(Article begins on next page)

Comparable corpora in cross-cultural genre studies: Tools for the analysis of CSR reports

Marina Bondi

1. Introduction

The chapter explores the potential of corpus tools in cross-cultural studies of genres. The discussion is based on important assumptions about corpus linguistics and cross-cultural genre analysis.

The first assumption is that a cross-cultural comparison of genres requires integration of lexical categories with semantic and functional pragmatic perspectives. Studies on specialized discourse have witnessed a constant interest in genre analysis over the past thirty years. Different approaches to genre in the 1990s (Swales 1990, Berkenkotter/ Huckin 1994, Christie/Martin 1997) were clearly marked by an interest in the definition of the organizational structure of genres. Identifying the rhetorical functions of a genre became a primary preoccupation, especially in the pedagogically oriented fields of ESP and EAP, largely influenced by Swales's two-layered move/step models. The early focus of genre analysis promptly extended to the linguistic features of the communicative functions and to cross-cultural comparison. Cross-cultural perspectives have rapidly caught the interest of scholars interested in different aspects of intercultural communication, often in relation to the expansion of English as the international language of knowledge and business. The need to communicate cross-culturally requires an awareness of convergences and divergences that may characterize both linguistic and rhetorical features of the discourses that identify the communicative event in various cultural contexts.

A second assumption regards corpus linguistics. Corpus-based and corpus-driven approaches have had enormous influence on all linguistic disciplines over the past thirty years or so. The most important contributions of corpus
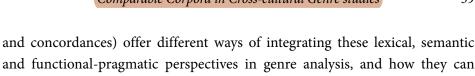
linguistics appear to me to lie in providing tools for the calculation of frequencies and for the study of the combination of elements, that is, phraseology in a wider sense. Corpus tools like frequency lists and keywords provide us with immediate and accurate access to quantitative data that may either substantiate our analysis or drive it. Corpus tools like concordancing have offered new perspectives to the study of words in combination and therefore to the study of meaning in text and of the interface between lexis and grammar. Sinclair's extended-units-of-meaning model (Sinclair 1991, 1996, 1998, Stubbs 2001, 2007, 2013), for instance, explores the phraseological tendency in language through four levels of analysis: collocation (words that occur regularly with the node word), colligation (grammatical categories that define the immediate context of the node word), semantic preference (the tendency to co-occur with words sharing the same semantic features) and semantic prosody (the tendency of the word to occur in specific pragmatic contexts, its connection to a specific speech act and/or evaluation).

The first corpus projects, like the Brown Corpus, were mostly monolingual. Although contrastive work emerged rather early, with pioneer work on a corpus of Serbo-Croatian translations of texts from the Brown Corpus (Filipović 1971), it was only in the early 1990s that seminal work by Johansson (Johansson/Hofland 1994) opened the way for further developments in the field (see Ebeling/Olsefjell Ebeling 2013 for an overview).

Parallel and comparable corpora have since been growing progressively, with different goals. Over the past three decades they have proved to be key training tools in machine translation and multilingual natural language processing, as well as essential elements in contrastive linguistics, translation studies and foreign language teaching. The burgeoning availability of multilingual parallel resources (different versions of the same text), for example within the EU, has stimulated increasing research on parallel corpora, especially with a view to corpus-based and corpus-driven approaches to translation (Saldanha 2009, Zanettin 2012, Laviosa 2015; see Doval/Sánchez Nieto 2019 for a recent overview). Comparable corpora, on the other hand, involve 'monolingual subcorpora designed using the same sampling frame' (McEnery/Xiao 2007: 132). These have provided an empirical basis for contrastive and cross-cultural studies (Johansson 2007), while drawing attention to the problems of assessing the comparability of their components, given the multiplicity of contextual factors in which texts are situated (Moreno 2008).

Using cross-cultural studies of Corporate Social Responsibility (CSR) reports as a case study, the chapter looks at how corpus tools (most notably keyness

and concordances) offer different ways of integrating these lexical, semantic and functional-pragmatic perspectives in genre analysis, and how they can contribute to different, complementary paths of integration (bottom-up and top-down). It also shows that the integration of bottom-up procedures (from lexical units to semantics and functional-pragmatic units) and top-down procedures (from functional-pragmatic units to semantics and lexical units) can actually take place at different levels of textual structure.

The chapter is structured as follows. Section 2 deals with the background literature on cross-cultural approaches to CSR reporting. Section 3 presents problems of corpus compilation. Section 4 provides an analysis of different ways of looking at the structure of CSR reports and at some of their key components. Sections 5 and 6 look at keyness and concordances as tools for the analysis of the semantic and pragmatic features of the genre.

2. Background: cross-cultural analysis of CSR reports

The concept of CSR, first introduced by Bowen (1953), has been paid increasing attention. The European Commission has defined it as the responsibility of enterprises for their impact on society (COM 2011). The expression refers to the company's strategies, activities and practices that are relevant to its various stakeholder groups, covering the economic (e.g., shareholders, investors), environmental (e.g., local communities) and social dimension (e.g., employees). In the past twenty years, with globalization and the increasing role of the Web, CSR communication has proved a key tool for organizational communication.

CSR communication has also attracted great attention in applied language studies. An obvious focus has been the key role that CSR communication plays in impression management and identity creation. This has created intensive work on issues of legitimation (Breeze 2012, Fuoli 2012), repurposing and interdiscursivity (Catenaccio 2010, 2011, 2012), evaluative language use (Malavasi 2007, 2011, 2012, Lischinsky 2011, Fuoli 2012, Bondi 2016b) and modality (Aiezza 2015, Bondi 2016a). In a critical perspective, Aditi Bhatia (2012) has highlighted the interplay of three types of discourses: a) the *discourse of promotion*, aiming at establishing credentials and building importance to (re)construct the corporate image; b) the *discourse of good will,* aiming at demonstrating that the company cares for society and is committed to CSR principles and action; c) the *discourse of self-justification,* used by companies to

legitimize their practices emphasizing the constraints forced on the companies 'by external conditions' (Bhatia 2012: 235).

CSR reporting has rapidly grown into institutionalized practice in the world of corporate communication (Contrafatto 2014, Tang *et al.* 2015). This has led to the development of reporting guidelines and standards regulating what is otherwise in most countries voluntary practice (Bondi/Yu 2018b). The Global Reporting Initiative (GRI), for example, provides a flexible framework for creating standalone sustainability (non-financial) reports or integrated reports. National and supranational strategic documents – such as national action plans, the EU *2011 CSR strategy* and directive *95, 2014* – may also influence the structure of reports. Beyond coercive and normative processes, however, what is most important is often 'mimetic isomorphism' (Bondi/Yu 2018a), that is, the tendency of an institution to imitate the structure or practice of more prestigious organizations (DiMaggio/Powell 1983).
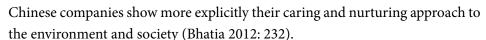
Studies in this field often aim to understand how CSR is communicated to different stakeholders in different cultural and linguistic contexts (Williams/ Aguilera 2008), highlighting elements of convergence and divergence. Scherer and Palazzo (2008) point at an interesting paradox in corporate reporting: corporate disclosure is clearly related to the global contexts and guidelines that influence the structure of reports, but it is also inevitably oriented to its immediate cultural context, that is, local orientations, business culture and corporate culture, rather than 'national culture' (see also Malavasi 2012).

On the one hand, it is true that companies all over the world are adopting the same standards and practices under the influence of globalization, converging national policies and international action (e.g. Fortanier *et al.* 2011, Tang *et al.* 2015). Many of the studies highlighting convergence focus on the themes and topics dealt with in reports, often drawing on the concept of legitimacy (e.g. Fortanier *et al.* 2011), that is, on the notion that the actions of an organization are only seen as appropriate within a socially constructed system of norms, values and beliefs (Suchman 1995: 574) and therefore tend to mirror them.

On the other hand, it is also possible to see the influence of specific cultural, social and political contexts (e.g. Bravo *et al.* 2013) with their values and norms. National cultural norms, organizational culture, professional culture and different types of stakeholders (communities, consumers, customers, regulators) may contribute to divergence patterns (Williams/Aguilera 2008, Tang *et al.* 2015). Bhatia (2012), for example, shows that US companies tend to articulate the discourse of goodwill in terms of company engagement with society, while

Chinese companies show more explicitly their caring and nurturing approach to the environment and society (Bhatia 2012: 232).

In line with what has been studied of globalizing and localizing trends in general, the two opposing forces may be seen to contribute to the coexistence of divergent local practices with convergent global trends (Einwiller *et al.* 2016, Kim *et al.* 2018) in forms of synergistic interaction that are often called forms of 'crossvergence' (Jamali/Neville 2011: 616, Conte *et al.* 2020: 9) or cross-cultural hybridity. A study of comparable corpora can help highlight these patterns from the point of view of cross-cultural genre analysis.

3. Corpora and issues of corpus compilation

The analytical pathways discussed in the next sections are based on two different but related comparable corpora for a cross-cultural study of CSR reports. Both corpora were created for cross-cultural analysis of the genre and were used for different kinds of analyses.

3.1. The Ba-CSR corpus

The first corpus is a comparable corpus of CSR reports in the banking sector comprising reports in English and in Italian. When looking at these two languages from a cross-cultural perspective, a number of observations are in order. A major question lies obviously in the comparability of the two languages in terms of use, rather than in terms of the two language systems. This issue is linked to the question of the varieties of English to be considered and the status of different banks.

First of all, in fact, it will be obvious that Italian tends to address a basically national audience, whereas English may be thought of as also addressing an international audience. The different international status of the two languages is undeniable, but this difference may be more or less relevant in different contexts. In many fields of academic discourse, for example, it is a matter of language choice, and ultimately of different degrees of competitiveness of the contexts addressing a national and international audience. In corporate reporting, the readership issue ties in to the type of company: companies may operate more locally or more internationally and may therefore produce the same document in both languages with a national or international audience of stakeholders in mind (employees, investors, competitors, analysts, trading

partners and activists). The CSR report is a high-stakes text anyway, for its role in communicating with stakeholders, as well as constructing and maintaining a positive corporate identity. This makes it a key document for smaller and large companies alike.

The choice of what varieties of English to consider is another important question. Rather than thinking of this in terms of native or non-native varieties, one could perhaps consider publication in English-speaking or non-English speaking contexts. Reports are often realized through different production processes, by national or multinational teams, but they are hardly ever the product of a single speaker and almost never of a definite speaker. This makes it difficult to establish systematically whether they were written by a native speaker, or by an international team, or rather translated or revised through language services, as there may be different production contexts globally and locally. In a way, from the point of view of professional communication in the field, the question is hardly relevant: the language variety that is relevant for professionals is primarily the English produced and disseminated by major companies in the field. It may thus be important to consider the use of English in an international context more generally, including both companies based in an English-speaking country, and companies based in non-English speaking countries.

The Ba-CSR corpus includes thirty standalone CSR reports from international banks who publish their reports in English and a 'comparable' corpus of thirty CSR reports in Italian from Italian banks, collected in the same time span of five years (before the introduction of integrated reports). The Ba-CSR-En component (Banking CSR in English) (about 855,000 words) consists of reports from six top international banks: three headquartered in the UK (Barclays, HSBC, Royal Bank of Scotland) and three headquartered in Spain, Germany and China, and thus more clearly representative of EIL (English as an International Language): Banco Santander, Deutsche Bank and ICBC. These are all major banks and, irrespective of the English-speaking or non-English speaking contexts, have a clear international standing, which makes their use of English representative of the English used in international banking communication. The Ba-CSR-It_ Component (ca. 1.5 million words) comprises reports in Italian from six major Italian banks producing a standalone CSR report at the time: BNL, Cariparma, Monte dei Paschi di Siena, Banca Intesa SanPaolo, Unicredit, UBI. The range of banks is much wider here in terms of size and international orientation: two banks have an international standing (Unicredit and IntesaSanpaolo), two operate within the influence of French groups (Cariparma and BNL) and the other two are more local (see Bondi 2016b for more information).

There are therefore obvious limits in the comparability of the two language-specific components, in terms of the dimensions of the companies and of their potential international readership. Comparability, however, is hardly exempt from limits in contrastive linguistics. What is important is to be aware of these limits when setting research questions and interpreting the data. A corpus of this kind can still be used to highlight features of cross-cultural rhetoric, in that it allows comparison of textual structures and pragmatic choices. The internal variation of each component can also be further explored to study variation within the same language system. It is always extremely important, in fact, to avoid oversimplification in cultural contrast and appropriate awareness of corpus design principles can help.

3.2. The CSR-ICE Corpus

The CSR-ICE Corpus (Yu/Bondi 2017, 2019) (about 3 million tokens) comprises ninety CSR reports published by companies in the banking sector and in the energy sector in one financial year. Rather than trying to trace diachronic development over a time span, a corpus of this kind takes a picture of a point in time but offers a wider picture in terms of companies involved (ninety altogether, though with the same number of CSR reports per language, i.e. thirty). Having more than one sector of course introduced more variables, but this was regarded as a plus when trying to study the structure of the genre, as it avoided the risk of describing a specific case and treating it as a general case.
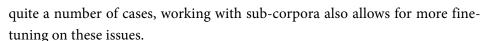
The CSR-ICE also offers a wider language span, in that it involves three languages: Italian, Chinese and English. The choice of three languages was deliberately taken in order to illuminate more clearly convergences and divergences of the three components, beyond easy oppositions (especially in terms of East and West). It was also interesting to have languages belonging to different families, so as to allow for a wider perspective when looking at more specific features.

Language and culture variation within the components was once again an issue: when looking for a definition of the generic structure of CSR reports (Yu/Bondi 2017), for example, it was important to consider a wide range of varieties and therefore we included British, American and Canadian companies. When interpreting the data, on the other hand, it may be important to keep in mind the different specific contexts and there may be different national influences in terms of legislation. Similarly, when looking at specific language questions, there may be norms that hold in one variety and do not hold in another. In

quite a number of cases, working with sub-corpora also allows for more fine-tuning on these issues.
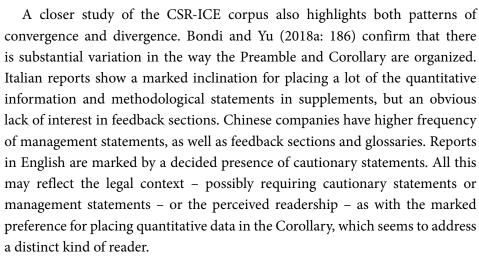
A combination of perspectives – working with the full corpus and with comparable subcorpora – is also often necessary for different reasons. When manually annotating a corpus in terms of rhetorical units or pragmatic functions, it is often expedient to reduce the scope of the analysis, while keeping the larger corpus as the appropriate tool for studying the frequency of occurrence of language features: a subcorpus of eighteen annotated reports (six per language) was thus created for the purpose. When the need is felt to have cultural homogeneity in the corpus from English-speaking contexts, it is also possible to select a homogenous subcorpus: Bondi/Yu (2018c), for example, used a subcorpus of twenty reports from China, Italy and the United States.

4. Exploring top-down pathways: Generic structure

The analysis of the rhetorical structure of reports is based on the tools of genre analysis (e.g. Swales 1990, Bhatia 1993, 2004). With longer documents like CSR reports, this may be done in different ways and stages. Bondi (2016b), for example, looks at the BA-CSR corpus trying to identify the highly complex nature of the whole document. The main report is introduced by a Preamble, consisting of part-genres like the front cover, the table of contents, the CEO's letter etc. and followed by a Corollary including guidelines, indicators, appendices etc. Borrowing the terminology from Parodi (2010), the complexity of these extended documents is characterized by a recursive structure that can be defined as a 'colony-in-loops', which appears to be based on the topics identified as relevant by the international guidelines of the GRI-Index. The structure also highlights two main sections aimed at self-presentation (with general issues of company history, corporate identity, governance etc.) and performance reporting (with a specific focus on economic, social and environmental performance) (Bondi 2016b: 177). The overall analysis highlights a homogeneous thematic structure across cultural contexts and marked repetitive patterns, with functional sequences repeated for each new topic and similar thematic organization. The organization of the report seems to be largely influenced by international standards, which lead to convergence in the nucleus of the macro-structure. There is variation in the satellites, however: Italian reports, for example, show a marked preference for including a conspicuous amount of facts and figures in the Corollary, with no explicit explanation for the lay reader.

A closer study of the CSR-ICE corpus also highlights both patterns of convergence and divergence. Bondi and Yu (2018a: 186) confirm that there is substantial variation in the way the Preamble and Corollary are organized. Italian reports show a marked inclination for placing a lot of the quantitative information and methodological statements in supplements, but an obvious lack of interest in feedback sections. Chinese companies have higher frequency of management statements, as well as feedback sections and glossaries. Reports in English are marked by a decided presence of cautionary statements. All this may reflect the legal context – possibly requiring cautionary statements or management statements – or the perceived readership – as with the marked preference for placing quantitative data in the Corollary, which seems to address a distinct kind of reader.

Closer move analysis (Yu/Bondi 2017) is based on a subcorpus of CSR-ICE consisting of eighteen reports (six per language), which was manually annotated (with Notetab) for the main moves used in CSR reports. Starting from Catenaccio's (2012) four-macro-move structure (presenting the company, with a focus on its role as responsible social actor; presenting the company's CSR strategy, reporting the company's activities and performance, situating the context), the study developed a model with fifteen moves, which were in turn analysed into steps/strategies used to realize them. The results certainly contributed to highlighting remarkable similarities in the general organization of the document and its main moves, while also confirming peculiarities, such as an Italian emphasis on performance reporting and a Chinese preference for presenting individual cases.

What is most relevant here, however, is that the process of manual annotation for rhetorical categories leads researchers to an exploration of the complexity of textual processes, as well as to the need for reliability checks. The identification of moves requires attention to lexico-grammatical, semantic and pragmatic features of the verbal and visual text, with a view to matching linguistic signals and units defined by content and function (see also Moreno/Swales 2018 for an overview of the problems).

Rhetorical tagging requires a number of steps for the identification of moves (see Yu/Bondi 2017: 277–81 for a more specific account of the annotation procedures). It is important to begin from the communicative purposes of the genre, but also to devise a procedure for segmenting the text into appropriate units. We aimed at defining the elements that influenced the inductive and cognitive process of move identification: the 'distinct linguistic boundaries that can be objectively analysed' mentioned by Biber *et al.* (2007: 24). In
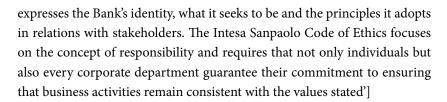
particular, we made use of formal realizations and elements related to the hallidayan metafunctions of language (ideational, interpersonal and textual, see e.g. Halliday 1985). By formal realizations we meant both elements of the verbal text (e.g. presence of future markers for the move 'Previewing future performance') and elements of the non-verbal text (e.g. paragraphing, highlighting, initial position etc.). A consideration of the metafunctions also helped us define units by a combination of their communicative purpose (e.g. presenting) and content-related themes (e.g. performance, values etc.). We also felt that in some cases it was important to grant the status of move to units that had a major textual function, like 'Introducing an aspect of CSR performance' (usually leading to 'Presenting performance' or 'Communicating strategies/ methods/ practices').

Identifying the moves first allowed measuring their extensiveness, showing for example what percentage of the report is actually devoted to a move. Chinese reports are characterized by greater extensiveness of moves showing commitment to CSR values, Italian reports by presenting performance moves and reports in English by presenting strategies/methods/practices (Yu/Bondi 2017: 282). The identification of moves and steps also helps to show typical sequences of moves or steps. For example, in the Ba-CSR corpus, where the English component shows a tendency to move from 'Identifying elements of vision/value' to 'Specifying strategy' and 'Stating strategies/ methods/practices', as in Example 1, the Italian corpus shows a tendency to proceed from strategy to values, as in Example 2.

1. Good governance. HSBC's governance structure is focused on delivering sustainable value to our shareholders. The strategy and risk appetite for HSBC is set by the Board, which delegates the day-to-day running of the business to the Group Management Board. Global businesses and functions have established operating, financial reporting and management reporting standards for application throughout HSBC. (HSBC)

2. I VALORI. Intesa Sanpaolo ha adottato nel luglio 2007 il Codice Etico, carta valoriale che esprime in primo luogo l'identità della Banca, quello che vuole essere e i principi che adotta nelle relazioni con i propri stakeholder. Il Codice di Intesa Sanpaolo è imperniato sul concetto di responsabilità e richiede, non solo ai singoli ma anche a ogni funzione aziendale, di garantire il proprio impegno perché le attività siano sempre coerenti con i valori dichiarati. (IntesaSanpaolo) ['OUR VALUES. In July 2007 Intesa Sanpaolo adopted the Code of Ethics, a charter of values that first of all

expresses the Bank's identity, what it seeks to be and the principles it adopts in relations with stakeholders. The Intesa Sanpaolo Code of Ethics focuses on the concept of responsibility and requires that not only individuals but also every corporate department guarantee their commitment to ensuring that business activities remain consistent with the values stated']

Finally, and most importantly, the identification of components of the text (whether sections or moves), allows further lexico-grammatical, semantic and pragmatic study. One can for example focus on the lexico-grammatical features of a specific move, beyond the constitutive elements that defined it or in terms of the specific patterns that characterized the constitutive elements. Yu and Bondi (2019), for example, focus on forward-looking statements in 'Previewing future performance' moves, by looking first at the moves and steps that characterize the two corpora and then at the language features that characterize the move. Looking at the frequency and distribution of moves helps to identify elements of convergence such as the constitutive steps of the move ('Committing to future actions' and 'Predicting future actions and results') and the predominance of commitment, while also showing cross-cultural divergences in the extensiveness of prediction, which is highest in English and lowest in Chinese. Looking at the language features that characterize the move, on the other hand, allows a study of the key role played by words of 'change' in the move and, conversely, the preference of words of 'change' for a semantics of futurity in their immediate context.

The next two sections look more closely at how to explore the language of CSR reports in relation to their functions and purposes.

5. Bottom-up pathways: Starting from frequency data

A preliminary analysis of frequency data can help illuminate new aspects or confirm previous hypotheses. This might mean looking at complete frequency lists or at selected frequencies that highlight what characterizes a text or a corpus. The study of Keywords, for example (as elaborated by *WS Tools,* Scott 2020) is the study of word forms whose frequency varies significantly across two corpora. With similar principles (and a tagged corpus) one can calculate the frequencies of the tags. *WMatrix* (Rayson 2009), for example, is a tool for corpus analysis and comparison that extends the keywords method to key grammatical categories and key semantic domains.

A study of keywords can be very useful in highlighting the peculiarities of the two main sections of the report. Using a version of Ba-CSR that separates the SP (Self-Presentation) section and the PR (Performance Reporting) section, the wordlist of the SP section can be contrasted with the one of the PR section in both languages (and vice versa, of course). The contrast produces a list of the words that are distinctively more frequent or less frequent.

The number and nature of words may vary according to the parameters used: basic frequency of each word, number of texts in which it is found etc. While an analysis of the whole set of lexical elements is obviously advisable (see Gabrielatos/Marchi 2011, 2012, Gabrielatos 2018, this volume), a brief overview of the top keywords can still provide a picture of the words that characterize each section. Table 3.1 reports the top fifty words characterizing the Italian SP section (using BIC as the key ordering principle, and arbitrarily restricting the words to those present in 40 per cent of the reports, in order to focus on the words that are used across a substantial range of reports). With these settings we obtain 573 word forms, 226 of which are positively key, that is, they occur more often than would be expected by chance in comparison with the reference corpus, while 347 are negative keywords, that is, they occur less frequently. The top fifty are all positive.

Table 3.1 Top fifty keywords in Italian

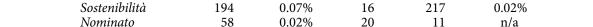| Key word	Freq.	%	Texts	RC. Freq.	Rc. %
Consiglio	902	0.31%	29	188	0.02%
Controllo	595	0.21%	29	288	0.03%
Comitato	528	0.18%	29	231	0.02%
Sorveglianza	364	0.13%	15	71	n/a
President	390	0.14%	25	94	n/a
Governance	342	0.12%	30	85	n/a
Stakeholder	521	0.18%	28	336	0.03%
Consigliere	258	0.09%	15	69	n/a
Direzione	410	0.14%	29	283	0.03%
Rischi	571	0.20%	30	553	0.06%
Amministrazione	354	0.12%	29	218	0.02%
Delegato	225	0.08%	27	61	n/a
Funzioni	315	0.11%	30	189	0.02%
Consiglieri	151	0.05%	17	13	n/a
Controlli	228	0.08%	26	105	0.01%
Codice	342	0.12%	29	279	0.03%
Governo	230	0.08%	28	115	0.01%
Compliance	215	0.07%	25	104	0.01%
Direttore	158	0.06%	27	41	n/a
Carica	117	0.04%	20	9	n/a
Statuto	146	0.05%	21	36	n/a

Key word	Freq.	%	Texts	RC. Freq.	Rc. %
Amministratore	110	0.04%	19	10	n/a
Vice	128	0.04%	20	24	n/a
CSR	237	0.08%	19	178	0.02%
Audit	134	0.05%	21	43	n/a
Piena	165	0.06%	21	79	n/a
Reputazione	150	0.05%	21	69	n/a
Responsabilità	299	0.10%	20	341	0.03%
Collegio	103	0.04%	20	32	n/a
Funzione	256	0.09%	27	267	0.03%
Generale	261	0.09%	29	291	0.03%
Soci	226	0.08%	23	231	0.02%
Controllate	97	0.03%	21	35	n/a
Societario	88	0.03%	23	27	n/a
Vigilanza	149	0.05%	26	108	0.01%
Nomina	90	0.03%	21	33	n/a
Presidio	153	0.05%	23	121	0.01%
Sindacale	114	0.04%	19	68	n/a
Supervision	73	0.03%	21	19	n/a
Incarichi	78	0.03%	19	25	n/a
Sportelli	150	0.05%	19	126	0.01%
PIL	56	0.02%	13	6	n/a
Nomine	56	0.02%	16	6	n/a
Organi	119	0.04%	26	80	n/a
Esecutivo	63	0.02%	19	13	n/a
Interno	177	0.06%	30	182	0.02%
Sostenibilità	194	0.07%	16	217	0.02%
Nominato	58	0.02%	20	11	n/a
Comitati	122	0.04%	26	93	n/a
Amministratori	95	0.03%	16	58	n/a

If these words are grouped around relevant lexico-semantic fields, such as mission and vision, company history and description, stakeholder consultation etc., we may be able to see that the Italian SP section is largely dominated by the lexis of corporate governance – *consiglio* ('Board'), *controllo* ('control'), *comitato* ('Committee'), *president* ('President'), *governance, consigliere* ('Councillor'), *direzione* ('management'), *delegato* ('delegate', 'Chief officer'), *amministrazione* ('administration') etc. – with great emphasis on control functions – *controllo* ('control'), *sorveglianza* ('supervision'), *rischi* ('risks'), *controlli* ('controls'), *audit*, *vigilanza* ('surveillance'), *supervisione* ('supervision') – whereas the role of 'mission and vision' is relatively reduced – *codice ('code'), compliance, CSR, reputazione ('reputation'), responsabilità ('responsibility'), sostenibilità* ('sustainability') – and the role of stakeholders is very limited and generically identified – *stakeholders*.

If we look at the same list for English, the picture is more varied, even if the list of positive keywords is much smaller – it is limited to fifty-five items. The lexis of corporate governance is obviously present there too, but not so largely

dominant as in the Italian corpus. The list shows greater attention to depicting stakeholders and their engagement, as well as a clear focus on banking services – for example, *deposits, equity, dividends* – and the presence of a number of metadiscursive nouns – for example, *issues, report, disclosure* – that highlight the previewing function of the first section of the report, pointing to various textual units.

Table 3.2 groups the fifty-five word forms (listed in decreasing order) according to topic or function areas.

The table shows there is much greater attention to describing financial issues in the English corpus than in the Italian one, an impression which is confirmed by looking at the whole list of 226 word forms that characterize the self-presentation section in Italian. In Italian reports the terms from the world of finance and economics are limited to a couple of references to financial services offered by the bank – *sportelli* ('branches')*, leasing, factoring* – and to the general economic context – *economie* ('economies'), *PIL ('GDP').*

The arbitrariness of thematic categorizations is obvious, but it can be countered by adopting specific principles of selection: in this case, for example, the choice was made with a view to the typical thematic elements identified in the functional analysis of SP sections of reports. It is also obvious that there are many cases where the attribution of lexical words to semantic areas necessarily requires looking at concordances: anything to do with accounting, for example, could relate both to the governance of the bank itself and to the services provided.

Table 3.2 Positive keywords in the English corpus

Groupings	Word forms
Corporate governance	*committee, board, directors, governance, executive, director, audit, chairman, regulators, meeting, committees, rules*
Vision	*responsibility, CSR, citizenship, citizen, reputation, integrity, vision*
Stakeholders	*stakeholders, stakeholder, shareholders, engagement, dialogue, engage, investor, views, surveys, media, opinion, discussions*
Bank history or general description	*founded, strength, owned, presence, limited*
Financial services	*tax, profit, remuneration, profits, deposits, profitable, ratio, profitability, equity, stability, dividends*
Labelling nouns	*issues, report, disclosure, page, highlights*
Function words	*You, what, before*

Negative keywords complete the picture, as they point at what is typically missing in this section when compared to the actual performance reporting. The most marked case in both components of the corpus is that of figures, representing 3.96 per cent of the tokens in the SP section vs 4.95 per cent in the PF section of the English corpus, and 4.94 per cent vs 5.93 per cent respectively in the Italian corpus. The data confirm the general structure of the genre, with the vast majority of quantitative data referring to CSR activities rather than to the bank in general; it also confirms a tendency of Italian reports to provide more figures overall.

Other negative keywords refer to specific sectors of CSR activity, with their practices – for example, *energy, programme, project, training, housing, electricity, art, service, cooperation, emissions; energia* ('energy'), *progetto/I (*'project/s'), *emissioni* ('emissions'), *formazione* ('training'), *iniziative* ('initiatives'), *ricerca* ('research') – and actions – *launched, help, support, reduce, continue, providing, supported; sostegno* ('support'), *riduzione* ('reduction'), *ridurre* ('reduce'), *realizzato* ('realized'), *contribuito* ('contributed') – as well as their specific stakeholders – *teachers, universities, youth, schools, children, students, employees; giovani* ('young'), *donne* ('women'), *clienti* ('clients'), *lavoratori* ('workers'), *fornitori* ('suppliers').

The convergence on topics is thus very strong. Divergences, on the other hand, appear to be mostly linked to well-known tendencies of writing in the two cultural contexts, with more frequent verbal elements in English and more frequent nominalizations in Italian. The most remarkable difference in English is shown by verbs – for example, *launched*, with fifty-two occurrences (0.04 per cent) in SP vs 533 (0.09 per cent) in PR – whereas in Italian it is represented by nominalizations of actions – for example, *sostegno* ('support', noun), with the same normalized figures, that is, 555 occurrences (0.04 per cent) vs 900 (0.09 per cent).

What is most important to notice, however, is that function words also deserve attention. Rather than pointing at the topics of the text, they point at other stylistic traits, such as the typical need to use specific structures, which can in turn be related to the meanings and functions of the text.

The most obvious function word in our list of positive keywords in the English corpus appears to be the use of *you*. This is often studied as a marker of establishing dialogue with the reader. What is interesting to notice here, however, is that in this case it does not manifest a higher degree of personalization in the SP section of the report. The higher incidence of the pronoun is rather due to a very high number of occurrences in a series of interviews reported by a single bank

(DB). The vast majority of the examples are either direct interview questions, or exclusive pronouns that refer impersonally to the activity of economic and financial agents, rather than the reader, for example:

(3) [Q:] Do you think that a focus on pure commerce and profit can achieve scale faster than a hybrid approach that combines commerce with a social mission?

    [A:] While interest rates matter to poor people, access to money is far more important to them. Because so many poor need money, it is true that they will accept higher interest rates. At the same time, I can't agree that you can grow fast only with a purely commercial approach that allows you to charge whatever interest you can. I don't think this will yield good results. (DB)

Rather than pointing at the reader-friendly nature of the section, the pronoun seems to point at the different uses that reports can make of a multiplicity of voices (cfr. Bondi/Yu 2017, Bondi/Yu 2018c).
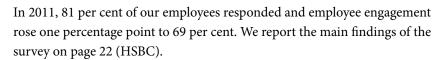
Similarly, the presence of *before* might suggest a key role of the temporal dimension in the company history. A look at the concordances shows however that 27/66 occurrences are in fact part of the financial expression *before tax*, thus leaving a more limited role to the temporal dimension of the company history or the economic context.

The presence of *what* is much more interesting in terms of qualifying the SP section. The expression is mostly used here as interrogative pronoun (66/127 occurrences), rather than a nominal relative pronoun (61/127). In the brief preview of issues that is usually offered in the SP section, direct and indirect questions can play an important cataphoric, introductory function by marking transition from one topic to the other. About a third of the total occurrences are topic-setting headings in the form of indirect questions – *What we do; What citizenship means to Barclays* (12/127) – or topic-setting direct questions – *What areas show social deficits? What would a road-map look like? What are the most important issues that HSBC's sustainability reporting should address?* (27/127). All the other indirect interrogative *what*-clauses (24/127) are found in passages that typically introduce elements of CSR by identifying an area to be explored. These dependent interrogatives introduce a new issue, often specified in the lines that follow, much in the same way as direct questions:

(4) HSBC's Global People Survey helps us to measure employee engagement in order to understand **what** actions are needed to increase business performance and to benchmark our progress both internally and externally.

> In 2011, 81 per cent of our employees responded and employee engagement rose one percentage point to 69 per cent. We report the main findings of the survey on page 22 (HSBC).

On the whole, the presence of these occurrences, combined with that of metadiscursive labelling nouns, would be largely interpretable in terms of the often-observed tendency of English writing to use more explicit forms of guiding the reader through the text.

Keyword analysis can then suggest items and groups of items for more focused analysis. The need to look at concordances in a lexico-grammatical perspective indeed remains the key contribution of corpus linguistics to cross-cultural genre analysis.

6. From the lexico-grammar to semantics and pragmatics

Concordance analysis is the key tool that allows, through the study of co-text, the integration of lexico-grammatical, semantic and pragmatic perspectives. Adopting Sinclair's extended-units-of-meaning model, the node word(s) can be studied by looking at collocation, colligation, semantic preference and semantic prosody. The exploration of context with a view to lexico-grammatical patterns and combinations of meanings also allows the study of local grammars (context-specific terminologies meant to analyze specific pragmatic or discourse acts (e.g. Sinclair/Hunston 2000, Su/Wei 2018) or semantic sequences (sequences of semantic elements that may reveal patterning even in contexts of formal variation, Hunston 2008). These can help reconstruct pragmatic functions and sequences.

Looking in particular at evaluative phraseology (Hunston 2011), Bondi (2016b) uses markers of importance to illustrate the procedure starting from a lexical set. The study shows that, when looking at syntactic patterns or patterns of semantic preference around markers of importance in importance claims, they can mostly be related to three main types of statements:

1. quality-attributing statements: for example, *è per noi di importanza cruciale* ('it is of crucial importance for us') /*.. is of fundamental importance*
2. statements involving mental processes (showing awareness of commitment to particular aspects): *essere consapevoli/convinti dell'importanza* ('to be aware of the importance'); *recognize/realize the importance etc.;*

3. more explicit references to verbal/argumentative processes: *ribadire/ sottolineare l'importanza* ('to reaffirm/ underline the importance'); *highlight/ underscore the importance.*

The three types can be illustrated in the terms adopted by Sinclair and Hunston's local grammar of evaluation. The table below provides an example (see Bondi 2016b for the full argument).

We can exemplify a similar procedure with a focus on the lexis used to refer to stakeholders, starting from the most general term (*stakeholder*) in the SP section of the report. Looking at collocation and colligation easily leads to semantic preference.

A preliminary overview of collocates of English *stakeholders* (228 occurrences, 1.7 ptw) shows that the top ten lexical words are: *key, business, engagement, dialogue, shareholders, various, important, interest, engage, range.* The top ten grammatical words, on the other hand, are: *our, and, the, with, its, are, for, that, all, from.*

The immediate co-text in the concordance of *stakeholders* clearly shows that the node word is typically identified by a possessive – *our* (65) or *its* (17) in L1 position – or qualified by importance – *key* (13)*, important* (3) – and range of

Table 3.3 Statements of understanding/belief ('Vision statements')

Evaluation framework		Entity/process evaluated
Evaluator 'WE' + expression of awareness/ belief	Evaluation IM – Importance marker	Element of 'Vision'
We place	great importance	on activities that stimulate economic growth (RBS)
The Bank has been attaching	great importance	to communication with shareholders (ICBC)
Siamo consapevoli ('We are aware')	dell'importanza ('of the importance')	di avere una chiara visione a supporto della gestione dei contributi a favore del territorio (Cariparma) ('of a clear vision for the management of contributions to the local community')
UniCredit è convinta ('UniCredit Group is fully aware')	dell'importanza ('of the importance')	della lotta ai fenomeni del riciclaggio e del finanziamento del terrorismo (Unicredit) ('of combating money laundering and terrorist financing')

Source: M. Bondi, 'CSR Reports in English and Italian', (Newcastle. Cambridge Scholars 2016b: 187).

the stakeholders – *various* (9), *different* (3), but also *broad/wider range of* (7). It is also often found in the vicinity of words referring to other, more specific categories – *customers* (8), *shareholders* (11), *clients* (5), *investors* (5) – and preceded by prepositions which in turn are related to various nominal, adjectival or verbal forms expressing either the notion of engagement – *engage with* (28), *dialogue with* (12), *listening to* (14) – or of relevance – *important* (11), *matter most to* (7), *of interest to* (9). The co-text on the right is perhaps less marked by dominant patterns, but a few trends emerge: for one thing, the word form is often found in collocations like *stakeholders' expectations* and *stakeholders' interests*; moreover, the expression is often followed by *and*, and often listed with *shareholders*; similarly, it is often followed by *are*, and different forms of the verb *include* (in turn followed by types of stakeholders).

It soon becomes evident that it is possible to subdivide all the occurrences, irrespective of the specific forms used, into three main types of semantic preference: forms of engagement/communication between the bank and its stakeholders (97/226, 43 per cent), expressions of relevance/concern on the part of stakeholders (93/226, 41 per cent), identification of types of stakeholders (37/226, 26 per cent).

If we look at *stakeholder* (plural) in Italian, excluding the loan expression *stakeholder engagement* (twenty-nine occurrences), we have 471 occurrences (1,6 ptw), quite similar in normalized frequency to the English form. A preliminary overview of collocates shows that the top lexical words are: *coinvolgimento* ('engagement'), *tutti* ('all'), *dialogo* ('dialogue'), *gruppo* ('group'), *aspettative* ('expectations'), *interessi* ('interests'), *attività* ('activities'), *relazioni* ('relations'). The top grammatical words, on the other hand, are: *degli (dei/l/lla/lle)* ('of the'), *gli* ('the'), *con* ('with'), *per* ('for'/ 'to'), *nostri* ('our'), *che* ('that'/ 'which'), *propri* ('own'), *nei (nel)* ('in the'), *una* ('a'/ 'an'), *agli(/lla)* ('to the').

Looking at patterns of collocation, colligation and semantic preference in the concordance of *stakeholder* in Italian, the co-text shows that the node word is typically identified by articles, but also by possessives – *nostri* ('our') (66), *propri* ('own') (15) and *suoi* ('their') (7) – and quantifiers and adjectives highlighting the range of stakeholders – *tutti* ('all') (42), *altri* ('other') (12), *diversi* ('several') (12), *vari* ('various') (4) – or their importance – *principali* ('main') (8), *chiave* ('key') (4). The figures confirm a limited use of personal reference and greater attention to generic identification of stakeholders in Italian. The node word is indeed found in the vicinity of words referring to more specific categories, but with markedly lower frequency than in the English corpus – *clienti* ('clients') (15), *azionisti* ('shareholders') (6). The node word is also preceded

by prepositions related to expressions of 'engagement' – *coinvolgimento degli stakeholder* ('engagement') (52), *dialogo con gli stakeholder* ('dialogue with') (29), *relazione con gli stakeholder* ('relations') (9) – or 'relevance' – *aspettative degli stakeholder* ('expectations') (12), *interessi di tutti gli stakeholder* ('interests') (12), *valore per tutti gli stakeholder* ('value') (5). When looking at semantic preference altogether, the approximate ratio of occurrences associated to what the bank does to engage the stakeholders is largely dominant in Italian reports (67 per cent of the occurrences), whereas both stakeholder identities and expectations/values are rather limited (around 16 per cent each).

Aiming to move towards larger patterns, it is possible to highlight recurrent combinations of meanings and thus identify the main functions of the sentences in which the expression is used. This points at functional units of the text which may correspond to the 'steps' in move-step analysis. In particular, in this case, they can be associated with steps that contribute to moves such as 'Stating values and beliefs', 'Stating strategies, methods, practices' or 'Showing commitment' to materiality assessment (i.e. identifying CSR priorities on the basis of stakeholders' interests).

Roughly speaking, when the node word is associated to specific references to stakeholders through linking verbs, expressions of inclusion and attributive complement constructions, we are often facing steps such as 'Identifying the nature/range of stakeholders':

(5) *'Stakeholders' are* groups and individuals who have an interest in, and may be affected by, our business activities. (RBS)

(6) *Riconosciamo come stakeholder* i soggetti portatori di un interesse rilevante rispetto all'attività della Banca, *quali* gli azionisti/soci, i dipendenti, la clientela, […]. ['We see stakeholders as holders of relevant interests in relation to the bank's business activities, like shareholders/partners, employees, our clientele, […]'] (MPS)

The step 'Presenting general methodological strategies' adopted by the company to establish contact with stakeholders is usually introduced by general verbs (*be/have*) attributed to the report or the company, verbs of reporting, expressions of intent or expressions of engagement:

(7) The Group's approach *is to work with* stakeholders to share knowledge and expertise (HSBC)

(8) *Our engagement with s*takeholders serves to inform and advise our decision-making process across our operations. *We are committed to developing*

constructive relationships with all groups impacted by our business as we progress though our five-year restructuring programme.

(9) *L'iniziativa* il Circolo di Sostenibilità di Bank Austria *organizza incontri frequenti* con stakeholder, ONG, mondo scientifico e pubblica amministrazione su temi come […] ['The UniCredit Bank Austria Sustainability Circle initiative organizes frequent meetings with stakeholders, NGOs, scientists and public administrators on topics such as […]'] (Unicredit)

(10) È stato attivato un *percorso di coinvolgimento* degli stakeholder che […] dà luogo alla costituzione di cantieri/laboratori permanenti che consentiranno di *integrare* in maniera sistematica il *punto di vista* degli stakeholder nelle attività aziendali […] nel quadro di un percorso *condiviso.* ['A structured stakeholder involvement program has been launched […]; this program will result in the constitution of permanent laboratories that will work on systematically integrating the points of view of the stakeholders in the Group's activities […] within the framework of a shared project'] (IntesaSanpaolo)

The combination of *stakeholders* with expressions of 'relevance' and reference to 'issues' characterizes steps 'Introducing material issues':

(11) the table below provides an overview of the engagement mechanisms and *key issues* for our main stakeholders at a Group-level. (RBS)

(12) Il risultato è un grafico di *materialità* multi-stakeholder che rappresenta i vari livelli di *rilevanza* che i nostri stakeholder (clienti, dipendenti e collettività) attribuiscono ai differenti *temi.* ['The result is the multi-stakeholder materiality chart that represents the varying levels of importance that our stakeholders (customers, employees and communities) place on different topics'] (Unicredit)
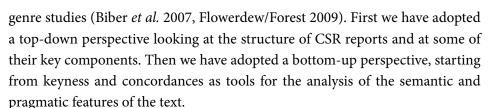
Looking at words in combination thus brings us back to potential functional contexts in which preferred combinations occur.

7. Conclusions

The chapter has explored the potential of corpus tools in cross-cultural genre studies, focusing on a cross-cultural analysis of CSR reports. Starting from a discussion of problems of corpus compilation, the chapter has provided examples of two main pathways in looking at the intersection of schematic structures of the genre and their lexico-grammatical realizations, a long-standing key question in

genre studies (Biber *et al.* 2007, Flowerdew/Forest 2009). First we have adopted a top-down perspective looking at the structure of CSR reports and at some of their key components. Then we have adopted a bottom-up perspective, starting from keyness and concordances as tools for the analysis of the semantic and pragmatic features of the text.
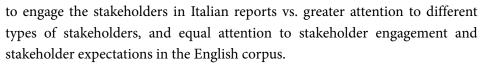
The top-down functional analysis has highlighted a homogeneous thematic structure across cultural contexts and a common set of basic moves; variation has been noticed in the organization of the corollary of the main report and in the sequences of steps. This perspective also helps analyze lexico-grammatical features building up the moves (Bondi 2016b, Bondi/Yu 2017, Yu/Bondi 2017). In more general methodological terms, I hope to have shown that a reliable identification of moves requires attention to definitions that include communicative purpose, topic, textual function, lexico-grammatical features and elements of the non-verbal text (Yu/Bondi 2017). Once the functional units have been identified, they can be more thoroughly described in terms of recurrent lexico-grammatical patterns and recurrent semantic patterns (Yu/Bondi 2019).

The bottom-up lexico-grammatical study presented here has focused on the distinctive features of the self-presentation section of the reports when compared to the performance-reporting section. Starting from keywords obtained with similar parameters in the two language components of the corpus, the preliminary overview grouped lexical word forms around relevant lexico-semantic fields and looked at how function words could be related to textual or interpersonal elements, to typical meanings and functions of the text. Common elements manifested convergent thematic, textual and interpersonal features of the report, together with peculiarities of the two cultural contexts – closer attention to metadiscourse and to presenting financial issues and banking services in the English corpus vs. greater attention to normative regulation and quantification in the Italian corpus.

Similarly, starting from individual words or lexical sets, it is possible to reconstruct their common lexico-grammatical associations (collocation, colligation, patterns) as well as their semantic and pragmatic associations (semantic preference and semantic or discourse prosody). This once again highlights both patterns of convergence – for example, the presence of common areas of semantic preference and of the same communicative functions around the node expressions – and patterns of divergence – for example, fewer personal patterns in identifying stakeholders and greater attention to what the bank does

to engage the stakeholders in Italian reports vs. greater attention to different types of stakeholders, and equal attention to stakeholder engagement and stakeholder expectations in the English corpus.

The distinction between the two pathways illustrated here is mostly heuristic. The movement can often be cyclically combined. The complexity of texts suggests the need for an integration of both perspectives, as well as of the different levels of analysis involved: lexico-grammatical, semantic, pragmatic.
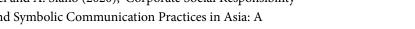
References

Aiezza, M. C. (2015), '"We May Face the Risks" … "Risks that Could Adversely Affect Our Face". A Corpus-assisted Discourse Analysis of Modality Markers in CSR Reports', *Studies in Communication Sciences*, 15 (1): 68–76.

Berkenkotter, C. and T. N. Huckin (1994), *Genre knowledge in Disciplinary Communication: Cognition/Culture/Power*, London: Routledge.

Bhatia, A. (2012), 'The Corporate Social Responsibility Report: The Hybridization of a "Confused" Genre (2007–2011)', *IEEE Transactions on Professional Communication*, 55 (3): 221–38.

Bhatia, V. K. (1993), *Analysing Genre: Language Use in Professional Settings*, Harlow: Longman.

Bhatia, V. K. (2004), *Worlds of Written Discourse: A Genre-based View*, London: Continuum.

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

Bondi, M. (2016a), 'The Future in Reports: Prediction, Commitment and Legitimization in CSR', in W. Cheng (ed.), *Pragmatics and Society, Special issue on The Pragmatics of Professional Discourse*, 7 (1): 57–81.

Bondi, M. (2016b), 'CSR Reports in English and Italian: Focus on Generic Structure and Importance Markers', in G. Garzone, D. Heaney and G. Riboni (eds), *Language for Specific Purposes. Research and Translation across Cultures and Media*, 168–99, Newcastle: Cambridge Scholars.

Bondi, M. and D. Yu (2017), '"I Highly Commend Its Efforts to Ensure Power Supply": Exploring the Pragmatics of Textual Voices in Chinese and English CSR Reports', in M. Janebová, E. Lapshinova-Koltunski and M. Martinková (eds), *Contrasting English and Other Languages through Corpora*, 218–47, Newcastle: Cambridge Scholars.

Bondi, M. and D. Yu (2018a), 'The Generic Structure of CSR Reports: Dynamicity, Multimodality, Complexity and Recursivity', in W. Giordano and G. Garzone (eds), *Discourse Communication and the Enterprise: When Business Meets Language*, 176–205, Newcastle: Cambridge Scholars.

Bondi, M. and D. Yu (2018b), 'CSR Between Guidelines and Voluntary Commitments', in G. Tessuto, V. K. Bhatia and J. Engberg (eds), *Frameworks for Discursive Contexts and Practices for the Law*, 40–68, Newcastle: Cambridge Scholars.

Bondi, M. and D. Yu (2018c), 'Textual Voices in Corporate Reporting: A Cross-cultural Analysis of Italian, Chinese and English CSR Reports', *International Journal of Business Communication*. Available online: https://doi. org/10.1177/2329488418784690 (accessed 6 June 2021).

Bowen, H. R. (1953), *Social Responsibility of the Businessman*, New York: Harper & Row.

Bravo, R., L. D. Chernatony, J. Matute and J. Pina (2013), 'Projecting Banks' Identities Through Corporate Websites: A Comparative Analysis of Spain and The United Kingdom', *Journal of Brand Management*, 20 (7): 533–57.

Breeze, R. (2012), 'Legitimation in Corporate Discourse: Oil Corporations after Deepwater Horizon', *Discourse & Society*, 23 (1): 3–18.

Catenaccio, P. (2010), 'Representations of Corporate Philanthropy. A Linguistic Approach', in P. Evangelisti Allori and G. Garzone (eds), *Discourse, Identities and Genres in Corporate Communication*, 121–42, Bern: Peter Lang.

Catenaccio, P. (2011), 'Green Advertising and Corporate CSR Communication: Hybrid Discourses, Mutant Genres', in S. Sarangi, V. Polese and G. Caliendo (eds), *Genre(s) on the Move. Hybridization and Discourse Change in Specialized Communication*, 353–72, Napoli: ESI.

Catenaccio, P. (2012), *Understanding CSR Discourse. Insights from Linguistics and Discourse Analysis*, Milano: Arcipelago.

Christie, F., J. R. Martin, eds (1997), *Genre and Institutions: Social Processes in the Workplace and School*, London: Cassell.

COM 2011. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Renewed EU strategy 2011–14 for Corporate Social Responsibility, COM/2011/0681. Available online: https://eur-lex.europa.eu/LexUriServ/ LexUriServ.do?uri=COM:2011:0681:FIN:en:PDF (accessed 6 June 2021).

Conte, F., A. Vollero, C. Covucci and A. Siano (2020), 'Corporate Social Responsibility Penetration, Explicitness, and Symbolic Communication Practices in Asia: A National Business System Exploration of Leading Firms in Sustainability', *Corporate Social Responsibility and Environmental Management*, 27 (3): 1425–35.

Contrafatto, M. (2014), 'The Institutionalization of Social and Environmental Reporting: An Italian Narrative', *Accounting, Organizations and Society*, 39 (6): 414–32.

DiMaggio, P. J. and W. W. Powell (1983), 'The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields', *American Sociological Review*, 48 (2): 147–60.

Doval, I. and M. T. Sánchez Nieto, eds (2019), *Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications*, Amsterdam and Philadelphia: John Benjamins.

Ebeling, J. and S. Oksefjell Ebeling (2013), *Patterns in Contrast*, Amsterdam and Philadelphia: John Benjamins.

Einwiller, S., C. Ruppel and A. Schnauber (2016), 'Harmonization and Differences in CSR Reporting of US and German Companies', *Corporate Communications: An International Journal*, 21 (2), 230–45.

Filipović, R. (1971), 'The Yugoslav Serbo-Croatian- English Contrastive Project', in G. Nickel (ed.), *Papers in Contrastive Linguistics*, 107–14, Cambridge: CUP.

Flowerdew, J. and R. W. Forest (2009), 'Schematic Structure and Lexico-Grammatical Realization in Corpus-based Genre Analysis: The Case of Research in the Phd Literature Review', in M. Charles, S. Hunston and D. Pecorari (eds), *Academic Writing: At the Interface of Corpus and Discourse*, 15–36, London: Continuum.

Fortanier, F., A. Kolk and J. Pinkse (2011), 'Harmonization in CSR Reporting: MNEs and Global CSR Standards', *Management International Review*, 51: 665–96.

Fuoli, M. (2012), 'Assessing Social Responsibility: A Quantitative Analysis of Appraisal in BP's and IKEA's Social Reports', *Discourse & Communication*, 6 (1): 55–81.

Gabrielatos, C. (2018), 'Keyness Analysis: Nature, Metrics and Techniques', in C. Taylor and A. Marchi (eds), *Corpus Approaches to Discourse: A Critical Review*, 225–56, Oxford: Routledge.

Gabrielatos, C. and A. Marchi (2011), 'Keyness: Matching Metrics to Definitions', *Theoretical-Methodological Challenges in Corpus Approaches to Discourse Studies - and Some Ways of Addressing Them*. Available online: http://eprints.lancs. ac.uk/51449 (accessed 6 June 2021).

Gabrielatos, C. and A. Marchi (2012), 'Keywords: Appropriate Metrics and Practical Issues', *CADS International Conference, Bologna, Italy, 13–15 September 2012*. Available online: https://repository.edgehill.ac.uk/4196 (accessed 6 June 2021).

Halliday, M. A. K. (1985), *An Introduction to Functional Grammar*, 1st edn, London: Edward Arnold.

Hunston, S. (2008), 'Starting with the Small Words: Patterns, Lexis and Semantic Sequences', *International Journal of Corpus Linguistics*, 13 (3): 271–95.

Hunston, S. (2011), *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*, New York: Routledge.

Jamali, D. and B. Neville (2011), 'Convergence versus Divergence of CSR in Developing Countries: An Embedded Multi-layered Institutional Lens', *Journal of Business Ethics*, 102, 599–621.

Johansson, S. (2007), *Seeing through Multilingual Corpora*, Amsterdam: John Benjamins.

Johansson, S. and K. Hofland (1994), 'Towards an English-Norwegian Parallel Corpus', in U. Fries, G. Tottie and P. Schneider (eds), *Creating and Using English Language Corpora*, 25–37, Amsterdam: Rodopi.

Kim, C., J. Kim, R. Marshall and H. Afzali (2018), 'Stakeholder Influence, Institutional Duality, and CSR Involvement of MNC Subsidiaries', *Journal of Business Research*, 91: 40–7.

Laviosa, S. (2015), 'Corpora and Holistic Cultural Translation', in M. T. Sánchez Nieto (ed.), *Corpus-based Translation and Interpreting Studies: From Description to Application*, 31–51, Berlin: Frank & Timme.

Lischinsky, A. (2011), 'In Times of Crisis: A Corpus Approach to the Construction of the Global Financial Crisis in Annual Reports', *Critical Discourse Studies*, 8: 153–68.

Malavasi, D. (2007), *Promotion in Banks' Annual Reports: An Integrated Analysis of Genre, Evaluative Lexis and Institutional Identity*, Modena: Il Fiorino.

Malavasi, D. (2011), '"Doing Well by Doing Good": A Comparative Analysis of Nokia's and Ericsson's Corporate Social Responsibility Reports', in G. Garzone and M. Gotti (eds), *Discourse, Communication and the Enterprise. Genres and Trends*, 193–212, Bern: Peter Lang.

Malavasi, D. (2012), 'The Necessary Balance between Sustainability and Economic Success: An Analysis of Fiat's and Toyota's Corporate Social Responsibility Reports', in P. Heynderickx, S. Dieltjens, J. Geert, P. Gillaerts and E. de Groot (eds), *The Language Factor in International Business. New Perspectives on Research, Teaching and Practice*, 247–64, Bern: Peter Lang.

McEnery, T. and R. Xiao (2007), 'Parallel and Comparable Corpora: The State of Play', in Y. Kawaguchi, T. Takagaki, N. Tomimori and Y. Tsuruga (eds), *Corpus-based Perspectives in Linguistics*, 131–45, Amsterdam: Benjamins.

Moreno, A. I. (2008), 'The Importance of Comparable Corpora in Cross-cultural Studies', in U. Connor, E. Nagelhout and W. Rozycki (eds), *Contrastive Rhetoric: Reaching to Intercultural Rhetoric*, 25–41, Amsterdam: John Benjamins.

Moreno, A. I. and J. M. Swales (2018), 'Strengthening Move Analysis Methodology towards Bridging the Function-form Gap', *English for Specific Purposes*, 50: 40–63.

Parodi, G., ed. (2010), *Academic and Professional Discourse Genres in Spanish*, Amsterdam: Benjamins.

Rayson, P. (2009), *Wmatrix: A Web-based Corpus Processing Environment*. Available online: http://ucrel.lancs.ac.uk/wmatrix/ (accessed 6 June 2021).

Saldanha, G. (2009), 'Principles of Corpus Linguistics and their Application to Translation Studies Research', *Tradumàtica: traducció i tecnologies de la informació i la comunicació*, 7. Available online: https://www.raco.cat/index.php/Tradumatica/article/view/154828 (accessed 6 June 2021).

Scott, M. (2020), *WordSmith Tools version 8*, Liverpool: Lexical Analysis Software.

Scherer, A. and G. Palazzo (2008), 'Globalization and Corporate Social Responsibility', in A. Crane, A. McWilliams, D. Matten, J. Moon and D. Siegel (eds), *The Oxford Handbook of Corporate Social Responsibility*, 413–31, Oxford: Oxford University Press.

Sinclair, J. (1991), *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

Sinclair, J. (1996), 'The Search for Units of Meaning', *Textus*, 9: 75–106.

Sinclair, J. (1998), 'The Lexical Item', in E. Weigand (ed.), *Contrastive Lexical Semantics*, 1–24, Amsterdam: John Benjamins.

Sinclair, J. and S. Hunston (2000), 'A Local Grammar of Evaluation', in S. Hunston and G. Thompson (eds), *Evaluation in Text: Authorial Stance and the Construction of Discourse*, 74–101, Oxford: Oxford University Press.

Stubbs, M. (2001), *Words and Phrases*, Oxford: Blackwell.

Stubbs, M. (2007), 'Quantitative Data on Multi-word Sequences in English: The Case of the Word *World*', in M. Hoey, M. Mahlberg, M. Stubbs and W. Teubert (eds), *Text, Discourse and Corpora*, 163–89, London: Continuum.

Stubbs, M. (2013), 'Sequence and Order: The Neo-Firthian Tradition of Corpus Semantics', in H. Hasselgård, J. Ebeling and S. Ebeling (eds), *Corpus Perspectives on Patterns of Lexis*, 13–33, Amsterdam: Benjamins.

Su, H. and N. Wei (2018), '"I'm Really Sorry about What I Said": A Local Grammar of Apology', *Pragmatics*, 28 (3): 439–62.

Suchmann, M. (1995), 'Managing Legitimacy: Strategic and Institutional Approaches', *Academy of Management Review*, 20, 571–611.

Swales, J. (1990), *Genre Analysis: English in Academic and Research Settings*, Cambridge: Cambridge University Press.

Tang, L., C. C. Gallagher and B. Bie (2015), 'Corporate Social Responsibility Communication through Corporate Websites: A Comparison of Leading Corporations in the United States and China', *International Journal of Business Communication*, 52 (2): 205–27.

Williams, C. A. and R. V. Aguilera (2008), 'Corporate Social Responsibility in a Comparative Perspective', in A. Crane, A. McWilliams, D. Matten, J. Moon and D. Siegel (eds), *The Oxford Handbook of Corporate Social Responsibility*, 452–72, Oxford: Oxford University Press.

Yu, D. and M. Bondi (2017), 'The Generic Structure of CSR Reports in Italian, Chinese, and English: A Corpus-Based Analysis', *IEEE Transactions on Professional Communication*, 60 (3): 273–91.

Yu, D. and M. Bondi (2019), 'A Genre-Based Analysis of Forward-Looking Statements in Corporate Social Responsibility Reports', *Written Communication*, 36 (3): 379–409.

Zanettin, F. (2012), *Translation-Driven Corpora. Corpus Resources in Descriptive and Applied Translation Studies*, Manchester: St Jerome.