36	STUDY	109	0,15	study[67] studied[3] studies[33] studying[6]
37	PRICE	103	0,14	price[46] prices[54] pricing[3]
38	AMERICAN	101	0,13	american[89] americans[12]
39	COME	101	0,13	come[47] came[13] comes[23] coming[18]
40	HIGH	100	0,13	high[67] higher[26] highest[4] highs[3]
41	SEE	99	0,13	see[69] saw[5] seeing[4] seen[15] sees[6]
42	TIME	98	0,13	time[60] times[37] timing[1]
43	REPORT	96	0,13	report[68] reported[11] reporting[10] reports[7]
44	PRODUCTION	92	0,12	
45	HYDRAULIC	91	0,12	
46	HEALTH	90	0,12	
47	EVEN	87	0,12	
48	NEED	85	0,11	need[34] needed[27] needs[24]
49	POWER	85	0,11	power[80] powered[3] powers[2]
50	CARBON	84	0,11	

Table 3.1.3. 100 Most frequent words in 'Magazines and Journals'. A stoplist and a lemmatization list were applied

N	Word	Freq.	%	Lemmas
	Wordlist-Magazines and Journals			
51	MANY	84	0,11	
52	RELEASE	84	0,11	release[31] released[37] releases[6] releasing[10]
53	DEEP	83	0,11	deep[76] deeper[7]
54	WAY	83	0,11	w ay[60] w ays[23]
55	CONTAMINATION	80	0,11	
56	NOW	80	0,11	
57	Â	79	0,11	
58	FUEL	79	0,11	fuel[54] fuelling[1] fuels[24]
59	TWO	79	0,11	
60	CAUSE	78	0,10	cause[32] caused[24] causes[8] causing[14]
61	CHANGE	77	0,10	change[58] changed[7] changes[6] changing[6]
62	KNOW	77	0,10	know [27] knew [1] know ing[1] know n[41] know s[7]
63	OPERATION	77	0,10	operation[17] operations[60]
64	UK	77	0,10	
65	GROUNDWATER	76	0,10	
66	LARGE	76	0,10	large[45] larger[5] largest[26]
67	PUBLIC	76	0,10	
68	RISK	76	0,10	risk[42] risks[34]
69	BIG	75	0,10	big[46] bigger[8] biggest[21]
70	WORK	75	0,10	w ork[41] w orked[4] w orking[21] w orks[9]
71	MOST	74	0,10	
72	AMERICA	73	0,10	
73	NORTH	73	0,10	
74	TAKE	73	0,10	take[30] taken[17] takes[6] taking[9] took[11]
75	FLUID	72	0,10	fluid[37] fluids[35]
76	IMPACT	72	0,10	impact[56] impacting[1] impacts[15]
77	JUST	72	0,10	
78	MILLION	72	0,10	million[58] millions[14]
79	COUNTRY	71	0,09	country[43] countries[28]
80	GET	71	0,09	get[46] gets[4] getting[11] got[9] gotten[1]
81	DRINK	70	0,09	drink[1] drank[2] drinking[64] drunk[3]
82	FIRST	70	0,09	
83	LAST	70	0,09	last[69] lasts[1]
84	SUPPLY	70	0,09	supply[29] supplied[7] supplies[32] supplying[2]
85	UNIVERSITY	70	0,09	
86	AREA	69	0,09	area[37] areas[32]
87	WORLD	69	0,09	w orld[68] w orlds[1]
88	YORK	69	0,09	
89	CLIMATE	68	0,09	
90	INCLUDE	68	0,09	include[12] included[8] includes[5] including[43]
91	SITE	68	0,09	site[33] sites[35]
92	AGENCY	67	0,09	agency[63] agencies[4]
93	REGULATION	67	0,09	regulation[20] regulations[47]
94	SCIENTIST	67	0,09	scientist[45] scientists[22]
95	TECHNOLOGY	67	0,09	technology[47] technologies[20]
96	GLOBAL	65	0,09	
97	PROCESS	65	0,09	process[58] processed[1] processes[4] processing[2]
98	SURFACE	65	0,09	
99	EPA	64	0,09	
100	GOVERNMENT	64	0,09	government[55] governments[9]

Table 3.1.3. 100 Most frequent words in 'Magazines and Journals'. A stoplist and a lemmatization list were applied

N	Word	Freq.	%	Lemmas
	Wordlist-Letters			
1	FRACKING	922	2,73	fracking[892] frack[20] fracked[10]
2	WATER	238	0,70	water[236] waters[2]
3	GAS	230	0,68	gas[226] gases[4]
4	WILL	154	0,46	
5	LETTER	139	0,41	letter[48] letters[91]
6	ENERGY	121	0,36	energy[120] energies[1]
7	WORD	119	0,35	word[5] worded[3] words[111]
8	WOULD	110	0,33	
9	COUNTY	94	0,28	county[85] counties[9]
10	INDUSTRY	93	0,28	industry[85] industries[8]
11	USE	89	0,26	use[29] used[34] uses[13] using[13]
12	CAN	85	0,25	
13	MORE	81	0,24	
14	NEW	78	0,23	new[77] newest[1]
15	CHEMICAL	76	0,23	chemical[11] chemicals[65]
16	PEOPLE	72	0,21	people[71] peoples[1]
17	STATE	72	0,21	state[42] stated[3] states[25] stating[2]
18	NEED	70	0,21	need[39] needed[9] needs[22]
19	OIL	68	0,20	
20	COMPANY	62	0,18	company[20] companies[42]
21	SHALE	62	0,18	
22	BAN	59	0,17	ban[42] banned[10] banning[5] bans[2]
23	CAUSE	59	0,17	cause[23] caused[21] causes[9] causing[6]
24	DRILL	59	0,17	drill[13] drilled[5] drilling[41]
25	ENVIRONMENTAL	57	0,17	
26	MAKE	57	0,17	make[27] made[17] makes[6] making[7]
27	COMMUNITY	56	0,17	community[32] communities[24]
28	PROCESS	56	0,17	process[53] processes[2] processing[1]
29	MANY	54	0,16	
30	CONCERN	53	0,16	concern[13] concerned[11] concerning[1] concerns[28]
31	HEALTH	53	0,16	
32	PUBLIC	53	0,16	
33	BUTTE	51	0,15	
34	COULD	51	0,15	
35	CALIFORNIA	50	0,15	
36	GOVERNMENT	50	0,15	government[46] governments[4]
37	LOCAL	50	0,15	
38	OPINION	50	0,15	opinion[47] opinions[3]
39	WELLS	50	0,15	
40	ENVIRONMENT	49	0,15	
41	ALSO	47	0,14	
42	AREA	47	0,14	area[33] areas[14]
43	SAY	47	0,14	say[23] said[16] saying[4] says[4]
44	WANT	47	0,14	want[34] wanted[5] wanting[1] wants[7]
45	NATURAL	46	0,14	
46	WASTE	46	0,14	waste[39] wastes[7]
47	EVEN	45	0,13	even[31] evening[14]
48	COUNCIL	43	0,13	council[25] councils[18]
49	GROUNDWATER	43	0,13	
50	TAKE	43	0,13	take[22] taken[3] takes[5] taking[8] took[5]

Table 3.1.4. Most frequent 100 words in 'Letters A stoplist and a lemmatization list were applied

N	Word	Freq.	%	Lemmas
				Wordlist-Letters
51	RISK	41	0,12	risk[22] risks[19]
52	SAFE	41	0,12	safe[40] safer[1]
53	CONTAMINATION	40	0,12	
54	FACT	40	0,12	fact[22] facts[18]
55	ISSUE	40	0,12	issue[22] issued[1] issues[15] issuing[2]
56	TIME	40	0,12	time[18] times[22]
57	MAY	39	0,12	
58	TOXIC	39	0,12	
59	CLEAN	38	0,11	clean[30] cleaned[2] cleaner[4] cleanest[1] cleaning[1]
60	COME	38	0,11	come[20] came[2] comes[9] coming[7]
61	KNOW	38	0,11	know[19] knew[4] knowing[1] known[8] knows[6]
62	WAY	38	0,11	way[37] ways[1]
63	SEE	37	0,11	see[24] saw[1] seeing[1] seen[10] sees[1]
64	MILLION	36	0,11	million[20] millions[16]
65	SUPPLY	36	0,11	supply[17] supplies[17] supplying[2]
66	VERY	36	0,11	
67	FIND	35	0,10	find[15] finding[2] finds[1] found[17]
68	PROTECT	35	0,10	protect[24] protected[6] protecting[5]
69	STUDY	35	0,10	study[20] studied[3] studies[12]
70	THINK	35	0,10	think[25] thinking[3] thinks[2] thought[5]
71	EDITOR	34	0,10	editor[33] editors[1]
72	FUEL	34	0,10	fuel[25] fuels[9]
73	GIVE	34	0,10	give[11] gave[2] given[13] gives[2] giving[6]
74	REPORT	34	0,10	report[21] reported[6] reporting[1] reports[6]
75	YEAR	34	0,10	year[11] years[23]
76	MOST	33	0,10	
77	SUPPORT	33	0,10	support[22] supported[3] supporting[6] supports[2]
78	AIR	32	0,09	
79	CONTAMINATE	32	0,09	contaminate[9] contaminated[20] contaminates[2] contaminating[1]
80	GET	32	0,09	get[23] gets[1] getting[6] got[2]
81	INCLUDE	32	0,09	include[7] included[1] includes[4] including[20]
82	JUST	32	0,09	
83	LAND	32	0,09	land[31] lands[1]
84	NATIONAL	32	0,09	
85	SUCH	32	0,09	
86	UK	32	0,09	
87	BENEFIT	31	0,09	benefit[12] benefited[1] benefits[18]
88	IT'S	31	0,09	
89	PRICE	31	0,09	price[11] prices[18] pricing[2]
90	AUGUST	30	0,09	
91	CHANGE	30	0,09	change[22] changed[2] changes[2] changing[4]
92	HERE	30	0,09	
93	HIGH	30	0,09	high[21] higher[4] highest[5]
94	JOB	30	0,09	job[4] jobs[26]
95	FRACTURE	29	0,09	fracture[5] fractured[1] fracturing[23]
96	LIFE	29	0,09	life[26] lives[3]
97	POTENTIAL	29	0,09	
98	ALLOW	28	0,08	allow[15] allowed[6] allowing[5] allows[2]
99	CASE	28	0,08	case[17] cases[10] casing[1]
100	CREATE	28	0,08	create[14] created[7] creates[7]

Table 3.1.4. Most frequent 100 words in 'Letters A stoplist and a lemmatization list were applied

word	% corpus	% newsp	% mag	% letters
FRACKING	2,260	2,516	1,060	2,730
GAS	0,921	0,866	1,296	0,681
SAY	0,836	1,041	0,497	0,139
WATER	0,549	0,550	0,439	0,705
WILL	0,416	0,434	0,332	0,456
ENERGY	0,361	0,340	0,439	0,358
SHALE	0,358	0,328	0,540	0,184
OIL	0,349	0,328	0,464	0,201
WOULD	0,325	0,358	0,231	0,326
DRILL	0,300	0,303	0,352	0,175
STATE	0,296	0,310	0,247	0,213
USE	0,290	0,306	0,247	0,264
NEW	0,281	0,265	0,336	0,231
COMPANY	0,241	0,255	0,195	0,184
MORE	0,237	0,209	0,340	0,240
CHEMICAL	0,225	0,211	0,236	0,225
COULD	0,224	0,225	0,280	0,151
ENVIRONMENTAL	0,216	0,231	0,181	0,169
CAN	0,213	0,191	0,283	0,252
NATURAL	0,209	0,213	0,259	0,136
FRACTURE	0,202	0,224	0,175	0,086
YEAR	0,194	0,176	0,305	0,101
INDUSTRY	0,191	0,163	0,236	0,275
PROCESS	0,186	0,218	0,087	0,166
ALSO	0,175	0,179	0,163	
GOVERNMENT	0,173	0,204	0,085	0,148
WELLS	0,170	0,147	0,255	0,148
PEOPLE	0,161	0,180		0,213
REPORT	0,159	0,178	0,128	0,101
HYDRAULIC	0,159	0,180	0,121	
CONCERN	0,158	0,180		0,157
COUNTY	0,157	0,188		0,278
BAN	0,157	0,174		0,175
MAKE	0,156	0,148	0,183	0,169
PUBLIC	0,151	0,157	0,101	0,157
NEED	0,149	0,152	0,113	0,207
COUNCIL	0,148	0,195		0,127
AREA	0,141	0,154	0,092	0,139
ROCK	0,141	0,149	0,164	
TAKE	0,140	0,154	0,097	0,127
GROUP	0,139	0,174		
ISSUE	0,138	0,159		0,118
ENVIRONMENT	0,132	0,151		0,145
PLAN	0,130	0,155		
TIME	0,129	0,129	0,131	0,118
STUDY	0,128	0,125	0,145	0,104
LOCAL	0,125	0,146		0,148
MAY	0,125	0,113	0,183	0,115
RELEASE	0,123	0,126	0,112	
CAUSE	0,122	0,126	0,104	0,175
HIGH	0,118	0,117	0,133	0,089
RISK	0,114	0,110	0,101	0,121
HEALTH	0,112	0,099	0,120	0,157
KNOW	0,112	0,115	0,103	0,113
SITE	0,112	0,120	0,091	
COMMUNITY	0,111	0,124		0,166
IMPACT	0,110	0,115	0,096	
INCLUDE	0,105	0,109	0,091	0,095
LAST	0,105	0,114	0,093	
COME	0,104	0,096	0,135	0,113
ANTI	0,104	0,133		
SUCH	0,103	0,088	0,155	0,095
FIND	0,103	0,082	0,165	
PLACE	0,100	0,117		
ALLOW	0,100	0,115		0,083
POTENTIAL	0,100	0,112		0,086
COAL	0,099		0,231	
CALL	0,099	0,117		
SEE	0,098	0,091	0,132	0,110
RESOURCE	0,097	0,111		
GET	0,096	0,101	0,095	0,095
COUNTRY	0,095	0,102	0,095	
WANT	0,095	0,104		0,139
NORTH	0,093	0,099	0,097	
SUPPORT	0,092	0,103		0,098
NOW	0,091	0,089	0,107	
EVEN	0,090			
FLUID	0,090		0,096	0,104
UK	0,090	0,086	0,103	0,095
BILL	0,089	0,098		

Table3.2. All words appearing in the 100-word lists of the three
sections with their percentage frequencies. The frequencies were
coloured in gradual shades from dark red (high) to light blue (low) .

word	% corpus	% newsp	% mag	% letters
MORATORIUM	0,088	0,109		
MANY	0,088			0,160
EARTHQUAKE	0,087	0,098		
PRESSURE	0,087	0,092		
IT'S	0,086	0,102		0,092
WASTE	0,086			0,136
CITY	0,086	0,110		
SUPPLY	0,086	0,083	0,093	0,107
CHANGE	0,085		0,103	0,089
GREEN	0,085	0,104		
UNDERGROUND	0,085	0,093		
REGULATION	0,084	0,086	0,089	
GIVE	0,084	0,097		0,101
WAY	0,083		0,111	0,113
MUCH	0,082		0,161	
METHANE	0,081		0,213	
OPERATION	0,081		0,103	
JOB	0,081	0,084		0,089
EXTRACT	0,081	0,084		
INVOLVE	0,081	0,093		
BUSINESS		0,097		
SOUTH		0,094		
CONTROVERSIAL		0,091		
RESIDENT		0,090		
NATIONAL		0,087		0,095
LAND		0,085		0,095
RULE		0,084		
PRACTICE		0,083		
EXPLORATION		0,083		
EMISSION			0,173	
PRICE			0,137	0,092
PRODUCTION			0,123	
EVEN			0,116	0,133
POWER			0,113	
CARBON			0,112	
MANY			0,112	
DEEP			0,111	
CONTAMINATION			0,107	0,118
FUEL			0,105	0,101
TWO			0,105	
GROUNDWATER			0,101	0,127
LARGE			0,101	
BIG			0,100	
WORK			0,100	
MOST			0,099	0,098
AMERICA			0,097	
JUST			0,096	0,095
MILLION			0,096	0,107
DRINK			0,093	
FIRST			0,093	
UNIVERSITY			0,093	
WORLD			0,092	
YORK			0,092	
CLIMATE			0,091	
AGENCY			0,089	
SCIENTIST			0,089	
TECHNOLOGY			0,089	
GLOBAL			0,087	
SURFACE			0,087	
EPA			0,085	
DEPARTMENT			0,081	
BUTTE				0,151
CALIFORNIA				0,148
OPINION				0,148
ALSO				0,139
SAFE				0,121
FACT				0,118
TOXIC				0,115
CLEAN				0,113
VERY				0,107
PROTECT				0,104
THINK				0,104
AIR				0,095
CONTAMINATE				0,095
BENEFIT				0,092
HERE				0,089
LIFE				0,086
CASE				0,083
CREATE				0,083

Table3.2. All words appearing in the 100-word lists of the three sections with their percentage frequencies. The frequencies were coloured in gradual shades from dark red (high) to light blue (low) .

N	Keyword	Freq.	%	RC. Freq.	RC. %	Keyness
	Wordsmith Keyword list-corpus vs BNC					
1	FRACKING	8439	2,17	0	0,0000	93791,92
2	GAS	3581	0,92	7434	0,0074	25918,93
3	SHALE	1394	0,36	168	0,0002	14402,61
4	SAY	3252	0,83	67228	0,0676	10262,75
5	DRILL	1168	0,30	1035	0,0010	9921,48
6	WATER	2136	0,55	34134	0,0343	7728,30
7	FRACTURE	785	0,20	401	0,0004	7194,67
8	OIL	1357	0,35	10158	0,0102	6785,28
9	ENERGY	1403	0,36	12098	0,0122	6653,83
10	HYDRAULIC	617	0,16	264	0,0003	5771,20
11	BYLINE	487	0,12	34	0,0000	5151,53
12	CHEMICAL	874	0,22	4454	0,0045	4975,59
13	WELLS	661	0,17	1472	0,0015	4703,95
14	LENGTH	884	0,23	7543	0,0075	4208,29
15	ENVIRONMENTAL	842	0,22	8411	0,0084	3765,86
16	PG	415	0,11	276	0,0003	3676,15
17	BAN	609	0,16	3128	0,0031	3458,03
18	ANTI	405	0,10	378	0,0004	3411,25
19	MORATORIUM	343	0,09	204	0,0002	3083,99
20	FRACK	257	0,07	0	0,0000	2850,91
21	NATURAL	814	0,21	14082	0,0142	2825,69
22	STATE	1152	0,30	37907	0,0381	2689,42
23	EARTHQUAKE	339	0,09	478	0,0005	2655,50
24	METHANE	317	0,08	332	0,0003	2619,78
25	GROUNDWATER	297	0,08	286	0,0003	2488,89
26	NEWS	750	0,19	14193	0,0143	2481,90
27	WORDS	893	0,23	23681	0,0238	2418,83
28	CONTAMINATE	244	0,06	66	0,0001	2386,18
29	ROCK	548	0,14	6676	0,0067	2251,72
30	CONCERN	615	0,16	10243	0,0103	2176,89
31	CONTAMINATION	301	0,08	637	0,0006	2166,74
32	SECTION	760	0,2	18725	0,0187	2156,87
33	EDITION	400	0,11	2470	0,0025	2138,71
34	COUNTY	612	0,16	10972	0,0110	2084,73
35	INDUSTRY	744	0,19	19242	0,0193	2047,76
36	FLUID	350	0,09	1633	0,0016	2047,07
37	KAROO	188	0,05	14	0,00001	1983,82
38	COMPANY	939	0,24	35947	0,0361	1950,87
39	RESOURCE	378	0,10	2617	0,0026	1942,81
40	RELEASE	479	0,12	6558	0,0066	1866,05
41	EXTRACT	315	0,08	1460	0,0015	1846,04
42	CUADRILLA	158	0,04	0	0,0000	1752,66
43	PROCESS	722	0,19	22499	0,0226	1752,53
44	UNDERGROUND	330	0,08	2221	0,0022	1712,98
45	WASTEWATER	157	0,04	4	0,0000	1704,14
46	EXTRACTION	242	0,06	564	0,0006	1703,86
47	USE	1130	0,29	62273	0,0626	1681,94
48	REGULATION	328	0,08	2554	0,0026	1615,72
49	COAL	386	0,10	4998	0,0050	1542,95
50	PENNSYLVANIA	196	0,05	326	0,0003	1485,82

Table 3.4.1. Wordsmith Tools-generated keyword list containing the words with the fifty highest keyness measures. From left to right, the table shows: rank, keyword, frequency in the corpus, percentage frequency in the corpus, frequency in the reference corpus (BNC), percentage frequency in the reference corpus, keyness.The basis wordlist of the corpus was built with a stoplist and a lemmatization list, which has probably inflated the keyness of lemmatized items. For the lemmatized items, the percentage frequency of the words in the BNC were not computed by the program and have been therefore added manually. They refer to the frequencies reported in this tab and are colored in blue. This table should help to have an approximate idea of the differences in frequency between corpus and reference corpus despite the lemmatization. Lines coloured in light grey indicate that the frequency of the word is altered by Lexisnexis headings and should not be considered.

AntCon keyword list-corpus vs BNC			
Rank	Keyword	Freq.	Keyness
1	be	10432	418,78
2	fracking	7993	330,62
3	gas	3456	132,44
4	say	3145	119,76
5	have	2805	105,93
6	water	2084	76,70
7	will	1659	59,57
8	shale	1336	55,26
9	energy	1392	48,88
10	drill	1141	47,20
11	oil	1327	46,29
12	would	1264	43,78
13	state	1216	41,87
14	use	1123	38,18
15	-	870	35,99
16	fracture	788	32,60
17	company	963	31,87
18	more	922	30,26
19	can	896	29,24
20	length	884	28,77
21	could	871	28,27
22	conjurer	682	28,21
23	chemical	853	27,56
24	wells	661	27,34
25	environmental	841	27,10
26	hydraulic	616	25,48
27	word	925	25,27
28	natural	794	25,27
29	industry	766	24,18
30	section	760	23,95
31	t	756	23,79
32	news	736	23,02
33	process	722	22,48
34	government	718	22,32
35	year	689	21,21
36	people	670	20,48
37	byline	487	20,14
38	county	639	19,29
39	council	626	18,79
40	report	625	18,75
41	concern	616	18,41
42	ban	607	18,07
43	public	589	17,38
44	pg	415	17,17
45	need	572	16,74
46	group	561	16,32
47	area	554	16,06
48	rock	545	15,72
49	take	540	15,53
50	issue	533	15,27

Table 3.4.2. AntConc-generated keyword list containing the words with the fifty highest keyness measures. The BNC has been used as reference corpus. From left to right, the table shows: rank, keyword frequency in the corpus, keyness.The basis wordlist of the corpus was built with a stoplist and a lemmatization list, which has probably inflated the keyness of lemmatized items.This table should help to have an approximate idea of the differences in frequency between corpus and reference corpus despite the lemmatization. Lines coloured in light grey indicate that the frequency of the word is altered by Lexisnexis headings and should not be considered. The red-coloured line indicates a word which de facto does not appear in the corpus and should not be considered.

AntConc keyword list-corpus vs COCA			
Rank	Keyword	Freq.	Keyness
1	be	419,04	10432
2	fracking	330,82	7993
3	gas	132,52	3456
4	say	119,84	3145
5	have	105,99	2805
6	water	76,75	2084
7	shale	55,30	1336
8	will	53,33	1659
9	energy	48,91	1392
10	oil	46,32	1327
11	would	43,81	1264
12	drill	38,92	1141
13	state	36,24	1216
14	-	36,01	870
15	use	32,71	1123
16	fracture	32,61	788
17	company	31,89	963
18	t	31,29	756
19	length	28,79	884
20	could	28,29	871
21	conjurer	28,23	682
22	wells	27,36	661
23	environmental	27,11	841
24	hydraulic	25,50	616
25	word	25,29	925
26	natural	25,28	794
27	more	25,18	922
28	can	24,21	896
29	industry	24,20	766
30	section	23,96	760
31	news	23,04	736
32	chemical	22,63	853
33	government	22,34	718
34	year	21,22	689
35	people	20,49	670
36	byline	20,16	487
37	county	19,30	639
38	council	18,81	626
39	process	17,88	722
40	pg	17,18	415
41	group	16,33	561
42	area	16,07	554
43	uk	15,89	384
44	take	15,54	540
45	environment	14,71	518
46	report	14,44	625
47	make	14,38	509
48	anti-fracking	14,28	345
49	moratorium	14,20	343
50	concern	14,12	616

Table 3.4.3. AntConc-generated keyword list containing the words with the fifty highest keyness measures. The COCA corpus has been used as reference corpus. From left to right, the table shows: rank, keyword frequency in the corpus, keyness.The basis wordlist of the corpus was built with a stoplist and a lemmatization list, which has probably inflated the keyness of lemmatized items.This table should help to have an approximate idea of the differences in frequency between corpus and reference corpus despite the lemmatization. Lines coloured in light grey indicate that the frequency of the word is altered by Lexisnexis headings and should not be considered. The red-coloured line indicates a word which de facto does not appear in the corpus and should not be considered.

AntConc keyword list-corpus vs COCA-no lemma
Rank	Keyword	Freq.	Keyness
1	fracking	114055,61	7890
2	is	68679,11	4751
3	said	34476,89	2385
4	are	33074,68	2288
5	has	29981,16	2074
6	gas	25643,43	3398
7	was	19385,12	1341
8	shale	19197,19	1328
9	been	15597,72	1079
10	words	12908,96	893
11	drilling	12533,11	867
12	chemicals	10899,61	754
13	wells	9540,77	660
14	were	9121,56	631
15	fracturing	8933,63	618
16	hydraulic	8904,72	616
17	had	8543,33	591
18	companies	8384,32	580
19	water	7611,63	2035
20	byline	7039,93	487
21	s	6837,55	473
22	energy	6630,10	1376
23	says	6432,79	445
24	pg	5999,12	415
25	oil	5857,38	1316
26	length	5486,58	880
27	anti-fracking	4958,31	343
28	years	4958,31	343
29	uk	4929,40	341
30	concerns	4828,21	334
31	it's	4813,75	333
32	moratorium	4813,75	333
33	underground	4784,84	331
34	methane	4539,10	314
35	contamination	4322,26	299
36	resources	4293,35	297
37	groundwater	4278,89	296
38	earthquakes	4004,23	277
39	states	3975,32	275
40	york	3830,77	265
41	california	3787,40	262
42	residents	3642,84	252
43	environmental	3618,77	840
44	jobs	3599,47	249
45	don't	3541,65	245
46	extraction	3426,01	237
47	communities	3382,64	234
48	operations	3339,27	231
49	frack	3324,82	230
50	natural	3214,13	794

Table 3.4.4. AntConc-generated keyword list containing the words with the fifty highest keyness measures. The COCA corpus has been used as reference corpus. From left to right, the table shows: rank, keyword frequency in the corpus, keyness.The basis wordlist of the corpus was built with a stoplist and was NOT lemmatized. Lines coloured in light grey indicate that the frequency of the word is altered by Lexisnexis headings and should not be considered.

Riassunto in italiano

La tesi analizza, unendo metodi di analisi dei corpora e metodi sociologici, il modo in cui un argomento controverso come l'uso delle tecniche di fratturazione idraulica (*hydraulic fracturing* o *fracking*) per l'estrazione di petrolio e gas da scisto è stato rappresentato e descritto nei mezzi di comunicazione di massa in lingua inglese tra gli anni 2010 e 2014.

Il dibattito riguardante le tecniche di fratturazione idraulica è nato nel contesto nazionale statunitense durante il primo decennio degli anni duemila e si è diffuso al di fuori dell'ambito statunitense allorché l'abbondante produzione di energia che queste tecniche consentivano ha dato adito all'ipotesi di un affrancamento energetico degli Stati Uniti rispetto alle importazioni di idrocarburi dall'esterno. La prospettiva di una "nuova era del fossile", unitamente alle preoccupazioni legate all'impatto della fratturazione idraulica su ambiente e comunità interessate dalla costruzione degli impianti di estrazione (contaminazione delle falde acquifere, attività sismiche anomale, fughe di metano sono solo alcune), hanno dato origine ad un' accesa controversia, riportata e sviluppatasi nei mezzi di comunicazione di massa, e largamente diffusasi anche al di fuori degli Stati Uniti, sebbene in misura diversa a seconda dei vari contesti nazionali, culturali e politici.

La tesi ha dunque per oggetto d'analisi un corpus di 928 articoli da giornali e riviste in lingua inglese (sia nelle loro edizioni cartacee che in quelle online) provenienti principalmente da Stati uniti, Regno Unito, Canada, Irlanda, Sudafrica, Australia, Nuova Zelanda, scritti tra il 2010 ed il 2014 ed aventi per oggetto, appunto, l'impiego delle tecniche di *fracking* ed il relativo dibattito sull'impatto ambientale. Si è cercato di considerare, a livello di tendenze su larga scala, in quali termini si sviluppa la controversia, come viene contestualizzata, quali sono i punti dibattuti, quali le preoccupazioni destate dall'uso di queste tecniche di estrazione, e quali gli attori menzionati dai mezzi di comunicazione analizzati. Vengono prese in considerazione inoltre alcune differenze interne al corpus selezionato, costituito principalmente da articoli di giornale (in particolare quotidiani), articoli di riviste (prevalentemente di divulgazione cosiddetta 'alta' quali *New Scientist* e *Scientific American*) e lettere all'editore, che costituiscono tre diverse 'sezioni' del corpus.

La tematica si inquadra nel contesto di problematiche energetiche ed ambientali ampiamente riconosciute ormai a livello globale, toccando temi come il cambiamento climatico e la produzione, la gestione ed il consumo delle risorse energetiche ed ambientali da parte dell'uomo. Tali problematiche si intrecciano con questioni relative al ruolo dell'innovazione tecnoscientifica[13] nelle società contemporanee, e di conseguenza anche alla produzione e alla comunicazione della conoscenza scientifica e tecnologica al di fuori della comunità scientifica.

La ricerca metodologica e teorica che ha preceduto la realizzazione dello studio ha rivelato che, sebbene negli studi linguistici l'analisi di corpora sia ampiamente consolidata e sviluppata in vari ambiti, spaziando dalla caratterizzazione dei generi all'analisi di testi scientifici a quella di testi giornalistici, sembra non essere approfondito l'aspetto che indaga la funzione del linguaggio in una situazione di controversia tecnologica e scientifica. Inoltre, sebbene la sociolinguistica abbia affrontato temi come la divulgazione scientifica e la comunicazione del cambiamento climatico, sembra mancare un'analisi dettagliata di tali contesti comunicativi dal punto di vista linguistico. Lo studio dei rapporti tra scienza, tecnologia, innovazione e società, anche attraverso la comunicazione del sapere scientifico e tecnologico, è invece oggetto di studio di alcune discipline sociologiche, come gli *Science and Technology Studies* (STS) o gli studi sulla comunicazione pubblica di scienza e tecnologia (*Public Communication of Science and Technology,* PCST); tuttavia, l'analisi dell'aspetto comunicativo si limita a considerazioni generali o ad esempi circoscritti, senza scendere nel merito delle strategie linguistiche e retoriche. Si è pensato quindi di applicare entrambi gli approcci disciplinari al corpus, integrando l'interpretazione dei risultati linguistici con le considerazioni e le teorie di ordine sociologico.

L'analisi linguistica è stata suddivisa in due parti: la prima è di tipo lessico-grammaticale, mentre la seconda è di tipo lessico-semantica.

L'analisi lessico-grammaticale è basata sul metodo dell'analisi multi-dimensionale, elaborato da Biber come sistema di classificazione e caratterizzazione della variazione linguistica dell'inglese. La base di questo metodo è un corpus generico di testi sia scritti che trascritti dal parlato, nel quale l'autore conteggiò, tramite un programma

---

[13] Il termine tecnoscienza si riferisce alla crescente prossimità tra contesti di ricerca scientifica e contesti di applicazione della tecnologia, che caratterizza la produzione e la percezione pubblica delle conoscenze relative a questi ambiti.

appositamente creato per il loro riconoscimento e calcolo, una serie di circa sessanta elementi linguistici appositamente selezionati, che spaziano dall'ambito grammaticale a quello sintattico e lessicale e furono considerati particolarmente significativi dal punto divista della funzione comunicativa. In seguito, una procedura statistica chiamata analisi fattoriale fu applicata ai conteggi di frequenza per individuare, attraverso schemi ricorrenti di co-occorrenza, gruppi di elementi linguistici (i fattori) le cui frequenze all'interno del corpus seguivano percorsi simili o complementari. Questi fattori, sei in tutto, vennero poi interpretati in base alle funzioni comunicative degli elementi linguistici che li componevano, ed intesi come espressione di particolari caratteristiche, misurabili lungo un continuum diverso per ogni fattore. Si individuarono, così, sei diverse 'dimensioni di variazione linguistica' (da cui la denominazione del metodo – analisi multi-dimensionale). In base ai conteggi di frequenza effettuati sul corpus generico, ogni elemento linguistico ha un livello di rappresentatività rispetto al fattore nel quale si trova. Il peso che ogni dimensione ha su un dato testo o gruppo di testi può inoltre essere quantificato a partire dalla frequenza degli elementi linguistici che la compongono ed al loro livello di rappresentatività. Infatti, una volta individuate le dimensioni, Biber ne calcolò il peso, espresso in punteggi, per vari generi e sottogeneri in cui il corpus di partenza era stato precedentemente suddiviso. Delle sei dimensioni individuate, quattro sono state utilizzate nel presente elaborato (principalmente per ragioni di rilevanza nei confronti del corpus analizzato). Le caratteristiche che esse permettono di individuare in un testo o gruppo di testi sono: la finalità informativa di un testo contrapposta all'intento di interagire con il lettore o l'interlocutore attraverso un linguaggio meno informativo, che coinvolge direttamente l'autore o parlante (dimensione 1); l'uso della narrazione contrapposto all'uso della descrizione o dell'esposizione (dimensione 2); la presenza o assenza di un esplicito intento persuasivo nei confronti del lettore o interlocutore (dimensione 4); il livello di complessità, astrazione ed elaborazione dello stile utilizzato (dimensione 5). Il corpus e le sue sezioni sono stati quindi analizzati seguendo questo metodo e confrontati con i generi e sottogeneri presenti nello studio di Biber.

L'analisi lessico-semantica è invece mirata a rilevare le principali tendenze presenti nel corpus in termini di scelta lessicale, tramite l'osservazione delle parole più frequenti, sia nel corpus in generale che nelle sue sezioni. Utilizzando programmi che

generano concordanze, liste di frequenza e parole-chiave, si sono prese quindi in considerazione le parole che caratterizzano il corpus e che lo distinguono rispetto ad un corpus generico della lingua inglese (utilizzando due corpora di riferimento: il *British National Corpus* ed il *Corpus of Contemporary American*). È stato inoltre analizzata la presenza di alcuni dei vari aspetti legati alla controversia sul *fracking*, ad esempio la spiegazione più o meno tecnica del procedimento di estrazione, le conseguenze su ambiente, animali e persone, e gli attori coinvolti nella controversia. Ulteriori considerazioni sono state proposte su valori statistici come la varietà lessicale e la lunghezza media di parole e frasi. È stato infine applicato al corpus il concetto di meta-discorso (*metadiscourse*) studiato da Hyland, ed utilizzato per indicare gli strumenti linguistici in possesso dell'autore di un testo per comunicare e negoziare con il lettore il significato del proprio messaggio. Data la natura estremamente vasta ed indefinita di tale concetto, lo studio di Hyland si sofferma in particolare su alcune classi di parole cui vengono attribuite particolari funzioni meta-discorsive. Sette fra di esse sono state adattate ai programmi di conteggio delle frequenze e analizzate all'interno del corpus. Esse indicano: l'esplicita espressione delle opinioni e attitudini dell'autore; l'accettazione o meno di possibili alternative a quanto viene espresso (quindi la presentazione di un certo contenuto come ipotetico ed aperto o come sicuro e chiuso); la presenza di definizioni; la presenza di citazioni a supporto di ciò che si scrive; l'intenzione di coinvolgere il lettore interagendovi attraverso il testo; i casi di esplicito riferimento all'autore stesso del testo.

I risultati ottenuti sono stati combinati, interpretati ed infine rivisti alla luce delle teorie sociologiche considerate: prevalentemente, alcuni studi inerenti all'ambito degli STS e della PCST, integrati da alcuni accenni alla mappatura delle controversie, metodo sviluppato nell'ambito della teoria sociologica denominata *Actor-Network Theory*.

L'analisi linguistica ha rivelato alcune caratteristiche che sembrano confermare la natura controversa dei contenuti espressi dal corpus. In particolare, i livelli di coinvolgimento ed esplicita persuasione sono più alti della media individuata da Biber per il genere giornalistico. La presenza di diverse opinioni ed un certo livello di conflittualità ed incertezza emergono dal vasto uso del tempo futuro (usato in previsioni caratterizzate dall'alto grado di certezza) e dei verbi modali di possibilità. I valori legati all'aspetto narrativo o descrittivo/esplicativo e a quello dell'elaborazione stilistica, non

sono invece marcati, attestandosi su livelli molto vicini ai valori individuati da Biber per il genere giornalistico. Dal punto di vista semantico, sono stati individuati tre 'campi' principali di parole molto frequenti, accomunate dall'aspetto della controversia al quale possono riferirsi: il processo vero e proprio della fratturazione idraulica, il suo aspetto controverso (vale a dire i possibili rischi e danni posti da questo processo) e gli attori coinvolti. Molte parole appartenenti al primo di questi campi, ed alcune di quelle appartenenti al secondo, si sono rivelate essere anche parole-chiave.

Alcune differenze sono inoltre state messe in evidenza tra le diverse sezioni. Gli articoli di riviste si distinguono per un'attitudine particolarmente distaccata ed informativa rispetto alle altre sezioni (minori livelli di coinvolgimento e di esplicita persuasione), e per una maggiore apertura a tematiche globali, nonché al contesto generale nel quale le vicende si inquadrano. D'altra parte, mostrano un interesse leggermente minore, rispetto alle altre sezioni, verso le problematiche immediate legate al *fracking*, e verso la loro dimensione locale, preferendo la descrizione del processo tecnico. Questo approccio sembra fare degli articoli di rivista la sezione che più, tra le tre, tende ad avvicinarsi ad un'esposizione che si ritiene scientifica. Le lettere all'editore si caratterizzano invece come le meno informative e le più 'coinvolte', dove la presenza dell'autore è maggiormente visibile come lo sono i tentativi di coinvolgere i lettore ed asserire con forza il proprio punto di vista. È anche la sezione che mostra maggiore interesse per le possibili conseguenze negative del *fracking* e per la dimensione locale della controversia. Gli articoli di giornale sono, per certi aspetti, in una posizione intermedia tra le altre due sezioni: questo vale per i valori informativi e di coinvolgimento, come anche per il peso attribuito ai vari campi individuati nell'analisi lessico-semantica. Questa sezione si occupa ampiamente della descrizione del processo di fratturazione idraulica, ma attribuisce anche una considerevole rilevanza alle possibili e preoccupanti conseguenze di questa tecnologia. Tra gli attori più frequentemente menzionati si ritrovano sia realtà locali che nazionali o internazionali.

Gli articoli di riviste e soprattutto quelli di giornale fanno largo uso di discorso diretto ed indiretto, aspetto tipico del genere giornalistico e che si presta a riferire i vari interventi e le varie opinioni che caratterizzano la controversia. In tutte e tre le sezioni, tra gli attori che compaiono più frequentemente figurano le compagnie di estrazione e il

complesso dell'industria petrolifera ed energetica, a dimostrazione dell'importanza di questa componente all'interno della controversia.

Per quanto riguarda l'omogeneità all'interno delle varie sezioni, mentre gli articoli di giornale e quelli provenienti da riviste appaiono relativamente omogenee, le lettere all'editore si rivelano invece la sezione con la maggiore variabilità interna, il che potrebbe derivare dal minor peso che le convenzioni e le norme di genere hanno su questi testi, scritti prevalentemente da autori esterni alle redazioni dei giornali.

Al momento di integrare l'interpretazione dei risultati linguistici con le teorie sociologiche, ci si è ovviamente soffermati sugli strumenti analitici ed interpretativi che queste teorie forniscono. Si è quindi notato che questi approcci tendono a delineare, nei contesti analizzati, una modalità 'tradizionale' di intendere la comunicazione pubblica della scienza. Questa modalità tradizionale attribuisce alla comunità scientifica uno status privilegiato di unica fonte primaria ed autorevole del sapere scientifico e la separa nettamente dal resto della società, dalle applicazioni conseguenti alle scoperte scientifiche effettuate (quindi dalla tecnologia, e dai 'tecnici') e dalla divulgazione scientifica. In questa visione, il pubblico è essenzialmente ignorante e passivo, e può ricevere la conoscenza scientifica solo attraverso la mediazione effettuata dal mezzo divulgativo. Questa concezione tradizionale è stata ampiamente criticata in ambito sociologico, ma rimane estremamente presente ed influente in molti contesti. La sociologia ha invece messo in discussione l'isolamento della comunità scientifica dai processi sopra descritti, affermando che gli scienziati, come membri della società vivono a contatto con essa, che le applicazioni delle teorie elaborate sono strettamente legate all'ambito di ricerca da cui provengono (si ricordi il significato del termine tecnoscienza), e che il confine tra attività di ricerca, produzione di testi scientifici e divulgazione non è netto come si vorrebbe far credere. La comunicazione pubblica della scienza, inoltre è stata definita non come un atto di mediazione o traduzione, ma come un cambiamento vero e proprio a livello contenutistico, rappresentabile come un continuum tra due estremi: mentre infatti la comunicazione scientifica a livelli specialistici è caratterizzata dalla presenza di ipotesi, e da interpretazioni presentate come provvisorie e mai come assolutamente certe, la comunicazione scientifica a livelli popolari 'confeziona' il sapere come un insieme di fatti consolidati ed universalmente accettati. Inoltre, in alcune circostanze, anche questo assetto è alterato, e la

comunicazione pubblica della tecnoscienza assume caratteristiche diverse. È il caso delle controversie, durante le quali l'incertezza e la pluralità delle interpretazioni caratterizza sia i livelli specialistici che quelli popolari. Simili situazioni sono state analizzate nel contesto della PCST, e definite da Bucchi 'deviazioni' verso il pubblico. La causa della deviazione viene attribuita ad un'iniziativa proveniente da una parte della comunità scientifica, che decide di aprire un dibattito verso il pubblico non esperto, al fine di favorire lo sviluppo e forse la soluzione del dibattito stesso.

La controversia sul fracking si inquadra nella comunicazione pubblica della tecnoscienza al livello divulgativo-popolare: la parziale trasformazione del sapere scientifico tipica dell'aspetto divulgativo può essere confermata dai valori relativi alle dimensioni 2 e 5, tipici del linguaggio giornalistico e non di quello scientifico. Tuttavia la presentazione consolidata del fatto scientifico è fortemente alterata dall'incertezza e dai valori di coinvolgimento e persuasione che pervadono il corpus. Tuttavia, non si è ritenuto di poter parlare di un caso di deviazione, in quanto la controversia è nata perché alcune famiglie hanno subito pesanti danni ambientali e di salute a causa della costruzione di siti di fratturazione idraulica all'interno delle loro proprietà. La controversia non è stata quindi generata da uno stato di crisi all'interno della comunità scientifica. Inoltre, nei casi di deviazione descritti da Bucchi, le procedure scientifiche relative all'oggetto del dibattito vengono spesso 'rivelate' al pubblico, in una pratica discorsiva che è stata definita di '*backstage*', preposta appunto a svelare i retroscena (per quanto anch'essi possano poi essere costruiti ad hoc) dell'attività scientifica. Nella rappresentazione della controversia sulla fratturazione idraulica invece, pochi dettagli tecnici emergono riguardo sia alle procedure tecniche di applicazione della tecnologia, sia al modo in cui gli effetti del *fracking* sull'ambiente vengono verificati. Ciononostante, la scienza ha la sua rilevanza all'interno del corpus, dove sono numerosi i casi in cui ci si riferisce a studi e rapporti il cui compito sarebbe quello di accertare l'impatto del *fracking*. Manca inoltre dalle liste di frequenza qualsiasi riferimento ad iniziative di coinvolgimento pubblico nel processo decisionale e conoscitivo della controversia (altro oggetto di studio in ambito PCST). Si è quindi concluso che, malgrado l'elemento del dibattito abbia contribuito a far emergere l'incertezza che circonda questo ambito tecnologico ed a mettere in contatto diverse opinioni provenienti da diverse parti della società riguardo ad un argomento altrimenti strettamente tecnico e

scientifico, negli articoli analizzati è ancora forte l'influenza del modello tradizionale di comunicazione pubblica di scienza e tecnologia. Da ciò deriva probabilmente la sostanziale separazione riscontrata nel corpus sia tra cittadini e comunità scientifica, sia tra cittadini ed autorità politiche (che sembrano semplicemente impegnate a bandire, posticipare o autorizzare esplorazioni e trivellazioni di fronte alle richieste dei cettadini).

Un altro aspetto emerso dall'analisi linguistica è l'ampio spazio dedicato alle possibili conseguenze negative della fratturazione idraulica, ed alle preoccupazioni che queste destano tra cittadini ed attivisti, nonché alle azioni collettive di protesta anti-*fracking* promosse da gruppi di residenti od organizzazioni ambientaliste. Questo aspetto infatti supera di gran lunga quello legato ai possibili vantaggi energetici ed economici apportati dall'estrazione di gas e petrolio da scisto. Non sembra però che questa sproporzione sia dovuta ad una propensione dei media verso la causa anti-*fracking*, o verso la causa ambientalista in generale. Piuttosto, la rilevanza attribuita agli aspetti negativi e più controversi della vicenda è stata attribuita ad un'esigenza di drammatizzazione e spettacolarizzazione della notizia tipica della copertura mediatica dei grandi problemi e delle questioni dibattute (anche questo criterio di valutazione delle notizie è stato approfondito dagli STS).

Gli studi di mappatura sulle controversie, inoltre, hanno messo particolarmente in evidenza sia la loro complessità sia il loro dinamismo, che si realizza nel continuo formarsi e scomporsi delle sue componenti. Anche le teorie legate agli STS e alla PCST individuano in questi fenomeni particolari momenti di cambiamento sociale. Si è tentato pertanto di individuare, attraverso la rappresentazione che il corpus dà della controversia, qualche elemento di negoziazione o rinnovamento di paradigmi scientifici e culturali. Pur sempre limitati dal fatto che l'analisi di corpora non permette un esame dettagliato dei vari testi e quindi preclude determinate operazioni di verifica, alcuni risultati sembrano affiorare. In particolare, si è osservato come la componente ambientalista (la cui presenza e le cui attività sono abbastanza documentate nel corpus) si faccia forza delle perplessità riguardanti il *fracking* per tentare di contribuire ad un cambiamento delle politiche energetiche verso fonti rinnovabili e a minor impatto ambientale (alle energie rinnovabili nel corpus si fa riferimento molto più spesso rispetto che alla dicitura '*non-renewable*', non rinnovabile). Tuttavia, si è constatato

166

anche che la posizione anti-*fracking*, a cui la causa ambientalista contribuisce fortemente, poggia su basi piuttosto fragili. In effetti, la legittimità dei crescenti consumi energetici mondiali non viene mai messa in discussione (almeno per quanto riguarda i risultati dell'analisi linguistica), così come succede anche per la crescente dipendenza delle attività umane da fonti limitate di energia, destinate a declinare ed esaurirsi (secondo alcuni, in un tempo relativamente breve). Senza avvalersi di tali basi, l'opposizione al *fracking* continua a lasciare intatta la legittimità del paradigma occidentale e capitalista di approvvigionamento e sfruttamento delle risorse ambientali, involontariamente giustificando la tesi per la quale, una volta trovato un rimedio agli spiacevoli effetti immediati del *fracking*, qualsiasi opposizione perderebbe di senso, perché risulterebbe irragionevole rinunciare all'abbondanza energetica promessa dal *fracking* di fronte all'enorme fabbisogno energetico degli Stati Uniti e dell'umanità. E non essendoci ancora un giudizio scientifico unanime circa la sicurezza delle tecniche di *fracking*, esso rimane in campo quale cospicua opportunità energetica.

Gli STS offrono un'ulteriore analisi del ruolo dell'innovazione tecnoscientifica, incentrata sulla nozione di rischio. Tale analisi parte dal presupposto che, malgrado la complessità e l'ambivalente incertezza legata all'innovazione siano chiaramente percepiti, si tenda spesso a razionalizzare l'idea di tecnologia, separando la tecnologia dalle proprie applicazioni, da cui viene poi fatto dipendere l'esito positivo o negativo dell'innovazione tecnologica. Questo porta ad un fondamentale dualismo tra tecnologie che vengono ritenute utili e tecnologie rischiose, dove per rischiose si intende necessariamente pericolose. L'appiattimento del significato di rischio alla sola componente negativa è stato infatti interpretato come espressione dell'incertezza percepita dalla collettività di fronte all'incapacità della scienza di prevedere la portata e gli effetti delle attività umane. Dunque, anche la rappresentazione dell'innovazione apportata dal *fracking* può essere letta attraverso questo dualismo. Nel breve termine, questa tecnologia si è rivelata in certi casi effettivamente dannosa; ma grazie all'argomentazione di cui sopra, la parte favorevole al *fracking* resta in gioco, cercando di riportare la tecnologia in questione dal lato positivo. Nel lungo termine, chi si oppone al *fracking* asserisce che esso porterà sia all'esacerbazione dei danni ambientali già provocati, sia ad ulteriori emissioni di gas serra, contribuendo al cambiamento climatico. Chi invece lo sostiene, afferma che il gas da scisto produce meno emissioni

di petrolio e carbone, e garantirà agli Stati Uniti sufficienti scorte energetiche per potere nel frattempo elaborare fonti alternative che si sostituiranno ai combustibili fossili e finalmente argineranno il riscaldamento globale.

Si constata che, rimanendo preponderanti un discorso tradizionale sulla comunicazione della conoscenza tecnoscientifica e soprattutto il paradigma occidentale di produzione, gestione e consumo delle risorse ambientali, dal corpus emerge una posizione pro-*fracking* particolarmente versatile ed influente, mentre l'argomentazione che ad esso si oppone sembra estremamente presente ma comunque in ultima analisi riconducibile allo stesso paradigma, pertanto caratterizzata da una certa inconsistenza di fondo.

L'analisi linguistica del corpus ha quindi evidenziato alcune caratteristiche in seguito interpretate come tendenze non solo linguistiche ma anche culturali e sociali. Malgrado le interpretazioni proposte siano fortemente limitate ed in certa misura soggettive, è importante ricordare che, così come le innovazioni tecnologiche avrebbero potuto evolversi in maniere diverse, anche le scelte linguistiche preposte a rappresentarle avrebbero potuto essere diverse. Questo studio si propone come un tentativo di ricercare le dinamiche sociali e culturali che hanno influenzato queste scelte linguistiche, ed il ruolo che esse hanno nel comprendere, codificare e far conoscere la controversia.