https://doi.org/10.1075/RESLA.19048.RAB

**Autores**: Rosa Rabadán, Isabel Pizarro, Hugo Sanjurjo-González

# Authoring support for Spanish language writers: A genre-restricted case study

Authoring support consists of (semi)automated aids to be used at different stages during the writing process. Language information, however, tends to be restricted to areas such as spelling and grammar checking or term banks, and text construction difficulties that writers face concerning the structure of particular genres, associated sentence formulations or genre-specific vocabulary have not received proper attention. An additional gap in the research is that this support is generally addressed to English language users. This paper addresses these concerns focusing on a particular genre: the company's directors' report, and on Spanish language writers writing in English. A custom-made monolingual corpus has been analyzed using Bhatia (1993, 2004) and Swales (1990, 2004) definitions of genre and move combined with theme characterization. Recurrent strings for each move/step, which are conventionally associated with each rhetorical unit, were identified and formulated as "meta-strings." The bilingual glossary includes domain-specific items as well as move/step or genre-specific lexical and phraseological options, i.e., elements used irrespective of the business, places or people involved. The results are valuable by themselves, as an analysis of the genre, but also as the empirical basis for the authoring support tool that we present here, and as language training materials.

**Keywords:** authoring support, corpus linguistics, genre analysis, move structure, directors' report

## 1. Introduction

Authoring support consists of a series of (semi)automated aids offering linguistic information to be used at different stages during the writing process. Methods in

authoring support include spell checking and grammar checking, style rules, terminology checking and sentence clustering, and the offer concerns principally English (Siegel, 2018, p. 89-90).

Linguistically, spelling and grammar checking rely on morphological, PoS-tagged corpora, from which relevant information is retrieved. Style rules focus primarily on sentence-level to ensure text flow, reduction of ambiguity and intelligibility. The aim is to optimize texts for hassle-free communication. Style rules are frequently part of controlled languages aimed at establishing pre-editing routines, e.g. the recommendation to avoid passive constructions in technical documentation, which intend to facilitate multilingual text generation using machine translation. The rules for Global English (Kohl 2008) are a good example. Terminology checking rests on term banks extracted from available corpora as defined by the user(s).

Both spelling checkers and grammatical checkers are fully integrated into writing routines (e. g. Windows revision complements), and more sophisticated options, including style checking, are already commercially available (e. g. Grammarly premium version). Predictive writing, as enabled in everyday apps (e.g. WhatsApp Messenger), can be considered a basic form of sentence clustering.

This brief review suggests that there are at least three aspects that are not addressed by available authoring support: rhetorical structuring for specific genres, sentence formulations associated with specific rhetorical parts, and also genre-specific vocabulary and phraseology, distinct from expert terminology, and which plays an essential role in successful text production (Author et al. 2019).

Additionally, authoring aids do not cater specifically to the needs of non-L1 English users, as they are generally addressed to either native speakers of English or proficient users of English as an L2 that may or not be conversant with genre conventions.

The aim of this paper is three-fold: (1) to provide a description and characterization of the internal makeup of a particular genre, the director's report (DR), a type of economic performance report issued by companies listed on the stock market; (2) to use findings to propose genre-based authoring support guidance for Spanish language

writers when producing their DRs, and (3) to incorporate these results into authoring support software GEDIRE©.[1]

Next, section 2 reviews genre and domain-related work and outlines the characterization of the genre; section 3 presents the corpus and the move-and-step annotation procedure. Section 4 presents the rhetorical database resulting from the move-and-step analysis, including obligatory, expected and optional moves as well as their respective steps. Genre-specific vocabulary and phraseology and their organization are addressed in section 5, while the "meta-string" database is the object of section 6. A further section deals with the authoring support tool and a tour of its features from the user's point of view. The conclusions include considerations on machine-aided text production and controlled natural languages (CNLs).

## 2. The case study: the genre Directors' Report

The directors' report (DR) is a type of document which reports a company´s performance about financial, accounting and corporate social responsibility standards. It is not a completely independent genre, as it belongs to the genre set Annual Report (AR), an overarching text type that includes multiple sub-genres (Jameson, 2000; Zanola, 2010), while being independent in the sense that each has its purposes, internal coherence and readers. The sub-genres are also interdependent as each one contributes to the composition of the complete AR.

DRs are produced on behalf of the board of directors to encourage transparency in corporate governance. They are legally required when operating in the UK, are a necessary document for large and medium-sized companies. Furthermore, as Ditlevsen (2010, p. 164) points out, "many non-English companies publish two versions of the annual report (…) to reach an international audience". A DR reports to stakeholders and banking supervisors. It has legal status as an accounting document, given that expanded audit requirements for companies include the scrutiny of Directors' Reports. The one overriding communicative purpose of a DR is to comply with the principle of providing an accurate and honest account, one that reflects reality and does not mislead. The DR provides a truthful account of the financial situation of the company

[1] The acronym stands for GEnerator of DIrectors' Reports/ *GEnerador de Directors' REports*

and is in accordance with the fourth anti-money laundering EU directive in several critical areas, including the fight against the financing of terrorists (EUR-Lex, 2015).

Despite its importance in international business, the genre DR does not seem to have received due attention. The available research on the AR genre set reflects the use of diverse approaches. Bhatia (2010) found that ARs are a combination of four different types of discourse- accounting discourse, the discourse of economics, public relations discourse and legal discourse- which show different weights in each sub-genre. DRs mainly use the discourse of economics and partially the discourse of public relations. Garzone (2004, 2005) analyzes discoursal features and cultural markedness in chief executive officers' (CEO) letters from Italian and foreign companies; Mobasher, Ali, Abdullah & Chan (2013, p. 140) review studies on Corporate Annual Reports (CAR) that were conducted between 1990 and 2012 and report the "need to consider the genre content and structure of different sections within CARs using corpus analysis tools." Their review shows that most studies focus on specific sub-genres including Presidents' Letters (Kohut & Segars, 1992), Letters to the Shareholders (Prasad & Mir, 2002), Management Statements (Thomas, 1997; Hyland, 1998), or, Corporate Profiles (De Groot, 2008).

There are also semiotic analyses of the images of ARs (David, 2001) and content matter approaches, e. g., to corporate history narrative (Delahaye et al., 2009). Another group of studies take a genre perspective and account for move and strategy identification within AR sub-genres other than DRs. These include Nickerson & De Groot's (2005) comparison of British and Dutch Chairman Statements and CEO's Statements, and De Groot's (2008) analysis of Management Statement, Corporate Profile and Operational Review.

However, few studies have explicitly discussed the DR as a genre. Among them, Hyland's work (1998, p. 232) which compares the metadiscourse used in the CEO's letter with that in the DR and characterizes the latter as typically including "a review of the development and activities of the company and its subsidiaries during the year, information on company training and employment policies, and a summary of acquisitions and disposals." Also, De Groot (2008) offers scattered comments about the DR as a compulsory part of British ARs in her comparison of ARs from the UK and the Netherlands. Here she defines DRs as "a factual record of board activities and

policies" (2008, p. 29) that "take a management approach to financial reporting" (2008, p. 84) for the benefit of those involved or interested in the governance of a particular company.

Hyland (1998, p. 227) noted that the DR "reviews the year and describes important events affecting the company, changes in fixed assets, details of directors, and so on" and British informants indicated that the DR aim is "formally informing readers" (De Groot, 2008, p. 74). It is also a marketing tool as it presents the company as an efficient, reputable business to prospective partners and clients.

Concerning the linguistic makeup of the genre, previous studies (Author et al., 2015) suggest that DRs share resources with other sub-genres of the AR genre set, e. g., the audit report. Work on other business-management genres, e. g., meeting minutes (Author, 2016), shows that DRs also employ common strategies, e.g., opening, closing. However, any useful characterization of the DR necessarily calls for an in-depth analysis of other genre-specific features, i.e., move structure and discourse organization, as well as syntactic and lexical choices. These will be offered in sections 4 and 5.

## 3. Method

This section describes the corpus, its compilation process and the annotation procedure in full detail.

### 3.1 The corpus

Our research is based on a custom-made monolingual specialized corpus (M-En-GEDIRE), collected by the authors, consisting of 120 directors' reports extracted from their annual reports and totalling 230,646 words. The majority of studies on other genres belonging to the genre set AR are based on smaller corpora, e.g. Rutherford's (2005) analysis of Operating and Financial Review (44 texts), Thomas's (1997) study of Letters to the Shareholders (5 texts), Dragsted's (2014) case study of Letters to Shareholders (10 texts), Flowerdew & Wan (2010)'s analysis of Audit Reports (25). The exception is Hyland's (1988) work on CEO's letters and DRs, which feature 137 and 110 texts, respectively.

DRs were extracted manually from their respective ARs, which were published between 2003 and 2012, and saved as txt files (UTF-8) compatible with linguistic software like Antconc, WordSmith Tools and the ACTRES Tagger and ACTRES Browser. The main problem at this stage was identifying DRs within the AR. ARs were recognized by their official name —Annual Report— which was used as the single seed term in the browsing and identification process, whereas DRs were identified by their specific headings in the AR table of contents, as either Directors' Report or Report of the Directors. In this process, image-only and scanned pdfs were discarded due to their laborious and time-consuming conversion into txt files and, following Rutherford (2005, p. 360), the elements which were not part of the main narrative flow text were computed but not considered for the study, i.e., images, graphs and drawings (Table 1). This decision did not make a significant qualitative difference to the corpus composition because, as already observed by Hyland (1998, p. 232), "the report contains accounts and tabular information, but it is largely composed of running text". Neither did it make a great difference quantitatively, as the linguistic contents included in these elements are below 1.5% of the total number of words in the corpus. As the fact that the texts were publicly available from the companies' official websites meant there was no need to preserve confidentiality, the texts were not anonymized concerning personal, place or company names and brands.

The corpus is made of full texts, including their headings, that were treated as regular text. Their size ranges from 5,183 to 387 words, with an average of 1,922 words per text, which differs from Hyland's (1998, p. 232) average of about 1,200 words. The sample size range is not considered a disadvantage for corpus design, as specialized small corpora (Henry & Roseberry, 2001) are appropriate for the analysis of discourse structure (Biber, Connor & Upton, 2007) because they show a higher concentration of recurrent patterns than larger, broad-ranging corpora. It helps to verify whether DRs, regardless of length, follow a well-established rhetorical scheme, and, if so, which parts are more needed to produce an exemplar of DR, and which are less so and can, therefore, be used discretionarily.

Although produced by international companies, 97.5% of our corpus texts are written in British English, which means that dialectal variation is not an issue. DRs were produced by major companies operating in a wide range of business sectors,

including energy, finance, food industries, insurance, mining, telecommunications, technology, among others. Variation in DR business sectors is an essential factor when seeking to establish genre homogeneity -or not- and also in providing empirical proof of whether the rhetorical structure is dependent on the topic. This variety was achieved by including just one DR per company to avoid author or company bias and by choosing companies from a range of sectors to ensure that there is no topic bias either.

As an analysis of the genre, the findings are valuable in themselves, but they are all the more so as the empirical basis for providing authoring support, and as part of a set of training materials, as pointed by Biber, Connor & Upton (2007) and Upton & Cohen (2009) among others.

### 3.2. Annotation: Moves and steps

In this paper, we have used Bhatia (1993, 2004) and Swales (1990, 2004) working definitions of genre and move combined with theme characterization. These authors define a move as a functional unit that has a specific communicative purpose that fulfills the general purpose of the genre. The general purpose of the DR (one which all genres in the AR share), is "to give a true and fair view of the state of the company's affairs (…); to provide a positive image of the company" (Ditlevsen, 2012, p. 97).

The procedure to identify the specific purposes of each segment is to qualitatively characterize the moves and steps (Biber & Conrad, 2009, p. 34). It has been done "using a functional-semantic approach (…) rather than a reliance on linguistic criteria, to identify the intention of the text and the textual boundaries" (Kwan, 2006, p. 36). This way, we avoid 'the form-function gap' (Moreno & Swales, 2018), i.e., this part qualifies as move x because it features items a, b. Instead, we take the opposite view: rhetorical constituents are defined notionally, then linguistic items are examined for recurrent associations with a particular move or step.

In the approach adopted here, themes are also taken into consideration (Gómez González & García Varela, 2014). There are two reasons for this: (1) themes frequently help clarify rhetorical segmentation (Upton & Cohen, 2009), and (2) except opening and closing, move functions in DRs (reporting, informing, presenting information, etc.) can be reduced to one, i.e., reporting. However, in a purely functional approach, most of such segmentation would be irrelevant.

Once the moves and steps have been defined, two researchers–coders established an initial tagset and jointly annotated 20% of the texts. This task was carried out using the ACTRES tagger[2]. The remaining 80% of the texts was split between the coders. Agreement in average percentage terms was very high (97%), but, as in Kwan (2006, p. 37), it turned out that there were discrepancies in the case of twelve indeterminate steps. The average Kappa coefficient was good (0.65) in one case, but rated below 0.50 in eight of the steps proposed initially. Thus, differences were discussed, and a joint decision made: as part of these initial steps did not show a distinct functional definition and were irrelevant in terms of frequency (below 10% occurrence; <0.6% of corpus words), the textual material was subsumed into other theme-related occasional steps. Once the final tagset was fully agreed upon (Table 1, moves; 2 to 4 steps), the same modifications were also applied to the batch of texts used in the pilot analysis. Likewise, minor discrepancies in step "embeddedness" resulted in no consequence for our research aims and were not reflected in the final annotation scheme. Expert knowledge consultation was readily available throughout the process from one of the researchers-coders.

To establish the degree of obligatoriness of the different rhetorical parts in the construction of a typical DR, once the corpus was annotated and ready for use, frequencies were analyzed. The higher the frequency, the more necessary a move/ step is deemed to be.

Authors such as Biber, Connor & Upton (2007), Parkinson (2017), Yang & Allison (2003) and Suter (1993), among others, have used scales of obligatoriness which share a minimum of two levels ('obligatory' versus 'optional'). More commonly, however, scales include more levels with a certain degree of arbitrariness in the weighting of the ratios. Lewin et al. (2001, p. 36) established that "all structures do not have to be present to realize an instance of the genre," and since it is rare to find a 100% move frequency, it is accepted that frequencies higher than 80% indicate that the

[2] This tool allows the user to upload a custom-made tagset and offers a user interface that greatly improves the time-consuming annotating experience. The annotated texts can be queried using the web-based ACTRES rhetorical browser. When querying a corpus such as M-En-GEDIRE the browser's built-in statistical features help to unveil move and step size and frequencies and to single out move/ step-bound linguistic patterns.

move/step is obligatory. In line with Henry & Roseberry (2001), only segments that occur in above 10% of the corpus DRs are considered rhetorically significant, but those with frequencies below 10% are classified as occasional and computed in the quantitative analysis. We adopt the following four-tier scale to define the rhetorical grid of DRs:

1. Obligatory (OB): appearing in between 100% and 80%,
2. Expected (E): between 79% and 50%,
3. Optional (OP): between 49% and 11%,
4. Occasional (OC): below 10%

Moves and steps were scored according to this scale, and their hierarchy established. They were also classified according to their size -measured in percentage of corpus words and mean word length- to verify whether this offers an insight into their relevance for genre construction (Parkinson, 2017) or, on the contrary, whether length is not related to move/step obligatoriness (Lewin et al., 2001, p. 87).

Move complexity is determined in terms of the number of embedded steps: low (< three steps), intermediate (four-to-five steps) and high (> five steps). Complexity is considered to ascertain whether it relates to size and degree of obligatoriness or whether the rhetorical pattern is unrelated to either or both.

The rhetorical annotation presented in this section was used to tag the corpus components (moves and steps) for analysis, which is addressed in the following section.

## 4. Move structure and rhetorical database

The following pages present the rhetorical structure of the genre and its corresponding database, which is part of the authoring support tool described in Section 7.

The move structure yielded by our analysis can be seen in Table 1. It includes move raw frequency, frequency as a percentage, degree of obligatoriness, complexity (number of steps per move), and move size logged in as a percentage of total corpus words. Linguistic elements frequently associated with each move/step are offered and illustrated in Figures 1 to 10 to substantiate our claim that these are not genre, move or

step-specific. No illustrative examples are offered for steps ranked as *optional* or *occasional*.

**Table 1.** DR move structure

| **Moves and steps:**<br><br>**functions and themes** | **Move raw frequency** | **Frequency as %** | **Obligatoriness** | **Steps per move** | **% of corpus words** |
|---|---|---|---|---|---|
| M1 Opening the DR | 111 | 92.5 | OB | 0 | 1.44 |
| M2 Reviewing the company activities | 117 | 97.5 | OB | 3 | 7.71 |
| M3 Reporting on financial matters | 120 | 100 | OB | 3 | 10.72 |
| M4 Presenting information related to directors | 120 | 100 | OB | 9 | 20.61 |
| M5 Accounting for shareholdings | 113 | 94.17 | OB | 6 | 14.53 |
| M6 Giving information on tangible property | 19 | 15.83 | OP | 0 | 0.5 |
| M7 Explaining company risks | 46 | 38.33 | OP | 3 | 4.86 |
| M8 Announcing company agreements | 31 | 25.83 | OP | 2 | 2.27 |
| M9 Describing employees' issues | 86 | 71.67 | E | 7 | 9.1 |
| M10 Presenting issues on corporate management | 86 | 71.67 | E | 6 | 11.78 |
| M11 Stating business projections | 58 | 48.33 | OP | 2 | 2.28 |
| M12 Giving information about auditing | 113 | 94.17 | OB | 5 | 5.46 |
| M13 Detailing AGM issues | 85 | 70.83 | E | 4 | 6.19 |

| | | | | | |
|---|---|---|---|---|---|
| M14 Noting post balance sheet events | 24 | 20 | OP | 0 | 0.33 |
| M15 Signing off the DR | 117 | 97.50 | OB | 4 | 0.73 |
| [Text in images, graphs, drawings] | | | | | 1.48 |
| Total | | | | | 100 |

4.1. Obligatory moves

Data show seven obligatory moves that occur in more than 90% of the reports, the lowest frequency being 92.5%. They constitute the core of the DR, account for 61.2% of the words in the corpus, and constitute the predictable core structure of the genre (Table 1). M3 *Reporting on financial matters* and M4 *Presenting information related to directors* show 100% frequency as they occur in all the corpus texts. The hierarchical organization of steps embedded in moves identified as obligatory can be seen in Table 2. Further results in the table include frequencies, degree of step obligatoriness and percentage of text associated with each step.

**Table 2**. Obligatory moves: Steps.

| **Obligatory moves: Steps** | **Step raw frequency** | **% frequency** | **Obligatoriness** | **Size** | **% of corpus words** |
|---|---|---|---|---|---|
| **Move M1: Opening the DR** | | | | | |
| **Move M2: Reviewing the company activities** | | | | | |
| Step 1: Reviewing the present company activities | 109 | 90.83 | OB | 7647 | 3.31 |
| Step 2: Reviewing the past company activities | 69 | 57.5 | E | 9403 | 4.08 |
| Step 3: Stating the future company activities | 15 | 12.5 | OP | 608 | 0.26 |
| **M3 Reporting on financial matters** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Step 1: Presenting results and dividends | 118 | 98.33 | OB | 11513 | 4.99 |
| Step 2: Listing charitable and political donations | 100 | 83.33 | OB | 4321 | 1.87 |
| Step 3: Stating supplier payment policy and creditor days | 113 | 94.17 | OB | 8764 | 3.8 |
| **M4 Presenting information related to directors** | | | | | |
| Step 1: Listing directors and biodata | 104 | 86.67 | OB | 5817 | 2.52 |
| Step 2: Listing appointments, re-elections and replacements | 101 | 84.17 | OB | 11348 | 4.92 |
| Step 3: Reporting Directors' conflicts of interests | 105 | 87.5 | OB | 11807 | 5.12 |
| Step 4: Reporting remuneration | 32 | 26.67 | OP | 1883 | 0.82 |
| Step 5: Indicating indemnity | 29 | 21.67 | OP | 1973 | 0.86 |
| Step 6: Indicating insurance | 25 | 20.83 | OP | 1270 | 0.55 |
| Step 7: Explaining powers of directors | 18 | 15 | OP | 1212 | 0.53 |
| Step 8: Notifying Board and committee meeting attendance | 9 | 7.5 | OC | 1024 | 0.44 |
| Step 9: Stating the Statement of responsibilities | 40 | 33.33 | OP | 10504 | 4.55 |
| **M5 Accounting for shareholdings** | | | | | |
| Step 1: Detailing substantial shareholding | 110 | 91.67 | OB | 8506 | 3.69 |
| Step 2: Stating share capital | 75 | 62.5 | E | 6201 | 2.69 |
| Step 3: Explaining purchase of own shares | 37 | 30.83 | OP | 4181 | 1.81 |
| Step 4: Describing allotment and issue of shares | 33 | 27.5 | OP | 5205 | 2.26 |

| | | | | | |
|---|---|---|---|---|---|
| Step 5: Stating voting rights | 31 | 25.83 | OP | 5085 | 2.2 |
| Step 6: Stating acquisitions and disposals | 15 | 12.5 | OP | 2996 | 1.3 |
| **M12 Giving information about auditing** | | | | | |
| Step 1: Appointing auditors | 106 | 88.33 | OB | 4192 | 1.82 |
| Step 2: Disclosing information to auditor | 44 | 36.6 | OP | 4155 | 1.8 |
| Step 3: Explaining the audit committee composition | 18 | 15 | OP | 3457 | 1.5 |
| Step 4: Stating accountability | 5 | 4.17 | OC | 256 | 0.11 |
| Step 5: Giving information on non-audit services | 3 | 2.5 | OC | 387 | 0.17 |
| **M15 Signing off the DR** | | | | | |
| Step 1: Adding the wrap up formula | 113 | 94.17 | OB | 693 | 0.3 |
| Step 2: Including signature and position | 117 | 97.5 | OB | 541 | 0.23 |
| Step 3: Stating the date | 107 | 89.17 | OB | 121 | 0.05 |
| Step 4: Stating the registered office | 33 | 27.5 | OP | 324 | 0.14 |

M1 *Opening* is a formulaic, stereotyped move. The key linguistic choice is 'the year ended,' used to avoid any potential ambiguity concerning the period, as shown in Figure 1.

**M1: Opening the DR**

The Directors present their report and financial statements for the year ended 31 December 2004. [014DRwsIH050228BsnsEn]

The Directors present their Annual Report and Accounts to shareholders for the year ended 30 September 2010. [089DRwsVX100906BsnsEn]

**Figure 1.** Examples of M1 Opening

M15 *Signing off* is also largely formulaic, and it functions similarly to the *Polite Ending* move identified by Henry & Roseberry (2001, p. 159) in their rhetorical analysis of application letters where "The writer signs his or her name in a respectful manner, thus claiming ownership of the letter." In DRs, both functions are conveyed by steps 1 to 3 (Table 2). A fourth option states the registered office; according to Bhatia's (2010), this can be considered as a manifestation of legal discourse.

| **M15: Signing off steps** |
|---|
| **Step 1: Adding the wrap up formula** |
| By order of the Board [015DRwsBC041126BsnsEn] |
| Approved by the board of directors and signed on behalf of the board [036DRwsPM050307BsnsEn] |
| **Step 2: Including the signature and position of DR's author** |
| C Gill |
| Secretary [020DRwsET040423BsnsEn] |
| MN Balchin |
| Company Secretary [048DRwsFA040930BsnsEn] |
| **Step 3: Stating the date** |
| 14 March 2011 [111DRwsKL110314BsnsEn] |
| 24 February 2005 [012DRwsAL050224BsnsEn] |
| Step 4: Stating the registered office |

**Figure 2.** Examples M15 Signing off steps

M2 *Reviewing the company activities*, M5 *Accounting for shareholding*, and M12 *Giving information about auditing* share a primary function: report on the company's performance. They make full use of the discourse of economics (Bhatia 2010, p. 39) as do other genres in the business-management textual repertoire (Author et al., 2015).

M2 main rhetorical function is to review the company's present activities (step 1, 90.83% frequency), past (step 2, 57.5%), and future (step 3, 12.5%). The frequencies corroborate the importance given to the company's present situation since the DR, as part of the AR, focuses specifically on the current financial year (Figure 3).

**M2: Reviewing the company activities**

**Step 1: Reviewing the present company activities**

The Group's principal activity is non-life insurance underwriting. [013DRwsBM040309BsnsEn]

[nn] Principal activities. The group owns and operates an estate of independent cash machines and prepaid mobile telephone airtime electronic top-up terminals in the UK and Germany. [016DRwsCA041122BsnsEn]

**Step 2: Reviewing the past company activities**

A review of the Group's business, and developments during the year, is included in the Chairman's statement, the Chief Executive's strategy review and the Operating and financial review. [013DRwsBM040309BsnsEn]

On 6 May 2004, AMEC announced that, following a strategic review, it would exit from the US construction management market. [060DRwsAM050310BsnsEn]

**Step 3: Stating the future company activities**

**Figure 3.** Examples of M2: Reviewing the company activities steps

M3 *Reporting on financial matters* shows 100% frequency and presents aspects of the company's financial performance that have not been certified by the public accounting supervisor.

Interestingly, this is the only DR move that shows a structure of necessary embedded steps with a highly predictable co-occurrence. Step 1: *Presenting results and dividends* is the most important thematically, as shown by the percentage of corpus words (Table 2); Step 2 *Stating charitable and political donations* has to do with the combination of the economics and public relations discourse, used to promote a positive image of the company (Bhatia, 2010, p. 39). Step 3 *Stating supplier payment policy and creditor days* adopts the accounting discourse of another AR genre, i.e., 'Financial Statements,' to present information about profit, loss, taxation, payment of dividend and company's policy on payments.

**M3: Reporting on financial matters**

**Step 1: Presenting results and dividends**

The results for the year are set out in the profit and loss account on page 20.

The directors do not recommend the payment of a dividend and the loss after tax of £3,056,000 (2003: £608,000) has been transferred to reserves. [016DRwsCA041122BsnsEn]

In the year ended 31st December 2003 the group made a loss of £ 758,866 (to 31st December 2002 loss £ 1,561,684). The directors do not recommend the payment of any dividend. The results for the year and progress since year end are described in the Chairman's statement. [025DRwsIA040505BsnsEn]

**Step 2: Listing charitable and political donations**

During the period, the Group made charitable donations amounting to £86,259 (2003: £22,566). No donations were made to any political organisations (2003: nil). [014DRwsIH050228BsnsEn]

Charitable donations in the year amounted to £50,007 in total (2010: £138,991). No political donations were made in either year. [094DRwsXP110221BsnsEn]

**Step 3: Stating supplier payment policy and creditor days**

The Company's payment policy is to ensure settlement of suppliers' invoices in accordance with the stated terms. The average creditors days outstanding was 30 days at the year end (2003 - 30 days). [015DRwsBC041126BsnsEn]

Suppliers are paid in accordance with the individual payment terms agreed with each of them. The number of Group creditor days at 30 April 2011 was 57 days (30 April 2010:88 days) which reflects the terms agreed with individual suppliers. There were no trade creditors in the Company's balance sheet at any time during the past two years. [094DRwsXP110221BsnsEn]

**Figure 4.** Examples of M3: *Reporting financial issues* steps

M4 *Presenting information related to directors* is also present in all reports (100%), and its name defines its function. It is the longest move of the genre, accounting for 20.61% of words in the corpus, and the more complex, as it features nine embedded steps, devoted to informing about the DR authors−or at least the initiators− i.e., the company directors. Table 2 shows that steps 1 to 3 are compulsory and present a similar frequency (between 87.5% and 84.17%), steps 4 to 8 are optional, and step 9 only qualifies as occasional. Optional steps frequencies range between 33% and 15% and depend on the weight companies may want to give to a diversity of themes: data suggest that most companies are unwilling to provide information on their directors' benefits (remuneration 26.67%, indemnities 21,67%, insurances 20,83%) and even

fewer (7.5%) on their attendance at Board and Committee meetings. Steps 1, 2 and 3 constitute the obligatory core of this move.

**Move 4: Presenting information related to directors**

**Step 1**: **Listing directors and biodata**

The following Directors held office during the year to 31 December 201 0:
Sir Michael Jenkins
Peter Hind
Neil Hodgson
Craig Burton (resigned 9 April 2010)
Gideon Tadmor
Bill Guest (appointed 2 June 2010) [079DRwsMT110323BsnsEn]

The names of the directors who served during the year, together with their biographical details and other information are shown on pages 102 to 105. [081DRwsRT110304BsnsEn]

**Step 2: Listing appointments, re-elections and replacements**

Anthony Fabrizi, Lindsay Smith, Patrick Claridge, Leo Knifton and Michael Feltham retire at the forthcoming Annual General Meeting of the Company, and, being eligible, offer themselves for re-election. [015DRwsBC041126BsnsEn]

Mr. Peter A. Hurst retires by rotation at the Annual General Meeting and, being eligible, offers himself for re-election by the members. [017DRwsCS040609BsnsEn]

**Step 3**: **Reporting directors' conflicts of interests**

William R Berkley and W Robert Berkley are Executive Directors of W. R. Berkley Corporation, which has an indirect interest in 20 percent of the issued share capital of the company. [029DRwsKL040510BsnsEn]

Other than as disclosed, no director of the Company had any interest (beneficial or non-beneficial) in any shares of the Company or its subsidiaries at the dates stated and there has been no change in such interests between 1 January and 7 March 2005. [036DRwsPM050307BsnsEn]

**Step 4: Reporting remuneration**
**Step 5: Indicating indemnity**
**Step 6: Indicating insurance**
**Step 7: Explaining powers of directors**
**Step 8: Notifying Board and committee meeting attendance**
**Step 9: Stating the Statement of responsibilities**

**Figure 5.** Examples of M4 Presenting information related to directors' obligatory steps

M5 *Accounting for shareholdings* (94.17% frequency) reports on the company's share capital and related issues, such as voting rights, acquisitions and disposals. It is possibly the most technical move in the DR and expert knowledge, at work throughout the process, was essential to define the steps. The functional-thematic breakdown of M5 structure includes an obligatory Step 1 *Detailing substantial shareholding*, an expected Step 2 *Stating share capital* and four optional steps with a frequency range between

27.50% (step 4) and 12.50% (step 6). Steps 1 and 2 detail or state the company's current situation and are therefore more prevalent than the rest, which deal with events that may or may not have happened in the year such as 'purchase of own shares,' 'issue of shares,' or 'disposals.'

**M5: Accounting for shareholdings**

**Step 1: Detailing substantial shareholding**

Substantial shareholders

In addition to some of those directors listed above, as at 14 November 2004 the company had been notified of the following interests in 3%, or more, of the company's issued share capital, pursuant to section 211 of the Companies Act 1985:
Number of ordinary shares - % of issued ordinary share capital
Artemis Investment Management Limited 1,869,418 3.08% [016DRwsCA041122BsnsEn]

Substantial shareholdings
The Company had been notified of the following additional substantial shareholdings at 23 June 2004:
% held
Schroder Investment Management Ltd 18.98
Deutsche Bank AG London 12.27
Compagnie Financiere Tradition 10.00
INVESCO 9.10
Framlington Investment Management Ltd 8.18
Gartmore Investment Management PLC 5.17
Fidelity International Ltd 3.70 [024DRwsIG040628BsnsEn]

**Step 2: Stating share capital**

As at 25 February 2004, the Company had received notifications from AXA S.A. and Aviva Plc of holdings of the Company's issued share capital amounting to 3.70% and 3.02%, respectively. [030DRwsLG040225BsnsEn]

Share Capital

Under the terms of the Park Group Sharesave Scheme options over 792,061 ordinary shares of 2p each were exercised. No other shares were issued in the year to 31 March 2004. [037DRwsPG040604BsnsEn]

**Step 3: Explaining purchase of own shares**
**Step 4: Describing allotment and issue of shares**
**Step 5: Stating voting rights**
**Step 6: Stating acquisitions and disposals**

**Figure 6.** Examples of M5: Accounting for shareholdings obligatory and expected steps

M12 *Giving information about auditing* appears in 94.17% of DRs and is made up of five concise steps, of which only Step 1: *Appointing auditors* is compulsory and is always realized with a positive or negative statement on the auditors' continuity, which needs to be put to the Annual General Meeting. Optional steps 2 and 3 provide further

information about auditing matters. Occasional steps 4 and 5 yield frequencies lower than 10%, which means that, in this study, they are not considered a constituent part of the genre's structure.

**M12: Giving information about auditing**

**Step 1: Appointing auditors**

PricewaterhouseCoopers LLP are not seeking re-appointment as auditors to the group. The board propose to appoint BDO Stoy Hayward LLP in their place and a resolution approving their appointment will be put to the Annual General Meeting. [015DRwsBC041126BsnsEn]

Ernst & Young LLP have indicated their willingness to continue in office and a resolution for their reappointment and on their remuneration will be proposed at the Annual General Meeting. [077DRwsUB070302BsnsEn]

**Step 2: Disclosing information to auditor**
**Step 3: Explaining the audit committee composition**
**Step 4: Stating accountability**
**Step 5: Giving information on non-audit services**

**Figure 7.** Examples of M12: Giving information about auditing steps

4.2 Expected moves

M9, M10 and M13 account for 27.07% of the words in the corpus. Expected moves focus on a wide range of themes that tend to appear in different formulations, which may indicate that the structure is not typified to the degree that it is in the obligatory moves. Corpus data suggest that some corporate management matters, e.g., employee policies, the annual general meeting (AGM), included in the DR can serve to project a positive image of the company.

Only three steps (see M9s2 and 3 and M13s1) show 'expected' frequencies (Table 3).

**Table 3.** Expected moves: Steps

| Expected moves: steps | Step raw frq | Frq as % | Obligation | Step size | % of corpus words |
|---|---|---|---|---|---|
| **M9 Describing employees' issues** | | | | | |
| Step 1: Cross-referring to employee incentives | 48 | 40 | OP | 4914 | 2.13 |

| | | | | | |
|---|---|---|---|---|---|
| Step 2: Summarizing employee policies | 73 | 60.83 | E | 6175 | 2.68 |
| Step 3: Stating employee involvement | 62 | 51.67 | E | 4518 | 1.96 |
| Step 4: Stating the training and development policy | 23 | 19.17 | OP | 1531 | 0.66 |
| Step 5: Describing health and safety at work | 21 | 17.5 | OP | 2387 | 1.03 |
| Step 6: Describing employee fundraising | 2 | 1.67 | OC | 649 | 0.28 |
| Step 7: Acknowledging employees' awards | 5 | 4.17 | OC | 264 | 0.11 |
| **M10 Presenting issues on corporate management** | | | | | |
| Step 1: Stating the corporate social responsibility | 33 | 27.5 | OP | 4657 | 2.02 |
| Step 2: Commit to corporate governance | 43 | 35.83 | OP | 10046 | 4.36 |
| Step 3: Describing environmental matters | 28 | 23.33 | OP | 4295 | 1.86 |
| Step 4: Stating research and development commitment | 25 | 20.83 | OP | 1682 | 0.73 |
| Step 5: Describing Board matters | 19 | 15.83 | OP | 3850 | 1.67 |
| Step 6: Describing community involvement | 10 | 8.33 | OC | 1272 | 0.55 |
| **M13 Detailing AGM issues** | | | | | |
| Step 1: Providing information about meeting call | 70 | 58.83 | E | 3216 | 1.39 |
| Step 2: Listing ordinary resolutions | 19 | 15.83 | OP | 3881 | 1.68 |
| Step 3: Listing special business resolutions | 39 | 32.5 | OP | 6454 | 2.8 |
| Step 4: Detailing information on electronic communications | 4 | 3.33 | OC | 485 | 0.21 |

*M9 Describing employees' issues* (71.67%) is made up of two expected steps, dealing with employee policies (step 2) and employee involvement (step 3). The

remaining steps deal with different types of bonuses and incentives. Corpus data show low, optional frequencies for steps 4 (19.17%) and 5 (17.5%) and very low frequencies for steps 6 (1.67%) and 7 (4.17%), making them occasional and therefore dispensable in the template of the genre.

**M9: Describing employees' issues**

Step 1: Cross-referring to employee incentives

**Step 2: Summarizing employee policies**

The Group is an equal opportunities employer and the importance of good communications with employees is recognised by the directors. [047DRwsBE241018BsnsEn]

It is the policy of the Group to provide a friendly but professional environment in which all members of staff are given opportunity to contribute to their best ability. The Board's objective is to ensure that the goals and ambitions of individuals and those of the Group are closely aligned. [041DRwsTH040714BsnsEn]

**Step 3: Stating employee involvement**

The Group has a policy of providing employees with information about the Group, its activities and performance, ensuring that the suggestions and views of employees are taken into account. The Directors hold meetings with staff on a regular basis. [018DRwsCH040504BsnsEn]

The Group seeks to achieve a common awareness among the staff of financial and economic factors affecting the business by consultation and by a comprehensive system of employee communication. [024DRwsIG040628BsnsEn]

**Step 4: Stating the training and development policy**
**Step 5: Describing health and safety at work**
**Step 6: Describing employee fundraising**
**Step 7: Acknowledging employees' awards**

**Figure 8.** Examples of M9: Describing employees' issues steps

*M10 Presenting issues on corporate management* (71.67%) is extremely complicated as it may cover a wide range of corporate management issues, including social responsibility, business ethics, governance, environment, research and development, and community involvement — this complexity results in a relatively large number of steps that do not show any predictable pattern. The consequence is an expected move composed entirely of optional steps (Table 3). This unpredictability makes it useless to try to establish the associations between resources and each of the steps. We can, however, say that taking M10 as a unit, common choices reflect step themes, e.g. 'environmental,' 'responsibility,' committed,' 'ethical' or 'sustainable.'

Lastly, the function of M13 *Detailing AGM issues* (70.83%) is to provide detailed information on the Annual General Meeting (AGM). Step 1 is expected and provides information on the meeting call, steps 2 and 3 are optional and devoted to ordinary and extraordinary resolutions, and there is an occasional step 4 that reports on electronic communications. Typical resources in this somewhat predictable step 1 are, obviously, 'AGM,' 'will be held' and 'notice' (Figure 9), some of which are also common to other genres in the business-management textual repertoire (Author, 2016).

**M13: Detailing AGM issues**

**Step 1: providing information about meeting call**

The notice governing the Annual General Meeting to be held on 29 September 2004 at 10.00am at 20-22 Curtain Road, London EC2A 3NF accompanies this report. [041DRwsTH040714BsnsEn]

The AGM of the Company will be held at the Holiday Inn Hotel, 500 Saxon Gate West, Central Milton Keynes, Buckinghamshire MK9 2HQ on Thursday 14 April 2011 at 12 noon. [107DRwsFI110224BsnsEn]

**Step 2: Listing ordinary resolutions**
**Step 3: Listing special business resolutions**
**Step 4: Detailing information on electronic communications**

**Figure 9.** Examples of M13: Detailing AGM issues, step 1.

4.3. Optional moves

Table 4 presents the rhetorical organization of optional moves: M6 *Giving information on tangible property* (15.83% frequency), M7 *Explaining company risks* (38.33%), M8 *Announcing company agreements* (25.83%), M11 *Stating business projections* (48.33%), and M14 *Noting post balance sheet events* (20%). All show low or very low levels of complexity and together account for just 10.24% of corpus words.

**Table 4**. Optional moves: steps

| Optional moves: steps | Step raw frq | Frq as % | Obligation | Step size | % of corpus words |
|---|---|---|---|---|---|
| **M6 Giving information on tangible property** | | | | | |
| **M7 Explaining company risks** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Step 1: Listing financial instruments | 24 | 20 | OP | 1388 | 0.6 |
| Step 2: Explaining general risks and uncertainties | 7 | 5.83 | OC | 3776 | 1.64 |
| Step 3: Stating specific risks | 14 | 11.67 | OP | 5531 | 2.4 |
| **M8 Announcing company agreements** | | | | | |
| Step 1: Disclosing significant company agreements | 26 | 21.6 | OP | 3785 | 1.64 |
| Step 2: Notifying change of control and takeover | 6 | 5 | OC | 1369 | 0.59 |
| **Move M11: Stating business projections** | | | | | |
| Step 1: Evaluating going concern | 57 | 47.5 | OP | 5031 | 2.18 |
| Step 2: Stating cautionary, forward-looking statement | 2 | 1.67 | OC | 153 | 0.07 |
| **M14 Noting post balance sheet events** | | | | | |

It should be noted that all the steps present low frequencies and qualify as *optional* (around 20-21.6%) – except M11s1, *Evaluating going concern* (47.5%), a borderline case and close to the figures for expected moves- or *occasional* (less than 10%). *Going concern* means to state whether it is reasonable to think that the company's financial resources allow it to continue operating in the future. This step is also characterized by the use of exclusive stereotyped formulae such as: 'a reasonable expectation,' 'adequate resources to continue in operational existence,' 'foreseeable future' and 'the going concern (basis)' (Figure 10).

Both the low occurrence and the relative rhetorical simplicity of these elements corroborate that they are not central, but somewhat peripheral, in the rhetorical organization of the genre.

**M6: Giving information on tangible property**

**M7: Explaining company risks**

Step 1: Listing financial instruments

Step 2: Explaining general risks and uncertainties

Step 3: Stating specific risks

**M8: Announcing company agreements**

Step 1: Disclosing significant company agreements

Step 2: Notifying change of control and takeover

**M11: Stating business projection**

**Step 1: Evaluating going concern**

After making enquiries, the directors have a reasonable expectation that the company has adequate resources to continue in operational existence for the foreseeable future as a going concern. [029DRwsKL040510BsnsEn]

Having reviewed the group's plans and available financial facilities, the Board has a reasonable expectation that the group has adequate resources to continue in operational existence for the foreseeable future. For this reason it continues to adopt the going concern basis in preparing the group's accounts. [113DRwsSP110225BsnsEn]

**M14: Noting post balance sheet events**

**Figure 10.** Examples of optional moves and their steps.

The quantitative analysis of the rhetorical structure was key to establish central and peripheral moves and steps of the genre, and to define the typicality thermometer offered to users when drafting a new DR (see Section 7.1, figures 14, 15).

## 5. The glossary database

As noted in the introduction (Section 1), in addition to the rhetorical database, L2 and L1 writers also need support guidelines concerning genre-specific vocabulary and phraseology. DRs include domain or field-specific items (terms) that relate to the objects of the companies, general phraseology, general lexical items (words), and move/step genre-specific items with a particular function within a move/step (Author,

2017). These latter elements give the genre cohesion and are crucial when writing or recognizing DRs as such; consequently, their correct usage is vital. Thus, the restricted glossary databases incorporate information related to move/step specific items, elements that are common to all DRs irrespective of the material object of the companies, places or people involved. In short, they are move or step-bound.

The identification process was carried out using lists of keywords and n-grams. Before the detection and extraction of these items, the different subcorpora of moves and steps were retrieved using the ACTRES browser. WST 5.0 (Scott 2008), AntConc 3.4 (Anthony 2014) and the ACTRES browser helped to obtain raw lists of candidate elements that were analyzed, manually cleaned, grouped and summarized for each move and step. Finally, the resulting data were fed to the restricted glossaries, which were specifically designed to complete and edit the variable gaps of the corpus-based meta-strings (see Section 5).

N-grams are recurrent word combinations calculated in terms of their distributional approach. There is no specific term to designate this concept, and widely used terms for this concept include lexical bundles, clusters or word clusters, multi-word combinations, chains, and n-grams. According to Scott and Tribble (2006, ch. ix), they refer to "any multi-word combination which occurs within frequency thresholds set by the researcher"; therefore, n-grams are identified by statistical methods where the only human intervention, other than choosing the size of the n-grams, is to set the threshold frequency.

Researchers have used different n-gram sizes. For example, Biber et al. (1999) analyzed three to six n-grams in a 40 million word general corpus (Longman Spoken and Written English corpus), Altenberg (1988) established the phraseological interest of three to five grams for a small corpus of the spoken mode, and Ebeling and Ebeling (2013: 68) did not quantify the size of the n-grams although they stated that this "is also much dependent on the size of the corpus." In our research, the maximum n-gram size was fixed at four following previous studies on written academic English (Cortes, 2004; Hyland, 2008; Kashiha & Chang, 2014), and the minimum size was set at two, as this is the default option when calculating n-grams.

Regarding the threshold frequency, there is a direct correlation with the size of the corpus so that the larger the corpus, the higher the threshold frequency to avoid

much noise. Biber et al. (1999) and Ebeling & Ebeling (2013) set different frequencies due to the different characteristics of their corpora: ten times per million for a 40 million word corpus (LSWE) and eight times for 5.3 million running words (English-Norwegian Parallel Corpus ENPC). As GEDIRE© moves and steps are tiny in terms of tokens (ranging from 245 to 11213), threshold frequencies should be much lower than those referred to above. On the other hand, our moves and steps are highly specialized, and we are only interested in the n-grams commonly used within each and not in comprehensive lists, so we decided to use the topmost frequent results for each. So, threshold frequencies for the n-grams were established according to the move/step size: ten when the move/step subcorpus includes more than 6000 tokens, seven if the size is between 2000 and 6000, and four if it is smaller than 2000 tokens. AntConc 3.4 was used to obtain the lists of two to four grams for each move/step.

The identification of keywords, i.e., words whose frequency is unusual in comparison with some norm (Scott, 2008, "Help menu"), in this case, a large reference corpus, is a statistical process that compares the frequencies of two-word lists to detect those words which occur more (or less) often in a particular corpus, such as M-En-GEDIRE, than in a reference corpus. This comparison helps define the specificity of a move/step vocabulary. The British National Corpus (BNC, 90.7 million words, written contents) was our chosen reference corpus because it fulfils Berber-Sardihna's requirement of being "five times the size of the study corpus" (2000, p. 7), threshold proportion that yields "similar amounts of keywords." Johnson & Ensslin (2006) describe two problematic issues when using the written part of the BNC as a reference corpus, namely the British English variety and its compilation date (1993), which leaves out recent topics. None of these concerns affect this research because DRs are written in British English, and we are interested in move-bound keywords and not in content or field keywords. WST 5.0 was used to identify move/step keywords with the default setting. As in n-grams, we used move/step top results, taking into account the move/step size: top 10 keywords for moves/steps below 2000 tokens, top 15 when the size is between 2000 and 6000, and top 25 when above 6000 tokens. Thus, for example, we analyzed the top 25 keywords for M3s3 *Supplier payment policy and creditor days* and the top 20 for M8s1 *Significant Agreements*.

Top keywords and raw lists of n-grams were manually cleaned and

summarized for each move and step. The resulting data served two purposes: first, to obtain concordance lines used to identify recurrent patterns leading to the formulation of "meta-strings" (see Section 6), and second, to feed the restricted glossaries. These are either genre-bound, i.e., entries are used across the whole DR, or move/step specific; examples of the former are the *Comité / Committee,* and the *Quién-1 / Who-1 restricted glossaries* (see Table 5). An example of a move/step specific glossary is the *M8s1 Qué-8 / What-8 restricted glossary* (see Table 6). These restricted glossaries include bilingual records that are used to fill the gaps in the meta-strings. As M-En-GEDIRE is monolingual, it was not possible to apply any method for extracting bilingual information; therefore, researchers identified Spanish equivalents based on parallel texts belonging to AR genres and had them manually reviewed by experts on the field.

**Table 5.** Comité / Committee and Quién-1 / Who-1 restricted glossaries

| **COMITÉ** | **COMMITTEE** |
|---|---|
| Comité de auditoría y riesgo | Audit and Risk Committee |
| Comité de auditoría | Audit Committee |
| Comité de beneficencia | Charities Committee |
| Comité de ética y cumplimiento | Compliance and Ethics Committee |
| Comité de gobierno corporativo | Corporate Governance Committee |
| Comité de operaciones societarias | Corporate Transactions Committee |
| Comité de rsc, Comité de rse | CSR Committee |
| Comité ejecutivo | Executive Committee |
| Comité de inversiones | Investment Committee |
| Comité de nombramientos | Nominations Committee |
| Comité de pensiones y prestaciones por jubilación | Nomination Committee |
| Comité de jubilación | Pensions and Retirement Benefits Committee |
| Comité de prestaciones por jubilación | Pensions and Retirement Benefits Committee |
| Comité de retribuciones | Remuneration Committee |
| Comité de riesgos | Risk Committee |
| Comité de gestión de riesgos | Risk Management Committee |
| Comité de evaluación de riesgos | Risk Review Committee |
| Comité de socio ambiental | Social and Environmental Committee (SEC) |
| Comité comisión permanente | Standing Committee |
| **QUIÉN-1** | **WHO-1** |
| Compañía | Company |
| Corporación | Corporation |
| Firma | Firm |

**Table 6**. Qué-8 / What-8 restricted glossary

| QUÉ-8 | WHAT-8 |
|---|---|
| contrato | commercial contract |
| acuerdo bancario | bank facility agreement |
| Aval | guarantee |
| acuerdo sobre la propiedad | property agreement |
| plan de participación de los empleados en el accionariado | employee share plan |
| contrato de préstamo bancario | bank loan agreement |
| contrato de arrendamiento | property lease arrangement |

This procedure yielded the genre-specific vocabulary of DRs that feeds the glossary database, also a component of the authoring support tool described in Section 7.

## 6. The "meta-string" database

Our third database contains the normalized grammatical and phraseological combinations typically used in the construction of DRs. Computational models based on machine learning vectors and feature extraction have been successfully used, for example, for the automated classification of content components in technical communication (Oevermann & Ziegler 2018, among others). This procedure, however, requires suitably large data collections and detailed characterization of the domain-specific content, as well as human-informed linguistic knowledge, including word order (Oeverman & Ziegler 2018:15), to avoid adverse effects on the automated classification process. For the time being, and given the characteristics of our empirical data, we have taken the human-informed path as an intermediate stage to higher degrees of automation with increased accuracy (see Section 8). To do so, "meta-strings" were identified, starting from keywords and n-grams obtained using WST 5.0 and AntConc 3.4. (see Section 5). The top keywords were considered 'string nodes' for inclusion in the authoring support database. These were logged into the ACTRES browser to obtain full strings in context. This procedural choice rests on the assumption that keyness is not

just quantitative (Scott and Tribble 2006), but also includes the qualitative relevance of the item(s) for the text (Bondi 2011). Our search identified recurrent strings for each move/step, which are conventionally associated with each rhetorical unit.

To illustrate the procedure, we focus again on step 3, move 3, (M3s3), which focuses on *Money matters*. In the structure of a typical DR, M3 is an obligatory move, as its frequency range in the corpus lies between 81% and 100%. First, M3s3 keywords (see Table 7) were logged into the ACTRES browser to obtain concordance lines, then recurrent constructions, including these top items, were identified and regularities classified into invariable chunks and variable gaps. Finally, this information was used to build "meta-strings," which can be recycled as needed, customized employing the restricted glossaries and then become part of the authoring support model (see Section 7).

**Table 7.** Move 3, step3 (M3s3) 25 top keywords

| Or | Keyword | Raw f | Or | Keyword | Raw f | Or | Keyword | Raw f |
|---|---|---|---|---|---|---|---|---|
| 1 | payment | 274 | 10 | creditor | 78 | 19 | average | 45 |
| 2 | Terms | 244 | 11 | agreed | 76 | 20 | december | 45 |
| 3 | Days | 219 | 12 | supplier | 67 | 21 | payments | 43 |
| 4 | suppliers | 203 | 13 | groups | 66 | 22 | companies | 36 |
| 5 | Policy | 193 | 14 | conditions | 63 | 23 | provided | 35 |
| 6 | creditors | 130 | 15 | accordance | 60 | 24 | outstanding | 34 |
| 7 | Group | 127 | 16 | year | 54 | 25 | code | 34 |
| 8 | company | 105 | 17 | purchases | 49 | | | |
| 9 | Trade | 100 | 18 | made | 46 | | | |

For example, if we take the top item, "payment" as a string-node, it appears in the following combinations and frequencies:

1. Terms of payment (64)
2. Payment policy 75 (15)
3. [Creditor] payment policy (36)
4. [Supplier] payment policy (24)
5. Payment terms (28)

Many strings feature one or more of these combinations, often showing

recurrent constructions, as in examples (1) to (5).

(1) File [042DRwsTG041007BsnsEn].xml:... SUPPLIER PAYMENT POLICY The company's payment policy in respect of all suppliers is to settle agreed outstanding accounts in accordance with…

(2) File [079DRwsMT110323BsnsEn].xml:... SUPPLIER PAYMENT POLICY The Group and Company's policy is that payments made to suppliers are made in accordance with those terms and conditions...

(3) File [015DRwsBC041126BsnsEn].xml:... Payment Policy The Company's PAYMENT POLICY is to ensure settlement of suppliers' invoices in accordance with the stated terms.

(4) File [040DRwsSC050427BsnsEn].xml:... Creditor PAYMENT POLICY It is the group's policy that payments to suppliers are made in accordance with the terms and conditions agreed between

(5) File [014DRwsIH050228BsnsEn].xml:... It is the Group's policy to agree appropriate terms and conditions in advance with its suppliers and to make PAYMENT in accordance with those terms and conditions, provided that the supplier has complied with them…

Example (5) contains all the elements present under slightly different formulations in the rest of the strings: agree; terms; conditions; payment(s); suppliers. Based on these (and other) corpus-based findings, the elements common to the identified recurrent strings were collated, and meta-string (3.3a) formulated.

(6) [Meta-string 3.3.a] It is the (QUIÉN 1)'s policy to agree [appropriate] {terms of payment / terms and conditions /payments} [in advance] with its {suppliers / creditors} and to {abide by / make payments in accordance with} those {terms / terms and conditions} [{provided that the supplier has complied with them / subject to satisfactory performance by the supplier}]

The brackets signal the slots where the user has to intervene: The round brackets activate a genre glossary, e. g. (QUIÉN 1, see Table 5); the square brackets signal that the option is available but not required, e. g. [appropriate] and the curly brackets that one option must be necessarily chosen, e. g. {terms of payment / terms

and conditions /payments}. Section 7 will deal with how to access these guided, controlled operations.

## 7. Authoring support

These corpus-based databases were converted into practical authoring support largely thanks to the web-supported tool GEDIRE©, designed to help users draft their own DRs. The three corpus-based linguistic databases – the move structure, 'meta-strings' and genre/move glossary databases (Figure 11) - can be modified or expanded at any time, and the software automatically processes changes without human intervention. It includes additional features addressed to Spanish language writers, which can be replicated for other English-Lx combinations. A hands-on demo version is now available at http://actres.unileon.es/demos/applications.html.[3]

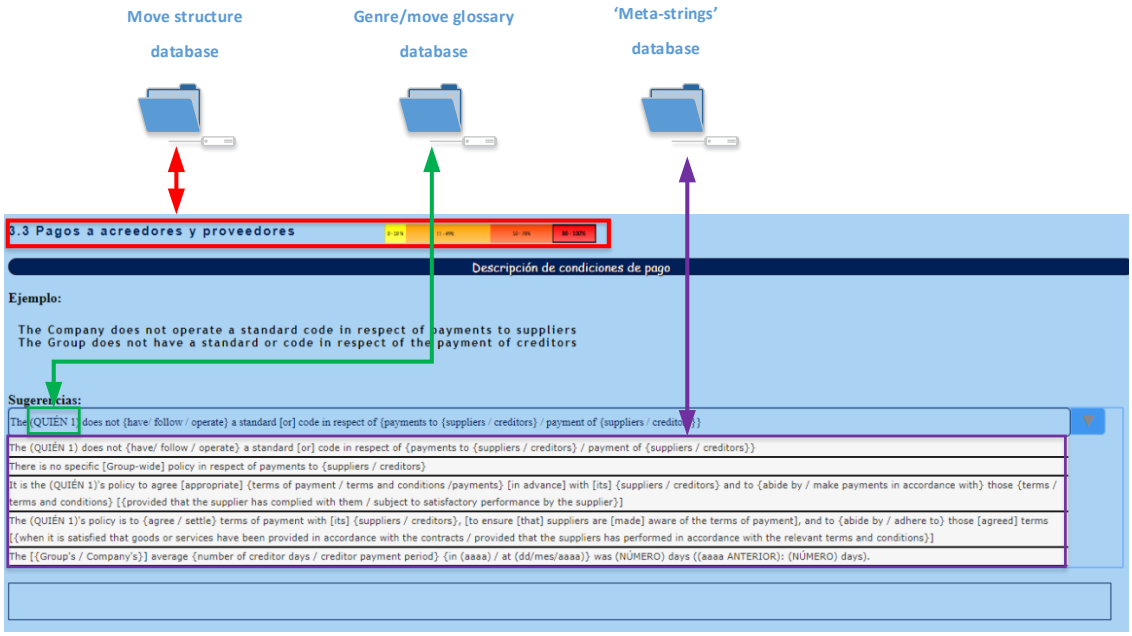


**Figure 11.** Relation between linguistic databases and the GEDIRE© interface

The tool has been designed in terms of usefulness and usability, i.e., it is user-centred to handle the writing of DRs in English effectively, and, it is also user-friendly

[3] The web-supported template (GEDIRE©, 00/2014/1532) copyright belongs to the University of León and is available *via* the Research Transfer Office <otri@unileon.es>.

as it is adapted to the average skills and typical computing performance of non-experts. It is web-based, i.e. hosted on a server and is accessible from any electronic device connected to the Internet. It does not require any installation, and it is multi-platform, which means that it runs in all operating systems such as Windows, Linux or Mac OS and with the most popular browsers: Firefox, Chrome, Safari, Opera and Internet Explorer. Mobile operating systems such as Android or iOS are also compatible through their Web browsers.

### 7.1. Using the tool

In its writing area, each move/step section offers one or several meta-strings to be filled in by the user. Each line presents recurrent structures with variable slots signalled by different types of brackets, as described in Section 6. Figure 12 displays the meta-strings and accompanying examples in M8s1 (move 8, *Undertakings and mergers*; step 1, *Significant agreements*).

```
<model_line>
    <str>
        The Company {is not party to / does not have} any [significant] agreements which [would] take effect,
        alter or terminate upon a change of control [of the Company] following a takeover bid.
    </str>
    <help>
        The Company is not party to any agreements which take effect, alter or terminate upon a change of
        control of the Company following a takeover bid. The Company confirms that there are no significant
        agreements to which it is party that would take effect, alter or terminate upon a change of control
        following a takeover bid.
    </help>
</model_line>
<model_line>
    <option>
        <str>
            [The Company confirms that] There are no [{circumstances connected with / other}] [significant]
            agreements to which the Company is a party {which / that} [would] take effect, alter or terminate
            {upon / in the event of} a change of control [of the Company] following a takeover bid [with the
            exception of {the Relationship Agreement referred to above / those disclosed below}].
        </str>
        <help>
            There are no circumstances connected with significant agreements to which the Company is a party
            that would take effect, alter or terminate upon a change of control following a takeover bid with
            the exception of the Relationship Agreement referred to above. The Company confirms that there are
            no significant agreements to which it is party that would take effect, alter or terminate upon a
            change of control following a takeover bid with the exception of those disclosed below.
        </help>
    </option>

</model_line>
<model_line>
    <str>
        There {are / exist} [a number of] [significant] agreements that take effect, alter or terminate upon a
        change of control of the Company [following a takeover bid], such as (QUÉ 8)^. [With the exception of
        (QUÉ 8)] [None of these are considered to be {essential / significant}] [{to the continuing operation of
        the Group / in terms of their potential impact on the business of the Group }].
    </str>
    <help>
        There are a number of agreements that take effect, alter or terminate upon a change of control of the
        Company following a takeover bid, such as commercial contracts, bank loan agreements, property lease
        arrangements and employee share plans. With the exception of the bank loan agreements noted above, none
        of these are considered to be essential to the continuing operation of the Group. There are a number of
        agreements that take effect after, or terminate upon, a change of control of the Company, such as
        commercial contracts, bank facility agreements, guarantees, property agreements and employee share
        plans. None of these are considered to be significant in terms of their potential impact on the business
        of the Group.
    </help>
</model_line>
<model_line>
    <str>
        The following [significant] {agreements / contracts} [contain provisions that] may terminate upon a
        change of control of the Company: - the (NOMBRE ACUERDO o CONTRATO) [dated (dd/mes/aaaa)] [{in relation
        to / between} (NOMBRE EMPRESA) [and (NOMBRE EMPRESA)] [described on page (NÚMERO)].</str>
    <help>
        The following contracts may terminate upon a change of control of the Company or its relevant
        subsidiary: - the Shareholders Agreement dated 22 June 2006 in relation to Foreland Shipping Limited. -
        the Singapore Submarine Rescue Service Agreement dated 17 October 2008, between James Fisher Defence
        Limited and First Response Marine pte Ltd. The following significant agreements contain provisions that
        may terminate upon a change of control of the Company: - three interest rate swap agreements, described
        more fully on page 88.
    </help>
</model_line>
```

**Figure 12.** Meta-strings and examples in M8s1 (Significant agreements)

Finally, genre/move glossaries provide semi-specialized Spanish-English repositories of lexical and phraseological items as used in DRs following bracket marking in the meta-strings. The entries contain a Spanish input item or combination thereof and one or more possible correspondences in English. Figure 13 offers an entry for genre/move glossary *QUÉ 8*.

```
<word type="que_8">
        <sp>contratos</sp>
        <en>
                <text>commercial contracts </text>
                <ex>-</ex>
        </en>
</word>
```

**Figure 13.** Genre/move glossary entry

Each drafting area (Figure 14) offers the following task-related features: (i) identification of rhetorical move/step and (ii) typicality thermometer to indicate whether this is an obligatory or optional part according to frequency. Next, (iii) a display of examples illustrating the chosen meta-string, (iv) a selector that shows available meta-strings featuring 'recyclable' materials and areas where variation is marked by brackets to be handled by the user. To end with, (v), a drafting area where the user completes his/her selected meta-string assisted by genre/move glossaries as required by his or her own DR (see Section 6).

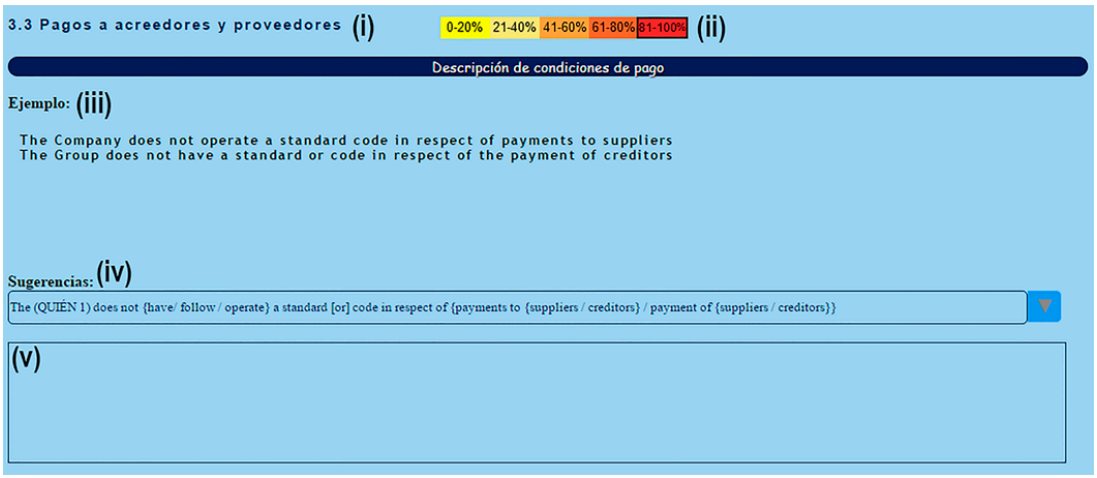


**Figure 14.** M3s3 drafting area on the screen

We use M8s1 to illustrate how to use GEDIRE (Figures 15 to 20). Figure 15 shows the beginning of the sequence of tasks applicable to the drafting of any move/step. Firstly, the user selects a text section (move/step) by clicking on the list on the left or directly scrolling down the main body of the interface. Secondly, as shown in Figure 16, on scrolling down the meta-string section, the corresponding examples pop up. When a particular model line is selected by clicking on it, the model line appears in the drafting area below to be completed and edited by the user, as shown in

Figure 17. In addition to language content to choose from, concrete guidelines on how to use the tool efficiently can be accessed at all times by clicking on the question mark at the right upper corner of the screen.


**Figure 15.** Selecting a text move/ step

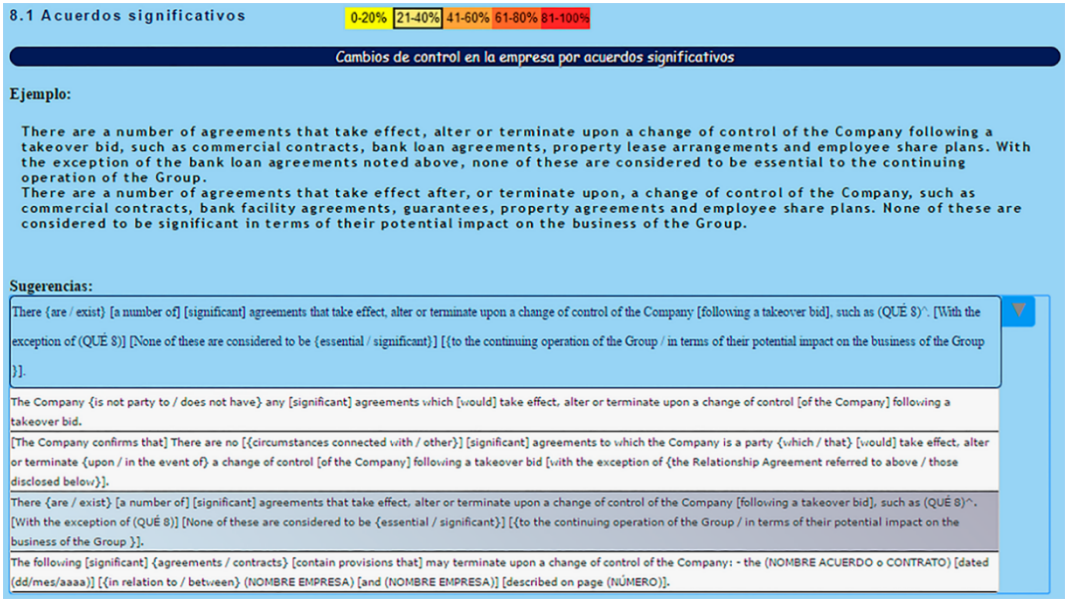


**Figure 16.** Selecting a meta-string

The next step is to edit the meta-string by modifying the content between brackets (round, square and curly). If the user does not know the English for a Spanish word in a round bracket slot, the glossaries described in Section 6 can be employed.

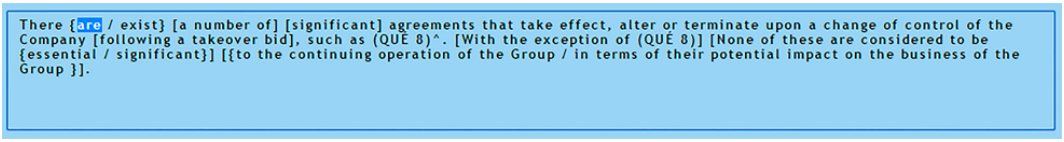


**Figure 17.** Completing and editing selected meta-string

For example, if the user needs the English expression for "*acuerdos bancarios*" (*bank facility agreements*) to fill in the slot signalled by *(QUÉ 8)*, *QUÉ 8* is selected on the string. Then, the glossary is activated, and the first letters of the Spanish expression are logged in, for instance, "*acu*," and all the possible options that match the query, among them "*acuerdos bancarios*," will be displayed as shown in Figure 18. When the user selects it, the English equivalent, *bank facility agreements,* is inserted automatically in the bracketed area. Once all the slots have been edited, the result is an acceptable and correct line, as shown in Figure 19.

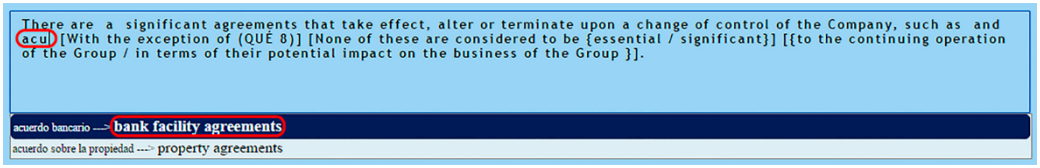


**Figure 18.** Using the genre/move glossaries

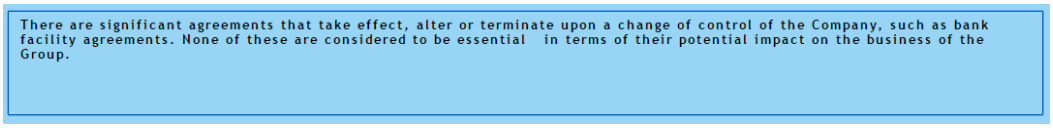


**Figure 19.** Final edited string

This task sequence is repeated for as many moves/steps as necessary until the DR is finished. During the process the user can (1) check the current state of the report using the preview tab (*vista previa*); (2) save an unfinished report at any time using the save button (*guardar*); (3) resume writing later using the import tab (*importar)*; (4) download (*descargar pdf, doc* or *Html*) (Figure 20). A video demo is available at https://actres.unileon.es/demos/generadores/applications.html.


**Figure 20.** Preview of text; save, import and download (pdf; doc and Html) features

Language professionals and economic experts informally tested the tool and reported it useful in two ways: it helps actually to write a DR and reduces the anxiety about the correctness of written communication. Both sets, however, said that they would like to see "less bracketing," as it made the process occasionally confusing (see footnote 4). Extensive work on the automated retrieval of the information in the databases is already being carried out at domain rather than genre-level to improve the tool and produce a robust protocol to be replicated for other genres/ domain/ language pair combinations.

## 8. Conclusions and further work

This study had two aims: first, to establish and describe the internal make-up of one particular genre, the DR, from its move structure to lexical preferences and, second, to show how these results can be used to formulate empirically-based guidelines for authoring support. These guidelines are addressed at Spanish language users, but, eventually, the procedure can be replicated for other language combinations Lx- English.

A custom-made corpus of DRs was collected and annotated for moves and steps to carry out the first aim. The tagset was defined by combining specific functions with themes, rather than relying exclusively on formal linguistic characteristics. The

annotated corpus was analyzed in terms of move frequency to determine the degree of obligatoriness, size (word count) and complexity (hierarchically embedded steps). Our move analysis has demonstrated that DRs have a rhetorical structure distinct from other members of the AR *genre set*. This rhetorical organization is highly typified and conventionalized as an overwhelming proportion of the information is organized and presented in obligatory and expected moves and steps, all of which rate above 50% of occurrence and make the genre structure highly predictable. The order of the rhetorical parts is not an issue, as sequence variation affects 4.16% of the texts and is not accompanied by functional or thematic differences. Corpus-based findings prove that the rhetorical structure of the DR is independent of the size of the DR and the company's business area.

In the process of defining the rhetorical structure of the DR, we also found that the move/step theme tends to be replicated in linguistic choices (see Section 4). While this is true of obligatory moves/steps (Section 4.1), it is not so with the optional ones, except M11s1 *Evaluating going concern*, that does not follow the trend. On the high end of the predictability scores are obligatory M1 and M15, where rhetorical and formulaic uniformity led us to consider them 'a fixture,' also present in other genres in the business textual repertoire (e. g., meeting minutes (Author, 2016)). On the lower end, optional moves/steps (Section 4.3) have not revealed any identifiable, predictable pattern.

Move size, measured in percentage of corpus words, can be taken as an indication of how much each particular rhetorical constituent is needed in genre construction. In our study, however, size does not necessarily correlate to move function or centrality in the rhetorical structure, which is in line with Lewin et al. (2001, p. 87) who did not find "an obligatory hierarchy for the amount of text which should be devoted to each move" and contrary to Parkinson's opinion (2017). Obligatory moves show both low (M1, M12, M15) and high (M3, M4, M5, M10) percentages of words (Section 4.1). A similar situation is found in the analysis of expected moves (Section 4.2). Optional moves, however, generally account for a low percentage of words (Section 4.3).

Move complexity was measured in terms of the number of embedded steps, which ranges from none in M1, M6 and M14 to nine in M4. Obligatory and expected

moves and their steps account for approximately 90% of corpus words, e.g., *M3 Reporting on financial matters* is obligatory and features three steps, which are also obligatory as they show frequencies above 80%.

Our results also indicate that move/ step linguistic (string and lexical) associations, as put forward in sections 4 to 6, are not distinctive and exclusive of DRs but are common to the business- management genre repertoire and, more specifically, to the AR set. They confirm our initial claim that difficulties in generic competence (Bhatia, 2004, p. 145), leading to poor textual performance, lie primarily in the rhetorical organization of the genre and not in the language choices.

Our findings have been used to prepare a dynamic template to assist writers in the composition of DRs and also learners and instructors in training sessions. The tool offers authoring support in the form of (1) a rhetorical grid in Spanish, with moves and steps sequentially and hierarchically organized, (2) the rank of obligatoriness for each rhetorical part, (3) embedded move-and-step 'meta-strings' to guide the drafting in English, together with examples from the corpus. Additionally, and depending on the level of expertise, the user can also activate a backup bilingual dictionary as required. Together or separately, they can function as a conceptual orientation or as a drafting aid.

Working on GEDIRE© has shown that the borders between bi/multilingual writing and translation are becoming increasingly blurred, with one activity or the other participating in one or more stage(s) of the cross-linguistic communication process. In both cases of professional text production, recyclability and customizability are already the norm rather than the exception and supporting apps play a significant role as time and resource savers. While "empty" apps such as templates or TMs depend on the skill of the writer or translator, standardized genre-specific, corpus-based aids could be the answer for companies big or small that do business multilingually. In addition to the obvious financial benefits (corpus-based apps require neither additional technical training nor extra equipment), they are readily received as they mirror real-life texts, and the language is fully recognizable. Furthermore, the guidelines provided in the user's L1 are easy to follow and guarantee the correctness and acceptability of the final text. Following them reduces drastically, and even eliminates, the need for human post-editing. If (s)he is to benefit fully from a

GEDIRE-type app, the user should have relatively high competence in English (B2), while time-intensive corpus-based contrast, as well as continuous updates, need to be provided by the builders in conjunction with extensive testing to identify and answer the needs of different user groups.

Designing GEDIRE© has raised the question of whether the language prototype would qualify as what is known in the industry as a controlled natural language (CNL). While working on establishing the multilevel organization intended to guide the writer in the production of a specimen DR, it became more and more evident that selecting and restricting language choices had much in common with CNLs. A CNL is usually developed to ensure effective communication within professional groups or to ease the cost of translation. Although there are many types of CNL (Kuhn, 2014), all of them share, at least, the following characteristics: (1) standardized terminology for both technical terms and core vocabulary, (2) regularized phraseological and sentence patterns, and (3) re-usable linguistic structures. According to this, GEDIRE© complies with all three.

Building upon a point briefly stated by Adolphson (1998), technical authors often learn to write by plagiarizing, to a great extent, the texts of their co-workers. If this is so, an accessible approach to repetitive writing would be to produce genre-specific, guided, hands-on writing aids to make CNL authoring attractive to both individuals and organizations alike. GEDIRE© provides genre-specific rhetorical information; meta-strings, including phraseology, and genre/move glossaries. It also offers implicit, built-in restriction rules that are handled automatically by the application. What the user can or cannot do has been previously defined, employing corpus-based statistics and forms an integral part of the program (Author & Author, 2016). In this sense, GEDIRE© shares some of the characteristics of SBVR Structured English (Kuhn, 2014, p. 137), but adds new dimensions and factors including restricted language support in Spanish (or the prospective users' L1) at all levels.

Further work on the GEDIRE© prototype includes more extensive language work to upgrade the meta-string module and to expand the genre/move glossaries. The glossaries will benefit from custom-made semantic tagging, which will allow researchers to both increase corpus size and lighten the task of collating glossary candidate items. These additions would result in an increased number of options

available to the user.

To help with this, a writing aid builder (Author et al., 2019) is already under development. Thus, linguists will be able to create a writing aid from scratch or modify data structures without technical development assistance. Lastly, and as suggested by the proofs of concept (PoC),[4] an upgraded version of GEDIRE© (2.0) is under development. It will include the automated retrieval of formal and content components, with improved visual and interactive features that will facilitate text production.

Studies of additional genres in this (and other) domains would constitute an empirically sound basis for genre-based CNLs, which can benefit from this protocol. The long term aim is to help to advance and promote our genre competence (and performance) and improve the quality of text production.

**Acknowledgements**

This research has been funded by the Spanish Ministry of Science and Innovation and ERDF [FFI2013-42994-R and FFI 2016-75672-R].

**References**

Adolphson, E. (1998). Writing instruction and controlled language applications: Panel discussion on standardization. In *Proceedings of the Second International Workshop on Controlled Language Applications* Adolphson, E. (1998). (p. 191). Pittsburgh, PA: Language Technologies Institute, Carnegie Mellon University.

Altenberg, B. (1988). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications*, 101-122. Oxford, UK: OUP.

Anthony, L. (2014). *AntConc* (Windows Version 3.4.3) [Computer software].

[4] A limited, but valuable PoC was carried out in consultation with colleagues at the School of Management. Some part-timers also work as consultants or in a variety of financial services. While they found the instructions to be clear enough, they suggested decluttering the bracketing, as this was not visually appealing.

Author (2017). Pizarro Sánchez, I. (2017). A corpus-based analysis of genre-specific multi-word combinations: Minutes in English and Spanish. In E. Egan & H. Dirdal (Eds.), *Cross-linguistic Correspondences. From lexis to genre* (pp. 221-252). Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.191.09san

Author et al. (2015). Izquierdo, M., Pizarro, I., & Rabadán, R. (2015). Keyness across business genres: Audit Reports and Directors' Reports. Paper presented at the *36 ICAME conference.* 27-31 May 2015. Universität Trier.

Author (2016). Rabadán, R. (2016). Proposals in meeting minutes An English-Spanish corpus-based study. *Languages in Contrast,* 16 (2), 213-238. https://doi.org/10.1075/lic.16.2.03rab

Author & Author (2016). Rabadán, R., V. Colwell & H. Sanjurjo-González. (2016). BiTeXting your food: Helping the gastro industry reach the global market. In A. Moreno Ortiz and C.Pérez-Hernández (Eds.), *CILC2016 - EPiC Series in Language and Linguistics, vol.1* (pp. 361-371). EasyChair. https://doi.org/10.29007/4xtp

Author et al. (2019). Sanjurjo-González, H., Alaiz-Moretón, H., Ramón, N., Labrador, B., and García, I. (2019). Using an Ontology-based Approach to Build Open Assisting Tools in Foreign Language Writing. *Journal of Information Systems Engineering & Management*, 4(2). https://doi.org/10.29333/jisem/5931

Berber-Sardinha, T. (2000). Comparing Corpora with WordSmith Tools: How large must the reference corpus be? In Kilgarriff, A. & T. Berber Sardinha (Eds.), *Proceedings of The Workshop on Comparing Corpora, Held in conjunction with The 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)* (pp. 7-13). New Brunswick, NJ: Association for Computational Linguistics. https://doi.org/10.3115/1117729.1117731

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. London: Continuum.

Bhatia, V. K. (2010). Intercursivity in professional discourse. *Discourse & Communication*, 21(1), 32-50. https://doi.org/10.1177/1750481309351208

Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511814358

Biber, D., Connor, U., & Upton, T. A. (2007). (Eds). *Discourse on the move. Using corpus analysis to describe discourse structure*. Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/scl.28

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited. https://doi.org/10.1162/089120101300346831

Bondi, M. (2011). Perspectives on keywords and keyness: An introduction. In M. Bondi and M. Scott (Eds.), *Keyness in Texts,* 1-18. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/scl.41.01bon

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes,* 23(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

David, C. (2001). Mythmaking in Annual Reports. *Journal of Business and Technical Communication*, 15(2), 195-222. https://doi.org/10.1177/105065190101500203

De Groot, E. B. (2008). *English annual reports in Europe: a study on the identification and reception of genre characteristics in multimodal annual reports originating in the Netherlands and in the United Kingdom*. Utrecht: LOT.

Delahaye, A., Booth, C., Clark, P., Procter, S. & Rowlinson, M. (2009). The genre of corporate history. *Journal of Organizational Change Management*, 22 (1), 27-48. https://doi.org/10.1108/09534810910933898

Ditlevsen, M. G. (2010). The diversity of the annual report in an international context, *Fachsprache* 3-4, 163-178. https://doi.org/10.24989/fs.v32i3-4.1396

Ditlevsen, M. G. (2012). Telling the story of Danisco's annual reports (1935 through 2007-2008) from a communicative perspective, *Journal of Business and Technical Communication*, 26(1), 92-115. https://doi.org/10.1177/1050651911421132

Dragsted, B. (2014). A case study of letters to shareholders in annual reports before, during and after the financial crisis. *LSP Journal*, 5(2), 84-104.

Ebeling, J., & Ebeling, S. O. (2013). *Patterns in Contrast*. Amsterdam / Philadelphia: John Benjamins.

EUR-Lex (2015). http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015L0849. Last accessed 18 July 2019.

Flowerdew, J. & Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes,* 29(1), 78-93. https://doi.org/10.1075/scl.58

Garzone, G. (2004). Annual Company Reports and CEO's Letters: Discoursal Features and Cultural Markedness. In C.N. Candlin & Maurizio Gotti (Eds.), *Intercultural Aspects of Specialized Communication* (pp. 311-341). Bern: Peter Lang.

Garzone, G. (2005). Letters to shareholders and Chairman's statements: textual variability and generic integrity. In P. Gillaerts & M. Gotti (Eds.), *Genre Variation in Business Letters* (pp. 179-204). Bern: Peter Lang.

Gómez González, M. A. & García Varela, A. P. (2014). Discourse-organizational patterns in English and Spanish. Some notes on the thematic management of news reports. *RESLA*, 27(1), 87 – 117. https://doi.org/10.1075/resla.27.1.04gom

Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: Letter of application. *English for Specific Purposes*, 20(2), 153-167. https://doi.org/10.1016/s0889-4906(99)00037-x

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K. (1998). Exploring Corporate Rhetoric: Metadiscourse in the CEO's letter. *The Journal of Business Communication*, 35(2), 224-245. https://doi.org/10.1177/002194369803500203

Jameson, D. A. (2000). Telling the investment story: A narrative analysis of shareholder reports. *International Journal of Business Communication*, 37(1), 7-38. https://doi.org/10.1177/002194360003700101

Johnson, S., & Ensslin, A. (2006). Language in the news: some reflections on keyword analysis using WordSmith Tools and the BNC. *Leeds Working Papers in Linguistics and Phonetics,* 11: 96-109.

Kashiha, H., & Chan, S. H. (2014). Using multi-word units to take a stance in academic lectures. *Journal of Language and Communication* 1(1). 31-40.

Kohl, J. (2008). *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market*. Cary, North Carolina: SAS Publishing.

Kohut, G. F. & Segars, A. H. (1992). The president's letter to stockholders: An examination of corporate communication strategy. *International Journal of Business Communication,* 29(1), 7-21. https://doi.org/10.1177/002194369202900101 1

Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics* 40(1), 121-170. https://doi.org/10.1162/COLI_a_00168

Kwan, B. S. C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics, *English for Specific Purposes*, 25, 30–55. https://doi.org/10.1016/j.esp.2005.06.001

Lewin, B. A., Fine, J. & Young, L. (2001). *Expository discourse: A genre-based approach to social science research texts.* London: Continuum.

Mobasher, A., Ali, A. M., Abdullah, F. S., & Chan, M. Y. (2013). Review of studies on corporate annual reports during 1990-2012. *International Journal of Applied Linguistics & English Literature,* 2(2), 133-141. http://doi.org/10.7575/aiac.ijalel.v.2n.2p.133

Moreno, A. I. & Swales, J. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, 50, 40-63. https://doi.org/10.1016/j.esp.2017.11.006

Nickerson, C. & De Groot, E. B. (2005). Dear shareholder, dear stockholder, dear stakeholder: The business letter genre in the annual general report. In P. Gillaerts & M. Gotti (Eds.), *Genre Variation in Business Letters*. (pp. 325-346). Bern: Peter Lang.

Oevermann, J. & Ziegler, W. (2018). Automated classification of content components in technical communication. *Computational Intelligence*, 34(1), 30-48. https://doi.org/10.1111/coin.12157

Parkinson, J. (2017). The student laboratory report genre: A genre analysis. *English for Specific Purposes*, 45(1), 1-13. https://doi.org/10.1016/j.esp.2016.08.001

Prasad, A. & Mir, R. (2002). Digging deep for meaning: A critical hermeneutic analysis of CEO letters to shareholders in the oil industry. *Journal of Business Communications*, 39(1). 92-116. https://doi.org/10.1177/002194360203900105

Rutherford, B. A. (2005). Genre analysis of corporate annual report narratives. A corpus linguistics-based approach. *International Journal of Business Communication*, 42(4), 349-378. https://doi.org/10.1177/0021943605279244

Scott, M. (2008). *WordSmith Tools* (Version 5). [Computer software]. Liverpool: Lexical Analysis Software.

Scott, M., and Tribble, C. (2006). *Textual Patterns: Keyword and Corpus Analysis in Language Education*. Amsterdam/Philadelphia: John Benjamins.

Siegel, M. (2018). Authoring support for controlled language and machine translation: A report from practice. In G. Rehm, F. Sasaki, D. Stein & A. Witt (Eds.), *Language technologies for a multilingual Europe* (pp. 85-102). Berlin: Language Science Press.

Suter, H.-J. (1993). *The wedding report. A prototypical approach to the study of traditional text types*. Amsterdam: John Benjamins.

Swales, J. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research genres: Exploration and analysis*. Cambridge: Cambridge University Press.

Thomas, J. (1997). Discourse in the marketplace: The making of meaning in annual reports. *Journal of Business Communication*, 34(1), 47-66. https://doi.org/10.1177/002194369703400103

Upton, T. A. & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example, *Discourse Studies*, 11(5), 585-605. https://doi.org/10.1177/1461445609341006

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions, *English for Specific Purposes*, 22(4), 365-385. https://doi.org/10.1016/S0889-4906(02)00026-1

Zanola, A. (2010). The annual report: an interdisciplinary approach to a 'contaminated' new genre. Paper presented at the conference *Genre on the move. Hybridization and discourse change in specialized communication*. Naples, December 9, 2009. https://www.unibs.it/sites/default/files/ricerca/allegati/Paper109.pdf Last accessed 18. July 2019.

## Translated abstract and keywords

Los asistentes de redacción son ayudas (semi)automatizadas que se activan en diferentes etapas del proceso de redacción. Sin embargo, la información lingüística que ofrecen suele limitarse a la ortografía, la revisión gramatical o los bancos terminológicos, pero no se ha prestado la atención necesaria a la organización retórica, las estructuras oracionales típicas o el vocabulario específico asociados a un género concreto. Una laguna adicional es que las ayudas se ofrecen en inglés a usuarios de inglés como primera lengua. Este artículo aborda ambos problemas centrándose en un género en particular: el directors' report y en los redactores de lengua española que escriben en inglés como lengua extranjera. Analizamos un corpus monolingüe compilado y anotado *ad hoc* utilizando las definiciones de género y movimiento retórico de Bhatia (1993, 2004) y Swales (1990, 2004), combinadas con la caracterización temática. El glosario bilingüe recoge terminología especializada y vocabulario específico del cada movimiento retórico o género. Los resultados tienen valor intrínseco, como un análisis de un género textual, pero, sobre todo, como base empírica del asistente de redacción que presentamos y como material de apoyo a la enseñanza y el aprendizaje.

**Palabras clave:** asistente de redacción, lingüística de corpus, análisis de géneros, estructura retórica, *directors' report*.