Researching evaluative discourse in Annual Reports using semantic tagging

Ruth Breeze
Universidad de Navarra (Spain)
rbreeze@unav.es

Abstract

This article uses the semantic tagging tools provided by Wmatrix3 to investigate the discourse of corporate annual reports to shareholders from leading UK-based companies in four sectors: pharmaceuticals, food, mining and finance. Six potentially interesting areas of commonality are identified (change, inclusion, size: big, important, cause and effect, and time: begin). Concordance lines from these areas in each subcorpus are then analysed qualitatively to identify the presence of shared value-systems in the discourse of the reports. A contrastive analysis is then conducted which reveals differences between the four sectors in the keyness of areas such as safety, strength, newness and focus, as well as colleague and client orientation. These findings are discussed in the light of previous research on business communication. Finally, some advantages of using semantic tagging over standard corpus linguistic tools are discussed.

Keywords: corporate discourse, annual reports, semantic tagging, corpus linguistics, business communication, values.

Resumen

La investigación del discurso evaluativo en Informes Anuales mediante etiquetado semántico

Este artículo utiliza herramientas de etiquetado semántico proporcionadas por Wmatrix3 para investigar el discurso de los informes anuales corporativos dirigidos a los accionistas de empresas ubicadas en el Reino Unido y pertenecientes a los siguientes sectores: sector farmacéutico, alimentación, minería y finanzas. Se identifican seis áreas de interés potencial (cambio, inclusión, tamaño: grande, importante, causa y efecto, y tiempo: inicio). Se analizan las líneas de concordancia de cada una de estas áreas en cada subcorpus

de forma cualitativa con el fin de identificar la presencia de sistemas de valores compartidos en el discurso de estos informes. Posteriormente, se lleva a cabo un análisis contrastivo, que revela diferencias entre los cuatro sectores respecto a las palabras clave utilizadas, de áreas tales como seguridad, fortaleza, novedad y enfoque, así como en orientación hacia colegas y clientes. Se comentan los resultados obtenidos a la luz de la investigación previa en el ámbito de la comunicación empresarial. Finalmente, se hace una valoración de algunas ventajas que conlleva la utilización de etiquetado semático frente a herramientas estándar empleadas en lingüística de corpus.

Palabras clave: discurso corporativo, informes anuales, etiquetado semántico, lingüística de corpus, comunicación empresarial, valores.

1. Introduction

Corporate discourse has long been a focus of interest for applied linguists, and a substantial volume of research exists into the way that language is used in the business world. One genre that is attracting increasing attention is the corporate annual report, which came into existence as a mainly factual document intended to inform shareholders about company performance, but which has now developed into a complex genre with a number of different communicative purposes (Bartlett & Jones, 1997; Stanton & Stanton, 2002; Ditlevsen, 2012; Breeze, 2015). According to Bhatia (2010), companies' annual reports generally combine at least four different discourses, namely accounting discourses, legal discourses, the discourse of economics, and public relations discourse. These are not evenly distributed across the report, but tend to be concentrated in specific sections, with a major division between the first and second half of the report. The second half, with its sober presentation and dense print, contains the factual information required by law. This is presented through accounting discourse (represented in technical data and auditors' statements), and legal discourse in the form of disclaimers (De Groot, 2014; Breeze, 2015), although the discourse of economics may also figure here in the interpretive paragraphs accompanying the numerical information. By contrast, the first half of the report, with its "magazine" format, has taken on a significant public relations role over the last thirty years or so (Ditlevsen, 2012), providing narratives which show "who the company is and what its values are, what its businesses are and how successful they have been" (Courtright & Smudde, 2009: 258). The promotional discourse of this part is also blended with the discourse of

economics, and it should be noted that the latter is deployed strategically to recontextualise and interpret the facts, sometimes with a considerable degree of licence (Bhatia, 2010).

Given the multifaceted nature of this complex genre, it is not surprising that researchers have often chosen to centre on one particular section or aspect. For example, considerable research efforts have focused on the "letter to shareholders" (Vázquez Orta & Foz Gil, 1995; Abrahamson & Amir, 1996; Hyland, 1998; Smith & Taffler, 2000; Bhatia, 2004; Craig & Amernic, 2004), while there has been a recent surge of interest in particular themes such as readability (Clatworthy & Jones, 2001), accuracy in corporate disclosure (Abraham & Shrives, 2014) or visual style (Davison, 2010), and multimodality (De Groot et al., 2006), as well as certain aspects of stakeholder response (Rutherford, 2005; De Groot, 2014) or specific sectors such as banking (Malavasi, 2010). However, perhaps because of the complex nature of the reports, or their presumed overlap with other areas of promotional discourse, so far less attention has been paid to the ethos or value system that runs through the public relations discourse of annual reports in general, and the possible differences between companies in different sectors.

The present paper centres on the first half of the report, which most clearly embodies the company's discursive self-presentation and is not constrained by legal requirements. As Sandell and Svensson state (2016: 9), annual reports "do much more than report financial performance; they partake in the symbolic production and reproduction of reality", and this is particularly evident in the first section, whose contents, wording and general presentation are designed to index particular values that enhance the company's image. Research into values in discourse in other areas has often been conducted using corpus linguistic tools which involve the identification of keywords and comparison of word frequencies across corpora (Hyland, 1997; Giannoni, 2010). However, it is clear that standard corpus data such as frequencies and keyness are only capable of providing part of the story. Innovative semantic tagging tools such as Wmatrix3 (Koller et al., 2008; Rayson, 2008) can go further, in that rather than identifying particular words that are salient, they can also pinpoint semantic areas that are especially important. By using such tools it has been possible, for example, to establish that doctor-patient communication contains frequent references to the semantic field of "violence", materialised in words such as "fight" or "battle" (Semino, 2008; Semino et al., 2015). Since many different words may

be used to convey the same idea, it is likely that on the level of simple word frequencies, or even keyness values, none of these words will attain significance individually. However, taken together as a semantic field, they can be visualised as an area that is surprisingly important.

Semantic tagging has already been applied to business discourse by authors interested in metaphor. For example, Kheovichai (2015) uses semantic tagging to identify potential metaphor scenarios in an academic business corpus, proceeding by scanning the semantic categories found for semantic domains "that seemed incongruent with the discourse of business science" (page 160), and then analysing the concordance lines manually. He did not specify any frequency cut-off point for this, and thus was able to include very low frequency items in his results. He identified scenarios involving war, sport, games, journeys, machines, living organisms, and buildings, all of which, in his interpretation, appeared to centre on the source domain of a bounded space. Importantly, he noted that source domains close to each other generate clusters of metaphorical expressions, so it is not necessarily appropriate to examine these at the lexical level, since exploration at the semantic level is likely to be more fruitful.

Semantic tagging also holds great promise for the study of values in discourse. In particular, semantic analysis should shed light on the underlying mind-set, that is, the values and assumptions which underpin what is considered to be effective communication in a given context. Previous discourse studies using corpus tools to uncover disciplinary value systems (Hyland, 1997; Giannoni, 2010; Breeze, 2011) have generally relied on mixed methods, using quantitative criteria to identify areas of interest, which are then followed up by qualitative analysis. The recent addition of semantic tagging to the analytical toolbox opens up new possibilities in this respect, as researchers go beyond mere recurrence of lexical items to examine salient areas of meaning.

In the present article, a mixed-methods approach to investigating values in discourse is extended and refined by the application of semantic tagging to a corpus of annual reports from four different areas: the pharmaceutical industry; supermarkets and the food industry; mining; and financial services/fund management. These sectors are all important for the UK economy but also provide some degree of contrast. Semantic tagging was used to identify key areas of meaning, and to compare the subcorpora both across semantic areas and within each area. The principal research questions

were therefore: i) can semantic analysis help identify the values present in annual reports; ii) what values are prominent; and iii) how do these values vary between the sectors chosen?

2. Material and method

A corpus was built from the 2013 edition of 16 Annual Reports published on the websites of major public limited companies listed on the London stock exchange, all of which had been listed on the FTSE 100 within the previous ten years. Four companies were selected from each of the following sectors: the pharmaceutical industry; food and supermarkets; financial services and fund management; and mining. The annual reports all fell naturally into two parts: the first consisted of general information about the company, its areas of business, performance and aspirations for the future; while the second contained the information required by law, including the audit reports, balance sheets, information about corporate governance, and so on. These parts were clearly distinguishable in terms of format, since the first part was visually attractive, making use of a variety of graphic and photographic techniques to create a magazine-like presentation, while the second part had smaller print and contained large quantities of numerical data in the form of soberly-presented tables or line graphs. For the purposes of this study, the first part of each annual report was saved in text format. The resulting texts were then assembled to constitute the four subcorpora (descriptive data are provided in Table 1).

	Pharmaceuticals	Food	Mining	Finance
Tokens	109,588	70,998	65,251	80,825
Standardised TTR	38.27	39.01	37.64	36.53

Table 1. Descriptive data from the four subcorpora.

Each subcorpus was then uploaded to Wmatrix3 (kindly provided by Dr. Paul Rayson, UCREL, University of Lancaster; see Rayson, 2008). Briefly, Wmatrix3 uses the UCREL semantic analysis system to tag corpus tokens according to 21 broad semantic fields (emotion, life and living things, entertainment, etc.), each subdivided into a large number of different subsections (Archer et al., 2002). Thus category I (money and commerce), for example, is subdivided into 22 different sections and subsections reflecting different aspects of money: for example, the term "invest" would

be tagged as I1.1 (money and pay), while "profitable" would be tagged as I1.1+ (money: affluence), "debt" or "loss" would be tagged as I1.2 (money: debt), and so on.

Since the first stage was to establish the key semantic areas in each subcorpus, it was necessary to select a reference corpus. The most obvious choices appeared to be the *BNC Business* or the *BNC Information* corpora available on Wmatrix3. To establish which corpus was likely to be most appropriate, lists of key semantic areas for the Finance subcorpus were generated using both potential reference corpora. As can be seen from Table 2, the differences in results were relatively small, since 23 of the top 30 categories were shared. Areas related to money, numbers and business were key when either reference corpus was used, that is, the reports were strongly oriented towards money, numbers and business even when compared with another business corpus. More interestingly, semantic fields such as "time: beginning" and "evaluation: good" were also salient in both, as were areas related to size and strength.

BNC Business only	BNC Business and BNC Information		BNC Information only
Giving	Business, general	Attentive	Knowledgeable
Interested	Business, selling	Cause-effect	Wanted
Geog. Terms	In power Money: pay	Numbers	Confident
Inclusion	Time: beginning	Drama	Able, intelligent
Measure: distance	Danger Quantities	Eval: good	Money, general
Constraint	Useful Important	Investigate	Degree
Change	Time: period Money: affluence		
	Ethical Size: big	Money: cost	
	Tough, strong	Belong group	
	Understanding		

Table 2. Top 30 semantic areas in finance subcorpus, using *BNC Business* and *BNC Information* as reference corpora.

Regarding the differences, by using *BNC Business* as a reference corpus, it was possible to pinpoint the ways in which these subcorpora were different from business documentation in general, i.e. in the frequency of geographical terms, references to change, etc., whereas when *BNC Information* was used, some key areas identified were related to common themes when discussing business activity (money, green issues). The decision was therefore made to use *BNC Business* as a reference corpus, since this would help to pinpoint any areas of value that were more important in these reports than in general business literature.

In order to identify values shared by the four subcorpora, a search was conducted for the top 30 key semantic categories for each subcorpus. Common potentially value-related areas were identified, and then analysed using quantitative frequency counts for lexical items, complemented by qualitative examination of concordance lines. Then, to explore possible differences between the value systems of the four subcorpora, all semantic areas found to have high keyness, measured as Log Likelihood >130, were identified in each subcorpus (Log Likelihood (LL) is the level of statistical significance of the difference between the frequencies of a particular semantic area in the two corpora; the larger the LL, the more certain we can be that the difference is not a coincidence, cf. Rayson, 2008). The potential values evoked by these in the subcorpora were then investigated.

3. Results

Analysis of the 30 areas with the highest keyness scores revealed considerable overlap between the four subcorpora. By way of illustration, the top 12 non-merged categories for each subcorpus are shown in Table 3.

Rank	Pharmaceuticals	Food	Mining	Finance
1	Medicine, science, tech	Food	Substances, solid	Money and pay
2	Numbers	Business, selling	Numbers	Numbers
3	Disease	Money and pay	Industry	Business generally
4	Money and pay	Farming, horticulture	Money and pay	In power
5	Business generally	Measurement: distance	Actions, making	Business: selling
6	Size: big	Numbers	Measurement: weight	Geographical names
7	Objects generally	Geographical names	Cause and effect	Time: beginning
8	Business: selling	Business generally	Money: cost and price	Useful
9	Time: new and young	Size: big	Substances, material	Drama
10	Cause and effect	Time: new and young	Geographical names	Ethical
11	Anatomy, physiology	Important	Business generally	Tough, strong
12	Quantities	Time: beginning	Belonging to group	Cause and effect

Table 3. Top 12 semantic areas for each subcorpus.

As might be expected, "medicine, science and technology" and "disease" ranked high in the subcorpus of pharmaceutical companies, while "substances, solid" was high on the list for mining companies, and "food" and "farming, horticulture" were important in food companies. In finance, predictably, "money and pay", "numbers" and "business generally" headed the list. Rather more interestingly, the area associated with "tough, strong" was salient in the finance subcorpus, while "time: new and young" was important in the food industry and in pharmaceuticals. On the other hand,

there was some degree of similarity between the lists for the four subcorpora, which all included the obvious categories "business generally", "numbers", "money and pay", but which also shared certain other categories (for example, "cause and effect" appears in the top 12 in three of the four subcorpora).

To explore the commonality between the four subcorpora further, I then identified the ten categories which occurred in the top thirty key semantic areas of all four subcorpora. Four of these categories ("money: pay", "numbers", "business, generally", and "danger") were examined in some depth, but were eventually excluded from the present study on the grounds that they almost always indexed items that referred to technical aspects of business: to profits and other aspects of company results, to financial risks, and to market behaviour in general. Although technical discourse is obviously far from free of ideological content, from a reading of the concordance lines it seemed that these areas were less closely related to the more general fields of meaning indexed in the other six ("change", "inclusion", "size: big", "important", "cause, effect", and "time: begin"), and would require a different analytical approach. The salience of these six semantic areas in the four subcorpora is illustrated in Graph 1, below.
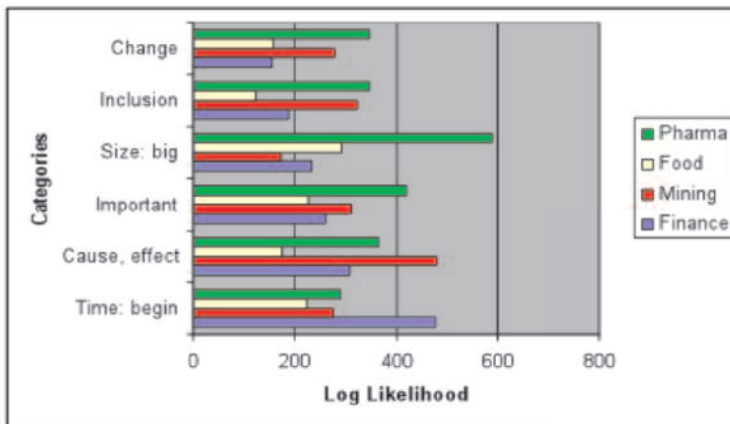
Graph 1. Keyness (Log Likelihood) of the six semantic areas selected for study in the four subcorpora.

As Graph 1 shows, these six semantic areas had high keyness values in all the subcorpora, but with some degree of variation between them. In what follows, I investigate the implications of this, looking at the semantic areas

in general, and then exploring the lexical items that are most typical of each in their particular context within the Annual Reports.
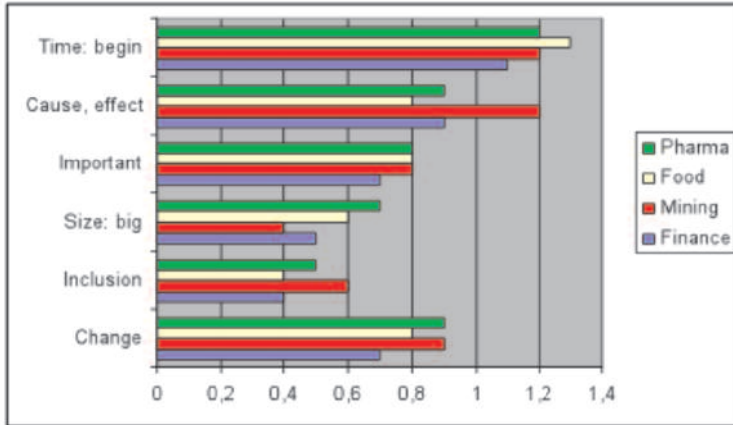
Graph 2. Percentage of tokens belonging to salient semantic categories in each corpus.

Regarding the actual frequency of tokens relating to each semantic area within each corpus, Graph 2 shows these data as percentages of the total number of words in the corpus. Interestingly, the percentages are fairly similar in all the data sets (Pearson correlation coefficient: R=1 significant at p<0.01). The largest differences between the percentages were for "cause, effect", which accounts for 1.2% of the tokens in the Mining corpus, but less than 0.9% of the tokens in all the other corpora.

In what follows, these six main categories are analysed in terms of their functions in the ideology of corporate reporting. Within each category, the frequency of salient lexical items is discussed, and examples of typical uses are provided.

3.1. Change

Under the semantic tag "change", the most popular words in all corpora were "develop" and "development".

Rank	Pharmaceuticals	Food	Mining	Finance
1	Develop (0.40)	Develop (0.25)	Develop (0.41)	Develop (0.23)
2	Change (0.10)	Adjust (0.13)	Occur (0.04)	Change (0.16)
3	Convert (0)	Change (0.11)	Volatile (0.01)	Volatile (0.01)
4	Volatile (0)	Volatile (0)	Become (0.01)	Convert (0.01)

Table 4. Top word families in each subcorpus for category "change" (relative frequency).

Table 4 shows the top word families in each subcorpus for the semantic area "change", with the relative frequency in brackets (the relative frequency is calculated by Wmatrix3 as a percentage of the total number of words in the corpus). This is not surprising in pharmaceuticals and food, where the development of new drugs or products is an important part of the company's activities:

1. The discovery and **development** of a new vaccine is a complex process that typically takes between 10 and 12 years (Pharma).

In mining, the use is rather broader. In Example 2, for example, development seems to refer to areas of company activity in general:

2. We have operations, exploration and **development** projects in Australia, Namibia, Mozambique and Canada. (Mining)

It is also used to index general areas of corporate activity as in the phrase "sustainable development":

3. …the approach that it has adopted to identify and prioritise its material sustainable **development** risks. (Mining)

However, "develop" is also often used more specifically (perhaps even euphemistically) to refer to mining itself:

4. Rio Tinto announced that all funding and work on underground **development** of Oyu Tolgoi would be delayed. (Mining)

The concept of development is also frequently applied to areas other than scientific experimentation:

5. Talent and leadership development. We aim to attract and retain the most talented people by investing in training and **development** that is tailored to individuals' needs and recognises the potential of our employees. (Pharma)

Interestingly, the noun "development/s" was at least twice as frequent as the verb forms in mining and pharmaceuticals, perhaps owing to the specialised use of the term "development" to mean creation of a new drug or exploitation of a new mine. By contrast, there were similar numbers of verb forms and noun forms in both the food and the finance subcorpora.

One interesting feature uncovered by this exploration was that the lexical items tagged as "change" often go beyond objective reporting, clearly moving into the area of promotional discourse

6. Our approach to R & D includes our strategy for open innovation for the diseases of the developing world, which seeks to stimulate innovation and enhance the productivity of our research process. This research has **transformed** our approach to intellectual property and external partnerships. (Pharma)

7. The Toronto office services our **burgeoning** Canadian client base. (Finance)

The evaluative associations of the items tagged as "change" are overwhelmingly positive. Only "volatility" is generally negative in its associations:

8. This is despite periods of **volatility** and weak market sentiment during the second half of the year, which saw a broad sell-off across global markets, with some emerging economies being particularly hard hit. (Finance)

3.2. Inclusion

The high salience of words related to the semantic area of inclusion is particularly interesting.

Rank	Pharmaceuticals	Food	Mining	Finance
1.	Include (0.32)	Include (0.22)	Include (0.35)	Include (0.19)
2.	Integrate (0.04)	Comprise (0.03)	Contain (0.03)	Integrate (0.03)
3.	Involve (0.04)	Involve (0.01)	Comprehensive (0.02)	Comprise (0.03)
4.	Comprise (0.02)	Comprehensive (0.01)	Involve (0.02)	Involve (0.02)

Table 5. Top word families in each subcorpus for category "inclusion" (relative frequency>0.02).

As Table 5 shows, the most frequent single headword in all four subcorpora was the verb "include", which seems to be used very often to convey the impression that more things could be mentioned, as in the following example from finance:

9. We have teams of skilled investment professionals across a range of investment strategies **including** equities, fixed income, property and solutions to serve our institutional and retail clients. (Finance)

The word "include" is patently an instance of vague language used to suggest that a comprehensive list would be much longer:

10. Nonetheless, we have not abandoned specialist products where there is

demand. These **include** our suite of high yield products, strengthened by the arrival of the Artio team. (Finance)

11. Westmill Foods specialises in supplying UK restaurants and wholesalers with high-quality ethnic foods **including** rice, spices, sauces, oils, flour and noodles. (Food)

Alternatives to "include", such as "encompass", "span", "involve", "comprise", and "integrate", are used less frequently, but to the same end.

12. The workforce **spans** multiple nationalities, ethnicities, languages and cultures in developing countries. (Mining)

13. The development of any pharmaceutical product candidate is a complex, risky and lengthy process **involving** significant financial, R&D, and other resources. (Pharma)

The frequent use of such terms is reminiscent of the category of "vague quantification" identified by Banks (1998) in scientific writing. The allusion is to multiple entities, some of which will be mentioned, others not, presumably owing to constraints of space. The notion that "run ons" like "etc." or "and so on" are symptoms of a careless writing style may explain why report writers opt for verbs to express vague notions of plurality. Interestingly, then, the semantic tag "inclusion" seems to point to the presence of what could be termed, adapting from Perelman and Olbrechts-Tyteca (1969), a "rhetoric of vague quantification".

3.3. Size: big

The frequency of the semantic area "size: big" evidently reflects a key aspect of the discourse of these annual reports. In particular, the concordance lines obtained indicated the values of "growth" and "expansion". All four subcorpora include many instances along the lines of: "enhance and *expand* our facilities", "excellent revenue and earnings *growth*", "address the rapidly *growing* demand for cardiovascular medication", "maintaining, upgrading and *expanding* our facilities".

Rank	Pharmaceuticals	Food	Mining	Finance
1	Grow (0.57)	Grow (0.43)	Expand (0.17)	Grow (1.36)
2	Expand (0.08)	Large (0.07)	Grow (0.15)	Large (0.09)
3	Large (0.06)	Big (0.07)	Large (0.04)	Expand (0.06)
4		Expand (0.06)	Substantial (0.04)	

Table 6. Most frequent word families in "size: big" (relative frequency >0.02).

In fact, "growth" and "grow" were among the most frequent lexical items associated with this area, followed by "large" and "expand" (see Table 6). In food, even when instances of "grow" referring to food crops were discounted manually, "grow" was still the most frequent size-related word family. In fact, "growth" is a topos in the discourse of economics (White, 2003), where it is one of the principal metaphors used to present quantitative progress, perhaps because – despite obvious problems with the organic source domain, the "growth" metaphor lends itself to narrative integration.

Interestingly, as Table 6 shows, "large" and "big" often occurred in the comparative or superlative forms in all the subcorpora (e.g. in finance, "large" (n=21), "larger" (11), "largest" (27); in food, "big" (n=18), "bigger" (4), "biggest" (7), while "large" (3), "larger" (4), "largest" (23); in mining "large" (21), "larger" (6), "largest" (32); in pharmaceuticals, "large" (27), "larger" (5), "largest" (31)). The abundant use of terms relating to large (but unspecific) size seems to be a further aspect of the rhetoric of "vague quantification" identified in section 3.3 above.

3.4. Important

Exploration of the semantic field that Wmatrix3 labels "important" in these corpora revealed considerable variety in the lexis used to reflect this idea.

Rank	Pharmaceuticals	Food	Mining	Finance
Nouns	Value (0.22) Priority (0.02)	Value (0.22) Priority (0.02)	Value (0.21) Priority (0.04)	Value (0.18) Priority (0.03)
Adjectives	Key (0.18) Major (0.08) Significant (0.08) Important (0.04)	Key (0.18) Major (0.08) Significant (0.08) Important (0.04)	Key (0.15) Significant (0.11) Major (0.7) Important (0.05) Central (0.02) Main (0.02)	Key (0.2) Significant (0.08) Important (0.04) Main (0.03) Major (0.03)
Adverbs	Significantly (0.02)	Significantly (0.02)	Significantly (0.05)	Significantly (0.02)

Table 7. Most frequent lemmas in category "important" (relative frequency >0.02).

Table 7 shows the most frequent lemmas found in the four corpora under this heading (verbs were considered, but did not reach a relative frequency of 0.02 in any instance). This seems to point to low lexical variation: the lemmas with the highest relative frequency far outstripped all the other items on the list. Word combinations were also rather limited.

Mining		Food	
Collocation	**Log Likelihood**	**Collocation**	**Log Likelihood**
Key locations	67.84	Key brands	78.30
Key indicators	58.38	Key part	42.55
Key performance	53.88	Key delivering	19.61
Key indicator	44.11	Key suppliers	17.16
Key management	39.48	Key success	17.16
Key financial	37.41	Key business	8.71
Significant costs	33.24	Significant year	32.94
Significant overrun	32.13	Significant progress	21.83
Significant incidents	25.39	Significant growth	19.60
Significant improvements	20.33	Significant sales	14.60
Significant cost	17.41		
Significant risks	15.51		

Table 8. Main collocations of "key" and "significant" in the mining and food subcorpora (LL>14).

Table 8 shows the main co-occurring pairs including "key" and "significant" in the mining and food subcorpora. As above in the case of "size" and "inclusion", the discourse of "importance" serves to enhance the company and its activities, often in a rather unspecific way. The following examples illustrate how words in this category are scattered through the text, heightening the tone of importance.

14. We have **significantly** enhanced our innovation capability by establishing numerous alliances and licensing opportunities. (Pharma)

15. Most **importantly** though, they share the common goal and belief that looking after our clients needs is our number one priority. (Finance)

The frequency of such words points to a discourse of urgency and significance, linked to discourses of efficiency (see section 3.5), while at the same time, the low lexical variation (Tables 5 and 6) also suggests that this way of writing has become conventionalised in this genre.

3.5. Cause and effect

Word families that group together in the semantic area "cause and effect" include "result", "generate", "impact", "base" (including "on the basis of"), "relate", "responsible", "cause", and so on.

Rank	Pharmaceuticals	Food	Mining	Finance
1	Result (0.16)	Result (0.21)	Result (0.20)	Result (0.17)
2	Impact (0.09)	Produce (0.07)	Due to (0.11)	Base (0.10)
3	Relate (0.06)	Impact (0.06)	Produce (0.10)	Impact (0.09)
4	Responsible (0.05)	Generate (0.05)	Base (0.09)	Generate (0.07)
5	Base (0.05)	Responsible (0.05)	Impact (0.08)	Relate (0.05)
6	Cause (0.04)	Effect (0.03)	Relate (0.06)	Responsible (0.04)
7	Effect (0.03)	Base (0.03)	Factor (0.04)	Due to (0.04)
8	Generate (0.03)		Attribute (0.04)	Attribute (0.03)
9	Determine (0.03)		Responsible (0.03)	
10			Effect (0.03)	

Table 9. Most frequent word families in "cause and effect" (relative frequency >0.02).

Table 9 shows the most frequent word families from this area in each subcorpus with their relative frequencies (<0.02). It might not seem surprising that "result" should be the most frequent word family in all the subcorpora, because annual reports are quintessentially about reporting results. However, the subcorpora do not include the "hard" financial data from the second part of the reports, where we might expect the word "result" to be used very often in a technical sense. In fact, the concordance lines for "result" (singular) are almost all about causality, rather than financial results:

16. The reorganisation of our balance sheet since the year end will lower interest charges and **result** in an improved dividend cover in the future. (Food)

Notably, 45.76% of all the instances of "result" in these subcorpora were part of the combination "as a result":

17. We were able to achieve this as a **result** of the experience and capability of our in-country team. (Food)

18. As a **result**, world demand for antibiotics and novel therapeutic approaches remains high and will continue to grow. (Pharma)

However, "results" (plural), which accounted for around 50% of the instances of the word family "result", was usually associated with financial results:

19. Management presents these **results** externally to meet investors' requirements for transparency and clarity. (Pharma)

The variety of lexical items used to express cause and effect was fairly wide:

20. **Due to** the change in product mix, we achieved double-digit growth in profitability. (Pharma)

21. We focus on those areas where scientific advances have opened up new opportunities that we consider most likely to **lead to** significant medical advances. (Pharma)

22. Our strategy … seeks to **stimulate** innovation and enhance the productivity of our research process. (Pharma)

In ideological terms, the frequent use of these words suggests that the writers of the reports habitually represent the things that happen in terms of direct causality: good decisions produce good results, adverse situations have a negative impact. The ethos of the annual report can thus be seen to reflect a worldview that is both utilitarian (seeking the maximum good in the most efficient way) and consequentialist (the rightness of the act depends on its consequences) (Stanford Encyclopedia of Philosophy, 2015). The seeming clarity of this utilitarian-consequentialist vision (good results are the consequence of our effective actions, while negative events are produced by external factors) is an important feature of company self-presentation in this genre.

3.6. Time: begin

The semantic area "Time: begin" is one of the more unexpected findings in the top 30 key areas of all four subcorpora. This area contains words related to continuity, and in these subcorpora is mainly accounted for by the high presence of verbs and adjectives relating to ongoing or sustained activity.

Rank	Pharmaceuticals	Food	Mining	Finance
1	Continue (0.34)	Continue (0.26)	Continue (0.26)	Continue (0.48)
2	Remain (0.10)	Remain (0.09)	Remain (0.14)	Remain (0.13)
3	Ongoing (0.03)	Ongoing (0.03)	Ongoing (0.05)	Ongoing (0.03)
4			Sustain (0.05)*	

* Not including sustainable, sustainability

Table 10. Most frequent word families in "Time: begin" (relative frequency >0.02).

It is noticeable from Table 10 that "continue" dominates this particular semantic field, which also includes items such as: "go on", "sustain", "sustained", "long-standing". It generally has a positive prosody here,

associated with sustained effort on the part of the company itself, presumably with a view to linking recent actions to past actions and showing continued purposeful activity:

> 23. We **continued** to manage our costs tightly and were pleased to deliver savings ahead of the targets we set out when we launched our major productivity initiatives. (Food)

The high keyness of this area not only suggests that the companies in question wish to present a dynamic view of time, fitting with an ideological framework favouring action, activity and consequentialism, but also indicates that they situate their present efforts within a long-term pattern, thus projecting what might be termed "proactive stability" or "sustained dynamism".

3.7. Areas of contrast between sectors

As was mentioned above, it was unsurprising that some of the most outstanding areas of salience were most strongly linked to the area of the companies' activities (e.g. for annual reports in the food sector, areas such as farming (LL 747.41) and food (LL 1214.41)), so these areas are not analysed here. However, categories like those outlined above ("Cause and effect", "Time: begin", etc.) which are not immediately related to the companies' activities offer considerably more interest. Other potentially interesting salient semantic areas which did not reach the top 30 in all four subcorpora (LL >130 in at least one subcorpus) are shown in Graph 3, which also displays their log likelihood in each subcorpus.
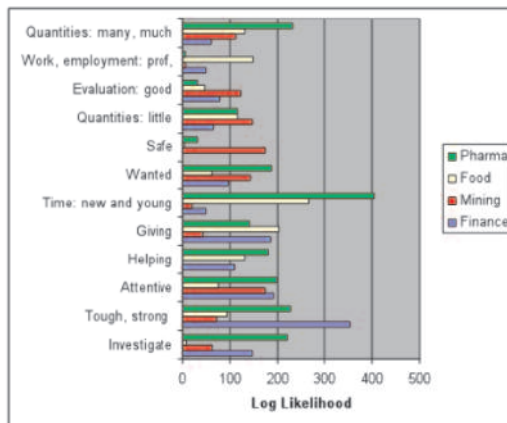
Graph 3. Semantic areas with keyness (LL >130) in at least one subcorpus.

One significant outlier in which the mining industry diverges from the other subcorpora is that of safety: the semantic area related to "safe" is not prominent in any of the other subcorpora. By the reverse logic of corporate communication, this hints at levels of danger associated with mining as compared with, say, managing money, and the heightened need to legitimise companies in this sector as responsible employers (see Breeze, 2012).

On the other hand, reports in the drugs and food industry are remarkable for their insistence on "newness", materialised in the keyness of "time: new and young". In the case of food, this "newness" refers to new business operations, but also notably to the freshness of the product, and, more frequently, the innovative nature of the production method or packaging.

> 24. We have long supported British farming and this year we achieved 100 per cent British sourcing for all our **fresh** pork. (Food)

> 25. The relaunch of Ryvita crispbread in **new** foil-fresh packaging drove increased sales. (Food)

In the pharmaceuticals industry, although "new" is still the most frequent word in this category, "innovate" and its derivatives come second. The stress here is on cutting-edge science, rather than on freshness and swift delivery:

> 26. People are still willing to pay for differentiated, **innovative** medicines that transform lives. (Pharma)

> 27. Omthera, a specialty pharmaceutical company based in the US, focused on the development and commercialisation of **new** therapies for dyslipidaemia. (Pharma)

Interestingly, while the pharmaceutical and food industries seem to have a preference for large quantities ("quantities: much and many"), small quantities seem to have a higher keyness factor in mining ("quantities: little"). It is evident that "much and many" point to use of the rhetoric of vague quantification (see above, in section 3.2), as in the following examples:

> 28. Associated British Foods is a diversified group of food, ingredients and retail businesses selling into **more** than 100 countries worldwide. (Food)

> 29. Ovaltine made **further** progress in its developing markets of Asia and South America. (Food)

The keyness of "quantities: little" would seem to point in the opposite direction, suggesting the smallness of things that are potentially negative, again, with a potentially legitimatory function:

30. During the second half of 2013, a consistent message from gold miners has been the need to **reduce** operating costs. (Mining)

31. 30 per cent **reduction** in the rate of new cases of occupational illness. (Mining)

Typical objects of "reduce" include costs, debt, emissions and energy consumption. In this, it seems that the ethos of mining companies is shaped by the need for control (perhaps related to the need to appear environmentally friendly with a view to legitimation) rather more than is that of the other sectors. However, "reduce" also appears with other objects, such as profits, revenues, values and share prices, which are undoubtedly indicative of negative results for the companies in the mining sector, perhaps linked to the financial crisis.

Another interesting feature of Graph 3 is that the semantic areas of "Wanted" and "Attentive" are prominent in pharmaceuticals, finance and mining, which is mainly due to the frequency of items such as strategy, aim, policy, programme and objective, on the one hand, and focus, on the other. It would appear that these three sectors are, at least explicitly, more highly strategy-driven than the food sector.

"Giving" is a prominent area in food and finance, which coincide in the prominence of lexical items related to "provide", "supply", "distribute" and "contribute". Notably, the financial sector appears to represent its activities rather as though it were providing a physical product like food:

32. They complement our organic efforts to broaden and strengthen our **distribution** channels and product mix. (Finance)

33. … by finding attractive and innovative investment opportunities globally to **provide** products which consistently meet our clients' investment needs both now and into the future. (Finance)

A similar dynamics seems to be operating in the semantic fields of "helping" (service, support, benefit, help, enable) and "giving" (provide, distribute, contribute, give). Here, particularly pharmaceuticals and food companies stress their social role, using lexical items related to support, help, care and service.

34. A number of non-governmental organisations and the World Health Organization, are leading efforts to **support** regions and countries in prioritising and introducing wider healthcare provision. (Pharma)

The finance sector also echoes this discourse – but here, the semantic area is dominated by the notion of "service":

35. …a highly regarded global asset management group founded on providing the highest levels of investment performance and client **service**. (Finance)

On the other hand, "giving" has a low keyness value in mining, and "helping" has none, which suggests that the companies in this sector do not portray themselves as providing benefits to society in this way.

Finally, the field "tough, strong" is also worthy of some attention. Most of the items tagged thus belong to the family of "strong", followed by "robust", "tough" and "resilient", and refer – in all the annual reports – to financial performance, as in the following example:

36. With the strength of the group's balance sheet and **strong** cash generation we have every reason to be confident in the continuing development of the group. (Finance)

This area is particularly salient in the annual reports from the financial sector, where "strong" seems to be used as a multipurpose positive evaluative adjective. In combinations such as "strong balance sheet", "strong investment performance", "strong track record" and "strong capital base", "strong" seems to be virtually synonymous with "good". The selection of this term points to a desire to create a more dynamic or "masculine" style (in the sense outlined by Hofstede, 1998), which seems particularly characteristic of the financial sector.

Interestingly, food is the outlier in the area of "investigate": the other subcorpora – particularly pharmaceuticals and finance – seem to be more research-driven.

In brief, the overall pattern that emerges from Graph 3 is as follows.

- Annual reports from the pharmaceutical industry stand out in their predilection for "Time: new and young", with an emphasis on innovation. This sector seems to blend high keyness for the

service-oriented values like "Giving" and "Helping" with the achievement-oriented values grouped under "Attentive" (focused, etc.), "Wanted" (strategies, aims, etc.) and "Tough, strong" (strong, strengthen, robust). Pharmaceuticals also gives greatest prominence to stressing large quantities, which ties in with its high keyness score on the category "Inclusion" (see above), and points again to the rhetoric of vague quantification discussed above.

- Food is the only sector with an emphasis on "Work and employment: professionalism" (colleagues, reputation), which suggests a greater emphasis on the human factor in its self-presentation. As is logical, "Giving" and "Helping" are also important here, since the food industry is traditionally concerned with providing goods for public consumption. One interesting point is the keyness of "Time: new and young", which is accounted for by the prominence of lexis such as "new" and "fresh", which have a particular relevance in this sector.

- Finance annual reports stand out from the others in their emphasis on strength ("Tough, strong"), and the importance of research ("Investigate"). They are also characterised by the prominence they give to "Helping" and "Giving" (service, distributing, providing), which, as in the case of the pharmaceuticals sector, is combined with achievement-oriented semantic areas grouped under "Attentive" (focus, etc.) and "Wanted" (strategy, aim).

- Mining annual reports stress superlative self-presentation ("Evaluation: good"), strategic action ("Wanted"), and high focus ("Attentive"). However, they also contain more safety-related content ("Safe") and make more reference to small as well as large quantities, perhaps reflecting a need to display control ("Quantities: many, much" and "Quantities: little").

4. Discussion

The annual reports of the four sectors analysed here coincide in reflecting a discourse of size, importance and power, with an underlying notion of time framed as sustained dynamic action that is both strategic and focused, and a deterministic philosophy of cause and effect that attributes success to the company's agency and difficulties to external factors. These results coincide

broadly with those obtained by other researchers such as Malavasi (2010), who researched banks' annual reports using standard corpus linguistic tools and found evaluative adjectives falling into the categories of efficiency, importance and client-orientation. Similar to our finding that time was conceptualised as dynamic, underpinned by strategic actions, her analysis also pointed to a strong future orientation, with many purposive verbs used to underscore the company's priorities. Like White (2003), we also found a strong reliance on the concept of growth, which can be best understood as an aspect of size and importance, and which fits with the rhetoric of vague quantification explained above. As in the texts discussed by Bhatia (2004, 2010), Craig and Amernic (2004), or Courtright and Smudde (2009), the prevailing ethos uncovered by our study is one that privileges positive actions and results, and thereby marginalises less satisfactory ones.

The findings of the present study fundamentally concur with previous research suggesting that business-related discourse is underpinned by a powerful, positivist rhetoric that sustains the symbolic order within the late capitalist system (Craig & Amernic, 2004), and that texts like the annual report reflect "the prevailing and hegemonic myths of the cultural and political environment within which the organisations operate" (Sandell & Svensson, 2016: 21). Moreover, the broad area of overlap between reports from the four sectors serves as evidence of the essentially repetitive, predictable nature of such texts, reflecting what has been termed "discursive isomorphism", that is, a tendency towards the homogenisation of corporate communication to reflect the prevailing rationalised concepts of what constitutes an appropriate or efficient practice (DiMaggio & Powell, 1983). Regarding Bhatia's discourses of public relations and economics (2010), even though the former ostensibly predominates in these corpora, it is clear that the latter ultimately sustains these values of size, growth, causality and sustained action: company self-representation feeds on wider macro-economic discourses. Taken together, the annual reports in this corpus rely on a common substrate in the ethos of utilitarian capitalism, which favours size, strength and competition, high focus and sustained dynamism.

Although the areas of commonality are more striking, the differences between sectors also warrant discussion. First, slight traces of discursive struggle are apparent if we look carefully at areas such as "Quantities: little" in the mining subcorpus. As previous research has shown, companies endeavour to legitimise their actions by pre-empting possible criticism, particularly when the sector they represent has been under fire (Craig & Amernic, 2004; Breeze,

2015). Drawing on institutionalised accounts that minimise blame and deflect criticism is "critical for the maintenance of organisational legitimacy and survival" (Sandell & Svensson, 2016: 21). To detect the presence of these discourses it is important to read between the lines, and to approach possible divergence from predicted patterns with sensitivity.

Other contrasts are equally interesting. While three of the sectors appear to embody a positive vision of their social role as givers and providers, it is only the food sector that noticeably pays lip-service to the importance of the human factor within the company. Moreover, the strength-focus-knowledge orientation of the financial sector can be seen to contrast with the ethos of working-serving in the food sector, or the strongly forward-pushing discourses of the pharmaceutical sector. Further research is needed to discover more about how the different business sectors construct persuasive discourses about themselves, and to explore possible differentiation between companies within one sector.

Regarding methodology, the use of semantic tagging rather than manual examination to locate areas of interest has both advantages and disadvantages. Analysis of individual items on word-lists, though time-consuming, might allow for greater sensitivity to polysemy, semantic prosody and ideological implications. Yet, once relevant single items have been isolated and scrutinised, the researcher is faced with the onerous secondary task of re-grouping them thematically in order to paint the broader picture. In the present study, the situation was in some sense reversed: reliance on a tagging system designed to find a pre-determined set of semantic categories allowed the initial search, and classification of the data, to be conducted more easily. On the other hand, since this method works with pre-established tags, it is not sensitive to complex interrelations between categories, and does not allow for the emergence of new categories in the intersection or overlap between existing ones. Further refinement of the semantic tagging system will probably resolve some of these issues, but it is likely that genre-specific or discipline-specific tagging will be necessary to provide optimum results in the long term.

5. Conclusions

Semantic tagging using Wmatrix3 has made it possible to identify and analyse the underlying value systems of annual reports across four sectors,

shedding light on a shared ideological adherence to values of size, dynamism, sustained action and causality. It has also pinpointed axiologically-loaded areas, such as service- or safety-orientation, in which sectors differ. Exploration of relevant concordance lines from key semantic areas has shed new light on the the promotional discourses of annual reports.

Acknowledgements

References

Abraham, A. & P.J. Shrives (2014). "Improving the relevance of risk factor disclosure in corporate annual reports". *The British Accounting Review* 46: 91-107.

Abrahamson, E. & E. Amir (1996). "The information content of the president's letter to shareholders". *Journal of Business Finance and Accounting* 23: 1157-1182.

Archer, D., A. Wilson & P. Rayson (2002). *Introduction to the USAS Category System*. URL: http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf [20/01/2018]

Banks, D. (1998). "Vague quantification in the scientific journal article". *ASp* 19-22: 17-27.

Bartlett, S. & M.J. Jones (1997). "Annual reporting disclosures 1970-1990: An exemplification". *Accounting, Business and Financial History* 7: 61-80.

Breeze, R. (2011). "Disciplinary values in legal discourse: a corpus study." *Ibérica, Journal of the European Association of Languages for Specific Purposes (AELFE)* 21: 93-116.

Breeze, R. (2012). "Legitimation in corporate discourse: oil corporations after Deepwater

Horizon." *Discourse and Society* 23,1: 3-18.

Breeze, R. (2015). *Corporate Discourse*. London: Bloomsbury.

Bhatia, V.K. (2004). *Worlds of Written Discourse*. London: Continuum.

Bhatia, V.K. (2010). "Interdiscursivity in professional communication". *Discourse and Communication* 21: 32-50.

Clatworthy, M. & M.J. Jones (2001). "The effect of thematic structure on the variability of annual report readability". *Accounting, Auditing & Accountability Journal* 14: 311-326.

Courtright, J.L. & P.M. Smudde, (2009). "Leveraging organizational innovation for strategic reputation management". *Corporate Reputation Review* 12: 245-269.

Craig, R.J. & J.H. Amernic (2004). "Enron discourse. The rhetoric of a resilient capitalism". *Critical Perspectives on Accounting* 15: 813-851.

Davison, J. (2010). "[In]visible [in]tangibles: Visual portraits of the business élite". *Accounting, Organizations & Society* 35: 165-183.

De Groot, E. (2014). "Corporate communication

and the role of annual reporting. Identifying areas for further research" in V.K. Bhatia & S. Bremner (eds.), *The Routledge Handbook of Language and Professional Communication*, 235-254. Oxford and New York: Routledge.

De Groot, E., H. Korzilius, C. Nickerson & M. Gerritsen, (2006). "A corpus analysis of text themes and photographic themes in managerial forewords of Dutch-English and British annual general reports". *IEEE Transactions on Professional Communication* 54: 1-17.

DiMaggio, P.J. & W.W. Powell (1983). "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields". *American Sociological Review* 48: 147-160.

Ditlevsen, M. (2012). "Telling the story of Danisco's annual reports from a communicative perspective". *Journal of Business and Technical Communication* 26: 92-115.

Giannoni, D.S. (2010). *Mapping Academic Values in the Disciplines: A Corpus-Based Approach.* Bern: Peter Lang.

Hofstede, G. (1998). *Masculinity and Femininity. The Taboo Dimension of National Cultures*. London: Sage.

Hyland, K. (1997). "Scientific claims and community values: Articulating an academic culture". *Language and Communication* 17: 19-31.

Hyland, K. (1998). "Exploring corporate rhetoric: Metadiscourse in the CEO's letter". *Journal of Business Communication* 35: 224-245.

Kheovichai, B. (2015). "Metaphorical scenarios in business science discourse". *Ibérica, Journal of the European Association of Languages for Specific Purposes (AELFE)* 29: 155-178.

Koller, V., A. Hardie, P. Rayson & E. Semino (2008). "Using a semantic annotation tool for the analysis of metaphor in discourse". *Metaphorik.de* 15: 141-160.

Malavasi, D. (2010). "The multifaceted nature of banks' annual reports as informative, promotional and corporate communication practices" in P.

Evangelisti Allori & G. Garzone (eds.), *Discourse, Identities and Genres in Corporate Communication*, 211-233. Bern: Peter Lang.

Perelman, C. & L. Olbrechts-Tyteca (1969). *The New Rhetoric: A Treatise on Argumentation*. Notre Dame: University of Notre Dame Press.

Rayson, P. (2008). "From key words to key semantic domains". *International Journal of Corpus Linguistics* 13: 519-549.

Rutherford, B. (2005). "Genre analysis of corporate annual report narratives". *Journal of Business Communication* 42: 349-378.

Sandell, N. & P. Svensson (2016). "The language of failure. The use of accounts in financial reports". *International Journal of Business Communication* 53: 5-26.

Semino, E. (2008). *Metaphor in Discourse*. Cambridge: Cambridge University Press.

Semino, E., Z. Demjén, J. Demmen, V. Koller, S. Payne, A. Hardie & P. Rayson (2015). "The online use of 'Violence' and 'Journey' metaphors by cancer patients, as compared with health professionals: A mixed methods study". *BMJ Supportive and Palliative Care* (Published online first).

Smith, M. & R. Taffler (2000). "The chairman's statement – A content analysis of discretionary narrative". *Accounting, Auditing & Accountability Journal* 13: 624-647.

Stanford Encyclopedia of Philosophy. (2015). URL: http://plato.stanford.edu/entries/

Stanton, P. & J. Stanton (2002). "Corporate annual reports: Research perspectives used". *Accounting, Auditing & Accountability Journal* 15: 478-500.

Vázquez Orta, I. & C. Foz Gil (1995). "The persuasive function of lexical cohesion in English: A pragmatic approach to the study of chairman's statements". *Estudios Ingleses de la Universidad Complutense* 3: 87-100.

White, M. (2003). "Metaphor and economics: The case of growth". *English for Specific Purposes* 22: 131-151.

Ruth Breeze has a Ph.D. in Applied Linguistics and has researched and published widely in the area of discourse analysis applied to media language and specialised language. She is a member of the GradUN Research Group in the Instituto Cultura y Sociedad at the University of Navarra, Spain. Her most recent books are *Corporate Discourse* (Bloomsbury Academic, 2015) and the edited volume *Interpersonality in Legal Genres* (Peter Lang, 2014).