**ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA**

**DOTTORATO DI RICERCA IN**

# TRADUZIONE, INTERPRETAZIONE E INTERCULTURALITÀ

**Ciclo XXX**

**Settore Concorsuale: 10/L1**
**Settore Scientifico Disciplinare: L-LIN/12**

## PROMOTIONAL ENGLISH ON THE WEB: NATIVE AND ELF UNIVERSITY WEBSITES

**Presentata da: Dott.ssa Erika Dalan**

**Coordinatore Dottorato**

**Prof.ssa Raffaella Baccolini**

**Supervisore**

**Prof.ssa Silvia Bernardini**

**Co-supervisore**

**Prof. Adriano Ferraresi**

**Esame finale anno 2018**

# Table of contents

# Chapter 1

# Introduction

In the context of European academic mobility and internationalisation policy for Higher Education (cf. European University Association (2013)), universities are placing a strong focus on marketing strategies and web-based promotional content produced in English, the *de facto* lingua franca in Europe (Cogo and Jenkins, 2010). Since university websites are a simple and effective way to reach the international audience, new trends are emerging to address these rising promotional needs that include the introduction of highly promotional genres. The question remains as to whether evaluation is also spread over more traditional academic genres, e.g. course descriptions and personal homepages.

Very much in line with the idea that self-promotion is becoming "part and parcel" of self-identity (Fairclough, 1995:140), these new promotional strategies aim to convey each university's distinctive qualities and their unique profile within a common institutional academic framework, and, crucially, a global educational market, thus requiring a significant effort to balance several different needs. On the one hand, the increasing need for universities to depict their most authentic image and create an identity that should be recognised across website contents and pages. On the other hand, EU and EHEA attention to a shared set of values aimed at building a European academic community (which implies student and staff understanding).

Attempts at striking a balance between these needs are further hindered in the case of universities where English is not a native language. English as a Lingua Franca (ELF) initially developed as a "means for communication" rather than a "means for identification" (cf. House (2003)) and is primarily aimed at negotiating meaning and achieving understanding, especially in spoken interaction. This may result in processes of accommodation that lead, for example, to preferring all-purpose verbs to other alternatives (Mollin, 2006:51), thus flattening English language structures. Yet, driven by this new urge of transferring

identities and values, ELF web communication may extend beyond the mere purposes of mutual understanding and start displaying more distinctive and recognisable features.

Against this background, this PhD thesis has two major goals:

1. contributing to current research on promotional English on the web, analysing how native English and English as a Lingua Franca (ELF) are used to express the promotional function of university webpages;

2. collecting language resources in native English and ELF, that could support ELF users in the production of promotional texts for university websites.

This study investigates promotional texts in institutional academic settings focusing on the language of evaluation. In its widest understanding, evaluation indicates any linguistic resources that express a social value, an opinion or a point of view, serving as an inherent constituent part of promotional texts (Hunston, 1989). As used in applied linguistics, it indicates any item of language that can express personal attitudes and feelings, through lexis, grammar or the text itself (Hunston, 2010).

In order to reach the main goals, a series of questions will be addressed:

- what promotional genres and text types can be identified in university websites?

- drawing on the literature, which language features of evaluation are (not) used to express promotional content in institutional academic English?

- how are evaluative features distributed in automatically identified promotional texts?

- from a quantitative and a qualitative point of view, which evaluative features better differentiate between native English and ELF texts?

The above questions are answered through a corpus-based methodology. Although corpus-based studies may fail to grasp the implied or hidden meaning of texts, they tend to better identify recurrent patterns of language as compared to traditional, close reading, especially when a high number of texts are involved, as is the case of this project.

The structure of the thesis is as follows: Chapter 2 reviews existing literature on institutional academic settings (Section 2.1), evaluation language (Section 2.3) and English as a Lingua Franca (Section 2.2). It addresses the controversies related to the concept of genre, text type and register, attempting to provide a working definition for each of these terms (Section 2.4), concluding with a brief account of Web-as-Corpus techniques and the different approaches to text classification (Section 2.5). Chapter 3 describes the procedure

for corpus construction and the internal composition of the corpus constructed specifically for the purposes of this analysis, while Chapter 4 focuses on automatic classification based on Functional Text Dimensions (FTDs). Chapter 5 presents the language features chosen as markers of evaluation to analyse university promotional websites. Features were selected from a wide range of empirical research studies and more theoretical works, in particular Biber (2006a), Hyland (2008) and Martin and White (2005). Chapter 6 presents the study findings: in order to characterise the promotional strategies of academic websites linguistically, features of evaluation are identified, quantified and analysed within the native English and ELF subcorpora. Finally, Chapter 7 summarises the study, presents its limitations and offers suggestions for future work.

# Chapter 2

# Review of the literature

Since the topic of this thesis – i.e. a corpus-based study of promotional language within university websites – stands at the crossroads of various disciplines, the present chapter provides a synthesis of the key concepts and previous studies related to each strand of research that are needed to understand the context and choices underlying this work. The main areas that form the background to this project include institutional language in academic settings, evaluation and its related notions (e.g. stance, appraisal and metadiscourse), English as a Lingua Franca (ELF), the concept of genre (and various related ones such as register, text type, style), and finally the web-based approach to corpus construction. Due to the large amount of work that has been produced in each of the areas involved, I will only focus on the key issues and concepts that are the most relevant to the thesis' final goals, and provide references to further readings for the topics that are only touched upon. This section is organised as follows: In Section 2.1, I will summarise previous studies on institutional academic language. I will then move on to the description of evaluative language (Section 2.3) and discuss relevant issues related to ELF (Section 2.2). In Section 2.4 I will focus on genre, while in Section 2.5, I will present the web-based approaches to corpus construction.

## 2.1 Institutional-academic language

Institutional-academic language may be seen as comprising a large amount of non-disciplinary communication between higher education institutions (mainly universities) and their target audience. It may take various forms, including spoken and written texts, such as service encounters between administrative staff and students – as an example of oral communication – and flyers, prospectuses, course catalogues – as examples of written communication. University informative material comes in electronic or paper form and can be addressed to a variety of people ranging from university local community to an international audience.

Within the digital and online environment, multimodal texts are becoming increasingly popular, including audio and video messages, pictures and animation. These different forms of communication share a common purpose of providing any interested parties with the information they need on how a specific university works and which services, facilities and opportunities it provides. Given this range of situations within the domain of institutional-academic language, the focus of this project is web-based written communication produced for international stakeholders (such as prospective students, institutional and professional partners, parents, alumni and so on), in line with the purposes stated in Chapter 1. Textbooks and classroom activities are produced for teaching purposes and are therefore excluded from this domain. However, academic style and specialist contents may be found in a few institutional-academic genres as well, for instance in the descriptions of courses and research activities, where they do not usually serve an explicit pedagogical purpose.

To provide the full picture of current research and findings in the domain of institutional language in academia, two approaches need to be mentioned here. The first has its roots in sociolinguistic theories that draw on language evidence to understand socio-political (and often undesirable) phenomena. Critical Discourse Analysis (CDA) is the preferred methodology to reach this goal, because it is based on the idea that society and language are inextricably linked and can exert mutual influence over each other. Fairclough (1993), which represents a milestone in the development of CDA theory and methodology, highlights the "value of critical discourse analysis as a method in *social scientific research*, and as a resource for *social struggle*" (Fairclough (1993:1), my emphasis). In Fairclough (2001:26), he restates that CDA "starts from social issues and problems, problems which face people in their social lives". Conversely, the second approach, which is more likely to be adopted by researchers who are interested in language variation and linguistic features, uses socio-political factors to provide explanations for observed language phenomena, offering a more detached overview of the research topic.

### 2.1.1 Intitutional-academic language and CDA

CDA is adopted by the majority of scholars dealing with institutional language and aiming in the first place to explore power relations between communication participants. The researchers' interest lies in exploring how hierarchical roles are constructed in conversation as well as society (Fairclough, 1993; Mayr, 2008; Thornborrow, 2002). As an instance of spontaneous discourse, spoken languag has been at the core of this extensive research work, which includes police interviews (e.g. Thornborrow (2002)), media discourse (e.g. Van Dijk (1981)) prison and courtroom discourses (e.g. Machin and Mayr (2012)). To a smaller extent, and mainly due to their uneven distribution of knowledge, interaction in the educational

and health systems are also presented as examples of asymmetric talk. This is the case, e.g. of classroom discussion (Thornborrow, 2002) and patient-physician communication (Wodak, 1997). As for work centred on written texts, news articles are typical examples of how the press uses its power to justify political choices, e.g. in Machin and Mayr (2012). Analysis of power easily applies to institutions where the relationship between dominant and subordinate groups is legitimised and where power inequalities are considered as social norms. The analysis of "power as dominance" is indeed one of the two main streams of research in the field of institutional discourse, alongside "power as persuasion" (Mayr, 2008). The first stream tends to analyse "institutional practices and discourses solely from the perspective of domination, oppression and exclusion" (Mayr, 2008:6), based on the idea that the strategic use of language enables institutions to contrast people's interests and stresses that the development of values and ideologies contributes to making these constraints look natural (Mayr, 2008:13). The second stream of research claims that power is not only exercised through oppression, but also through persuasion and complicity. This second perspective revolves around the concept of "consent" and implies that at least some institutions are considered to be "universally beneficial and commonsensical" (Mayr, 2008:14) leading to a mutual construction of relations. This second perspective is closely associated with the value of knowledge. Because knowledge is considered to be at the basis of our society, the work of experts is appreciated and respected. As Scott (2001) puts it: "Expertise has become an important feature of disciplining populations and is central to the dynamics of power in modern societies and their institutions" (Scott, 2001:92).

The case of higher education institutions can be viewed as an example of how power is reached through complicity and persuasion, since universities build their identity and reputation upon common shared values, such as personal freedom, knowledge, and cultural diversity. This view is also shared by specialists in the marketing sector that examined the evolution of universities from the old academic model to the modern business model (e.g. Foskett and Lumby (1999); Mazzei (2002)). Against the general concept of "university marketization" that tends to associate higher education with a profit-oriented business exchanging goods and selling products to customers, modern universities aim at building "a relationship which is based on partnership, mutual trust and confidence" (Foskett and Lumby, 1999:37-38). This need for universities to reshape their relationship with prospective students and other institutions arise from socio-political events in the 1970s during the conservative British government headed by Margaret Thatcher. Especially in the UK, a series of reforms were introduced that cut funding to educational institutions and induced universities to rely heavily on external funding sources. The Government's decision was justified by the need to increase occupation among young people through a strong relation with enterprises, which

eventually led to a shift in the goals of educational institutions at any level, i.e. from the delivery of knowledge per se, to a sort of provider of human resources for companies and enterprises. These reforms have influenced higher education institutions in the rest of Europe as well. It is not by chance that CDA applied to the educational environment developed in the 1990s, when researchers showed a growing interest in unfolding hidden practices put in place by universities to raise their reputation, improve their image and thus attract more students and funding.

Fairclough (1993) is one of the first studies adopting CDA to explore the evolution of British universities and their discursive practices through the genre of academic job advertisements. His three-dimensional analysis based on "text", "discourse" and "social practice" led him to note a shift from traditional institutional and professional/academic figures to more entrepreneurial ones. He eventually interpreted his findings within the more general social and market changes affecting the UK at the end of the millenium, albeit stating that the few examples used in his study "can hardly be said to be properly representative of the complex order of discourse of a modern university" (Fairclough, 1993:157). Similarly, Mayr (2008) explores the relationship between academia and business through the analysis of four texts produced by British universities and published on their websites. From the point of view of genre, all texts can be categorised as either promotional or business-related genres, since they include a message from the Vice-Chancellor, an Employability Strategy document, a job description and some written material to train university staff on "negotiation skills". Her contribution offers examples of the "increasingly marketised and managerialist orientation of many British universities" (Mayr, 2008:43). Mayr's results point in this direction since she states that "the University presents itself as a marketable commodity" while students "are turned into marketable products" and constructed as potential customers (Mayr, 2008:33). From a linguistic point of view, she notes that texts contain high amounts of managerialist vocabulary and business-related lexis, which emerges from most studies in this sector.

Other studies have combined CDA with existing methods for the analysis of language use (e.g. corpus linguistics) and rhetorical strategies (e.g. move analysis, along the lines of Swales (1990)). The work of Askehave (2007) grows from socio-political reasons and focuses on international student prospectuses, which should "reflect the values and forces of the free market". Taking for granted the ongoing process of marketization, he conducts two analyses combining traditional genre theories and CDA-inspired methods. The first is aimed at identifying a set of rhetorical moves in four international prospectuses, whereas the second are used to explore language features related to participants' roles in a British university prospectus. He notes that the university very often appears as an agent, performing a set of actions, for instance *offering* or *providing* a service of some kind. The same is true

for clauses containing a subject and a beneficiary, where the former (the university) *enables* or *encourages* the latter (students) to do something. Another interesting point relates to the description of the university itself. The analysis of this UK prospectus shows that more weight is given to university contextual features (such as size, location, facilities) rather than good qualities in the academic domain (Askehave, 2007:733). Finally, students may also be represented as actors, especially when they are asked to choose a type of service, in which case they perform actions like *accepting*, *applying*, *choosing*, *contacting*, *receiving* (see Askehave (2007) for the full list of verbs).

A more recent study inspired by CDA and rhetorical analysis is that of Saichaie (2011), which considers twelve colleges in the United States, seeking to understand in what ways higher education institutions use visual and textual language to promote themselves in their websites. Saichaie focuses on *Homepages*, *About us* pages, and other written web-based documents related to admission, academic programs, financial aid and student life, reaching conclusions for each of these website sections. For instance, homepages seem to prefer short words and active voice, and third person is also used to increase 'brand awareness' (Saichaie, 2011:163). US homepages are also characterised by the presence of news articles, which seem to have acquired a new function of delivering academic merits and students' success, thus reinforcing institutional prestige. The author agrees with many others in claiming that the use of language tactics from private advertising is "indicative of the growing consumerism in higher education" (Saichaie, 2011:139). Finally, the study suggests that the websites under investigation tend to be rather homogenous in content, despite their distinctive institutional features:

> colleges and universities [...] utilize promotional discourse en masse to market rather systematic representations of "higher education" despite the fact that they vary widely a number of institutional characteristics (e.g., type, control, geographic region, selectivity). (Saichaie, 2011:173-174)

Similar results have been found in Morrish and Sauntson (2013), which combines CDA and appraisal analysis (Martin and White, 2005) for exploring mission statements in a sample of ten UK universities and investigating how "evaluative language used in the statement embodies the values of the universities" (Morrish and Sauntson, 2013:61). In terms of promotional language strategies, they note that universities tend to use rather abstract words referring to an undefined uniqueness:

> the self-proclaimed "uniqueness" of the universities was actually undermined by the fact that all of them made the same claims to uniqueness using a very similar and restricted set of linguistic signifiers. (Morrish and Sauntson, 2013:62)

Their study also serves as a source of appraising items that are frequent in UK universities, such as *challenging*, *stimulating*, *supportive*, *committed to* and many more (see full list of items in Morrish and Sauntson (2013:69-74)).

Issues related to higher education institutions and their change towards a new business model have been addressed from a historical perspective too in Wodak and Fairclough (2010), describing how two European countries, namely Romania and Austria, have implemented EU regulations and international processes differently. Within the larger framework of CDA, and adopting a discourse-historical approach, this study discusses once again the question of EU standardisation policies and the Bologna Process. Apart from recognising the existence of homogenisation strategies as others had already done, this contribution embraces the concept of *glocalisation* and highlights the contradictions that emerge when global processes meet local specificities. From a discourse perspective, their main results reveal that very often change is legitimised "in terms of the need to avoid risks and dangers" (Wodak and Fairclough, 2010:34).

Mautner (2005) is one of the first works to combine corpus linguistics approaches to critical discourse analysis with the purpose of exploring socio-political matters – i.e. the marketization process of higher education – through language. In her study, she focuses on a key phrase that clearly embodies the current developments of higher education, i.e. *entrepreneurial university*, and analyses dominant semantic prosodies related to that phrase in university websites and a reference corpus of English, namely the COBUILD corpus. As regards university websites, the phrase under investigation collocates exclusively with positive lexis, such as *strong*, *modern*, *dynamic*, *active* and *competitive*, as well as with lexis transferring the concept of movement, e.g. *dynamic*, *nimble*, *active* (see full list in Mautner (2005:16-17)). In terms of representation and identity, she notes that universities tend to assume a passive role, e.g. by *meeting* the needs of their stakeholders, especially companies, with which they maintain a relation of partnerships (Mautner, 2005:17). On the other hand, the term *entrepreneur* (and related words such as *entrepreneurial* and *enterprising*) are surrounded by a "negative aura" in the COBUILD corpus, which raises issues on the ability of universities to resist encroachment from the private sector:

> It seems hard to understand why universities, with hundreds of years of tradition under their belt and a formidable assemblage of intellect under their roofs, should not be able to pursue a reform agenda independently, without playing to the rules set by economically powerful external constituents, and without deliberately appropriating the language of the commercial sector. (Mautner, 2005:30)

Driven by a stated interest in the rhetorical construction of relationships between universities and students, Gesuato (2011) adopts corpus-based methods to survey the case of

academic course descriptions (ACDs) taken from the web. In this study, no attempt is made to select texts from specific countries, which means that 45 out of 60 ACDs come from US universities. After a detailed description of ACDs' structure, she identifies the main contents of courses and examines all the entities involved (course, content, participant and planned activities) to analyse how they relate to each other. The study attempts to reveal power relations among the main participants, i.e. the institution, lecturers and students, observing for example that teachers are represented as third parties, thus hiding their duty to give tasks and assignments.

More recently and once again from a sociolinguistic perspective, early career researchers have investigated promotional discourse in higher education, inspired by the aforementioned approach. Xiong (2012) surveys discourse and marketization of higher education in China looking at 48 exemplars of Chinese-language advertisements for academic posts. Findings have highlighted occasional instances of promotional strategies, such as the use of *we* as a way to reduce the asymmetric gap between universities and their potential applicants, alongside more traditional types of bureaucratic discourse (Xiong, 2012:331). As in Wodak and Fairclough (2010), a paradoxical picture emerges in which universities attempt to keep their "status quo" while occasionally introducing promotional elements. Mokhtarnia and Ghafar-Samar (2015) also adopts CDA to compare American and Iranian universities' About pages, confirming that socio-cultural factors influence the composition of this web genre. American universities tend to include larger amounts of information about themselves, their goals, history and so on, since they produce a higher number of pages and subpages under each *About us* section. According to Mokhtarnia and Ghafar-Samar (2015), this may be an indicator of a more transparent and information-sharing culture, as well as a direct consequence of the larger numbers of international students in the US as compared to Iran. Finally, another recent study is Hoang and Rojas-Lizana (2015), which analyses promotional discourse in two Australian university websites, the first being an institution of long tradition and the second a younger one. The analysis reveals that both websites exhibit promotional elements, for instance in building a personal and peer-to-peer relationship with their audience, although Melbourne University puts more emphasis on its international role as compared to the more recent Macquarie University (Hoang and Rojas-Lizana, 2015:17).

### 2.1.2 Institutional-academic language in Corpus Linguistics and Translation Studies

The second approach to investigating university language relevant for this thesis takes a diametrically opposite view as compared to CDA-driven research. Instead of exploring

socio-political issues through the analysis of specific aspects of language, researchers seek to identify language patterns and determine whether there is any correlation between language variation and external or contextual variables, from genre to social factors, geography, history, institutional practices and the like. This process is particularly relevant in fields where language use is at the core of research, in a time where information and communication are driving forces for economy and society. Corpus linguistics – already mentioned above in combination with CDA – is generally considered a suitable methodology for revealing patterns and trends that would be otherwise missed.

A major project in this area is the one by Biber et al. (2004), conducted on the T2K-SWAL corpus and considering a whole range of academic genres and registers including a small selection of institutional texts (e.g. syllabi). Biber et al's work does not directly address the issue of self-promotion, since the first goal of this project is pedagogical in nature, i.e. it aims to help students understand (institutional) academic tasks and develop the language skills required to fulfil them (Biber et al., 2004:2). In this study of language use at university, the authors perform a multi-dimensional analysis of texts based on Biber (1988), identifying five main dimensions, where a dimension represents the co-occurrence distributions of 67 linguistic features (Biber et al., 2004:46). Dimension 4 is relevant to this thesis since it is interpreted as "overt expression of persuasion" (Biber et al., 2004:72), particularly explicit in two genres, namely classroom management and office hours (spoken genres). Institutional written material, instead, tends towards the non-overt extreme, meaning that it contains a lower degree of promotional features, apparently against what many CDA studies have extensively found.

Yet, these results are not surprising, since the T2K-SWAL corpus is unlikely to include the texts that, according to the aforementioned research, are likely to contain those features (e.g. *About us* pages, mission statements, job advertisements). As already mentioned, the T2K-SWAL corpus was mainly compiled for pedagogical purposes and includes a large amount of "internal" communication addressed to current students (e.g. classroom management and office hours). This explains a) why persuasive language seems to be exclusive to spoken and disciplinary registers and b) why persuasive contents are aimed at engaging students in their daily learning activities, rather than attracting prospective students Biber et al. (2004:72). In a later study, Biber (2006b) interprets linguistic features as indicators of language functions in academic settings for research purposes across a set of genres (textbooks, syllabi, classroom teaching) and registers (spoken versus written). In this second project too, he does not directly address the issue of promotional functions, albeit questioning, as a minor comment, the accessibility of university language by students (Biber, 2006b:199). Stance features, generally considered as a resource for promotional language functions, "are frequent and

pervasive in the university context" (Biber, 2006b:199), although they are interpreted as an attempt of involving students in teaching and giving them some advice of any kind, for the same reasons discussed above with reference to Biber et al. (2004). On the one hand, these results are somewhat marginal to the questions posed in this thesis since Biber's corpus addresses the US context exclusively and does not account for a significant part of written web-based material. On the other hand, Biber's findings show a general tendency of universities in the US to "incorporate a range of communicative purposes beyond the conveyance of informational content" including "the expression of personal attitudes and evaluations" (Biber, 2006b:222) that apparently encompasses the teaching environment as well.

Influenced by the same underlying motives inspiring CDA research on this subject, but taking an essentially corpus-based approach, Sanigar (2013) focuses on the keyword *research* in the UK academic settings. By exploring its collocational patterns in a corpus of UK university websites, she notes that collocates such as *quality*, *higher quality*, *world* and *internationally* are extensively used, implying that research needs to be packed with positive adjectives and nouns to fulfil institutional promotional needs.

Caiazzo (2011) makes a central contribution to the analysis of university language variation in different institutional settings adopting corpus-based techniques. Her study compares the use of the personal pronoun *we* in British and Indian university webpages, focusing on the genre of *About us* pages. She observes that British websites make a larger use of highly evaluative language compared to the Indian ones and suggests that *we* plays a different function in each country. In British websites, it can be conceived of as an "enticing we", thus creating a corporate identity. On the other hand, Indian websites tend to use it as a "seeping we", since it plays a marginal role in the representation of the university. This suggests that the Indian corporate identity is still "work in progress" (Caiazzo, 2011:257). In line with Wodak and Fairclough (2010), she also notes a general tendency to associate global trends with different local perspectives (Caiazzo, 2011:255).

Costales (2012) is one of the few research projects in the field of Translation Studies investigating language differences in university websites. He conducted a corpus-based analysis on a sample of 215 EU university websites, addressing issues such as the translation and/or localization of website contents, as well as the implementation of European guidelines. The main results show that universities in the EU tend to offer homogeneous contents in English, rather than multilingual localised versions and that most universities assign translation tasks to amateur translators.

Finally, two studies are especially relevant to this thesis, i.e. Bernardini et al. (2010) and Ferraresi and Bernardini (2013), both based on acWaC-EU, a very large corpus of university

websites (around 90 million words) in native English and ELF, including 27 European countries. The acWaC-EU project fills a gap within research on institutional language in academia in two ways: i) it provides researchers in this field with primary data for analysing university language, ii) it allows an extensive analysis of the current European situation, since the corpus includes both Anglophone and ELF countries.

Bernardini et al. (2010) compares native English web-based texts to their lingua franca variety in Italy. Results indicate that English texts are more formulaic, as they make larger use of short phrases or lexical bundles (i.e. typical sequences of words). Futhermore, a few UK/Irish lexical bundles also express positive evaluation, e.g. "one of the most", "one of the largest", which imply reference to the now very familiar competitive market. Conversely, Italian online documents seem to contain next to no promotional elements of this kind. Both in Bernardini et al. (2010) and Ferraresi and Bernardini (2013), another significant difference is found in the use of modal and semi-modal verbs, which are known to be one of the strongest stance markers. Native texts, compared to ELF ones, seem to favour indirect modals to express obligation, such as *should* and *will*, whereas ELF texts tend to convey obligation directly, using *must* and *have to* (Ferraresi and Bernardini, 2013).

### 2.1.3 Summing-up

Research on institutional-academic English has mainly adopted CDA techniques or has combined CDA with other methods or approaches (e.g. genre analysis, appraisal theory, historical perspectives, corpus linguistics) seeking to investigate the recent development of higher education institutions in their race for money, students and status. The points shared by these studies share some main points that can be summarised as follows:

- use of loaded positive vocabulary, as well as other terms that originally developed for the business sector;
- use of personalised forms, such as the personal pronoun *we*, aimed to reduce asymmetrical power relations and create a peer-to-peer relationship between university and students;
- tendency to embed different local manifestations into global processes.

These findings apply not only to strictly promotional genres such as mission statements, *About us* pages and job advertisements, but also to more academic types of texts such as student prospectuses and course descriptions. On the other hand, very few studies have investigated the use of institutional language in universities from a linguistic perspective.

As Caiazzo (2011) states as a concluding remark after her investigation on the use of *we*, it would be interesting to focus on other verbal environments (such as the collocational analysis of the word *university*) and other personal pronouns, also with reference to the cultural dimension (e.g. Hofstede (2001)). Her final thoughts seem to suggest that the lexico-grammatical features related to the representation of the university have not yet been extensively investigated, which is the challenge that this project aims to address. Besides the need for a large-scale descriptive study of the different ways in which universities and/or cultures promote their image, this second linguistic approach is likely to accomplish practical purposes as well, in that it may help anyone involved in university communication (administrative staff, lecturers, translators, institutional representatives, among others) to keep abreast of new trends and enhance the quality of information sharing, which represents a crucial ability in our society.

The case of ELF speakers is central to this issue since ELF users may need support, at least in the first stages of learning, to adopt communicative strategies efficiently. Surprisingly, most of the existing work focuses exclusively on countries where English is a native language, despite the growing number of international users of English, especially in Europe. This is arguably due to the fact that native English is very often regarded as a standard of reference, not just for its native status, but because British and US universities have a longer tradition in attracting economic resources, as is also noted by Askehave:

> My reason for considering this prospectus is the fact that UK universities are among the leading players in the higher education export industry (with international students bringing in about £4 billion a year to British universities and some £10 billion to the economy as a whole), and as such play a significant role in setting the scene and tone for the discursive practices associated with the industry. (Askehave, 2007:730)

Although examples from native English contexts can definitely be used as a resource for inspiration and learning, it is imperative for universities to transfer a genuine image – if this is going to be the new standard of institutional websites – once learning has progressed to the point where original ideas get their start.

The next section therefore focuses specifically on the issue of ELF that has been introduced very shortly in these last paragraphs and provides a discussion of current research on how ELF is evolving throughout Europe. Given the scarcity of work on promotion in institutional academic ELF, insights on settings other than the institutional academic one, and strategies other than promotional ones, will also be surveyed.

## 2.2 English as a lingua franca

Despite the spread of numerous *lingua francas* over the centuries (cf. House (2003); Meierkord and Knapp (2002)), accompanied by – at least – an intuitive understanding of the meaning and functions of these contact languages, the case of English as a Lingua Franca (ELF) is filled with uncertainties. Although lingua francas can be regarded as a natural and recurrent phenomenon resulting from the need of overcoming language barriers, the impressive global development of ELF and the variety of contexts in which it is used makes it rather different from similar examples of *lingua francas*, as Dröschel (2011) notes with reference to pidgins:

> While pidgins are used as lingua francas for restricted purposes only (Zima 1977, Muhlhausler 1986) and are more radically simplified varieties of the source language with regard to phonology, grammar and vocabulary, Lingua Franca English affects a large variety of domains, such as international communication and trade, technological information, tourism, banking, politics, international organizations etc. and covers a range of more or less simplified varieties, depending on the sociolinguistic background of its speakers. (Dröschel, 2011:41)

Similarly, Garzone (2007) describes the successful development of English as "unique in its unprecedented rapidity and proportions" and states that "the widespread use of English in international communication is not limited to business and politics, but extends to all other sectors of contemporary life" (Garzone, 2007:10).

Partially due to these diverse and often hybrid backgrounds, researchers are left with a number of unanswered questions. One of the major concerns surrounding the debate on ELF is related to whether or not the term *lingua franca* includes native speakers in its definition. A common misconception is the overlapping of two very different communication elements, i.e. the primary function of contact languages (which serve as a *lingua franca*) and its speakers' status (native vs. non-native speakers). Dröschel (2011) notes that "there seems to be some general consensus that lingua franca communication typically is non-native/non-native communication" (Dröschel, 2011:39), and argues that this view does not correspond to reality since there are a number of situations in which lingua franca communication comprises native speakers as well, especially in some African countries. This second interpretation has finally been accepted by those researchers who initially were more inclined to include non-native speakers only. For instance, Jenkins (2004) has eventually agreed that "ELF interactions can be said to include NSs, although only as a small minority within the totality of the world's ELF users" (Jenkins, 2004:2). Dröschel (2011) provides her own perspective on this issue, which seems to mediate between different views. She suggests that English as

a Lingua Franca (ELF) might be differentiated from Lingua Franca English (LFE), where the latter is a "an additionally acquired form of English even for native speakers" (Dröschel, 2011:40) and goes on claiming that Lingua Franca English "can be understood as a contact language used by native and non-native speakers alike but which functions as an independent system which as such has no real native speakers" (Dröschel, 2011:41).

On the one hand, suggesting a (new) term for each specific context may help diminish vagueness; on the other hand, it may also exacerbate the "terminological quandary" framing the ongoing discussion on ELF. *English as a foreign-language*, *English as a second-language*, *English as an International language*, *English as a Lingua Franca*, *Lingua Franca English*, *Global English*, *New Englishes* (in addition to other terms which do not specifically refer to English, for instance *Interlanguage*) partially or completely refer to the use of English in Kachru "Expanding Circle" (Kachru, 1985). As the term itself suggests, this third "expanding" circle identifies geographical areas where English has not fully developed yet as a standalone variety. For this reason, it is likely that many communicative situations will escape any terminological categorization, especially if one considers that none of these terms is associated with an institutionalised form of English. Dröschel (2011) claims that institutionalised, non-native varieties of English have to be distinguished from localised forms of LFE that have spread in the Expanding Circle (Dröschel, 2011:47). Considering that it may take a long time before these localised varieties can be uniquely identified, suggesting an umbrella term for any possible contexts within the Expanding Circle might be more straightforward. Further subcategories based on geographical, historical or socio-linguistic specificities may naturally arise afterwards.

In this thesis, the term ELF covers the European area exclusively, and more specifically, European countries where English is not an official language (cf. Section 3). Although some native English webpages may be addressed to an international community and specifically written for them – which would theoretically include native speakers in lingua franca communication – native speakers of English will most likely generate native-like patterns of English, as far as language production is concerned. Even if native speakers are asked to produce specific language patterns addressing their international audience, how do we know whether these patterns are closer to a lingua franca variety than their native (and more natural) counterpart? The rationale behind this thesis is that any type of content which is written by native speakers of English in a native English context, and directed to either international or local recipients cannot be treated as ELF, although it may differ from native-to-native communication. This is particularly true for written, distant, monologic communication, where there is next to no interaction or sentence co-construction, as compared to conferences and business meetings.

Other major and still critical issues involve the construction of global and local identities among groups of people who use English as a Lingua Franca – closely interwoven with the development of a shared culture and set of values – as well as the question as to whether ELF should be regarded as a transactional language variety only. A further point of interest is the identification of typical linguistic aspects of ELF communication, e.g. the list of Lingua Franca Core (Jenkins, 2000), which might indicate that a new variety of English is emerging, besides representing a practical (embryonic) alternative to native English grammars.

In this regard, the VOICE (Vienna-Oxford International Corpus of English) and ELFA (English as a Lingua Franca in Academic Settings) projects have taken significant steps towards creating resources for identifying ELF patterns. VOICE is a corpus of spoken and naturally-occurring ELF interactions, mainly compiled for linguistic research and aimed at describing the use of (especially European) ELF. The corpus is made of one million words and is classified into text types – interviews, press conferences, service encounters, seminar discussions and many others. The ELFA project is a one-million-word collection of spoken ELF in academic settings, which among other purposes seeks to find any ELF features that deviate from native English varieties and provide a reference model for teaching international English.[1] "Academic" here refers to disciplinary content such as lectures, presentations, seminars, thesis defences and conferences.

More recently, a written subcorpus named WrELFA was created as part of this project. Designed as the written equivalent of the oral ELFA corpus and serving as a practical resource for academic writing, this corpus is made of disciplinary (mainly research-oriented) text types – research papers, PhD examiner reports and research blogs – and does not include any institutional texts, such as information to potential students and other external parties, as does the corpus built for this thesis (see Chapter 3). Both projects and their relevant publications contribute to investigating many of the research issues that have been briefly profiled above and that include – but are not limited to – ELF definition, terminology, culture and identity and lexico-grammatical features.

To avoid the risk of over-simplifying the ELF debate and reducing it to a number of controversial topics, these issues need to be examined from different perspectives, each of them matching a specific category of ELF users. Four groups of users can be identified that are involved in the development of English as a Lingua Franca, namely:

1. teachers and learners
2. professionals in the business sector
3. institutional representatives

[1] http://www.helsinki.fi/englanti/elfa/project.html

These are covered in the next subsections (2.2.1 – 2.2.3).

### 2.2.1 ELF in language teaching

One of the most widely discussed paradoxes of ELF is the need for a model for ELF teaching and a concurrent lack of ELF standards. The latter is associated with a very low production of ELF-oriented materials (Jenkins, 2012:493). Models are a priority for both teaching and learning, since instructors need a reference for teaching and students arguably need a standard to imitate, especially during their first years of learning. Students are indeed very often satisfied and motivated by their ability to follow, reproduce or creatively apply the yardstick provided in order to evaluate their progress. Therefore, a question arises as to which model is appropriate for teaching. As noted by Penny (2010) there are many reasons why any native-speaker model is now obsolete. Some of these reasons include the fact that native speakers of English are a minority as compared to the number of ELF speakers in the world, and that native English language has its own varieties as well, which would raise similar questions regarding which native variety to choose (Penny, 2010:86). However, most teachers are allegedly "reluctant to disassociate notions of correctness from nativeness" (Jenkins, 2007:141). Although some may relate this to 'a physiological leap of faith' (Jenkins, 2007:141) towards ELF, it is widely recognised that ELF is currently too indefinite to offer a standard for teaching. As Mollin (2006) notes, "the perspective that ELF is no stable variety as such would make it a bad teaching standard, as indeed most protagonists agree" (Mollin, 2006:52). Either way, it might not be the time yet for ELF to be fully implemented in schools and universities as a "monolithic" teaching model; however, raising awareness of how ELF can vary across countries is highly recommended in order to provide students with a realistic portrayal of modern English and let them speak freely and fluently. Yet, focusing on variability issues is easier to achieve with intermediate and advanced students than elementary ones. In practical terms, Penny (2010) notes that:

> If the teacher has to teach such variants every time a new item is introduced, then his or her students will never have time to acquire a minimal useful syllabus of the language they need. (Penny, 2010:89)

To sum up, heterogeneity seems to be an intrinsic value and at the same time a major obstacle for ELF teaching. For this reason, new and more flexible types of reference models might be envisaged that could integrate traditional, rule-driven reference books, especially at higher levels of instruction. "Discovery learning" with corpora (Bernardini, 2016) has already been suggested as an aid for advanced, university-level learners. Besides providing

evidence on how language is used in real life, ELF corpora may supplement conventional norm-providing materials. As an alternative, Penny (2010) suggests that the model for ELF teachers should be the *fully competent ELF user*, with no need for such a speaker to be originally a native speaker in the traditional sense (Penny, 2010:87). Many ELF users have indeed reached such a high mastery of (whichever variety or form of) English that they can express themsleves clearly and fluently in any situations and, because of that, can be considered themselves as "flexible models" for ELF teaching. Even if these ELF users have acquired a native-like form of English – which is very likely considering that, even today, there are (next to) no ELF schools or ELF course books – we need to keep in mind that ELF has developed thanks to those same speakers who were probably aiming at a native-like standard and proficiency level. If we accept that English is naturally developing into a new language variety because of its use as a lingua franca in real-life contexts, any "academic" model will eventually be overcome in favour of more adaptable and diversified forms of English. As Jenkins (2012) has observed referring to her 1980s teaching classrooms, students tended to use language freely in natural(istic) conversation and not to apply the rules they were taught, without hindering their mutual understanding (Jenkins, 2012:487). It is also worth remembering that these alternative ways of combining the same language features into different patterns do not necessarily lead to simplified forms of English. Although there are approaches to English which recommend and promote simplification,[2] research findings have already demonstrated that "ELF is as rich as any other English, including that of NESs" (Jenkins, 2012:491).

### 2.2.2 ELF in business settings

The question of simplification applies especially to ELF in business settings, where mutual intelligibility is a priority over language proficiency. Whereas instructors in the field of foreign languages place emphasis on concepts like correctness and language quality, the central concerns for the business domain are "efficiency, relevance and economy" (Dröschel, 2011:64).

This is probably one of the reasons why Globish and similar approaches to English that promote fast and easy learning are gaining so much popularity, although language simplification does not necessarily imply language fluency. Any experienced user of English will certainly be able to reduce its vocabulary and language complexity to a minimum level and, still, speak fluently. Conversely, beginners can strive to produce a clear and complete

[2]Global English, or Globish, is based on the idea that anyone can learn English and speak fluently with just 1,500 words, especially in business settings. Source: http://www.globish.com/ [last consulted on 15 December 2017].

sentence, no matter the number of words they have learnt. Since learning is firstly a personal experience, there should be no need to force users to adopt a specific and limited set of words since that does not necessarily improve proficiency in terms of language command and intelligibility, besides representing an obstacle to the natural development of ELF.

The general assumption is that in order to transfer plain information, basic skills in language understanding and production are needed. In other words, the informational function of a language does not imply any strategic choice and can be fulfilled by limited complexity from a lexical and structural point of view. As trade languages show, this is not totally wrong: "when communities begin to trade with each other, they communicate by adopting a simplified language" (Crystal, 2012:11). Since the term 'lingua franca' dates back to medieval times when Latin and Greek acted as *linguae francae* in Mediterranean countries, it feels natural to associate *lingua francas* in business settings with simplified languages, although this seems to reflect the past rather than current trends. As already mentioned above, today ELF is widely thought to be a complex and heterogeneous phenomenon:

> Although past research has made a point about lingua francas being simplified, reduced languages, it seems most probable that the heterogeneity attested to the term 'lingua franca' and the types of language forms and situations of language use it covers will be reflected in the characteristics that can be observed in the linguistic form as well in the structure of interactions that take place in lingua franca. (Dröschel, 2011:42)

One of the reasons why ELF has developed into a complex and heterogeneous system may be related to a change in the way of doing business. As compared to early forms of trading – usually compared to 'transactional' forms of business – modern business is more of a relational activity and is very focused on communication strategies. Whereas transactional marketing results in a one-time transaction, relational commerce aims at creating long-term loyalty for future selling (Mazzei, 2002). Hence, the strategic importance of communication and the construction of shared values through language. As Meierkord and Knapp (2002) point out, "this common language, originally serving as a tool for mere communication, now obtains a further function: that of a tool for construction and expressing a common identity" (Meierkord and Knapp, 2002:13). To succeed in modern business, it becomes essential to communicate with clients efficiently. This implies using language strategically in order to produce a specific action, either buying or, most importantly, creating a brand *image*, thus enabling companies to be uniquely identified among thousands of similar others in the world.

The increasing use of English is largely documented, especially in promotional genres (Nickerson, 2007), where positive evaluative features are frequently used as part of a strategy

for representing brand identity and its status worldwide (e.g. Poncini (2007)). Because of this shift in the use of English in various social domains, the term *lingua franca* seems to have lost its meaning as a mere vehicle for communication for specific purposes (Meierkord and Knapp, 2002:15), and gradually extended to a wide range of social and professional domains. No other *lingua franca* had ever fulfilled such a high and diversified range of purposes and functions, that extends beyond traditional business genres (for a definition of genre see Section 2.4.1) and technical documentation.

Some authors have highlighted the importance of focusing on specific genres in the business context for which a "shadow structure" exists (Nickerson, 2007). Familiarising with typical structures and phraseology may indeed help in producing spoken and written texts according to conventional norms. This easily applies to genres such as emails, reports, business letters, complaints, presentations, which benefit from various real-life examples of specific language strategies. However, when it comes to delivering content that is not specifically tied to any genre category, things become far more complex.

As an example, promotional content can be delivered either through explicitly persuasive genres (such as *why choose us* pages), or across many other text types that are not associated with any conventional forms. While the former tend to follow standard criteria and patterns that can serve as a model for reproducing similar content, the latter are more flexible in their nature, thus requiring a greater ability to use language freely and creatively. Evaluation, which is "a key tactic of persuasion" (Hunston, 1989:2) is realised through a series of pervasive, contextualised features of language, including emphatics, attitude and relational markers, hedges and even hidden expression of stance (for an overview of evaluative language see Section 2.3). The "sometimes false assumption that everyone speaks English" (Nickerson, 2007:359) may result in misinterpretations of what ELF proficiency actually means. On the one hand, the erroneous idea that proficiency is measured against a native model (for the reasons outlined in section 2.2.1) has widely spread and needs to be set aside. On the other hand, if economy of language and simplification are given priority over competence, many professionals may be led to think that acquiring a simplified or reduced version of a language is enough to express anything they need or like. Although learning how to use ELF core grammar and lexicon effectively does provide simple solutions to daily business tasks (e.g. writing emails, holding a presentation and so on), this will not necessarily imply a full understanding and command of English in different contexts. Consequently, the production of oral and written texts in ELF can range from rich and beautifully designed texts to highly simplified or even badly written and potentially unsuccessful documents.

The highly diversified picture of ELF in the business context is attested by research studies in this field, for instance Nickerson (2007). After comparing the use of English as a

Lingua Franca against multilingualism, she noted that "not only can the (indiscriminate) use of English lead to practical difficulties, it may also create situations in which some employees are more equal than others", adding that "there is the potential for a catastrophic failure of communication" (Nickerson, 2007:356). Whether or not ELF can be thought as an alternative to multilingual communication, there is an evident and urgent need to inform professionals that the advantages of using ELF are necessarily limited by poor or inadequate language practice.

Similar issues also apply to written texts within higher education institutions that have started promoting their image in ELF through websites and social media. In the wake of internationalisation of higher education and promotion of academic mobility (cf. Askehave (2007)), higher education institutions are becoming more business-like and are applying business strategies to promote their values and cultural background internationally. However, language tends to receive relatively little attention, especially in Italy, where drafting in ELF and/or translation are often accomplished as secondary tasks by members of administrative staff:

> It is not common for a university to have a dedicated translation service. For communication in English a number of universities make use of staff employed in their own language centres. [...] Other language centres (perhaps the majority) provide translation services on a more loosely organised basis, with various forms of collaboration with the marketing and web communication teams of the university. (Palumbo, 2013:99)

As Palumbo (2013) notes, despite the increasing volume and range of texts that have to be made available in ELF, "no unified or authoritative resource has been created to provide support for translation or foreign-language drafting" (Palumbo, 2013:101). In most cases, translation services are provided through collaboration with university members of staff, showing that ELF language matters tend to be underestimated, even though the creation of shared values through communication is at the core of any institutional authority. The relation between ELF and the development of an institutional identity is central to the European Higher Education Area (EHEA) and will be discussed in the next section.

### 2.2.3 ELF in EU institutional settings

Institutional language policies and practices constitute a critical factor for the development of ELF in Europe, especially as concerns the spread of common European values, identities and beliefs through language(s). A European flag, anthem, TV-channel, as well as the organisation

of European cultural events (such as the Euro-festival) and the development of institutional procedures and practices are just some instruments that the EU has put in place to generate a European identity. The promotion of forms of "high culture" are generally seen as "means for clarifying and reinforcing identity at both the national and the European levels" (European Commission, 2012:16). As compared to other similar institutions, the EU has embraced a peculiar vision of *union*, since it is founded on and carries together "two antagonistic perspectives such as Unity and Diversity, which constitute its motto" (Bellier, 2002:85). *Diversity* is a core value of European identity and is rooted in institutional processes as well.

In terms of language, the EU has always favoured multilingualism, allocating substantial resources to services of translation and interpretation that guarantee the right of EU members to communicate in their own mother tongue. Yet, and somewhat paradoxically, a parallel process has been observed over the years. Although the European Union has always been an advocate of language and culture diversity, and has promoted multilingualism for international communication, an underlying trend is currently emerging in which English as a Lingua Franca is extensively being used across Europe when addressing the European community as a whole. The use of English in the EU has expanded particularly at an unofficial level, to the point that "it has become indispensable for communication outside meeting rooms and for networking purposes: the famous corridor talks predominantly take place in this lingua franca" (Reithofer, 2010:146). On the written side, "it has become the *de facto drafting language* for most texts elaborated in the EU institutions" (Reithofer, 2010:146).

This had a domino effect on many other EU-related associations and organisations, such as the European Higher Education Area (EHEA), currently composed of 48 countries, mostly in Europe. During the first Ministerial Conference in 1999, the *Bologna Process* was launched, marking the foundation of the EHEA. This common space was created as a collective effort of European public authorities and universities (including the European Commission) to make higher education more compatible and promote common values, such as "freedom of expression, autonomy for institutions, independent students unions, academic freedom, free movement of students and staff."[3] Increasing mobility and employability are still major goals of the EHEA, as well as university modernisation and worldwide promotion: "the Bologna reforms help to make European universities and colleges more competitive and attractive to the rest of the world".[4] In this context of internationalisation, English has acquired a "unique position" among Higher Education Institutions (HEIs), raising issues on how to reduce tension between multilingualism and English, as claimed in a recent document produced by the European Language Council:

[3] www.ehea.info

[4] http://ec.europa.eu/education/policy/higher-education/bologna-process_en

> In a European context, individual and societal multilingualism has traditionally been promoted; at the same time an increasing number of higher education programmes is now being offered through English, which has created an unfortunate dichotomy between multilingualism and English Medium Instruction (EMI). However, the time seems to have come when it would make a lot of sense to reconsider the concept of multilingualism and strike a new balance between individual multilingualism on the one hand and English, including EMI, on the other. (European Language Council, 2013)

In a context where language and cultural heterogeneity has gained such a fundamental role, the *the facto* special status of English (Seidlhofer, 2010:10) is reasonably perceived as an obstacle to diversification and as a possible threat to multilingualism: "While it may be a very well-considered decision to use English as the lingua franca or lingua academica, more often than not, it seems to be to the detriment of multilingualism and linguistic diversity" (European Language Council, 2013:1-2). In fact, most researchers agree that this is a false assumption and multidisciplinary EU projects that have been carried out to tackle this issue confirm it. Referring to the use of English globally, Seidlhofer (2010) stated that "World English tends to establish itself alongside local languages rather than replacing them, and so contributes to multilingualism rather than jeopardize it" (Seidlhofer, 2010:9). More recently, the LINEE project (Languages in a Network of European Excellence),[5] conducted between 2006 and 2015, claimed that English is used as a "vibrant and complex lingua franca" that has emerged "as a neutral common language with only a marginal national connotation" (European Commission, 2012:19) concluding that "English is not necessarily a threat to linguistic diversity" (European Commission, 2012:19). LIMEE contributors have referred to ELF as "an authentic and legitimate English"' and made some policy recommendations, such as encouraging the recognition and promotion of the authenticity of English as a lingua franca (European Commission, 2012:20).

Despite these encouraging trends and results, the Brexit vote in the UK has put into question the role of English within the EU. With the UK exit from the EU, discussion is now taking place as to whether English should maintain its special status and role. According to Danuta Hubner, a senior Member of European Parliament, it will not: "We have a regulation [...] where every EU country has the right to notify one official language," Hübner said; "The Irish have notified Gaelic, and the Maltese have notified Maltese, so you have only the U.K. notifying English".[6] The reasons why Ireland and Malta have chosen a language different from English, as well as the political consequences arising from that choice, go

[5] www.linee.info

[6] http://www.politico.eu/article/english-will-not-be-an-official-eu-language-after-brexit-senior-mep/

beyond the scope of this thesis. However, this political issue indicates that the use and legitimisation of English is still widely linked to and dependent on Anglophone countries. The use of English *as a lingua franca* within European institutions is in fact loosely related to British English (cf. Section 2.2.1), especially from a cultural point of view – including culture-laden words and phrases.

### 2.2.4 Summing-up

To conclude this review of ELF, the ongoing debate on the need to dissociate ELF teaching from native standards (Section 2.2.1), supported by the studies that have compared native English and ELF (Section 2.1), as well as the general low concern of ELF business communication for native models (Section 2.2.2) may lead to think that ELF is indeed taking its own, albeit rather indefinite, way. Linguistically, it is crucial to study what ELF looks like as compared to British English and whether or not there are language indicators showing the evolution of ELF. This might develop either as a standalone variety in Europe or as a mosaic of varieties matching country-specific peculiarities. Results from a PhD thesis on these issues, and subsequent publications (Mollin, 2006), showed that "there were hardly any common features that united lingua franca speakers, even in a context such as the EU, where speakers use English frequently with each other and do have the opportunity to negotiate a common standard" (Mollin, 2006:48).

Moving now to Section 2.3 below, an outline of evaluative language will be presented. In particular, the next Section explores four approaches to the analysis of evaluation emerging from the literature.

## 2.3 Evaluative language

*Promotion*, *persuasion* and *evaluation* are inter-related terms. English dictionaries describe *promotion* as a synonym of *publicity* and define it as the activity of increasing sales or public awareness of a product or organization.[7] More generally, *promotion* consists in bringing something to a higher position (even in terms of social value). Persuasion is used in linguistic settings to designate the functions and strategies to promote something, a meaning that is also shared by the more encompassing term *evaluation*. Persuasive communication aims to convince the audience to perform an action and coincides with one of the primary functions of language, namely the "conative" function in Jakobson's model (Jakobson 1960). *Evaluation*, instead, is used in applied linguistics and related fields to indicate any item of language that

[7] https://en.oxforddictionaries.com/definition/promotion [last consulted on 15 December 2017].

can express personal attitudes and feelings, through lexis, grammar or the text itself (Hunston, 2010). Therefore, evaluation can be thought of as a set of language strategies serving as a means to create promotional content and fulfil the persuasive function of language.

However, evaluative features can fulfil many other functions besides promotion. For instance, directives are used in institutional academic settings for promotional reasons (e.g. *Discover our courses*; *Visit our website*; *Get involved*), since they create a personal relation with recipients. Yet, directives accomplish many other purposes as well, prototipically giving orders. Similarly, positive vocabulary (e.g. *dynamic*, *active*, *stimulating*), epistemic structures (e.g. *I think that*, *It is clear that*) and modals (e.g. *may* and *might*) can express evaluation, although they do not primarily fulfil promotional functions. Rather intuitively, evaluative language shows judgment over a person or a thing. In its widest understanding, it indicates any linguistic resources that express a social value, an opinion or a point of view, serving as an inherent constituent part of promotional texts:

> Another reason for the importance of evaluation to linguists is that its use is a crucial component of other discourse features and functions, [...] as point of view, attitude, persuasion, the monitoring of interaction. (Hunston, 1989:2)

For ease of presentation, I will distinguish lexical from grammatical features used to express evaluation in written texts, and refer to four approaches to the analysis of evaluation language emerging from the literature. The four approaches, summarized in Hunston (2010) are:

**Stance**: although this term is adopted by various authors with different meanings, I follow Biber's view, which is extremely relevant to this PhD project since it uses a quantitative approach to identify the grammatical marking of stance.

**Appraisal theory**: the term refers to the study by Martin and White (2005) in discourse semantics. They developed a complex theoretical system, which focuses on the meaning of words. Evaluation or 'appraisal' is categorised into a tree-structured system composed of semantic (sub-)categories.

**Metadiscourse**: the term, which in this thesis refers in particular to Hyland and Tse (2004) and Hyland (2008), identifies language resources (including both lexical and grammatical features) connected to interaction between the author of a text and its recipients.

**Evaluation**: this term is used in a rather broader sense by Hunston (2010). Besides providing her own definition of the concept of evaluation, her approach is central to

this project as it shows how corpus-based analyses favour the identification of recurrent aspects of evaluation.

The next four sections will briefly review these approaches and describe how they can apply to this project as concerns the identification of promotional strategies in university websites.

### 2.3.1 Stance

Within corpus linguistics, the term 'stance' is principally associated with Conrad and Biber (2000) and Biber (2006a). Conrad and Biber (2000) adopt traditional quantitative methods of corpus research to analyse the distribution of stance adverbials across spoken and written academic registers. The authors prioritise a grammatical approach to stance, including epistemic markers (dealing with the assessment and modulation of a proposition), attitudinal markers (broadly expressing the speaker's opinion) and style markers (dealing with the way in which information is delivered).

Biber (2006a) further develops the original classification by taking account of a broader range of language patterns, including modal verbs and a combination of lexical and grammatical features (e.g. stance nouns followed by that-clauses). The full list of stance markers is presented below (adapted from Biber (2006a)):

1. Modal and semi-modal verbs
    - Possibility/permission/ability: *can*, *could*, *may*, *might*
    - Necessity/obligation: *must*, *should*, *have to*, *got to*, *ought to*
    - Prediction/volition: *will*, *would*, *shall*, *be going to*
2. Stance adverbs
    - Epistemic certainty: *actually*, *certainly*, *in fact*
    - Epistemic likelihood: *apparently*, *perhaps*, *possibly*
    - Attitude: *amazingly*, *importantly*, *surprisingly*
    - Style: *according to*, *generally*, *typically*
3. Stance complement clauses controlled by verbs
    (a) Stance verb + that-clause
        - Epistemic certainty verbs: *conclude*, *determine*, *know*

- Epistemic likelihood verbs: *believe*, *doubt*, *think*
- Attitude verbs: *expect*, *hope*, *worry*
- Speech act and other communication verbs: *argue*, *claim*, *report*, *say*

(b) Stance verb + to-clause

- Likelihood verbs: *appear*, *happen*, *seem*, *tend*
- Mental verbs: *believe*, *consider*
- Desire/decision verbs: *intend*, *need*, *want*
- Effort/facilitation: *attempt*, *help*, *try*
- Speech/communication verbs: *advise*, *remind*, *request*

4. Stance complement clauses controlled by adjectives

(a) Stance adjective + that-clause

- Epistemic certainty: *certain*, *clear*, *obvious*
- Epistemic likelihood: *un/likely*, *possible*, *probable*
- Attitude adjectives: *amazed*, *shocked*, *surprised*
- Evaluation adjectives: *essential*, *interesting*, *noteworthy*

(b) Stance adjective + to-clause

- Epistemic certainty: *certain*, *like*, *sure*
- Attitude adjectives: *happy*, *pleased*, *surprised*
- Evaluation adjectives: *able*, *eager*, *willing*
- Ease or difficult adjectives: *difficult*, *easy*, *hard*

5. Stance complement clauses controlled by nouns

(a) Stance nouns + that-clause

- Epistemic certainty: *conclusion*, *fact*, *observation*
- Epistemic likelihood: *assumption*, *claim*, *hypothesis*
- Attitude nouns: *hope*, *view*
- Communication nouns: *comment*, *proposal*, *report*

(b) Stance nouns + to clause: *failure*, *obligation*, *tendency*

Biber's study has taken a focussed approach to the analysis of stance, isolating lexico-grammatical markers in academic registers and providing examples of this phenomenon. Even though the analysis is limited to explicit language patterns, the study shows that the expression of stance in academic settings is pervasive, especially in the spoken registers (Biber, 2006a:115).

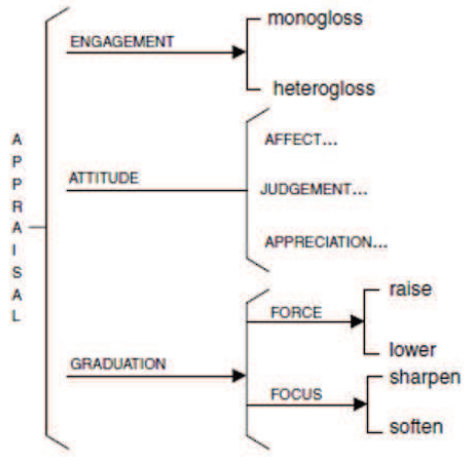


Fig. 2.1 An overview of appraisal, taken from Martin and White (2005:38)

### 2.3.2 Appraisal theory

This theory refers to the field of discourse semantics, as it is concerned with the meaning of words 'beyond the clause', mainly adopting a lexical perspective to evaluation language. Although evaluation can be realised through lexical and grammatical categories, meaning provides a way to get to the subjective level of language (either explicit or implicit), overcoming any type of 'formal' realisation. Martin and White (2005:34) describe appraisal as "one of three major discourse semantic resources construing interpersonal meaning" (Martin and White, 2005:34). As indicated in the diagram (Figure 2.1), appraisal consists of three interrelated elements, namely 'engagement', 'attitude' and 'graduation'. 'Attitude' stands at the core of this system because it expresses emotions or emotional reactions, whereas 'engagement' refers to the interaction between participants and 'graduation' deals with resources that can either intensify or weaken feeling.

Each of these dimensions are further developed as a discourse semantic system. The subclassification of appraisal is outlined in Figure 2.1, taken from Martin and White (2005:38). Attitude covers three emotive dimensions, namely 'affect', 'judgment' and 'appreciation'. Affect includes resources expressing positive and negative feelings (e.g. *un/happy*, *(in)secure*, *(di)satisfied*). Judgement involves ethics and the evaluation of people behaviour, which can be positive or negative. In general terms, judgement can be divided into meanings registering either social esteem (e.g. *powerful*, *educated*, *fortunate*) or social sanction (e.g. *moral*, *humble*, *arrogant*, *dishonest*). Lastly, appreciation has to do with aesthetics and the way natural phenomena are evaluated (e.g. *engaging*, *exciting*, *irregular*, *boring*). While affect includes resources that describe human feelings, judgment and appreciation can be viewed

as institutionalised feelings, since they depend on social rules and values (Martin and White, 2005:45).

Engagement is the second dimension of appraisal (see Figure 2.1) and can be defined as 'intersubjective stance' (Martin and White, 2005:97). It refers to words and phrases indicating viewpoints, in particular how speakers or writers position themselves linguistically towards other voices. The taxonomy of engagement includes 'disclaim', 'proclaim', 'entertain' and 'attribute'. Within the 'disclaim' category, speakers disagree with something or someone (e.g. *not* and *although*). Meanings associated with the 'proclaim' class are used to present a statement as valid and generally reliable (e.g. *naturally*, *obviously*, *as X has shown*). 'Entertain' deals with individual subjectivity and possible positions (e.g. *in my view*, *it seems*, *probably*), while 'attribute' refers to the expression of stance by an external voice that can be either endorsed or rejected (e.g. *X believes*, *X claims that*).

Graduation includes meanings which in the literature have been labelled as 'hedges', 'downtoners', 'boosters' and 'intensifiers' (see Section 2.3.3 on metadiscourse), such as *somewhat*, *slightly*, *rather*. These lexical forms enable speakers and writers to modulate their degree of alignment with their statement (Martin and White, 2005:94). Graduation can be divided into 'force' and 'focus', where the former operates on graduation (e.g. *slightly* and *extremely*), whereas the latter operates on the degree of prototypicality (e.g. *true*, *effectively*).

Summing up, the system of appraisal acts on the level of meaning and refers to a) attitudes towards ourselves, other people and the external world, b) engagement between participants and their voices and c) the graduation of attitudes and other phenomena. This complex taxonomy is developed in a tree structure summarised below:

1. Attitude
    (a) Affect
        - Happiness (for example *like*, *dislike*)
        - Satisfaction (for example *impressed*, *involved*)
        - Security (for example *assert*, *declare*)

    (b) Appreciation
        - Balance (for example *consistent*, *contradictory*)
        - Complexity (for example *clear*, *detailed*)
        - Impact (for example *engaging*, *boring*)
        - Quality (for example *appealing*, *beautiful*)
        - valuation (for example *authentic*, *effective*)

    (c) Judgement

- Esteem
  - Capacity (for example *expert*, *clever*)
  - Normality (for example *celebrated*, *eccentric*)
  - Tenacity (for example *cautious*, *constant*)
- Sanction
  - Propriety (for example *abiding*, *cruel*)
  - Veracity (for example *blunt*, *credible*)

2. Engagement
   (a) Attribute (for example *according to*, *in her view*)
   (b) Disclaim (for example *although*)
   (c) Entertain (for example *I suspect that*, *I believe that*)
   (d) Proclaim (for example *admittedly*, *naturally*)
3. Graduation
   (a) Force (for example *very*, *slightly*)
   (b) Focus (for example *kind of*, *real*)

This framework is based on close reading of individual texts and has been adopted in, e.g., Fuoli (2012), to analyse BP's and IKEA's 2009 social reports, with the ultimate goal of exploring how these two companies construct their identities. Appraisal theory has been applied to the academic context as well, for instance in Morrish and Sauntson (2013). The authors examined a sample of ten UK universities from two different groups (the research-driven Russell Group and the business-engaged University Alliance) and investigated how evaluation resources offered by the English language are employed to represent university values (Morrish and Sauntson, 2013:61). Besides noting that all universities use the same abstract words to transfer the idea of 'uniqueness', the authors observe that the Russell Group appraises the university itself, while in the Alliance Group local industries are also the object of positive appraisal, in accordance with the specific values of each group.

### 2.3.3 Metadiscourse

Ken Hyland has contributed significantly to the field of metadiscourse by developing an analytical framework for describing internal and external relations between the text and its contributors. He defines metadiscourse as "the linguistic resources used to organize a

discourse or the writer's stance towards either its content or the reader" (Hyland 2000:109). Therefore, this term gives emphasis to the interactive aspect of evaluation, which is established between the author and his or her audience and between the author and the content of the text itself. Hyland and Tse (2004) distinguish between "interactive" and "interpersonal" metadiscourse resources. The former "set out an argument to explicitly establish the writer's preferred interpretations" (Hyland and Tse, 2004:168) and have to do with the way in which discourse is organised (e.g. *however* and *therefore*), while the latter 'involve readers in the argument by alerting them to the author's perspective towards both propositional information and readers themselves' (Hyland and Tse, 2004:168), including resources like hedges and boosters.

More specifically, interactive resources comprise five subcategories:

1. transitions, mainly composed of conjunctions and aimed at signalling a change in discourse (e.g. *in addition*, *but*, *thus* and *and*);
2. frame markers, which indicate a change in the structural sequence of texts (e.g. *finally*, *to conclude*, *my purpose here is to*);
3. endophoric markers, which refer to other points in the text (e.g. *noted above*, *see Fig*, *in section*);
4. evidentials, used to refer to an external source of information (e.g. *according to X*, *X states*);
5. code glosses, which provide an explanation of ideational statements (e.g. *namely*, *such as*, *in other words*).

Interpersonal resources are also divided into subcategories aimed at involving the reader in different ways:

- hedges, used to moderate the commitment to a proposition (e.g. *might*, *perhaps*, *possible*, *about*);
- boosters, used to give emphasis to a proposition (e.g. *in fact*, *definitely*, *it is clear that*);
- attitude markers, which assess a proposition (e.g. *unfortunately*, *I agree*, *surprisingly*);
- engagement markers, which actively engage the reader through imperatives, question marks and the like (e.g. *consider*, *note that*, *you can see that*);

- self-mention, which enables the author the explicitly mention herself or himself (*I*, *we*, *my*, *our*).

This framework has been used, e.g., in Hyland (2008) to explore interactive persuasion in research writing, showing how textual practices connected to metadiscourse can be employed to produce agreement on the way real data are interpreted.

### 2.3.4 Evaluation

The term 'evaluation' tends to acquire a more comprehensive meaning compared with the notions reviewed so far. As compared to 'stance' and 'appraisal', which can be easily associated with either a lexical or grammatical level of analysis, 'evaluation' is perceived more like an umbrella term, subsuming all the various approaches. As Hunston and Thompson (2000) point out, evaluation "is the broad cover term for the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about" (Hunston and Thompson, 2000:5). From a functional point of view, evaluative language is an essential aspect of communication since it enables users to:

- express an opinion, relying on subjectivity, a set of shared values and ideological assumptions;
- establish and maintain relations between conversation participants. This can be used to manipulate or persuade readers about something;
- organize the discourse. The connection between writers and readers is construed through discourse markers as well, for instance by signalling specific 'points' in discourse.

There is consensus that evaluation at a linguistic level can be expressed through lexis, grammar and text, even if it is frequently construed on the social and implicit level of discourse. Hunston claims that, despite the implicit nature of evaluation, corpus linguistics is a suitable method to investigate evaluative language and can do more than just "adding quantitative detail" (Hunston, 2010:4). In particular, she suggests that "phraseology as broadly defined plays a number of roles in the study of evaluative language" (Hunston, 2010:4).

Indeed, Hunston (2010) adopts qualitative and quantitative perspectives to analyse evaluation. Concordance lines are used to show that some English words and phrases frequently co-occur with modal-like expressions or modal phrases (e.g. *it is necessary to*). From a quantitative perspective, she compares the frequencies of specific patterns and suggests that

wh-words (such as *whether*) frequently co-occur with verbs of modality (e.g. *I don't know whether*). As Hunston argues, a corpus-based approach to the analysis of modal-like expressions is interesting because "in looking at verbs frequently modified in this way we see the importance of phraseology to that evaluative meaning" (Hunston, 2010:85). In other words, the analysis of collocates and concordance lines enables researchers to consider evaluation as an 'integral' aspect of clauses.

### 2.3.5 Summing-up

The above sections have provided a review of the various approaches to the analysis of evaluation (lexical and grammatical approaches, implicit or explicit evaluation) and provided an inventory of lexico-grammatical features that are listed in the relevant literature as an important device for expressing evaluative meaning.

These features represent a basis for identifying evaluation strategies in academic webpages, with the final goal to compare how European universities construe promotional web-based genres. Lexico-grammatical features can be translated into queries (see Section 5) and used to analyse the variety and quality of evaluation over academic webpages.

## 2.4 Genre in theory and practice: an overview

### 2.4.1 The concept of genre

Since an important part of this thesis includes manual and automatic classification of university web-based genres – with the ultimate goal of examining ELF and native English promotional features – a short investigation into the "jungle" of genres (Lee, 2001) is necessary. Therefore, this section reviews existing definitions and interpretations of the concept of genre (including sub- and macro-genre) and related ones such as *register*, *text type*, and *style* and suggests that, despite some terminological confusion, there is in fact a more extensive common ground than one might expect. This common ground will be used as a basis for describing how university websites are structured, and for classifying their texts accordingly.

Theories on genre have been developed within many different disciplines, four of which are immediately relevant to this thesis and will be presented along with their most influential authors. The four disciplines covered are New Rhetoric (Campbell and Jamieson, 1978; Miller, 1984), Systemic Functional Linguistics (Halliday, 1978; Hasan, 1989; Martin, 1984), mainly drawing upon the review offered by Eggins (1994), English for Specific Purposes (Bhatia, 2004; Swales, 1990) and corpus linguistics (Biber, 1988; Biber and Conrad, 2009; Lee, 2001). Some authors have confined their attention to one term only, while others have

attempted to distinguish and classify some or all of them. This review will primarily focus on the terms *genre* and *register* since they seem to be the most controversial ones, whereas *text type* and *style* are less subject to diverse interpretations, thus their meaning can be more easily sketched once the two other terms have been clarified.

#### 2.4.1.1 Genre and register

As obvious as it might seem, the first and probably most important concept that links the terms *genre* and *register* across the disciplines listed above is the powerful correlation between language and context.[8] Both terms have been used across time and theories to identify some connection between internal features of language and situational or contextual factors influencing linguistic choices. The quotes below illustrate that, in spite of some subtle nuances differentiating between each personal contribution – e.g. the Hallidayan focus on semantics and Swales's emphasis on communicative purposes – both terms tend to associate language (words and structures) to the situation/context in which lexicogrammar (language) is actually used, as clearly stated by Miller within the study of genre in rhetoric:

> What we learn when we learn a genre in not just a pattern of forms... We learn more importantly, what ends we may have: we may eulogize, apologize, recommend one person to another, instruct customers on behalf of a manufacturer, take on an official role, account for progress in achieving goals. We learn to better understand the situations in which we find ourselves. (Miller, 1984:165)

Within Systemic Functional Linguistics (SFL), Halliday has mainly focused on the term *register*, described as follows:

> A register can be defined as the configuration of semantic resources that the member of a culture typically associates with a situation type. It is the meaning potential that is accessible in a given social context. (Halliday, 1978:111)

His thoughts were largely shared amongst the group of linguists influenced by SFL, such as Hasan, Martin and other followers who gave their own interpretation of register and extended their theory to include the notion of genre:

> [Register is] a useful abstraction linking variations of language to variations of social context. (Gregory and Carroll 1978:64)

[8] The word context is here used to identify any non-linguistic elements and will be discussed more deeply below.

> The register range of a language comprises the range of social situations recognized and controlled by its speakers – situations for which appropriate patterns are available. (Ure 1982:5)
>
> Register variation is the linguistic difference that correlates with different occasions of use. (Ferguson 1994:16)
>
> A genre is a staged, goal-oriented, purposeful activity in which speakers engage as members of our culture. (Martin 2001:155)
>
> Genre theory is a theory of language use. (Ghadessy 1993:31)

The relation with the external environment is then found within the framework of genre analysis in Swales's approach:

> The definition [of genre] may be based on the communicative purposes, the situational context(s) in which it is **generally** [my emphasis][9] used, and some distinctive textual characteristics of the genre-text or some combination of these. (Swales 2004:164)

Finally, contributions by corpus and computational linguists have combined previous traditions with their own understanding of *register*, *genre* (and other similar concepts such as *text type* and *style* that will be examined further in this section):

> The register perspective combines an analysis of linguistic characteristics that are common in a text variety with analysis of the situation of use of the variety. (Biber and Conrad, 2009:2)
>
> The genre perspective is similar to the register perspective in that it includes description of the purposes and situational context of a text variety. (Biber and Conrad, 2009:2)
>
> Register [is] a way of classifying texts according to non-linguistic criteria, such as the purpose for which a text was produced, the intended audience, the level of formality, whether its purpose is narration or description and so on. For most purposes, the term genre can be considered equivalent. (McEnery and Hardie, 2012)

The above quotes may seem very confusing and one would not be completely wrong in suggesting that they are expressing similar concepts. This is certainly true since, as already

[9] Note that the use of *generally* in Swales's definition is crucial to distinguish between genre and register when referring to the relation between language and context, since *generally* implies *typical* contexts. This aspect will be further discussed in Sections 2.4.1.1.1, 2.4.1.1.2, 2.4.1.1.3 and 2.4.1.1.4.

mentioned, all these definitions convey the idea that linguistic elements are in some way related to non-linguistic features. Yet, if one analyses these terms within their relevant fields of studies, some differences emerge in the *type* of relation between language and context. As will be shown in the next sections (2.4.1.1.1 through 2.4.1.1.4), genre and register share a common ground, but perhaps more crucially, they have been used in a rather consistent way across the various theories when considering their individual characteristics and major differences.

**2.4.1.1.1 Rhetoric and New Rhetoric.** Rhetorical scholars interested in classifying texts turned their attention to the concept of genre, inherited from the long-standing tradition of genre classes in ancient Greek literature (e.g. epic, lyric, drama). The concept of genre evolved within this discipline from early classifications mostly based upon "aims" (e.g. Kinneavy's (1971) classification of expressive, persuasive, literary and referential texts) to theories that give emphasis not only to form, but to context as well. Such is the case of Campbell and Jamieson (1978), who view genres as "groups of discourses which share substantive, stylistic, and situational characteristics" (Campbell and Jamieson, 1978:20). The meaning of discourse does not derive from linguistic aspects only; social context also contributes to shaping that meaning, together with form (grammar) and substance (semantics). Although the authors do not provide detailed information on situational characteristics, they do recognise that "external factors, including human needs and exposure to antecedent rhetorical forms, create expectations which constrain rhetorical responses" (Campbell and Jamieson, 1978:21). This last observation implies that not only does a genre consist of "a constellation of forms that recurs in each of its members" (Campbell and Jamieson, 1978:16), but such constellation of forms is also viewed as recognizable by readers.

Miller (1984) agrees with Campbell and Jamieson's interpretation of genres as recognizable and recurrent forms of daily life:

> To consider as potential genres such homely discourse as the letter of recommendation, the user manual, the progress report, the lecture, the white paper [...] is to take seriously the rhetoric in which we are immersed and the situations in which we find ourselves. (Miller, 1984:155)

Despite offering an extensive list of potential genres, Miller refuses to develop a taxonomy since "genres change, evolve, and decay" (Miller, 1984:163) in time and place: "the number of genres current in any society is indeterminate and depends upon the complexity and diversity of the society" (Miller, 1984:163). Yet, it may be argued that the fact that labels are not stable across time and place does not mean that those labels do not exist, or are not worthy of attention.

Miller further claims that "a rhetorically sound definition of genre must be centred not on the substance or form of discourse but on the action it is used to accomplish" (Miller, 1984:151), thus reasserting the importance of social action as a pragmatic effect of intention. This aspect is directly relevant to subsequent discussions about genre, such as the analysis provided by Swales (1990) and his focus on communicative purposes that will be presented below (Section 2.4.1.1.2). Finally, Miller associates genres to "cultural artefacts", understood as the textual representation of a cultural way of reasoning and fulfilling purposes; in Miller's words, "a genre embodies an aspect of cultural rationality" (Miller, 1984:165). Approximately ten years later, the correlation between genre and culture is reinforced within New Rhetoric (cf. Freedman and Medway (1994)). While adopting traditional theories of genres as "types" or "kinds" of discourse characterised by textual regularities, new rhetoric approaches "focus on tying these linguistic and substantive similarities to regularities in human spheres of activity" (Freedman and Medway, 1994). According to new rhetoricians, the traditional concepts of genre had eventually turned into a set of 'formal requirements' while genre represents "a system of actions and interactions that have specific social locations and functions as well as repeated or recurrent value or function" (Freedman and Medway, 1994:59).

Apart from the word *register*, it is worth noting that a very similar term was circulating among rhetoricians, i.e. *strategy*, defined as "substantive and stylistic forms chosen to respond to situational requirements" (Campbell and Jamieson, 1978:14). This term seems to refer to linguistic choices (or indeed strategies) functionally related to the situation of use, as clearly exemplified by the following excerpt commenting on a speech delivered by J. F. Kennedy:

> The most evident form is repetition, a strategy implying a situation in which a key idea must be established and emphasized. In this case, the refrain not only repeats the theme, it also functions as refutation. (Campbell and Jamieson, 1978:15).

Therefore, the term strategy is more related to how speakers and writers use language according to immediate external circumstances and the choices of language they make to adapt to that specific situation, without necessarily satisfying expectations in terms of rhetorical actions and cultural conventions.

**2.4.1.1.2 English for Specific Purposes.** Within genre analysis in linguistics, in particular within English for Specific and Academic Purposes (ESP/EAP), Swales focused on the term genre, building on previous research from non-linguistic fields. As is already well known and described in the literature, Swales's primary criterion for identifying genres is

a "set of communicative purposes" (Swales, 1990:46), similar to what Martin calls "social action". Swales indeed seems to agree with Martin's "exceptional work" that "reinforces the concept of genre as a means of social action", and works "not only as a mechanism for reaching communicative goals but also of clarifying what those goals might be" (Swales, 1990:44). Thus, both of them insist on the centrality of either communicative or pragmatic goals.

Despite the privileged role of communicative purposes, other properties are essential for recognizing genres such as schematic structure, content, style and intended audience. All of these elements may vary across different exemplars of the same genre, thus meeting different levels of expectations. Hence, "if all high probability expectations are realized, the exemplar will be viewed as prototypical by the parent discourse community" (Swales, 1990:58). Genre prototypes are traditionally associated with Swales's theory, albeit closely connected to the principle of genre conventions, which is used by various other authors (e.g. Miller (1984); Campbell and Jamieson (1978); Biber and Conrad (2009)).

The CARS model for research article introductions well exemplifies structural and linguistic prototypical features of genres by providing non-experts with a list of "moves" to fulfil, i.e. communicative purposes, as well as some examples of typical language patterns used to introduce and connect "moves" (e.g. *there has been wide interest in*; *however, the use of*). Unlike Martin, who rejects taxonomies for the varying nature of genre labels, Swales believes that "nomenclature is an important source of insight" (Swales, 1990:52), though acknowledging that "there must be genres without a name" (Swales, 1990:57).

From what has been stated so far, some aspects of genre theory emerge that are common to both rhetoric studies and ESP:

- the focus on either social action or communicative purposes that makes genre interesting for reasons that go beyond its internal forms;
- the typicality of genre, usually referred to as conventions of forms and content;
- the cultural dimension of genres that makes them recognizable by the members of a specific community.

I will now move on to theories that have worked with both the notions of genre and register, starting with Systemic Functional Linguistics (SFL), one of the most influential and comprehensive language theories.

**2.4.1.1.3 Systemic Functional Linguistics.** The concept of *register* was developed in the sixties, when, according to Halliday (1978), Reid (1956) adopted that term for the first

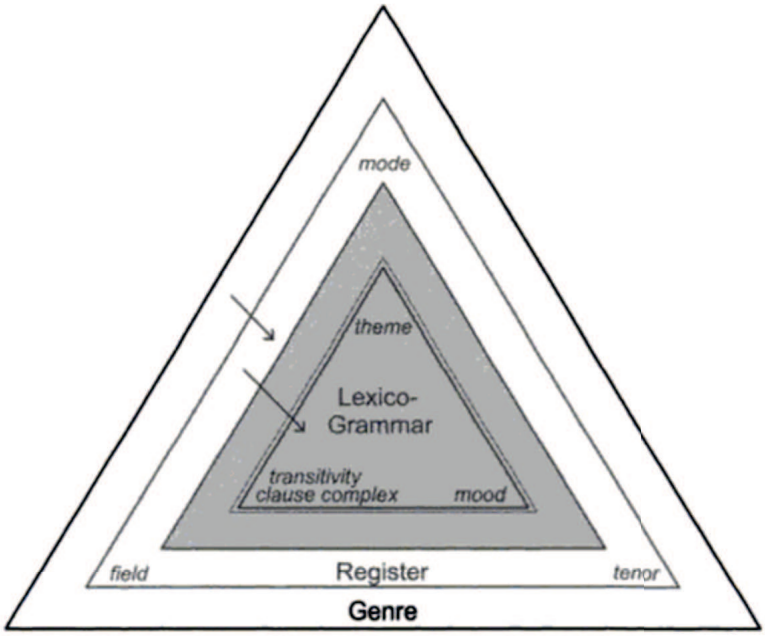


Fig. 2.2 Relation between language and context in SFL. Reproduced from Eggins (2004:111).

time. However, it was Halliday (and the group of systemic functional linguists working with him and inspired by him), who adopted the term within his approach to the study of language. According to systemic functional linguists, register and genre belong to two different levels of context: register refers to the context of *situation*, i.e. "the immediate situational context in which the text was produced" (Eggins, 1994:26), whereas genre belongs to a broader level, the context of *culture*, i.e. "wherever language is being used to achieve culturally recognized and culturally established purposes" (Eggins, 1994:47). The traditional image associated with this concept is a triangle in which language is placed at the very centre (Figure 2.2)

The way language is used is highly dependent on the external situation, represented by the second layer in Figure 2.2, as well as on some cultural aspects, represented by the third layer that also includes the context of situation (since cultural aspects are much more general than real-life circumstances). More precisely, systemic functional linguists have identified three variables characterising the context of situation and affecting language: *field*, *tenor* and *mode*, which broadly correspond to topic, participants and channel. The selection of lexico-grammatical features depends on what the author of a text is talking about (the level of technicality), the relationship existing between participants (their power relationships, the affective involvement between them and the frequency of their contact) and the distance between language and situation (written versus spoken situations). In other words, describing

the register of a text means describing its relationship with the particular situation in which it occurred, as reflected in the language choices it displays.

On the other hand, describing a text from the point of view of the *genre* extends the analysis of the immediate context of situation, since that specific text (or indeed genre) is a "recognizable social activity in our culture" (Eggins, 1994:27). Swales shares very similar ideas by stating that if all genre characteristics are transposed, one will still recognise the same genre "occurring in an atypical location" (Swales, 1990:39). SFL theory has other points of contact with Swales's genre theory as regards the schematic structure used to transfer social meaning and purposes, which are indeed realised through typical constituent parts or stages. A very simple example of constituent stages are the *Beginning*, *Middle* and *End* parts of a text (Eggins, 1994:37), although there may be more specific steps, such as the ones suggested by Swales in research article introductions[10] – i.e. *establishing the territory*, *creating a niche*, *occupying the niche*. There are two types of criteria for labelling stages: functional criteria are used to express the distinctive functions of each stage, as occurs in the examples given above, whereas formal criteria just refer to formal units making up the whole text, such as *chapter 1*, *chapter 2* etc. (Eggins, 1994:37-38). Finally, stages are usually ordered, e.g. by following a logic sequence; they can be either obligatory or optional, and they can also be recursive (when they occur more than once in the same genre).

Some systemic functional linguists, especially Martin (1992), have accounted for schematic stages within longer texts by introducing the concept of macro-genre. Martin uses the term *macro-genre* with reference to more complex texts "within which it is possible to identify a range of other genres being used" (Eggins, 1994:47). The example provided by (Eggins, 1994) is particularly relevant to the topic of this thesis:

> A university department's handbook is itself a macro-genre (a staged, goal-oriented use of language), but it typically contains sections exemplifying the genres of exposition (why you should study the subject, what it is about), description (course outlines) and regulation (rights/responsabilities/penalties). (Eggins, 1994:47)

To summarise, the key concepts of SFL that are useful to identify an area of agreement among the various disciplines are the following:

- register is constrained by the immediate situational context in which texts are created;
- genre is realized and recognized within the context of culture;

[10] Swales CARS model is in line with SFL in the definition of stages (or moves, in Swales' terms).

- the relation between genre and the context of culture is stronger than the relation between genre and the context of situation;
- genres are (culturally) structured and labelled.

The next section discusses the distinction between register and genre as conceived by some authors within corpus linguistics, especially Biber and Conrad (2009), Lee (2001), McEnery and Hardie (2012).

**2.4.1.1.4 Corpus linguistics.** As claimed by Biber and Conrad (2009), the terms genre and register (as well as style and text type) do not indicate different texts; rather, they refer to different perspectives on text varieties (Biber and Conrad, 2009:2). Lee (2001) shares the same idea by stating that "the two most problematic terms, *register* and *genre*" can be regarded as "essentially two different points of view covering the same ground" (Lee, 2001:46). The register perspective is perfectly summarised in the following quotation:

> The register perspective combines an analysis of linguistic characteristics that are common in a text variety with analysis of the situation of use of the variety. The underlying assumption of the register perspective is that core linguistic features like pronouns and verbs are functional, and, as a result, particular features are commonly used in association with the communicative purposes and situational context of texts. (Biber and Conrad, 2009:2)

A key element expressed in the above excerpt refers to the connection between language features and situation that has already been noted at the beginning of this section. Furthermore, and in line with SFL, this relation is functional, which means that the choice of verbs, pronouns, words and so on depends on the situation in which each communicative act is produced and received. However, the communicative purposes do not play such a significant role as they do in SFL. In Biber and Conrad (2009), purpose seems to be included among situational factors, whereas SFL emphasises purpose by relating it to a meaningful use of language in the culture in which it is produced (cf. Eggins (1994)), i.e., by assigning it to the realm of genre analysis. According to Biber and Conrad (2009:6), the description of a register includes three major components: "the situational context, the linguistic features, and the functional relationships between the first two components", where both external situational characteristics (e.g. participants, settings, channel etc.) and the internal linguistic features to which they are functionally related are key factors discriminating among registers. A direct consequence is that any text can be regarded as belonging to a register since any

text (or even text extract) can be described according to production circumstances (authors, recipients, settings and so on) and lexico-grammatical features.

Moving to the genre perspective, the authors state that:

> The genre perspective is similar to the register perspective in that it includes description of the purposes and situational context of a text variety, but its linguistic analysis contrasts with the register perspective by focusing on the conventional structures used to construct a complete text within the variety, for example, the conventional way in which a letter begins and ends. (Biber and Conrad, 2009:2)

Thus, genre and register perspectives diverge in two fundamental ways. From the point of view of context, the difference is comparable to that outlined by SFL. Registers vary according to the immediate situation, whereas genres also depend on cultural factors. From the point of view of language, lexico-grammatical elements analysed through a register perspective are distributed all along the text to fulfil its communicative functions; such linguistic elements, functionally related to the immediate external context, are clearly pervasive. On the other hand, genre markers are conventional rather than functional – i.e. they meet primarily social or cultural expectations, not functional purposes. Consequently, genre markers can only be spotted either by immediate recognition from the group of experts or by comparing exemplars of the same genre (e.g. *the study of*, *in recent years*, *has a great importance for* are examples of genre markers in academic research articles).[11] Lee's position is similar to that of Biber and Conrad and, with regards to the conceptual issues of genre and register, he hopes not to 'muddy the waters' any further (Lee, 2001:45). In his view, a register is used with reference to "the instantiation of a conventionalised, functional configuration of language tied to certain broad societal situations, that is, variety according to use". As for genre, he takes up Miller's definition as "a culturally recognised artefact, a grouping of texts according to some conventionally recognised criteria, a grouping according to purposive goals, culturally defined" (Lee, 2001:46). As Lee observes, "there is a superficial terminological difference in the way *genre* is used by some theorists, but no real, substantive disagreement" (Lee, 2001:45).

This survey suggests that there is indeed a sense of continuity, despite terminological inconsistencies, for example in the functional relation between language and situation (commonly referred to as register) and the cultural/social recognition of some texts according to a number of recurrent (non)linguistic elements (very much associated with the term genre). At least three elements seem to synthesise the various theoretical contributions:

[11] Examples are taken from Swales (1990:144).

- both *genre* and *register* identify a link between language and non-linguistic or external elements;
- register refers to how language choices are functionally related to the situational context;
- genre is defined by cultural conventions on language choices, text structure and genre labels.

Now that the main similarities and differences between genre and register have been discussed, we may tackle the more straightforward concepts of text type and style.

#### 2.4.1.2 Text type

Paltridge (1996) notes that the term *text type* has not received as much attention as *genre* and *register*. By providing a comprehensive review of existing theories, he illustrates how the term is used in a rather ambiguous way and sometimes interchangeably with the term *genre*. Lee (2001) agrees with Paltridge by stating that "the term *text type* [...] can be used in a very vague way to mean almost anything" (Lee, 2001:38). Hasan (1989) provides some examples of text types that may be easily confused with both registers and genres:

> If presented with a text in displacement—in isolation from the situation in which it was produced—we are still able to say what type of text it is. Is it a sermon? Is it a lecture? Is it a buying and selling transaction? Is it to control a child? Is it a story? (Hasan 1989:58)

De Beaugrande and Dressler (1981) also provide a rather vague definition of text type:

> A set of heuristics for producing, predicting, and processing textual occurrences, which hence acts as a prominent determiner of efficiency, effectiveness, and appropriateness. (De Beaugrande and Dressler 1981:184)

Following this definition, they list seven types of texts: descriptive, narrative, argumentative, scientific, didactic, literary and poetic. Werlich (1976) instead suggests five text types that partially overlap with Beaugrande and Dressler's categories: narration, description, exposition, argumentation and instruction. Whereas the examples just given are similar to the meaning of 'functions', Biber (1988) provides a very different definition of the term *text type*:

> Text types are defined such that the texts within each type are maximally similar in their linguistic characteristics, regardless of their situational/genre characteristics. (Biber et al., 2007:172)

Referring to Biber's Multi-Dimensional analysis (Biber, 1988), McEnery and Hardie (2012) also explain how register differs from text type:

> Text types are distinct from registers in a crucial respect: while a register is a group of texts defined on the basis of language-external features (i.e. context, medium, purpose), a text type is a group of texts defined on the basis of linguistic similarity, with no necessary implication that they are from similar contexts of use.

In other words, Biber et al. (2007) claims that *text type* is the only term used for text classification that makes no reference to external situation or context, thus describing texts in terms of their co-occurring linguistic features. Since Biber's analysis is based on the assumption that linguistic features can be interpreted in functional terms, different co-occurrence patterns are analysed functionally as dimensions of variation (cf. Biber (1988) and Biber et al. (2007)).

#### 2.4.1.3 Style

The concept of style as used within linguistics raises similar issues to those outlined above for text types, since the term has been used in a variety of fields for many decades, thus contributing to expanding its numerous and often contradictory meanings. To give a very few examples, Campbell and Jamieson (1978) used *stylistic forms* to refer to some undefined internal features of language (Campbell and Jamieson, 1978:14). Halliday's three elements of discourse initially included *field*, *style* and *mode*, where *style* referred to the relation among participants; later on, he replaced the term *style* with *tenor*, taken from Gregory and Carroll (1978). Within English for Specific Purposes, Swales (1990) counts it among criteria for recognizing genres:

> In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience. (Swales, 1990:58)

Finally, the EAGLES guidelines on corpus typology (Sinclair and Ball 1996: unpaginated) acknowledge the lack of an "agreed set of values", but strive to provide some recommendations on style and only argue that style is mostly defined internally than externally; the authors also relate style to the broad categories of "formal" and "informal" language, drawing

on Joos (1961). Within corpus linguistics, Biber and Lee have attempted to achieve coherence among divergent viewpoints, also trying to avoid overlapping of concepts, especially with terms that had already been defined. Lee's own view is that "*style* is essentially to do with an individual's use of language" (Lee, 2001:45), in line with Biber and Conrad (2009) who believe that style is just another perspective on the analysis of texts, along with register and genre, but more focused on the authors' preferences:

> The style perspective is similar to the register perspective in its linguistic focus, analyzing the use of core linguistic features that are distributed throughout text samples from a variety. The key difference from the register perspective is that the use of these features is not functionally motivated by the situational context; rather, style features reflect aesthetic preferences, associated with particular authors or historical periods. (Biber and Conrad, 2009:2)

Thus, Biber and Conrad (2009) and Lee (2001) provide a sufficiently clear and unambiguous definition of style. Style is indeed comparable to register in the correlation between language and context; however, style describes the relation between language and one aspect of situation, i.e. individual writers' aesthetic preferences. Since style becomes more relevant in literary genres, this thesis does not analyse texts according to that parameter. Apart from the fact that university websites rarely provide details on webpage writers, and that webpages may well be written collectively, the main research interests behind the study of institutional (academic) communication lie in the way institutions portray themselves through language, rather than in the stylistic preferences of single authors. One may argue that university website authors adapt their style to institutional guidelines on web writing, thus expanding the concept of style to include "ideological" preferences on language. Yet, not to blur the boundaries any further, I will not use the term *style* when describing language features that convey institutional images and values.

### 2.4.2 Classifying genres

Texts can be classified in many ways, e.g. by author, language, topic, and obviously, genre. While some of these criteria provide a rather simple means to distribute texts into classes, others are definitely more complex and genre is one of them. Due to its slippery definition(s) and related concepts – such as register, text type and style – (cf. Section 2.4.1), classifying documents by genre can become extremely challenging. On the other hand, genre categorisation has several practical applications, from natural language processing (NLP) to applied tasks for second-language learners, translators and many other professionals. In

corpus-based studies, coding large amounts of texts with information on genre/text types, or details about the situational context in which texts have been created (e.g. authors, receivers and purposes), helps to control "external" variables, thus providing insights for a deeper understanding of the language(s) under investigation. For the purposes of this thesis, classification represents an important advantage, since it enables the analysis of comparable subcorpora in terms of functions, as well as the analysis of promotional language markers across genres.

As was argued in Section 2.4, genres can be identified and grouped on the basis of internal and external features. External criteria refer to "features of the non-linguistic environment or society in which the texts occurred" (Sinclair and Ball 1996: unpaginated): these were used to compile the first electronic corpora to account for external factors that might influence language variation. Lee (2001:38) refers to external criteria as "properties other than lexical or grammatical (co-)occurrence features". Information concerning the speaker/writer and audience tends to be included among external criteria, as well as sociocultural aspects. For example, university web pages aiming at recruiting prospective students are "situationally" different from content that is aimed at building or increasing confidence among newly enrolled students. Furthermore, universities with different institutional traditions may be assigned to separate classes. For instance, British universities have been using web-based marketing strategies for recruiting students long before other European universities, which may lead to different practices in text writing.

As opposed to external features, internal criteria are defined linguistically, taking into account the presence and distribution of language features. In other words, internal criteria are "features of the language", based on internal choices and patterns (Sinclair and Ball 1996). For instance, spoken registers are allegedly known to contain more adverbs than written registers (Biber, 2006a). Assuming that external features are reflected internally – since context is in text (Eggins, 1994:49) – one can easily deduce a hypothetical situation from reading a piece of an utterance, or conversely, extracting language features and patterns that are typical of specific contexts.

In order to analyse how features of language interact with situational factors, two approaches are traditionally mentioned in corpus-based studies, namely top-down and bottom-up methods (cf. Biber et al. (2007)). In the top-down approach, the first step is to create a functional framework (e.g. Swales's moves), while linguistic analysis comes afterwards. The functional framework provides an interpretative background for any systematic linguistic characteristics found in the corpus. Conversely, linguistic analysis is primary in any bottom-up approach, which aims at grouping linguistically similar categories and interpret them afterwards according to situational or functional "types". As summarised in Biber et al.

| | | |
|---|---|---|
| **Move 1:** | **Establishing a territory** | |
| | Step 1 | Claiming centrality and/or |
| | Step 2 | Making topic generalization(s) and/or |
| | Step 3 | Reviewing items of previous research |
| **Move 2:** | **Establishing a niche** | |
| | Step 1A | Counter-claiming or |
| | Step 1B | Indicating a gap or |
| | Step 1C | Question raising or |
| | Step 1D | Continuing a tradition |
| **Move 3:** | **Occupying the niche** | |
| | Step 1A | Outlining purposes or |
| | Step 1B | Announcing present research |
| | Step 2 | Announcing principal findings |
| | Step 3 | Indicating RA structure |

Fig. 2.3 Swales's CARS model for research article introductions (taken from Biber et al. (2007:25))

(2007:241), top-down approaches are not necessarily corpus-based and are very often labour-intensive since they require analysis of texts on a qualitative basis. In contrast, bottom-up approaches can be automated and applied to a corpus of any size (cf. Biber et al. (2007)).

Genre-based move analysis (Swales, 1990) is a typical example of top-down approaches as it aims at categorising texts into structural units, or moves, according to the specific function(s) of each unit. Assuming that different functions are realised using different language elements (e.g. vocabulary, phraseology, lexico-grammatical features), move analysis provides a functional framework for identifying conventions of content and style, as well as more creative realisations. Swales's moves for research article introductions, also known as the CARS model, is presented in Figure 2.3.

Figure 2.3 shows the very first stage of move analysis, i.e. a user-based development of genre purposes and functions. This first stage is followed by examples of how each move is performed linguistically. Examples of Move 1, Step 1 include:

- The study of... has become an important aspect of...
- A central issue in... is the validity of... (Swales, 1990:144)

As opposed to move analysis, Multi-Dimensional (MD) analysis (Biber, 1988) represents a typical example of a bottom-up approach for investigating interaction between internal and external features. MD analysis is based on statistical techniques (factor analysis) that measure the distribution of linguistic features and group them according to their co-occurrence in texts. Once features have been grouped, human interpretation is needed to assign a functional

meaning to each group. For instance, Biber (1988) concludes that the co-occurrence of first and second person pronouns, direct questions, and imperatives is related to interactive texts (i.e. spoken-like types), while past tense verbs co-occurring with perfect aspect verbs, matched to a negative occurrence of present tense verbs, are all related to prototypical fictional narratives. As is clear from this brief description, top-down and bottom-up approaches imply similar but inverted processes, i.e. a) the development/interpretation of text functions and b) the analysis of linguistic features.

In between these two distinct approaches, there are examples of studies that have adopted a complementary view, for instance Egbert et al. (2015) and Sharoff (2018). The former has been proposed as the first attempt to develop a user-based, bottom-up classification framework for English web registers, moving away from standard procedures that had always combined bottom-up approaches with machine-learning methods. In Egbert et al. (2015) *bottom-up* refers to the empirical process adopted for creating the register rubric, since functions are gleaned from reading texts – i.e. through an empirical analysis of language. Web users were asked to read a text and answer a set of multiple choice questions. This process ended up with coding 1,625 random internet pages. The framework consists of a two-level system composed of eight general registers – narrative, opinion, description, discussion, lyrical, instructional, informational persuasion, spoken – and several sub-registers (e.g. description: abstract, FAQ, research article).

Similarly to Egbert et al. (2015), Sharoff (2018) takes a user-based perspective for classifying large web corpora into genres. In Sharoff (2018), human annotators were asked to evaluate a range of texts according to twelve Functional Text Dimensions (FTDs), identified in a previous study (see references in Sharoff (2018)). FTDs can be grouped into four general functional categories (Sharoff, 2018:6-7):[12]

**information** General functional categories aimed at informing the reader:

**A7** instruct Tutorials, FAQs, manuals, recipes;

**A8** news Newswires, newsletters;

**A9** legal Laws, contracts, small print;

**A16** info Specifications, CVs, encyclopedic articles, abstracts;

**discussion** General functional categories aimed at discussing a state of affairs:

**A1** argum Editorials, columns, argumentative blogs, political debates;

[12]Numbers idenitfying FTDs range from 1 to 20 and are in an arbitrary order. Due to a series of subsequent adjustments, some of them have been removed.

**A14** research Research articles, essays;

**A17** eval Reviews of products and services;

**narration** General functional categories aimed at presenting a story

**A4** fiction Novels, stories, verses;

**A11** person Diary-like blogs, personal letters, traditional diaries;

**promotion** General functional categories aimed at promotion of information:

**A12** commpuff Advertising and commercial texts for promotion;

**A13** ideopuff Propaganda, manifestos;

**A20** appell Requests, small ads.

While Egbert et al. (2015) does not aim at describing documents internally and leaves linguistic analysis for future work, Sharoff (2018) goes a step further by using machine-learning tools to extend genre classification to similar texts. Methodologically, he combines traditional top-down approaches that start from a text-external perspective with automated procedures that are more typical of bottom-up, empirical approaches. Once a reliable external, functional framework is established, correlation with any text-internal features is investigated through automatic classifiers. Human judgement serves as a frame of reference for training classifiers. The key point in Sharoff (2018) is its reliance on human assessment of text functions. Whereas purely computerized, bottom-up procedures – such as MD analysis – rely on human interpretation of any recurrent linguistic markers found, Sharoff (2018) puts emphasis on readers' perception and their initial interpretation of text functions, attempting to answer questions as to the extent to which a text can be interpreted as belonging to a generalised functional category.

Classification based on readers' perception of functions, as compared to interpretation of co-occurring lexico-grammatical features, is particularly relevant to this PhD thesis for one main reason. The promotional function of language is rather "liquid" and subject to different theories and contexts. This translates into multiple ways of creating promotional content linguistically. Although there are some widely recognised promotional strategies (see Section 2.3), promotional and evaluative language is often perceived on a discourse level, since evaluation is contextual, cumulative and implicit (Hunston, 2010). For this reason, (shared) human subjectivity can be considered as an advantage rather than an obstacle. Furthermore, this approach enables comparison of different language varieties or even different languages. If participants of a discourse community are able to recognise text functions despite the

varying micro-level language realisations, then one can exploit computerised techniques in order to observe whether there are significant differences between languages or language varieties in fulfilling the same function(s). In other words, ELF and native English texts might be perceived as being similarly promotional, despite their different ways of performing promotion internally. In this specific case, adopting exclusively bottom-up approaches does not provide any clues as to whether co-occurring linguistic markers are indeed serving the same function. Therefore, Sharoff (2018)'s complementary view represents an important advantage for this study, in which user-based/qualitative analyses are applied to large web-based corpora (see Chapter 4 for a classification experiment).

## 2.5 Web-based approaches to corpus construction

Since this study explores university websites' promotional language adopting corpus linguistics' techniques, several theoretical issues associated with corpus-based methods require attention in this Section.

Any introductory text on corpus linguistics will at some point mention a core principle of corpus-based studies, which is representativeness.[13] Balance is the second "desideratum" of corpus linguists, which is very much related to the first principle since balanced sampling is more likely to produce representative results when there are different levels or variables to be considered. Other principles are listed as the basis of research in corpus linguistics, such as reproducibility and comparability, both relying on similarity of design criteria and content data.

McEnery and Hardie (2012) express a viewpoint probably shared by the majority of corpus linguists when they claim that representativeness is more of an ideal than a reality. Leech (2007) calls it the "Holy grail"; Kilgarriff and Grefenstette (2003) open their section on representativeness using the "Pandora's box" metaphor. Since it is very unlikely that a corpus can include all and only the texts addressing a specific research question, sampling data is considered a viable alternative, provided that the sample reflects the whole population. Collecting representative data that match a research question is a prerequisite for obtaining reliable results and making generalisations. This is true of any type of study, regardless of whether one is dealing with a spoken or a written corpus, whether the corpus is compiled manually or (semi-)automatically or whether it is a web- or paper- based collection.

In this respect, Fletcher (2004) makes a crucial distinction between two approaches in which the web can be exploited for corpus-based studies, *web as corpus* and *web for*

[13] The term "representativity" may also be used as an alternative to "representativeness" (cf. Leech (2007:140)).

*corpus*. The first term quite obviously refers to the use of the web as a corpus in itself, i.e. an enormous source of linguistic data. This approach has raised a number of issues related to representativeness, for at least two reasons. First, language content on the web is still a grey area. Although statistics show that content in English covers more than half of all webpages,[14] there is in fact a lack of evidence and sound arguments for ascertaining that the web reflects English language usage in general. Even if it did, another issue would be related to the variety of language adopted within those pages. Do they reflect spoken or written English? Native or non-native? The answer remains largely unknown. Second, the most widely used tools for querying the web (as a whole) rely on search engines' results, which are not ranked according to linguistic parameters, thus potentially skewing results. Some examples of these tools include WebCorp (Kehoe and Renouf, 2002) and KwicFinder (Fletcher, 2004), online concordancers that enable users to query the web and analyse concordance lines. The second term, i.e. *web for corpus*, tends to be regarded as a safer choice in which a controlled selection of webpages is performed (Hundt et al., 2007:2). The feeling behind this tension, particularly at the very beginning of this millennium, was that automatic techniques for text mining were replacing traditional methods for corpus building – such as careful human selection of texts – thus hindering corpus quality, representativeness and other fundamental principles of corpus linguistics:

> All this adds up to the rather uncomfortable impression that in the web-as-corpus-approach, the machine is determining the results in a most "unlinguistic" fashion over which we have little or no control. (Hundt et al., 2007:3)

As much as this is true, all the debate emerging from the use of the web for language purposes and very much centred on the *methods* for corpus building may have diverted the focus from the reasons why one resorts to web data, the purposes for using a specific web-based corpus and ultimately the scope of one's study. Whether one decides to crawl the whole web, a small portion of it, or instead, to manually select a sample of pages, what really matters is how that corpus will be used afterwards and whether the use of that corpus is warranted by the building procedures used. There are two main reasons for selecting the web as a source of linguistic knowledge:

1. to increase the quantity of available data, as an alternative to more traditional sources, when existing data are not sufficient for the scope of a study;

2. to address a certain research question for which the web is the ideal source of data.

[14] http://w3techs.com/technologies/overview/content_language/all [last consulted on 15 December 2017].

As far as the first reason is concerned, two possible scenarios are envisaged that reflect a different usage of the web as a medium to:

- store, copy and share existing material that was originally produced for a different medium (e.g. paper or television);
- create and disseminate new content, for example running text within companis' and institutions' webpages, blogs and forums, twitters, posts and the like.

This distinction is not always clear-cut, since webpages may be drawn from paper-based material, especially for encyclopaedic genres, which means that more caution is required when the first aim is involved. Yet, archives and text collections are often restricted to a recognizable portion of the web, thus enabling researchers to exploit it through automatic techniques and, ultimately, increase their data sets. Examples include Hoffmann (2005), who compiled a large corpus of approximately 100,000 CNN transcripts and Nesselhauf (2005), who collected 23 fiction texts (1M words) in electronic format dated from the 19th century. Their experiments are often counted among successful research studies using the web *for* corpus building (Hundt et al., 2007:3).

However, it is mainly the second scenario that showed its unlimited potential in supplementing traditional corpora with present-day, freely accessible linguistic resources. The World Wide Web is continuously growing in size and represents an extensive source of modern language varieties. This has triggered the development of web platforms such as KWiCFinder and WebCorp, as well as other tools that let users crawl large portions of the web for research and teaching/learning purposes. The WaCky project (Baroni and Bernardini, 2006) developed a set of tools for web crawling, as well as language resources addressed to people with different backgrounds and goals in the field of languages.[15] Although this project is listed among the web *as* corpus approaches (Hundt et al., 2007:3), web-crawled corpora diverge from online concordancers (e.g. KWiCFinder and WebCorp) since the former do not heavily rely on search engines' indexes. By crawling specific web subdomains (e.g. .co.uk) and querying them offline via specific software, it is possible to have a higher control on the selection of webpages, as well as on retrieved queries. UkWaC – an English web corpus compiled within the WaCky project – and enTenTen – a similar project compiled at Masaryk University – can be employed by learners of English and language professionals, as suggested by Bernardini et al. (2010). Translators, for instance, often turn to the web as their primary source of information for documentation and terminological work. Besides being suitable for pragmatic reasons, the use of UkWaC has also proven successful as a reference corpus of general English in lexicographic studies (e.g. Ferraresi et al. (2010)).

[15] http://wacky.sslmit.unibo.it/doku.php [last consulted on 15 December 2017].

Therefore, using the web for the first reason, i.e. to compensate for a lack of (sufficient) data from other sources should not be regarded as an undesirable procedure, even when dealing with larger domains, as long as this decision is taken after careful reflection as to whether a detailed breakdown of the type of linguistic data is needed, and whether or not the influence of the digital medium on language will skew results. This scenario also includes choosing the web as an alternative to "expensive" corpora. Relying on existing well-balanced, large corpora is sometimes not possible due to restrictions on freely accessible corpora, as claimed by Schäfer and Bildhauer (2013:1), who made a list of corpora that cannot be downloaded as a whole.

The second reason for the use of internet data is related to the development of new genres and text types. The web has reshaped traditional text types – for example by repurposing news articles into blogs – and has given rise to an altogether new range of genres, for instance forums and chat rooms, social network posts, homepages, thus inspiring new strands of research specifically aimed at studying web language and content.[16] Research addressing specific web genres in corpus linguistics has adopted various methods, depending on the scope of analysis and the desired level of granularity. Gesuato (2011) performed a hand-picked, qualitative selection of academic course descriptions (ACDs) aimed at examining a prototypical sample of web ACDs and ultimately exploring genre structure and communicative functions. Morrow (2006) described discourse features in forums about depression: he gathered three types of messages (problems, advice and thanks) for an overall number of 85 messages in order to look at how advice-givers use discourse strategies to soften their advice.

As an example of research methods with wider scope and higher granularity, Thelwall (2008) focused on the MySpace website and automatically crawled a number of profile homepages based on a range of similar URLs. Pak and Paroubek (2010) collected a large corpus of Twitter text posts for sentiment analysis purposes, by querying Twitter for types of emoticons through the Twitter API. Bhosale et al. (2014) extracted approximately 30,000 English articles from Wikipedia to analyse promotional content. Finally, the acWaC project[17] provides a large resource for exploring a neglected but strategically important domain, i.e. the institutional academic one (Bernardini et al., 2010). Although the authors use caution in deepening their focus to internal subgenres for which they have limited control, this resource can be said to reflect English language delivered by university websites at the time of crawling and can thus be used to compare university language across countries.

As this very brief outline shows, one can employ automatic techniques to download "noisy" data from the web that serve general lexicographic purposes, or to fetch a limited por-

[16] For a complete overview of web genres see Mehler et al. (2010).

[17] http://mrscoulter.sslmit.unibo.it/acwac [last consulted on 15 December 2017].

tion of the web to match tightly focused research questions addressing a specific web domain (that may reflect a specific genre). Crawlers (either homemade crawlers or commercial/open-source programmes) provide a flexible strategy that can be arranged to address multiple goals, from general use of British English – by crawling the .co.uk domain and use it as a reference corpus (e.g. UkWaC and enTenTen) – to specific domain-related tasks.

## 2.6 Summing-up

The present Chapter provided a synthesis of the key concepts and previous studies related to each strand of research underlying this work. Section 2.1 provided a definition of institutional academic language and presented two main approaches to the study of this domain, namely Critical Discourse Analysis (CDA) and corpus-based studies. Section 2.3 offered a review of the various approaches to the analysis of evaluation (lexical and grammatical approaches, implicit or explicit evaluation) and produced an inventory of lexico-grammatical features that are listed in the relevant literature as an important device for expressing evaluative meaning. Section 2.4 took a double focus on genre: on the one hand, it reviewed existing definitions and interpretations of the concept of genre (including sub- and macro-genre) and related ones such as *register*, *text type*, and *style*. On the other hand, it presented top-down and bottom-up approaches to genre classification (e.g. Swales (1990) and Biber (1988) respectively), as well as examples of studies that have adopted a complementary view, for instance Egbert et al. (2015) and Sharoff (2018). Finally, Section 2.5, presented web-based approaches to corpus construction. Given the framework and the extensive research about web-based principles, defining the target population is the first step for building a corpus that ideally matches the initial research question. Corpus design criteria and corpus building will be described in Chapter 3.

# Chapter 3

# Corpus design and construction

## 3.1 Introduction

As pointed out by Biber (1993:1), at least two elements have to be taken into account in order to define a target population for corpus building:

1. the boundaries of the population — what texts are included and excluded from the population;
2. hierarchical organization within the population — what text categories are included in the population.

The first element is determined by the specific research questions of the study and the reasons why it is being conducted. Three types of boundaries can be drawn that reflect the general aims of this thesis:

1. macro-genre boundaries, namely academic websites;
2. geographical/institutional boundaries, i.e. European universities;
3. language boundaries, i.e. English language used as a means of communication between universities and their stakeholders.

As regards genre, academic websites are typically recognizable by a unique domain name, which can be used as a macro-genre delimiter. For instance, most URLs produced by the University of Bologna are in the `http://www.unibo.it/example` format, where "unibo" is the domain name associated with the University of Bologna and the same applies to most universities in Europe. Therefore, the whole population comprises any webpage produced by

universities (using an institutional domain name) to deliver informational, instructional, promotional content through various web (sub-)genres. Considering the extensive population to be included in the corpus, texts are collected automatically through web crawling techniques (Section 3.3).

The second type of boundary is a geographical one and coincides with European borders. As already discussed in this thesis Introduction (Chapter 1), the reasons for choosing European universities are related to EU efforts to foster mobility and harmonise degree programmes and web-based communication. Yet, it would be inaccurate to choose EU-only countries – thus excluding European countries that are not currently an EU member state – since many of these non-EU countries in Europe have indeed signed specific agreements with the Union to support mobility and provide exchange grants (e.g. the Swiss-European mobility Programme). Moreover, they are generally closely associated with the EU because of political and economic agreements (e.g. Free Trade Association and European Economic Area). The reasons that have kept some countries out of the EU are not preventing these same countries from creating a multicultural education environment, especially when geographical proximity favours this process. Therefore, this project compilation process includes any country in the European geographical area that is involved in the European international scene, e.g. by taking part in the Bologna Process and European Higher Education Area (EHEA). Information on country status (EU vs. non-EU) is recorded within metadata and can thus be used to either include or exclude specific countries. Finally, at this stage no attempt is made to include countries outside the European geographical area. Conducting a global-scale study would be hardly feasible at this point due to many varieties of native English language, as well as incomparable education systems and internationalisation policies outside Europe.

Finally, language boundaries are a crucial element in this project, which aims at analysing English language in its native and lingua franca varieties. Pages that are produced in other languages will have to be excluded through automatic techniques for language identification (Section 3.3.3).

These three external boundaries draw a line between data to be included in the corpus and data to be excluded from the corpus and represent a benchmark to define criteria for corpus building (Sections 3.2 and 3.3). On the other hand, hierarchical organisation within the population refers to a set of internal categories that will group texts according to specific parameters, such as the variety of English and genre. Corpus composition and its internal organisation are described in Section 3.4.

## 3.2 Corpus design

### 3.2.1 Language

As already discussed in the Introduction to this Chapter (Section 3.1), population boundaries are defined by English language texts exclusively, in their native English and ELF varieties. In this thesis, the term ELF covers the European area, and more specifically, European countries where English is not an official language (cf. Section 2.2). Web contents produced in the UK and Ireland are very likely to be produced by native speakers of English and are therefore coded as native English language.[1] Although some native English webpages may be addressed to an international community and specifically written for them – which would theoretically include native speakers in lingua franca communication – native speakers of English will most likely generate native-like patterns of English, as far as language production is concerned, which is the very focus of this PhD thesis. Even if native speakers are asked to produce specific language patterns addressing their international audience, how do we know whether these patterns are closer to a lingua franca variety than their native (and more natural) counterpart? If ELF ever turned into a fully-fledged language variety, English speakers would be able to speak and write in their own English language variety, as well as to "translate" into a lingua franca variety of English, just as British and American English speakers do. The rationale behind this thesis is that any type of content which is written by native speakers of English in a native English context, and directed to either international or local recipients cannot be treated as ELF, although it may differ from native-to-native communication. This is particularly true for written, distant, monologic communication, where there is next to no interaction or sentence co-construction, as compared to conferences and business meetings.

While it is very likely that most webpages delivered by British and Irish academic websites are produced by native speakers of English, the same might not be true of continental websites. However, this thesis shares Palumbo's perspective according to which any text produced in a non-native country and addressed to an international audience will have undergone some kind of translational or mediation process (Palumbo, 2013:98). This is arguably true if one considers that universities are deeply rooted in their cultural and environmental settings and that there are no apparent reasons for reproducing native language models, especially when the majority of international students are ELF speakers. The general tendency of adopting the word *course* in ELF, instead of the British word *module* is a clear example of this (Palumbo, 2013:104).

[1] In applied linguistics, the concept of "native speaker" is generally considered as a useful fiction due to its slippery definition (cf. Davies (2004)).

Therefore, texts coming from a British or Irish academic domain will be automatically coded as native (NAT), whereas English texts downloaded from ELF countries will be automatically coded as lingua franca (ELF). Although there is no way of excluding the possibility that texts coded as "native" are written by ELF speakers and that texts coded as "ELF" are written by native speakers of English, this technique is thought to produce a useful and acceptable generalisation.

### 3.2.2 Number of universities

To maximise chances that communicatively effective texts are included in the corpus, only high-ranked universities are considered for inclusion in the corpus, in the attempt of obtaining a gold standard. It is indeed very likely that leading universities invest substantial amounts of money in communication to keep their online image at high standards. Furthermore, since these universities are evidently involved in the international scene, texts in the gold standard are more easily comparable and are more likely to adopt European conventional promotional strategies. For these reasons, this study relied on the list of countries and universities provided by international rankings. The three most influential global rankings are allegedly said to be:

- The Academic Ranking of World Universities (ARWU);[2]
- QS World University Rankings;[3]
- Times Higher Education World University Rankings.[4]

Considering that variation over rankings tends to be quite low among top positions, it was not deemed important to select the most accurate ranking in terms of indicators for measuring academic quality. As for data mining, QS World University Rankings enables users to filter universities by geographical region (e.g. Europe), while ARWU and THE World University Rankings allow country filtering exclusively. QS World University Rankings also provides contextual details, such as the number of international students, university size and status in terms of public/private funding, which might be used to control extra-linguistic variables as well. Therefore, QS World University Rankings was used for university selection.[5] Criteria for corpus building include the full list of European countries and a selection of universities

[2] http://www.shanghairanking.com

[3] http://www.topuniversities.com/qs-world-university-rankings

[4] https://www.timeshighereducation.com/world-university-rankings

[5] The Webometrics Ranking of World Universities (http://www.webometrics.info/en) was also considered a viable option since a) it had already been used in the acWaC project and b) it includes parameters on web contents, visibility and impact of universities on the Web. However, Webometrics was undergoing major updates and was not working efficiently at the time of corpus building (February 2016).

based on the total number of universities per country listed in the ranking. The top 30% of universities in each country were chosen, setting a maximum of ten universities per country.

### 3.2.3 Website structure level

As corpus building heavily relies on websites' internal linking structure (Section 3.3.2), one of the criteria for collecting pages includes how deep to get in the website architecture. Downloading the full website is not an ideal solution in terms of time and resources employed since the deeper you go the higher the probability of retrieving pages that have already been collected. Furthermore, top-level pages are viewed as the most valuable pages of any website because they have the potential to attract new visitors. Based on experiments in the acWaC Project (Ferraresi and Bernardini, 2013), two levels were deemed to be an acceptable trade-off for covering a sufficient number of webpages as well as a wide range of genres (e.g. *How to apply*, *Why choose us*, *FAQs*, *Personal Homepages*).

## 3.3 Corpus building

The procedure for text collection relies on the pipeline described in the acWaC Project (Ferraresi and Bernardini, 2013), including post-processing techniques developed in the WaCky Project (Baroni and Bernardini, 2006). Corpus building consists of three steps:

1. scraping the web to retrieve a list of seed URLs, i.e. university English homepages, along with relevant metadata;
2. crawling university websites starting from the list of URLs;
3. post-processing data, annotation and indexing.

Each of these steps will be briefly described in the next sections.

### 3.3.1 Scraping

As concerns the first step, web scraping is needed to retrieve a list of URLs – i.e. university homepages – that will then be used as seeds to perform a crawl. Many other relevant metadata can also be collected in this phase, for instance university world ranking, the number of international students, university size and status (public/private). To perform this task, one can either use available scraping programs or write their own scripts, as did Ferraresi and Bernardini (2013) to retrieve the full list of English university homepages from Webometrics

Rankings (Ferraresi and Bernardini, 2013:55). As this project resorted to a different ranking – i.e. QS University Rankings – *Web Scraper*[6] was used to collect the list of seed URLs and relevant metadata. This tool navigates a site according to a precompiled site map and retrieves contents accordingly. The main advantage of this open-source software is that it simulates human behaviour and can thus download dynamic pages, as is the case of QS World University Rankings.

A set of metadata was defined at this point, in order to account for internal categorisation and to record contextual information. Some metadata are collected during the scraping process, while others can only be acquired during crawling (Section 3.3.2) or even afterwards, as is the case of genre (see Chapter 4). The full list of metadata comprises:

- webpage URL, i.e. the web address of each webpage included in the corpus. The best practice in SEO is to use words in URLs so that visitors can have a clear idea of what to expect on a webpage. This implies that words can be exploited to elicit webpage categories, such as topic or genre;
- university homepages in English, i.e. the homepage of each university website in its English version, which is needed to start the crawl (Section 3.3.2);
- level of crawling (from 0 to 2, where 0 is the homepage in English);
- university extended name and main domain, e.g."University_of_Bologna" and "unibo", which are useful to easily identify universities in the corpus;
- QS World University overall ranking and QS World University score associated with the number of international students;
- university ranking according to its presence on the web. This relates to the Webometrics Ranking since data were retrieved from existing sources created in the acWaC Project (Ferraresi and Bernardini, 2013);
- university status (public/private) and size (s/m/l/xl), as registered in the QS World University Ranking. This information may help to compare communication strategies adopted by different "types" of universities and give a better picture of the institutions involved in the study;

[6] A free Chrome browser extension, available at: https://chrome.google.com/webstore/detail/web-scraper/jnhgnonknehpejjnehehllkliplmbmhn?utm_source=chrome-app-launcher-info-dialog [last consulted on 15 December 2017].

Table 3.1 An example of URLs used as seeds to run the crawl together with some metadata.

| | | | | | |
|---|---|---|---|---|---|
| https://www.univie.ac.at/en/public | University_of_Vienna | univie | 153 | 93.4 | 118 |
| | xl | at | germanic | eu27 | elf |
| http://www.au.dk/en/public | Aarhus_University | au | 107 | 47.1 | 88 |
| | xl | at | germanic | eu27 | elf |
| http://www.aalto.fi/en/public | Aalto_University | aalto | 139 | 38.7 | 110 |
| | l | dk | germanic | eu27 | elf |

- country and family of the country official language (e.g. Germanic in Norway and Romance in Italy). This information can be used to group texts according to the native language family;
- country status (EU-27 or other);
- variety of English (either native or ELF);
- functional categorisation of each webpage (see Chapter 4 for a full description).

The scraping process generated a CSV file that was then processed and converted into a tab-separated file to be used as input data in the next stage. The input file is made of lines like those in Table 3.1.

### 3.3.2 Crawling

This is the core phase when collecting raw data for corpus building. The list of URLs pointing to English homepages is used to run a crawl of university websites. Crawlers are programs built to download text from the web by following hyperlinks from one page to another (Hundt et al., 2007). This crawler is set to download internal links exclusively, which means that hyperlinks pointing to external webpages (i.e. not belonging to a university website) are ignored. The crawl ran for about five days and retrieved 11 GB of text, including any internal link between the English version of university homepages and level two in the website structure. This does not ensure that all the pages retrieved are written in English, especially if websites are not available in at least two languages. Furthermore, raw data include HTML tags, duplicates and further noise that have to be removed in order to obtain a "clean" corpus.

Therefore, the subsequent step includes automatic procedures developed in Baroni and Bernardini (2006) to clean the corpus from HTML code and boilerplate (such as templates), as well as duplicate pages and pages that are not written in English.

### 3.3.3 Post-processing

Raw data collected from university websites need to be post-processed in order to remove HTML code and any other unwanted document. This third step uses the same tools used in the WaCky pipeline (Baroni and Bernardini, 2006) and includes boilerplate removal, de-duplication and language identification.

**Boilerplate and code removal**. HTML code and "boilerplate" (e.g. code and text needed to create webpage templates) should not be included in the final corpus as it may inflate counts during linguistic analysis. In order to discard unwanted code, the ratio of tokens to HTML tags is used to spot and remove tag-rich sections of pages that have to be stripped off. Conversely, portions of pages containing no tags signal coherent text that need to be included in the final corpus (refer to Baroni and Bernardini (2006) for further details).

**Language identification**. The strategy adopted to filter out non-English pages is based on the idea that any English text should contain a high number of English function words as compared to the total amount of words in the document. Therefore, all documents that do not meet this condition can be discarded from the corpus. Still, some pages may include contents in other languages as well – for instance if original texts and their translations are in the same page – provided that there is a sufficient proportion of English connected text.

**Duplicate and near duplicate removal**. This step consists in identifying and discarding perfect and near duplicates. There are dozens of reasons that cause duplicate content on the web and most of them are technical, for instance if there are more subdomains or there are different versions of the same page (printer-friendly pages, mobile versions and so on). While perfect duplicates are easily recognisable, near duplicates are more difficult to detect. Following the procedure used in Baroni and Bernardini (2006), near duplicates have been detected by comparing the number of shared n-grams for each pair of documents (see Baroni and Bernardini (2006) for further details).

This led to a corpus of 34,945 texts overall, about half belonging to the native English variety, and half to the ELF one (Table 3.2). After post-processing data, the corpus has to be annotated and indexed in order to be used for linguistic analysis. Before going on with these final stages, another major step was carried out, namely functional classification. The whole procedure and post-hoc evaluation is described in Chapter 4, while the next Section provides a brief account of the annotation and indexing procedure.

```
<text id="manchester12726" url="http://www.manchester.ac.uk/study/international/why-
manchester/" crawl_level="1" eng_status="nat" instruct="4" hardnews="3" legal="6" compuff="1"
academ="10" info="8" narrate="4" set="test" home="http://www.manchester.ac.uk"
name="The_University_of_Manchester" rankall="33" intscore="97.1" webpresence="353"
unistatus="public" size="xl" country="uk" country_langfam="germanic" country_status="eu27">
<s>
Why             WRB         why
Manchester      NP          Manchester
?               SENT        ?
</s>
<s>
International               JJ          international
students        NNS         student
play            VVP         play
a               DT          a
huge            JJ          huge
part            NN          part
at              IN          at
The             DT          the
University      NP          University
of              IN          of
Manchester      NP          Manchester
--              :           --
our             PP$         our
student         NN          student
population      NN          population
hails           VVZ         hail
from            IN          from
160             CD          @card@
countries       NNS         country
.               SENT        .
</s>
<s>
```

Fig. 3.1 Example of an annotated sentence in the corpus, together with the full set of metadata.

### 3.3.4 Annotation and indexing

After classifying texts according to Functional Text Dimensions (see Chapter 4), the corpus was POS-tagged, lemmatised and indexed, so as to be accessed through a concordancing tool. Part-of-speech tagging and lemmatization were performed using the TreeTagger tool.[7] Figure 3.1 illustrates an example of the annotation output, together with the whole set of metadata.

Indexing was carried out through a library provided in the NoSketch Engine project,[8] an open-source corpus management system, which also performs corpus building and indexing. NoSketch Engine is a limited version of the full-featured Sketch Engine service.[9] Both

[7] http://www.cis.uni-muenchen.de/ schmid/tools/TreeTagger/ [last consulted on 15 December 2017].
[8] https://nlp.fi.muni.cz/trac/noske [last consulted on 15 December 2017].
[9] https://www.sketchengine.co.uk/ [last consulted on 15 December 2017].

versions are suited to handle large corpora, and support the CQL query language to perform complex searches for POS-tags and regular expressions. Due to the uncertain status of copyright for the texts in the corpus, at the moment the corpus is password-protected but can be consulted through the NoSketch Engine platform hosted by the Department of Interpreting and Translation.[10]

The next Section provides a brief account of corpus composition.

## 3.4 Corpus composition

This section provides a description of internal corpus composition, with respect to design criteria introduced in Section 3.1.

The final corpus contains approximately 20M tokens and 35K texts (see Table 3.2). The number of texts is equally split between ELF and native English countries (17,383 texts in ELF vs. 17,562 texts in native English), even though there are 27 countries in the ELF subcorpus against 2 countries in the native English one. This might be due to the fact that the English language webpages found in English speaking countries are clearly many times more numerous than those found in ELF countries. As for the number of tokens, the native subcorpus is slightly larger than its ELF counterpart, which may suggest that (some) ELF texts are somewhat shorter.

As for the number of universities per country, a sample of 100 universities overall was selected in the first place, on a proportional basis according to the number of universities listed in QS World University Rankings (cf. Section 3.2). Approximately 10 universities overall were discarded during crawling and post-processing, either because their website homepage was not fetched or because the retrieved webpages were discarded during language identification processes. The final corpus includes 91 universities, 78 of which come from ELF countries whereas 13 come from European countries where English is a native language. A map is also provided in Figure 3.2, displaying all universities included in the corpus. From a geographical point of view, it is not surprising that most universities tend to coincide with capital cities, which host two or even three universities, while leaving other regions completely uncovered. Yet, geographical distribution within single countries is not a priority of this project, which aims at analysing English texts produced by leading universities in Europe. All of them are indeed ranked among the top positions worldwide.

In addition to the ELF vs. NAT perspective, the corpus can be split by language family too. Language families of each country's official language may serve to understand whether and

[10]The corpus is available at: https://corpora.dipintra.it/ [last consulted on 15 December 2017]. Permission can be requested from the author.


Fig. 3.2 Map of universities included in the corpus.

Table 3.2 Corpus statistics by English language variety (ELF and native English).

| | ELF | NAT | Total |
|---|---|---|---|
| Tokens | 9,375,739 | 11,813,692 | 21,189,431 |
| Texts | 17,383 | 17,562 | 34,945 |
| Universities | 78 | 13 | 91 |
| Countries | 27 | 2 | 29 |

how any difference or similarity in the use of English is related to the authors' native language. Table 3.3 offers detailed information on corpus composition, while Figure 3.3 illustrates the percentage of tokens in the corpus by language family. ELF content produced in a country whose official language is of Germanic origins account for 20% of the whole corpus, while textual content produced in a Romance-language country accounts for approximately 10% of the corpus. Uralic and Slavic languages are represented by 5% of corpus tokens each, while Hellenic (Greece) and Baltic (Lithuania and Latvia) regions include less then 1% of tokens in the corpus.

Table 3.3 Detailed corpus statistics. Native texts are highlighted in bold.

| **Language Family** | **Country** | **Status** | **N. of texts** | **N. of tokens** |
|---|---|---|---|---|
| Baltic | Lithuania | ELF | 39 | 36,552 |
| | Latvia | ELF | 161 | 72,568 |
| Germanic | Austria | ELF | 352 | 185,224 |
| | Germany | ELF | 2,674 | 1,269,884 |
| | Denmark | ELF | 1,059 | 779,139 |
| | Netherlands | ELF | 1,845 | 801,244 |
| | Norway | ELF | 657 | 283,059 |
| | Sweden | ELF | 941 | 680,928 |
| | **United Kingdom** | **NAT** | **13,773** | **9,069,383** |
| **Germanic-Celtic** | **Ireland** | **NAT** | **3,789** | **2,744,309** |
| Germanic-Romance | Belgium | ELF | 722 | 408,088 |
| Hellenic | Greece | ELF | 30 | 14,881 |
| Romance | Spain | ELF | 1,155 | 603,882 |
| | France | ELF | 1,258 | 633,523 |
| | Italy | ELF | 1,263 | 620,940 |
| | Portugal | ELF | 234 | 117,919 |
| | Romania | ELF | 111 | 58,915 |
| Romance-Germanic | Switzerland | ELF | 1,767 | 807,456 |
| Slavic | Belarus | ELF | 81 | 46,291 |
| | Czech Republic | ELF | 299 | 183,370 |
| | Poland | ELF | 96 | 63,443 |
| | Serbia | ELF | 83 | 40,606 |
| | Russia | ELF | 554 | 530,522 |
| | Slovenia | ELF | 123 | 95,309 |
| | Slovakia | ELF | 18 | 7,905 |
| | Ukraine | ELF | 44 | 30,632 |
| Uralic | Estonia | ELF | 324 | 176,162 |
| | Finland | ELF | 1,382 | 771,860 |
| | Hungary | ELF | 111 | 55,437 |

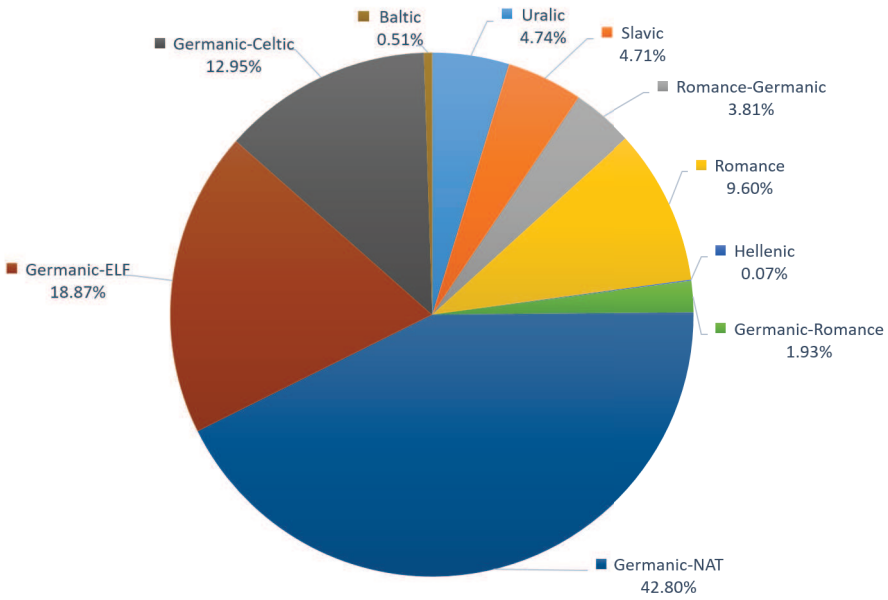


Fig. 3.3 N. of tokens (%) by language family.

The next Chapter describes automatic classification of texts according to Functional Text Dimensions (FTDs) and reports on the post-hoc evaluation that was carried out for texts scoring high on the promotional dimension.

# Chapter 4

# Automatic classification

## 4.1 Introduction

This Chapter deals with the classification of webpages in the corpus by measuring their linguistic distance according to pre-defined criteria that performed well in previous experiments (Forsyth and Sharoff, 2014). In particular, it aims at applying Sharoff's method for quantifying text similarity using human judgment as a reference standard (Sharoff, 2018). The classification is meant to increase the usability of the corpus, as well as to provide input for the analysis of promotional language described in Chapter 6.

Traditional approaches for document classification adopt either internal (linguistic) criteria or external (situational) criteria. As already discussed in Section 2.4.2, the former are associated with bottom-up classifications – e.g. the Multi-Dimensional analysis conducted by Biber (1988) – the latter with top-down procedures where texts are categorised according to non-linguistic parameters such as author, recipient, field, and genre category. This latter approach is frequently chosen within sociolinguistics and discourse/rhetorical studies (e.g. Swales (1990), see Biber et al. (2007) for a full account). Experiments conducted in Forsyth and Sharoff (2014) and Sharoff (2018) have moved into a new direction by adopting statistical measures to quantify linguistic (dis)similairty across documents and comparing the output with a text-external standard, i.e. human judgement. This approach is particularly relevant to this PhD project, which has a strong focus on readers' perception and on how language impacts upon readers' choices – rather than investigating how language reflects university power structures and their underlying ideologies. Therefore, the same methodology and criteria used in Sharoff (2018) can be applied to university webpages, so as to identify text types as perceived by website users. The main stages include:

1. development of a set of relevant descriptors that can represent the main text types of university websites (Section 4.2);
2. manual annotation of a random sample of academic pages serving as a training set for automatic classification (Section 4.3);
3. automatic classification of the remaining pages (test set) on each functional dimension (Section 4.4).

Each of these stages will be described in the next Sections.

## 4.2 Developing descriptors

The webpages in the corpus can express several functions at the same time. For example, typical *About us* pages include informative descriptions ("Description of a thing" according to the Web text classification scheme developed by Egbert et al. (2015)) as well as promotional materials ("Informational Persuasion"). In order to deal with such variation, the typology based on Functional Text Dimensions (FTD) (Forsyth and Sharoff, 2014) has been adapted to this specific case of academic webpages. The following dimensions are considered to be especially relevant for describing academic websites:

**A7, instruct** To what extent does the text aim at teaching the reader how something works?

**A8, hardnews** To what extent does the text appear to be an informative report of events recent at the time of writing?

**A9, legal** To what extent does the text lay down a contract or specify a set of regulations?

**A12, compuff** To what extent does the text promote a product or service?

**A14, academ** To what extent does the text serve as an example of academic research?

**A16, info** To what extent does the text provide information to define a topic?

**A21, narrate** To what extent does the text describe a chronologically ordered sequence of events?

This list of descriptors offers an approach to text classification in which the texts are described in terms of their similarity to prototype genres. For examples, **A7** (instruct) is strongly represented by a tutorial or a FAQ, whereas descriptor **A9** (legal) includes copyright

notices as the ideal prototype genre. Application of this procedure leads to a compact description of each text as scoring high or low on some of the dimensions. For example, some *About us* webpages are strictly informational (**A16**),[1] some are narrative (**A21**),[2] while others combine information with promotion.[3]

This framework is used to conduct a manual annotation of a sample of pages to be used as training set for automatic classification.

## 4.3 Manual annotation

A subset of about 900 webpages was randomly sampled from the main corpus. Sampling was done by selecting ten pages for each university randomly. To balance the lack of information required to perform a stratified sample and the need for a representative sample of most text types, URLs have been manually analysed to make sure that specific portions of the website did not dominate over other portions. If URLs were skewed towards a portion of a website (e.g. `www.bg.ac.rs/en/bodies/`), more pages were taken from other uncovered sections.

Each webpage has been annotated using a scale from 0 to 2, with 0 meaning that the descriptor is not present at all, 0.5 meaning that it is present to a small extent, 1 meaning that it is partly present and 2 meaning that the text is strongly characterised by a specific descriptor. This four-value scale has proven successful in a number of experiments (Forsyth and Sharoff, 2014) and was deemed to provide an acceptable trade-off between precision and confidence for annotation. In order to get cleaner text types for training purposes, pages containing two or more separate text types were split into different texts. On the other hand, proper hybrid pages, i.e. those fulfilling multiple functions simultaneously, were given a strong value in each applicable attribute. This resulted in a training corpus of 931 texts. Based on experience from several annotation experiments, this number was deemed sufficiently large to contain a representative picture of variation in academic webpages.

Due to limited resources, annotation was done by one annotator only. However, other studies which used the FTD annotation categories listed above demonstrated reasonable inter-annotator agreement levels, with Krippendorff's $\alpha$ ranging from 0.78 to 0.97 for different FTDs (Sharoff, 2018).

The annotation process produced a numeric data matrix in which each row corresponds to an observation and each column corresponds to a functional descriptor. Many texts score medium-to-high on several dimensions at the same time due to some degree of genre

[1] https://www.cam.ac.uk/public-engagement/about-us [last consulted on 15 December 2017].
[2] http://www.sci.u-szeged.hu/english/brief-history/about-us [last consulted on 15 December 2017].
[3] http://wwwf.imperial.ac.uk/business-school/ [last consulted on 15 December 2017].

hybridity. Legal and instructional texts tend to be more recognizable, whereas informative, promotional and narrative pages show a higher degree of overlapping. Texts dealing with academic research tend to score one or two points on the hardnews dimension as well, since they are often presented in the form of "news bites". The annotation matrix is converted into a set of positive and negative examples for each FTD, which are eventually used as a training set for experimenting automatic classification of the entire corpus. Table 4.1 provides excerpts taken from texts scoring high on single descriptors and used as positive examples, while Table 4.2 indicates the amount of positive examples for each FTD in the training set.

## 4.4 Automatic classification

Classification of texts according to their genres can be achieved by extracting a range of higher-level features, such as combinations of POS tags, parse trees or rhetorical relations (Santini, 2007). However, lower-level features based on character n-grams have been found to offer a surprisingly efficient method for detecting genres without requiring heavy linguistic resources (Kanaris and Stamatatos, 2007). In a comparative evaluation, their performance can exceed what is achieved by resource-heavier approaches. For example, pure n-grams can successfully generalise dates (.***day* for *yesterday, today, Friday*), which are typical in reporting, nominalisations (.***tion*) or passives (.***ed by*), which are typical of scientific discourse (Forsyth and Sharoff, 2014).

The frequencies of character n-grams can be directly used as features in algorithms of Machine Learning. Support Vector Machines (Smola and Schölkopf, 2004) or Relevance Vector Machines (Tipping, 2001) can be used to experiment with classification parameters. The advantage of RVM is the ability to produce a small number of Support Vectors, leading to better learning generalisation in the case of relatively sparse data – for instance, only 25 positive examples were identified for **A9** (legal texts). The task is to predict whether a webpage scores high in each FTD. The commonly used F1 measure is reported in Table 4.2 with cross-validation for detecting the FTDs.

After producing reliable classifiers for each dimension, these classifiers are then applied to the entire corpus of academic webpages, excluding texts that were used in the training set. Each classifier predicts a score ranging from 0 to 1 for each page in the test set. A crucial part here is to translate a numerical output into meaningful data. In other words, classifiers produce a score that has to be interpreted in order to establish which pages score on each dimension with minimal noise outside the training set. The closer to 1, the more likely that a file scores highly on that dimension, but each dimension may have its own threshold for

Table 4.1 Excerpts from positive examples for each FTD in the training set.

| FTD | Positive example |
|---|---|
| A7 – instruct | Enrollment procedure. Before departure Latin American students who wish to apply to the Inclinados hacia América Latina must submit a pre-enrollment application to the Italian diplomatic authority competent for their geographic region, specifying an interest to participate in the project. |
| A8 – hardnews | Deeyah Khan, artist and champion of women's rights, is awarded the University of Oslo's Human Rights Award. Deeyah Khan has shed an important light on women's rights and freedom of speech. As a young Norwegian-Pakistani musician in Norway she experienced being threatened to silence by conservative forces in the Pakistani environment and had to leave Norway at the age of 17. |
| A9 – legal | III. Requirements, assignment and admissions procedure. Art.2. Those students to which these regulations apply may not be first year students at their home university, nor join the first course that leads to no UAM qualification. Art.3. Students from overseas centres must have at least an intermediate level of Spanish in order to be able to study at UAM, except for those disciplines where the centre to which they wish to be assigned considers basic knowledge sufficient. |
| A12 – compuff | The University of Freiburg is one of only six universities in Germany to be distinguished in the "Excellent Teaching" competition organized in October 2009 by the Stifterverband and the Standing Conference of the Ministers of Education and Cultural Affairs.The reasons for the university's success in this competition are the quality of its existing course offerings and its overall concept for instructional development, "Windows for Higher Education." |
| A14 – academ | Trends in biodiversity dynamics are studied in a broader context of geological history, with a special emphasize on the Quaternary period. Phenomena such as speciation, hybridization, genome evolution, phenotypic plasticity, developmental processes, and changes of behavioural or morphological traits are investigated using modern - omics techniques in combination with morphological, ecological and behavioural approaches. |
| A16 – info | The Managing Board is an administrative body that primarily decides on economic matters and ensures the smooth material operations of the university.The Managing Board has a classification committee and may set up other committees and working bodies as required. The Board consists of nine members. |
| A21 – narrate | Geology was first introduced at Moscow State University at the beginning of the 19th century. In 1804 the Chair of Mineralogy and Rural Home Economics was established at the Department of Physical and Mathematical Sciences. Within the same year the Department of Natural History and the Mineralogy museum were founded. |

| | A7 | A8 | A9 | A12 | A14 | A16 | A21 |
|---|---|---|---|---|---|---|---|
| % in training set | 8.4 | 5.0 | 3.2 | 8.5 | 6.3 | 13.6 | 5.5 |
| F-measure | 0.95 | 0.92 | 0.96 | 0.85 | 0.93 | 0.79 | 0.94 |

Table 4.2 Manual annotation of the training set and F-measure.

establishing confidence. For example, the A7 classifier (instruct) assigned a 0.99 value to the text below, which is indeed instructing users on how to apply to university on paper:

> Please send an email to graduateadmissions@admin.cam.ac.uk requesting a copy of the paper application form. Please include a brief explanation of why a paper application form is required; in many cases the Graduate Admissions Office may be able to suggest a more preferable application method.[4]

However, there is no straightforward way of establishing which specific value differentiates between pages that highly represent that dimension and pages that do not include that dimension at all. There are two possibilities to perform this task. One can experiment with thresholds to achieve the desired precision – for instance through a post-hoc evaluation of texts – or fix an arbitrary threshold with no subjective interpretation at all. The former implies a second round of human evaluation, which may lead to a circular process, since automatic classification includes human ratings in the first place. The latter can be accomplished e.g. by choosing the top *-n* values as highly representative of any dimension. Although this second approach may not account for natural variation in the distribution of text types over university websites (e.g. it is likely that legal texts account for a very low proportion of pages as compared to instructional texts), it can be viewed as an unbiased way of establishing reliable thresholds. Therefore, the scores produced for each dimension have been divided into deciles where the first decile (1) identifies pages scoring highest on that dimension, whereas the last decile (10) identifies pages scoring lowest on that dimension. Each decile accounts for about 3,000 texts. Functional Text Dimensions (FTDs) and their corresponding values (translated into deciles) were encoded in the corpus as metadata (Figure 4.1).

Section 4.5 provides an exploratory post-hoc evaluation of top-ranked pages in the promotional dimension and a few examples of low-scoring pages.

[4]Full text available at: http://www.graduate.study.cam.ac.uk/applying-paper [last consulted on 15 December 2017].

| | |
|---|---|
| The Graduate School plays a key role in delivering the postgraduate student experience as well as with postgraduate education, policy and strategy development. The Graduate School enriches the postgraduate student experience by delivering a tailored programme of professional skills training which enhances the professional impact and helps to ensure personal ambitions are realised. | text.instruct 10<br>text.hardnews 8<br>text.legal 5<br>text.compuff 1<br>text.academ 9<br>text.info 2<br>text.narrate 9 |

Fig. 4.1 Example of classification output. An extract of a longer text is displayed.

## 4.5 Post-hoc evaluation

As described in Section 4.4, the automatic classifier produced a numerical value ranging from 0 to 1 for each page in the corpus. Values approximating 0 indicate a lower probability that a document scores highly on that dimension, while values approximating 1 indicate a high probability that a text fulfills a specific functional dimension. In order to determine whether the first decile is truly representative of promotional texts, top-ranked pages and low-ranked pages within the first decile are assessed. Both native English and ELF texts will be selected, although this stage does not aim at comparing the two varieties of English; rather, it provides examples of texts scoring high on the A12 dimension to assess the classification output. Pages that were manually coded in the training set are excluded from this analysis.

Table 4.3 provides examples of texts associated with higher values in the promotional dimension. Both native English texts, namely Examples (1) and (2) promote their respective university by showing their leadership in university rankings (*consistently rated among the best in the world*; *international leadership*; *ranked 7th overall among UK research universities*). Excerpt number (3) from Finland and excerpt number (4) from France do not refer specifically to international rankings, but they are clearly promoting excellence and international prestige (*internationally distinguished research centres*; *excellence in its training*; *extensive international coverage*).

In order to understand whether the first decile is fully representative of the promotional dimension, lower-ranked pages in the first decile are observed as well (Table 4.4). Attention was paid to select texts coming from different countries, even if no comparison is carried out at this stage. Examples (5) and (6) come from an Irish university, while Examples (7) and (8) have been produced in an ELF country (Greece and the Netherlands respectively). All the texts are at least partially fulfilling a promotional function. Text number (5) is quite similar to a tourist brochure which portrays the advantages of visiting Cork (*great for shoppers*; *many beautiful attractions*; *excellent bus service*). Text number (6) from University College Cork in Ireland reports the positive review obtained by an internationally renowned magazine. Excerpt number (7) comes from a Greek university and aims at persuading

Table 4.3 Extracts from top-ranked pages in the promotional dimension.

| | Text extract | Classification value |
|---|---|---|
| (1) | Consistently rated among the best in the world, the Faculty of Engineering provides international leadership in engineering research and education, and inspires its staff and students in a dynamic and stimulating environment. The Faculty of Engineering is one of three faculties within Imperial College London and is led by the Dean, Professor Jeff Magee FREng. | 1.150322 |
| (2) | In the past 50 years, we have earned an enviable reputation for the wide-reaching impact of our research. We were ranked 7th overall among UK research universities in the UK government's 2014 Research Excellence Framework (REF), with 87% of our research rated either "internationally excellent" or "world-leading". | 1.071497 |
| (3) | The research at the Faculty of Education focuses on the future educational challenges. Research at the Faculty concentrates around two nationally and internationally distinguished Research Centres. In addition, research is conducted on the two Departments of the Faculty. | 1.039180 |
| (4) | Centrale Paris is today known for: the excellence in its training: Centralien engineer, Master, Advanced Master, Doctorate; a pluridisciplinary research centre which has created numerous academic and industrial partnerships and collaborations both in France and internationally; very close links to the business world; extensive international coverage. | 0.951514 |

Table 4.4 Extracts from lower-ranked pages in the promotional dimension.

| **Text extract** | **Classification value** | |
|---|---|---|
| (5) | Cork is a friendly welcoming and negotiable city. It is small enough to get around on foot with many cinemas, restaurants and pubs (check out The Long Valley Bar around since 1842) and is great for shoppers. If you prefer the suburbs it has an excellent bus service (if you don't have a car) and there are many beautiful attractions to visit like Fota Wildlife Park and Blarney Castle to name just a few. | 0.405944 |
| (6) | The inaugural issue of the Global Entrepreneurship Network (GEN) Magazine distributed by the Financial Times worldwide highlighted Ireland and UCC as a hub for innovation, new ideas, start-ups and creativity. The GEN initiative is a new voice for promoting entrepreneurial growth and support start-ups around the world and featured a foreword by U.S. President Barack Obama. | 0.405939 |
| (7) | The scientific staff in the Schools, together with post-graduate researchers, apart from their teaching and related educational activities, conduct research work assisted by post-graduate students and a considerable number of external collaborators; the amount and the high standards of this research are proved by the numerous publications in International Scientific Journals and Proceedings of International Conferences as well as by the prominent place of NTUA among all Europeans Universities, due to the increasing number of research projects financed by the EU and other Greek and foreign organizations of the public and the private sector. | 0.405933 |
| (8) | Separately and collectively Media studies has as its objects of study and conceptual concerns a rich variety of cultural objects and practices, as seen through both disciplinary and cross-disciplinary lenses. [...] Media studies in Amsterdam is unique and well situated in its combination of disciplines, approaches and expertise to train students and researchers for intellectual work across the creative industries and beyond. | 0.405872 |

readers of the excellent research carried out at the university (*high standards of this research*; *numerous publications in international scientific journals*; *prominent place of NTUA among all European Universities*). Finally, example number (8) describes a course programme on Media Studies in Amsterdam and proclaims its uniqueness and its close link with the entrepreneurial world.

As these examples show, both top-ranked pages and low-ranked pages in the first decile are very likely to fulfil a promotional function in many ways. Some of them provide evidence of university excellence through international rankings, others list university or departmental strengths in the field of education and research, and positively describe their services or other student needs.

## 4.6 Summing-up

The classification of European university webpages serves the aim of increasing the usability of the corpus, as well as to provide input for the analysis of promotional language described in Chapter 6. The classification was conducted by measuring the linguistic distance between texts according to pre-defined criteria that performed well in previous experiments (Forsyth and Sharoff, 2014). In particular, classification applies Sharoff's method for quantifying text similarity using human judgment as a reference standard (Sharoff, 2018). The main stages were described in this Chapter, including the development of a set of relevant descriptors (Section 4.2), manual annotation of a random sample of academic pages serving as a training set (Section 4.3), and finally automatic classification of the remaining pages (test set) on each functional dimension (Section 4.4). Functional Text Dimensions (FTDs) and their corresponding values were encoded in the corpus as metadata (Figure 4.1).

We will now turn to Chapter 5 which presents the method used to search for evaluation features based on a survey of the relevant literature. A set of queries was developed specifically to carry a corpus-based analysis of evaluative language used in promotional texts within academic websites.

# Chapter 5

# Searching for promotional features

Section 2.3 of this thesis reviewed the most significant studies that have categorised and/or analysed linguistic features expressing evaluation in English. This section builds on that review to build a set of corpus queries aimed at detecting explicit evaluation features in the corpus. The initial selection of features was expanded by adding further evaluative elements expressing the same function. For instance, Biber (2006a) lists the major lexico-grammatical features used for the analysis of stance in the LGSWE corpus. In an effort to be as inclusive as possible, this list of stance markers was integrated with further instances of the same type, by selecting a greater range of nouns, adjectives and verbs that express the same function from Biber et al. (1999).

In this sense, there are some very straightforward search queries that reflect a specific part of speech – e.g. personal pronouns – and more complex structures combining grammatical patterns and semantic categories – e.g. verbs expressing mental processes. The latter can be expanded by collecting similar lexical elements from various sources (previous studies, dictionaries, grammars and the like).

Search queries constructed for this thesis are divided into three groups that reflect different approaches to evaluation, namely grammatical forms, lexico-grammatical forms and attitudinal lexis. The first group includes any feature of English marking stance because of its grammatical function exclusively (i.e. pronouns, imperatives, questions and modal verbs). The second group of search queries includes features that are classified as stance markers because of their grammatical function and their lexical meaning (e.g. adverbials expressing stance). The third group includes attitudinal lexis exclusively, without any regard to their grammatical function.

Note that linguistic features are likely to recur across these three groups. For instance, *may* and *might* are included in the first group since the grammatical category of modal verbs is widely considered a feature of evaluation. In addition, and from a lexico-grammatical

perspective, *may* and *might* can act as hedging devices and serve the same purpose as uncertainty devices (Section 2.3). This is not seen as a duplication of information, but as a way of accounting for the different functions that lexico-grammatical items may have. All queries are constructed using the Corpus Query Language, a widely used query syntax to search annotated corpora (Evert, 2016).

## 5.1 Grammatical forms

This set of queries includes any grammatical device that expresses evaluation according to the consulted literature (cf. Section 2.3). This group includes personal pronouns, imperatives, questions and modal verbs. Queries can be easily developed by exploiting corpus part-of-speech tagging.[1] Contracted and negative forms are also included in the POS-tag search.

### 5.1.1 Pronouns

As will be remembered from Section 2.3.3, Hyland argues that self-mention (including first-person pronouns and possessive adjectives) is a key strategy to present a particular authorial identity (Hyland, 2008:8), while reader pronouns are the most overt expression of engagement (Hyland, 2008:9).

Two queries were built that reflect these grammatical forms:

1. Personal, reflexive, possessive pronouns referring to the first person

   ```
   [word="I|me|myself|mine|my|we|us|ourselves|our|ours"&tag="PP.*"]
   ```

2. Personal, reflexive, possessive pronouns referring to the second person

   ```
   [word="you|your|yours|yourself|yourselves"&tag="PP.*"]
   ```

### 5.1.2 Imperatives

According to Hyland (2008), imperatives engage readers in textual, physical or cognitive acts, thus representing a crucial linguistic feature of stance. Imperatives and obligation modals fall into Hyland's "Directives" category. Since the POS-tagger used does not provide a specific label for imperative forms, the following is a "hand-made" query, through which the typical structure of imperative forms is approximated:

```
[tag="SENT"][(tag="VV|VB") & (!lemma="do")]
```

[1] The tagset summary is available here: http://www.terminologia.it/index.php/english-tagset [last consulted on 15 December 2017].

### 5.1.3 Questions

Similarly to reader mention and directives, questions are an element of engagement, leading readers to the writer's viewpoint (Hyland, 2008). Below is a very straightforward query to retrieve any question in the corpus:

[word="\?"]

### 5.1.4 Modal and semi-modal verbs

Modal verbs are widely considered as stance markers (e.g. Biber (2006a); Hyland (2008)). A straightforward query allowing for the retrieval of any modal verb is the following one:

```
[tag="MD"]
```

However, in order to increase search precision, one can also split modals into their typical semantic categories (e.g. ability, necessity, prediction). Semi-modals are also included in this list. Specifying the "modal" tag (i.e. MD) is needed to exclude instances of *can*, *may*, *must*, *will* used as nouns.

1. **Permission, possibility, ability**

   ```
   [lemma="can|could|may|might" & tag="MD"]
   ```

2. **Necessity, obligation**

   Four queries were constructed for this category. However, retrieved instances can be counted altogether.

   ```
   [lemma="must|should|ought" & tag="MD"]

   [lemma="have|need"] [tag="PP.*"]0,1 "to" [tag="V.*"]

   [tag="VB.*|VH.*"][word="got|better"]"to"

   [tag="VB.*"][word="supposed"]"to"
   ```

3. **Volition, prediction**

   ```
   [lemma="would|shall|will" & tag="MD"]

   [tag="VB.*"][word="going"]"to"
   ```
   [2]

[2] "Be going to" is included in the list of modals expressing volition and prediction, following Biber (2006a). However, according to Biber et al. (1999:490) "be going to" is a semi-modal verb marking time distinctions rather than stance.

The next section provides search queries used to detect lexico-grammatical forms of stance.

## 5.2 Lexico-grammatical forms

Lexico-grammatical forms are mostly grammatical patterns that include lexical units expressing evaluation. This set includes certainty/uncertainty adverbials, amplifiers or intensifiers, attitude adverbs, style adverbs and stance complement clauses. The latter are further subdivided into different classes, reflecting Biber's framework (Biber, 2006a).

### 5.2.1 (Un)certainty markers

This section includes any grammatical device expressing certainty and uncertainty stance markers. In particular, certainty and actuality devices include:

- boosters, defined as features that allow writers "to express certainty in what they say" (Hyland, 2008:7)
- epistemic certainty adverbials, which "show the speaker's certainty about a proposition" (Biber et al., 1999:854)
- actuality and reality adverbials, which "comment on the status of the proposition as real-life fact" (Biber et al., 1999:854).

This group also includes adverbials commenting on the source of knowledge, especially those that give evidence to a proposition, e.g. "as X notes".

Several queries are needed to search for items in this group. In particular, the adverbial phrase "of course" has to be searched for separately in order to exclude instances where "course" is used as a noun modifier (e.g. *of course completion*, *of course agenda*, *of course applicants* and the like). Occurrences are added up at a later stage.

```
"as""a""matter""of""fact"|"actually"|"apparently"|"certainly"|"clearly"
|"decidedly"|"definitely"|"evidently"|"for""sure"|"infact"|"incontrover
tibly"|"indeed"|"no""doubt"|"obviously"|"really"|"sure"|"surely"|"undou
btedly"|"likely"
```

```
"of" "course" [!tag="N.*"]
```

```
"according""to"|"reportedly"|"reputedly"|"as"[]0,1[(lemma="note|state")&
(tag="V.*")]
```

Uncertainty and likelihood markers include: hedges (Hyland, 2008), epistemic likelihood adverbs (Biber, 2006a), stance adverbials marking doubt and imprecision (Biber et al., 1999) and approximators, i.e. hedging forms used with numbers, measurements and quantities (Biber et al., 1999:557). Degree adverbs, such as *slightly*, *somewhat* and *rather* are also counted among hedging devices since they can diminish precision or accuracy of what is being reported (Biber et al., 1999:555).

```
"apparently"|"approximately"|"arguably"|"may"|"maybe"|"might"|"nearly"|
"perhaps"|"possible"|"possibly"|"probably"|"rather"|"roughly"|"slightly
"|"somewhat"|"kind""of"|"sort""of"|"suggest.""that"|"it""seems""that"
```

### 5.2.2 Amplifiers/intensifiers

Even though amplifiers (e.g. *very*, *so*, *really*) are not included among stance devices in Biber's 2006 paper, Biber et al. (1999:970) considers adverbs of degree as premodifying stance adverbs ("I'm **so** happy for you. Honestly, I'm **really** happy for you"). Furthermore, Partington (1993:178) argues that intensification "is a vehicle for impressing, praising, persuading, insulting, and generally influencing the listener's reception of the message". The list of amplifiers is taken from Biber (1988) and Biber et al. (1999). Since not all the lexical items follow the same structure or belong to the same POS-tag, several queries are needed to search for this feature.

```
[word="absolutely|altogether|completely|definitely|enormously|entirely|
extremely|fully|greatly|highly|intensely|perfectly|strongly|thoroughly|
totally|utterly|very"&tag="RB"]
[word="just|more|most|really"&tag="RB"]

"such" "a"|"a" "lot"

[word="real|so"] [tag="JJ"]

[(lemma="do")&(tag="VV|VVD")][tag="VV"]
```

### 5.2.3 Attitude markers

Attitude markers (Hyland and Tse, 2004) and attitude adverbs (Biber, 2006a) are counted among stance or involvement devices expressing writer's attitude towards a proposition. As Biber states, "attitude adverbs convey an evaluation or assessment of expectations" (Biber,

2006a:100). Due to the various grammatical forms involved, three search queries are needed to retrieve this category of features:

```
"amazingly|appropriately|astonishingly|conveniently|curiously|disturbingly|fortunately|hopefully|importantly|inevitably|ironically|sadly|sensibly|surprisingly|unfortunately|wisely"
```

```
"as"[tag="PP"]0,1[tag="MD"]0,1[tag="VB"]0,1[word="guess|think|expect(ed)?"]
```

```
"to" [tag="PP.*"] "surprise"
```

### 5.2.4 Markers of perspective

This category is made of all the adverbs that tend to either limit or generalise the information given. It includes style adverbials (Biber, 2006a) such as *frankly* and *generally*, which comment on the way of conveying a message. Two queries are needed to retrieve this stance feature, due to the variety of grammatical forms involved:

```
"generally"|"largely"|"mainly"|"typically"|"in""general"|"in""most""cases"|"frankly"|"honestly"|"literally"|"personally"|"simply"
```

```
[word="from|in"] [tag="PP.*"] [lemma="perspective|opinion|view"]
```

### 5.2.5 Stance complement clauses

According to Biber (2006a), grammatical stance structures with first-person pronouns are the most overt expressions of stance, combined with a stance verb, adjective or noun. Examples of stance verbs, adjectives and nouns controlling a complement clause are taken from Biber (2006a) and Biber et al. (1999). Attention was also paid to include the different forms of English spelling, especially the -ize/-ization endings.

1. **Stance verbs + that-clauses**

   ```
   [lemma="conclude|determine|know|recognise|recognize|assume|feel|argue|claim|report|say|believe|doubt|guess|think|expect|hope|worry"&tag="V.*"][word="that"&tag="IN"]
   ```

   The above stance devices can be divided into classes based on their meaning (epistemic, attitude and speech or other communication verbs):

(a) **Epistemic certainty verbs + that-clauses**

```
[lemma="conclude|determine|know|recognise|recognize"&tag="V.*"
][word="that"&tag="IN"]
```

(b) **Epistemic likelihood verbs + that-clauses**[3]

```
[lemma="assume|believe|doubt|feel|guess|think"&tag="V.*"][word
="that"&tag="IN"]
```

(c) **Attitude verbs + that-clauses**

```
[lemma="expect|hope|worry"&tag="V.*"][word="that"&tag="IN"]
```

(d) **Communication verbs + that-clauses**

```
[lemma="argue|claim|report|say"&tag="V.*"][word="that"&tag="IN
"]
```

2. **Stance verbs + to-clauses**

```
[lemma="appear|happen|seem|tend|believe|consider|love|hope|intend|n
eed|want|wish|allow|attempt|help|try|advise|remind|request"&tag="V.
*"]"to"[!tag="DT|N.*|JJ|PP.*"]

"would" "like" "to"
```

The above stance structures can be further classified according to their meaning. Stance categories include:

(a) **Probability verbs + to-clauses**

```
[lemma="appear|happen|seem|tend"&tag="V.*"]"to"[!tag="DT|N.*|J
J|PP.*"]
```

(b) **Mental verbs + to-clauses**

```
[lemma="believe|consider|like|love"&tag="V.*"]"to"[!tag="DT|N.*
|JJ|PP.*"]
```

(c) **Desire verbs + to-clauses**

```
[lemma="hope|intend|need|want|wish"&tag="V.*"]"to"[!tag="DT|N.*
|JJ|PP.*"]

"would" "like" "to"
```
[4]

[3] *Assume* and *feel* were added to this category because Biber et al. (1999) lists them among "mental verbs" used to express stance, similarly to the verb "to think".

[4] Note that *would* is also included in modal verbs.

(d) **Effort verbs + to-clauses**

```
[lemma="allow|attempt|help|try"&tag="V.*"]"to"[!tag="DT|N.*|JJ
|PP.*"]
```

(e) **Communication verbs + to-clauses**

```
[lemma="advise|remind|request"&tag="V.*"]"to"[!tag="DT|N.*|JJ|
PP.*"]
```

3. **Stance adjectives + that-clauses**

As Biber et al. (1999:671) explains, stance adjectives that control a *that*-complement clause fall into three semantic domains: degrees of certainty (*certain*, *confident*, *evident*), affective states (*annoyed*, *glad*, *sad*) and evaluation of events and situations (*appropriate*, *odd*, *good*, *important*, *advisable*). The list of adjectives provided does not differentiate between adjectives that control post-predicate that-clauses (e.g. *I am glad that*) and adjectives that occur more frequently with extraposed that-clauses (e.g. *it is nice that*). The full list of stance adjectives controlling that-clauses is taken from Biber et al. (1999:671-675).

```
[word="acceptable|accepted|adamant|advisable|alarmed|amazed|amazing
|annoyed|annoying|anomalous|apparent|appropriate|astonishing|aware|
awful|awkward|careful|certain|clear|concerned|conceivable|confident
|convinced|correct|critical|crucial|curious|desirable|disappointed|
disappointing|dissatisfied|doubtful|dreadful|encouraged|essential|e
vident|extraordinary|fair|false|fortunate|frightened|funny|glad|goo
d|great|hopeful|horrible|imperative|important|incidental|inconceiva
ble|incredible|inevitable|interesting|ironic|irritating|likely|luck
y|natural|neat|necessary|notable|noteworthy|noticeable|obligatory|o
bvious|odd|okay|peculiar|pleasant|pleased|positive|possible|prefera
ble|probable|proud|relieved|right|sad|satisfied|sensible|shocked|so
rry|strange|sufficient|sure|surprised|surprising|thankful|tough|tra
gic|true|typical|unacceptable|unaware|understandable|unfair|unfortu
nate|unlikely|unlucky|unpleasant|untypical|unusual|vital|wonderful|
worried"&tag="JJ"][word="that"&tag="IN"]
```

The above stance adjectives can be classified semantically. Here are the queries created for each semantic category:

(a) **Ease adjectives + that-clauses**

```
[word="awkward|pleasant|tough|unpleasant"&tag="JJ"][word="that
"&tag="IN"]
```

(b) **Epistemic certainty adjectives + that-clauses**

```
[word="certain|clear|obvious|true|accepted|apparent|confident|
convinced|correct|doubtful|evident|false|inevitable|positive|r
ight"&tag="JJ"][word="that"&tag="IN"]
```

(c) **Epistemic likelihood adjectives + that-clauses**

```
[word="likely|possible|probable|unlikely"&tag="JJ"][word="that
"&tag="IN"]
```

(d) **Epistemic evaluation adjectives + that-clauses**

```
[word="acceptable|adamant|advisable|alarmed|amazed|amazing|ann
oyed|annoying|anomalous|appropriate|astonishing|aware|awful|ca
reful|conceivable|concerned|critical|crucial|curious|desirable
|disappointed|disappointing|dissatisfied|dreadful|encouraged|e
ssential|extraordinary|fair|fortunate|funny|good|great|hopeful
|horrible|imperative|important|incidental|inconceivable|incred
ible|interesting|ironic|irritating|lucky|natural|neat|necessar
y|notable|noteworthy|noticeable|obligatory|odd|okay|peculiar|p
leased|preferable|proud|relieved|sad|satisfied|sensible|shocke
d|sorry|strange|sufficient|sure|surprised|surprising|thankful|
tragic|typical|unacceptable|unaware|understandable|unfair|unfo
rtunate|unlucky|untypical|unusual|vital|wonderful|worried"&tag
="JJ"][word="that"&tag="IN"]
```

4. **Stance adjectives + to-clauses**

   Considering that the POS-tagger used does not include a tag for "to" used to introduce an infinitival clause, and in order to ensure a high level of precision, this query excludes any occurrences of *to* followed by a noun, pronoun, adjective or determiner. Also note that there might be a certain degree of repetition in the grammatical or lexical features included in this category, with respect to other queries (e.g. *able to* is counted among both modal verbs and stance adjectives + to-clauses).

   ```
   [word="able|acceptable|afraid|amazed|angry|annoyed|anxious|apt|asha
   med|bad|best|better|bound|brave|careless|certain|competent|convenie
   nt|crazy|desirable|determined|difficult|disgusted|disposed|doomed|d
   ue|eager|easy|eligible|embarrassed|essential|expensive|fit|free|fur
   ```

```
ious|glad|good|grateful|greedy|guaranteed|happy|hard|hesitant|impat
ient|important|impossible|inclined|indignant|keen|liable|likely|luc
ky|mad|necessary|nervous|nice|permissible|perturbed|pleased|possibl
e|prone|proud|puzzled|reasonable|relieved|reluctant|right|silly|sma
rt|sorry|sure|surprised|unable|unlikely|unwise|useless|willing|wise
|wonderful|worried|worse|wrong"&tag="JJ"]"to"[!tag="DT|N.*|JJ|PP.*"
]
```

As for previous lexico-grammatical stance structures, this feature of stance can be split into a set of more specific semantic categories:

(a) **Ability adjectives + to-clauses**

```
[word="able|anxious|bound|competent|desirable|determined|dispo
sed|doomed|eager|eligible|fit|greedy|hesitant|inclined|keen|re
luctant|unable|willing"&tag="JJ"]"to"[!tag="DT|N.*|JJ|PP.*"]
```

(b) **Attitude adjectives + to-clauses**

```
[word="afraid|amazed|angry|annoyed|ashamed|disgusted|embarrass
ed|free|furious|glad|grateful|happy|impatient|indignant|nervou
s|perturbed|pleased|proud|puzzled|relieved|sorry|surprised|wor
ried"&tag="JJ"]"to"[!tag="DT|N.*|JJ|PP.*"]
```

(c) **Certainty adjectives + to-clauses**

```
[word="apt|due|guaranteed|liable|prone"&tag="JJ"]"to"[!tag="DT
|N.*|JJ|PP.*"]
```

(d) **Ease adjectives + to-clauses**

```
[word="difficult|easy|hard"&tag="JJ"]"to"[!tag="DT|N.*|JJ|PP.*"
]
```

(e) **Epistemic adjectives + to-clauses**

```
[word="certain|impossible|likely|possible|reasonable|sure|unli
kely"&tag="JJ"]"to"[!tag="DT|N.*|JJ|PP.*"]
```

(f) **Evaluative adjectives + to-clauses**

```
[word="acceptable|bad|best|better|brave|careless|convenient|cr
azy|essential|expensive|good|important|lucky|mad|necessary|nic
e|permissible|right|silly|smart|unwise|useless|wise|wonderful|
worse|wrong"&tag="JJ"]"to"[!tag="DT|N.*|JJ|PP.*"]
```

5. **Stance nouns + that-clauses**

According to Biber (2006a), nouns controlling a that-clause are less common than verbs or adjectives, and are restricted primarily to academic registers like textbooks and classrooms. A search for lemmas has been developed in this case, in order to fetch inflected forms.

```
[lemma="hope|view|conclusion|fact|observation|comment|proposal|report|assumption|claim|hypothesis|notion"&tag="NN.*"][word="that"&tag="IN"]
```

The above query can be split into four semantic categories, based on the meaning of nouns (attitude, epistemic certainty, communication and likelihood nouns):

(a) **Attitude nouns + that-clauses**

```
[lemma="hope|view"&tag="NN.*"][word="that"&tag="IN"]
```

(b) **Epistemic certainty nouns + that-clauses**

```
[lemma="conclusion|fact|observation"&tag="NN.*"][word="that"&tag="IN"]
```

(c) **Communication nouns + that-clauses**

```
[lemma="comment|proposal|report"&tag="NN.*"][word="that"&tag="IN"]
```

(d) **Likelihood nouns + that-clauses**

```
[lemma="assumption|claim|hypothesis|notion"&tag="NN.*"][word="that"&tag="IN"]
```

6. **Stance nouns + to-clauses**

```
[lemma="assumption|claim|hypothesis|notion"&tag="NN.*"]"to"[!tag="DT|N.*|JJ|PP.*"]
```

In the next Section we turn to a set of queries representing lexical forms of evaluation.

## 5.3 Attitudinal lexis

This section presents queries reflecting English words that have some type of evaluative meaning, according to previous studies. The taxonomy by Martin and White (2005) was used as a basis to build search queries in this Section. Lexical units are grouped into their three main emotive dimensions, namely ‘affect’, ‘judgment’ and ‘appreciation’ (see Section

2.3.2). Although this is a lexis-oriented classification, authors state that 'attitude realisations diversify across a range of grammatical structures' (Martin and White, 2005:45). Therefore, in order to be as inclusive as possible, lemma queries were used, aiming at including different inflected forms. As the authors themselves admit, the list of lexical items they created is by no means exhaustive, though they made an effort to include a wide set of meanings from Roget's Thesaurus (Martin and White, 2005:50).[5]

1. **Affect: (un)happiness**

   ```
   [lemma="abhor|abuse|adore|broken-hearted|buoyant|cheer|cheerful|cheerless|chuckle|cry|dejected|depressed|despondent|dislike|downcast|dreary|embrace|gloomy|grief-stricken|happiness|happy|hate|heart-broken|heavyhearted|hug|joyless|jubilant|laugh|like|love|melancholy|miserable|rejoice|revile|sad|sorrowful|tearful|unhappiness|unhappy|wail|weepy|whimper|woebegone"]

   "in""tears"|"sick""at""heart"|"fond""of"
   ```

2. **Affect: (dis)satisfaction**

   ```
   [lemma="absorbed|angry|attentive|bored|busy|castigate|caution|charmed|chuffed|compliment|cross|displeasure|engrossed|fidget|furious|impressed|industrious|involved|jaded|pleased|pleasure|reward|satisfaction|satisfied|scold|stale|thrilled|yawn"]

   "fed""up""with"|"sick""of"|"tune""out"
   ```

3. **Affect: (un)security**

   ```
   [lemma="anxious|assured|astonished|comfortable|commit|confidence|confident|entrust|faint|freaked|jolted|restless|security|staggered|startled|surprised|together|trust|trusting|twitching|uneasy"]
   ```

4. **Appreciation: (un)balance**

   ```
   [lemma="amorphous|balance|considered|consistent|contradictory|curvaceous|discordant|disorganised|distorted|flawed|harmonious|irregular|logical|proportioned|shapeless|shapely|symmetrical|unbalanced|uneven|unified|willowly"]
   ```

[5] Metaphorical expressions of stance (e.g. cut-up), general or context-dependent words (e.g. down, in, low, flat, cross, okay) and paralinguistic expressions of stance (e.g. shake hands, pat on the back) were removed from the list of attitudinal lexis, due to their high dependance on context for determining their attitudinal value.

5. **Appreciation: complexity**

   ```
   [lemma="arcane|clear|detailed|elegant|extravagant|intricate|lucid|monolithic|ornate|plain|precise|pure|rich|simple|simplistic|unclear|woolly"]
   ```

6. **Appreciation: impact**

   ```
   [lemma="arresting|ascetic|boring|captivating|dramatic|dry|dull|engaging|exciting|fascinating|intense|lively|monotonous|notable|pedestrian|predictable|remarkable|sensational|tedious|uninviting|unremarkable"]
   ```

7. **Appreciation: quality**

   ```
   [lemma="appealing|bad|beautiful|enchanting|fine|good|grotesque|lovely|nasty|off-putting|plain|repulsive|revolting|splendid|ugly|welcome"]
   ```

8. **Appreciation: valuation**

   ```
   [lemma="appropriate|authentic|common|conventional|creative|dated|deep|derivative|effective|everyday|exceptional|fake|genuine|helpful|ineffective|inimitable|innovative|insignificant|landmark|longawaited|original|overdue|penetrating|priceless|pricey|profound|prosaic|real|reductive|shallow|shoddy|timely|unique|untimely|useless|valuable|worthless|worthwhile|write-off"]
   ```

9. **Judgement: esteem: capacity**

   ```
   [lemma="accomplished|adult|balanced|childish|clever|competent|crippled|dreary|droll|dull|educated|experienced|expert|fit|flaky|foolish|gifted|grave|healthy|helpless|humorous|ignorant|illiterate|immature|incompetent|inexpert|insane|insightful|learned|literate|mature|mild|naive|powerful|productive|robust|sane|sensible|shrewd|sick|slow|sound|successful|thick|unaccomplished|uneducated|unproductive|unsound|unsuccessful|vigorous|weak|whimpy|witty"]
   ```

10. **Judgment: esteem: normality**

    ```
    [lemma="celebrated|cool|dated|eccentric|familiar|fortunate|hapless|lucky|natural|normal|obscure|odd|peculiar|predictable|stable|unlucky|unpredictable"]
    ```

11. **Judgement: esteem: tenacity**

    [lemma="accommodating|adaptable|brave|careful|cautious|constant|dependable|despondent|disloyal|distracted|faithful|flexible|hasty|impatient|impetuous|inconstant|loyal|meticulous|obstinate|patient|persevering|plucky|rash|reckless|reliable|resolute|stubborn|thorough|timid|tireless|undependable|unfaithful|unreliable|wary|weak|wilful"]

12. **Judgement: sanction: propriety**

    [lemma="abiding|altruistic|arrogant|avaricious|bad|caring|charitable|corrupt|cruel|discourteous|ethical|fair|generous|good|greedy|humble|immoral|insensitive|irreverent|just|kind|mean|modest|moral|polite|respectful|reverent|rude|selfish|sensitive|snobby|unassuming|unfair|unjust|vain"]

13. **Judgement: sanction: veracity**

    [lemma="blunt|candid|credible|deceitful|deceptive|devious|direct|discrete|dishonest|frank|honest|lying|manipulative|tactful|truthful"]

## 5.4 Summing-up

This Chapter has presented the process through which linguistic features of evalution were searched for in the literature. It grouped them according to linguistic/functional categories, namely grammatical forms, lexico-grammatical forms and attitudinal lexis. A set of queries were constructed to search the corpus for evaluative features. Queries are not meant to be exhaustive but an attempt was done to strike a balance between precision and recall. Furthermore, queries were checked and refined by trial and error. Getting an exhaustive list of queries is however a minor issue, consdiering that the two varieties of English will be compared using the same set of queries.

The next Chapter will present the corpus-based analysis carried out in the promotional native English and ELF subcorpora.

# Chapter 6

# Evaluation in university webpages

Chapter 6 presents research results and is structured into three Sections. Section 6.1 analyses promotional webpages from the point of view of genre, and tries to determine whether sampled pages match a typical promotional genre or text type. Section 6.2 focuses on the same promotional subset and examines it from an internal perspective. Section 6.3 discusses the main findings, attempting to provide a global view on the similarities and differences in the production of promotional strategies within native English and ELF academic websites.

## 6.1 Assessing genres in the promotional dimension

This Section aims at analysing pages scoring high on the A12 dimension, which represents promotional texts. In particular, it aims at assessing documents ranked in the first decile (see Section 4.4) and determine whether they match a (series of) typical promotional genre(s) or text type(s). This stage serves two purposes: on the one hand, it provides further insights into the relation between internal and external criteria for classifying texts, by answering the question of whether texts classified automatically as highly promotional can be linked to socially recognised promotional genres as well. On the other hand, it assesses whether there are any common promotional features that can be used to write promotional genres within university websites, as well as any differences between native English and ELF texts in terms of the genres used for self-promotion. Considering that automatic classification did not take into account the distinction between English varieties, native English and ELF texts can be compared to determine whether their promotional function is fulfilled through similar or different genres and text types.

Since each decile is composed of approximately 3,000 texts (including native English and ELF pages), it is not possible to check every webpage manually. Therefore, a random selection of 50 pages for each variety of English was performed in the first decile of the

Table 6.1 First decile in the promotional FTD.

| | N. of texts | N. of tokens |
|---|---|---|
| **NAT** | 1,658 | 550,329 |
| **ELF** | 1,832 | 587,347 |
| **Total** | 3,490 | 1,137,676 |

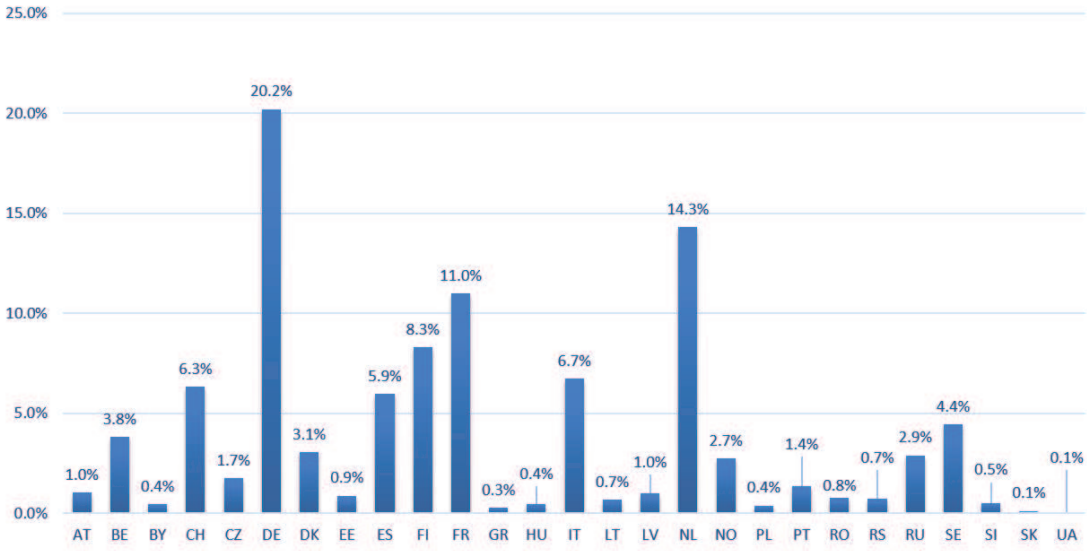


Fig. 6.1 Percentage of ELF texts in the first decile of the promotional dimension: breakdown by country.

promotional dimension, in order to assess whether these pages belong to a promotional academic genre. The first decile includes 3,490 texts and 1,137,676 tokens overall. In quantitative terms, there is not much difference between the number of texts and tokens in the two varieties of English (Table 6.1). A breakdown by country is also available in Figure 6.1 (ELF countries) and Figure 6.2 (native English countries), which show the percentage of texts included in the first decile by country. Note that each country in the corpus is represented in this set of promotional subset, although four countries (Germany, the Netherlands, France and Finland) account for the largest portion of ELF texts in the first decile (i.e. 53% of ELF documents).

The analysis was conducted by opening each URL using a web browser and checking whether it could be referred to a specific genre or text type. Each webpage was assigned to a genre based on both external criteria, such as the page title and the URL, and linguistic criteria, such as content and register (Section 2.4). As Figure 6.3 illustrates, many native English and ELF documents randomly selected belong to a specific institutional-academic genre,

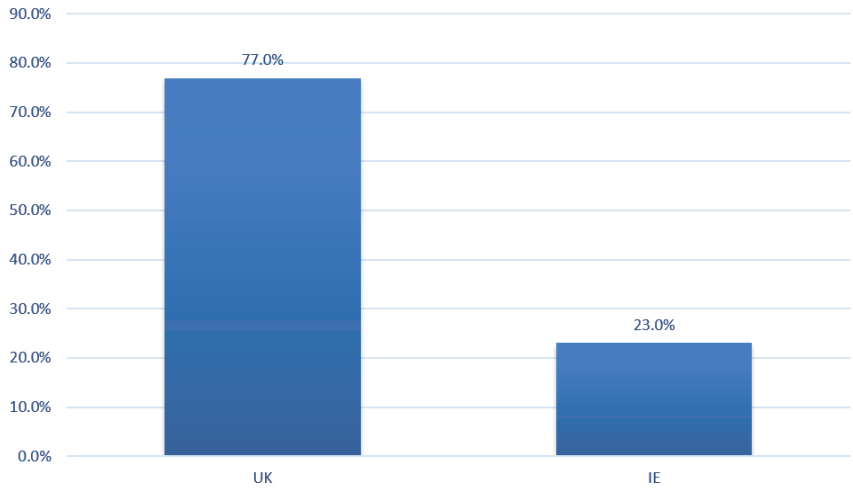


Fig. 6.2 Percentage of native English texts in the first decile of the promotional dimension: breakdown by country.

including Homepages, About us, Why choose us, news, course/programme descriptions, staff profiles. Although not belonging to a structured and widely recognised *genre*, there are further promotional webpages related to the description of the university (university bodies, facilities, services) and others regarding description of third mission activities (description of business partnerships, fundraising, social responsibility, university strategy and mission). Let us turn to a brief overview of these text sets, with examples from the corpus (native English examples will be provided in the first place while ELF examples will follow).

Homepages are used to facilitate users' navigation and guide them through the whole website. As occurs with corporate websites, university websites too can have multiple homepages. There is a general homepage referring to the whole university and there might be further department/school/faculty homepages, or homepages created of specific research centres. The example below shows an extract from the homepage of a centre within a school:

> Welcome to UCD Centre for Safety and Health at Work. The UCD Centre for Safety and Health at Work (CSHW) is a centre of teaching and research activity within the School of Public Health, Physiotherapy and Sports Science. The aim of the Centre is to strengthen our proven record of success in promoting the highest standards of Occupational Safety and Health through excellence in Teaching, Research and Administration.[1]

*About us* pages in the native context provide a brief presentation of the university, similarly to what the same genre accomplishes within other contexts, for instance in corporate websites

[1] http://www.ucd.ie/cshw/ [last consulted on 15 December 2017].

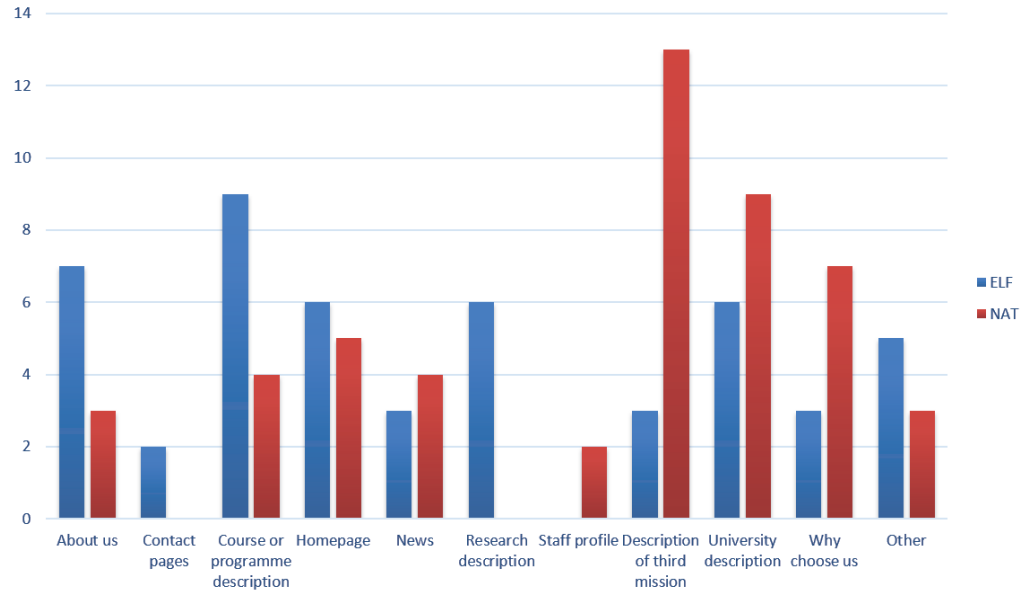


Fig. 6.3 Distribution of ELF and native English texts by genre (random subset).

(Casañ Pitarch, 2015). The text extract below provides university 'facts and figures' in the form of a list [emphasis as in the original]:

> Imperial is home to **16,882 students** and **7,547 staff**. Over **6,700 degrees** are awarded by Imperial College London every year. We are an international community, attracting undergraduates from more than **125 countries**. The College focuses on the four main disciplines of **science, engineering, medicine and business** and is renowned for its application of these skills to industry and enterprise.[2]

*Why choose us* texts are a typical promotional genre imported from the business sector (Casañ Pitarch, 2015) and provide reasons for choosing a university or a specific department:

> The Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care offers a wide range of postgraduate study including some unique programmes supporting career development for practitioners and health professionals. [] The Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care is regarded as a centre of excellence for nursing and midwifery, achieving high scores in research and teaching quality assessments. [3]

[2]https://www.imperial.ac.uk/about/introducing-imperial/facts-and-figures/ [last consulted on 15 December 2017].

[3]https://www.kcl.ac.uk/nursing/study/pg.aspx [last consulted on 15 December 2017].

Staff profiles are a typical academic genre in modern universities, although they serve analogous functions in company profiles, which typically include company policies and the most relevant staff (Casañ Pitarch, 2015:91). Sometimes referred to as Personal Home Pages PHPs (Santini, 2007), staff profiles may contain information about lecturers and their career, and aim at demonstrating the high academic standards achieved by university members:

> Dominic is Director of the Intelligent Efficiency Research Group (IERG) an award winning team of PhD and Masters Researchers specialising in the advancement of the next generation of energy efficiency technologies. [...] Dominic formerly led an energy consultancy business and has supported a number of multinationals across Ireland, Europe and the US in reducing energy consumption.[4]

News pages are rather common in native English academic websites, too. They can be used as a means to provide key information on initiatives, activities and other events, especially if they contribute to increasing university prestige, as the one below:

> 50 start-ups emerge from UCC-led EU-XCEL accelerator. The best of the EU-XCEL start-ups will then be short-listed to pitch to an international team of judges at the EU – XCEL Challenge Final in Cork, Ireland on 2 and 3 November 2015, which coincides with UCC's George Boole bi-centennial celebrations. [...] Boole was UCC's first professor of mathematics and is often called the forefather of the information age.[5]

Couse or programme descriptions differ from *About us* or *Why choose us* pages since they are not explicitly designed to express self-promotion as corporate genres do. Although not necessarily regarded as promotional texts, many course descriptions adopt promotional language strategies, along with the offer of plain information, as the following text exemplifies:

> The English course at St Peter's covers the entire history of English literature, from *Caedmon's Hymn to Catch 22*, and it includes literature in English from around the world. We encourage our students to read as broadly as possible, [...] we want them to make the most of the enormous opportunities available to them in Oxford, through the flexibility of the course, the richness of the libraries and museums, and the variety of expertise on offer.[6]

[4]https://www.ucc.ie/en/civileng/people/drdominicosullivan/ [last consulted on 15 December 2017].

[5]https://www.ucc.ie/en/news/archive/2015press/eu-xcel-creates-50-start-ups-.html [last consulted on 15 December 2017].

[6]http://www.spc.ox.ac.uk/study-here/subjects/english [last consulted on 15 December 2017].

As previously mentioned, the analysis of a random subset of promotional pages shows that many texts are not associated with a specific genre. However, they do deliver positive and engaging messages, and present the university in a good light. These pages include university descriptions (e.g. institutional bodies, services, facilities) and pages about third mission (business relationships, fundraising initiatives, social responsibility, mission and strategy). Examples are provided below:

> The Secretary's Office is a central administrative office which supports the College and its community across a wide range of activities from communications to legal and governance. It has a key role in ensuring that the College complies with legal and statutory requirements and provides the secretariat for Board, Council, Executive Officers, and a number of other College committees and working groups. We hope you find the information you require on these pages, or in our Annual Report.[7]

> The impact of our social responsibility initiatives. We strive to make a positive difference to the life and future of our region by taking socially responsible decisions that have real, beneficial, measurable impacts on the people and the world around us. We make a significant contribution to our communities and society as a whole through our research, our education and a wide range of activities undertaken by our staff, students and alumni.[8]

Webpages randomly extracted from the ELF counterpart point to some analogous genres and pages (Figure 6.3). An example for each genre is provided in the next paragraphs.

The homepage below is taken from a Greek university and presents the School of Childhood Education. As already stated above with reference to native English academic settings, numerous web-based texts provide a mixture of actual information and promotional content aimed at increasing reputation by demonstrating one's excellence in teaching and research:

> The School aims to promote research in early childhood education, offer courses linked to labour market demand, provide students with training directly related to their main field of study, and cooperate with foreign institutions and agencies. The young faculty members of the School and their innovative research methods contribute to the School's reputation for excellence in research and teaching. The

[7] https://www.tcd.ie/Secretary/ [last consulted on 15 December 2017].

[8] http://www.manchester.ac.uk/discover/social-responsibility/ [last consulted on 15 December 2017].

> School of Early Childhood Education offers high-quality education and prepares students for a successful career in early education.[9]

*About us* pages fulfil similar functions. They present universities or one of their internal structures in a positive light, often claiming uniqueness and superiority:

> The Technical University of Munich (TUM) combines top-class facilities for cutting-edge research with unique learning opportunities for students. It is committed to finding solutions to the major challenges facing society as we move forward: Health & Nutrition • Energy & Natural Resources • Environment & Climate • Information & Communications • Mobility & Infrastructure. The university thinks and acts with an entrepreneurial spirit. Its aim: to create lasting value for society. All this combines to make it one of Europe's leading universities.[10]

*Why choose us* pages are quite standardised in terms of their informative and promotional text. As was noted for native English texts, explicit reference is made to a ranking of some type, as proof of university excellence:

> Why choose Aarhus University? TOP 100 UNIVERSITY. AU is consistently ranked as one of the world's top universities. It was ranked number 73 in the 2015 Shanghai ranking and number 81 among the largest universities in the world in the 2015 Leiden Ranking.

News scoring high in the promotional dimension report special events which contribute to raising university prestige and international reputation, as shown in the extract below:

> Porto and Vigo strengthen cooperation in unprecedented summit. The rectors, vice-rectors and pro-rectors of Porto and Vigo sat together for the first time to discuss how to deepen cooperation between the two universities, laying the groundwork for future partnerships in the fields of education, research and governance. The creation of joint research projects was one of the first opportunities identified by the two executive boards, particularly in R&D fields considered strategic by both universities, namely in engineering, energy, bioengineering, biomedical, telecommunications and marine sciences.[11]

[9] https://www.auth.gr/en/nured [last consulted on 15 December 2017].

[10] https://www.tum.de/en/about-tum/ [last consulted on 15 December 2017].

[11] https://noticias.up.pt/porto-and-vigo-strengthen-cooperation-in-unprecedented-summit/ [last consulted on 15 December 2017].

As mentioned above with reference to native English universities, ELF university websites as well make use of traditional academic genres (e.g. course/programme descriptions) for advertising purposes. The example below was published in a Swiss university website:

> In view of changing life styles and an ever increasing aging society, sustaining and improving health has taken on a pivotal role. [...] The interdisciplinary study programme "Health Sciences and Technology" was introduced in autumn 2011. This study programme aims to communicate basic knowledge on the human body.[12]

ELF texts show a small proportion of research project descriptions, which did not appear at all among native English documents. Research project descriptions appearing in the random sample aim to promote research activities internationally and were published in five different academic websites located in Austria, Finland, German, Switzerland and the Netherlands. The example below is taken from the University of Vienna:

> The University of Vienna is a research university enjoying high international visibility. Its profile reflects the characteristics of the area it is located in, and understands research as a global challenge.[13]

The last genre found in the random sample of lingua franca promotional texts concerns 'Contact pages'. Although not usually regarded as a promotional genre, these contact pages may provide a series of initiatives and events where prospective students can meet university students and staff:

> Experience the university in person or online. Visit Open Days. This will give you a broad introduction to the city, our programmes, and extra-curricular activities. Do you already know the programme that interests you most? Attend an Experience Day – an in-depth look at a programme in the faculty location where you will study. Meet students, staff, experience a lecture and more. Become a Student for a Day. Shadow a student in a specific programme, attend lectures, see what to expect on a day-to-day basis.[14]

[12] https://www.hest.ethz.ch/en/studies/health-sciences-and-technology.html [last consulted on 15 December 2017].

[13] https://www.univie.ac.at/en/research/research-overview/research-at-the-university-of-vienna/ [last consulted on 15 December 2017].

[14] https://www.universiteitleiden.nl/en/education/bachelors/information-activities [last consulted on 15 December 2017].

Summing up, this analysis of a sample of highly promotional pages coming from native English and ELF contexts serves two purposes. First, it provides an overview of highly promotional content pages and helps to understand if these pages belong to a specific promotional genre. Second, it provides a way of approaching the comparison of native English and ELF institutional academic pages, to detect any similarities and differences in the use of web-based promotional genres. The analysis showed that texts written in either variety of English belong to two different groups. The first group comprises typical promotional genres or text types coming from the advertising sector and largely used in corporate websites to attract new clients and persuade stakeholders through a nice and positive image. These genres include Homepages, Why choose us pages, About us pages, staff profiles and news. Other texts can be associated with this category, although they are not as standardised as the ones mentioned above (descriptions related to social responsibility, corporate strategy and mission statements, indicated in Figure 6.3 as third mission). A second group of texts is composed of institutional and/or academic genres that were not initially created for marketing purposes, which nonetheless seem to feature promotional elements. Course and programme descriptions are the clearest examples of such pages. The same happens with less standardised texts, such as research project descriptions, which seem to be more typical of the ELF context (especially countries with a Germanic language background), where they fulfil both informative and promotional functions. Summing up, the native English subcorpus includes more examples of staff profiles and *Why choose us* pages, while the ELF subcorpus shows more *About us* pages, course descriptions and research descriptions, suggesting that ELF websites include less standardized genres and more "generic" promotional texts.

## 6.2 Evaluation markers in promotional texts

This second analysis complements Section 6.1 by offering a linguistic perspective on pages scoring high on the A12 dimension, which represents promotional texts (Section 4.4). While section 6.1 analysed a small subset of promotional documents from the point of view of genre, and tried to determine whether sampled pages match a typical promotional genre or text type, this second part focuses on the same promotional subset, but examines it from an internal perspective. Furthermore, while in section 6.1 the analysis was carried out on a random sample of 100 pages, in this section the full set of texts scoring highest on the promotional dimension was considered. Lexico-grammatical markers of evaluation previously collected from the relevant literature (see Chapter 5) are used to query the corpus in its native and ELF varieties in order to detect any differences or similarities in the use of evaluation markers, with the ultimate goal to compare promotional strategies in the two varieties of English.

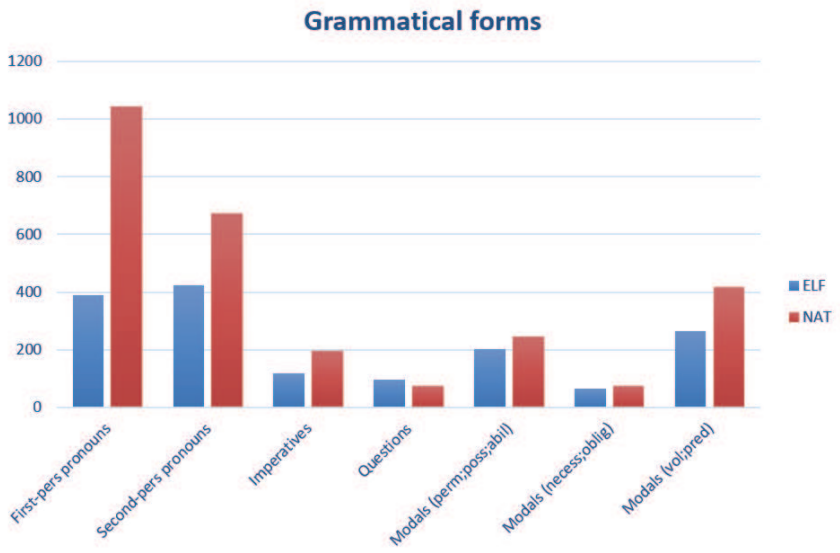


Fig. 6.4 Distribution of grammatical markers of evaluation over native English and ELF promotional pages (top decile in the A12 dimension).

Methods of investigation in this section include examination of concordance lines, collocates and any other patterns that arise from a corpus-based approach. Searches were carried out using the SketchEngine tool and results were normalised per 100K words. The analysis keeps the same structure as Chapter 5, where queries of evaluation were divided into three groups – grammatical forms, lexico-grammatical forms and attitudinal lexis – together with their corresponding language features.

## 6.2.1 Grammatical forms

Figure 6.4 displays the full set of grammatical markers of evaluation, including personal pronouns, imperatives, questions and modal verbs (divided into the most common semantic classes).

### 6.2.1.1 Pronouns

Pronouns are one of the most explicit signals of interaction between textual voices, serving as a tool to investigate how native English and ELF universities engage in a dialogue with their web audience (Hyland, 2008). In quantitative terms, it is clear from Figure 6.5 that first and second person pronouns are generally more typical of the native English setting, especially *our*, *you*, *your* and *we*, which make texts similar to face-to-face communication. Furthermore, native English texts make a large use of the pronoun *our*, addressing readers in

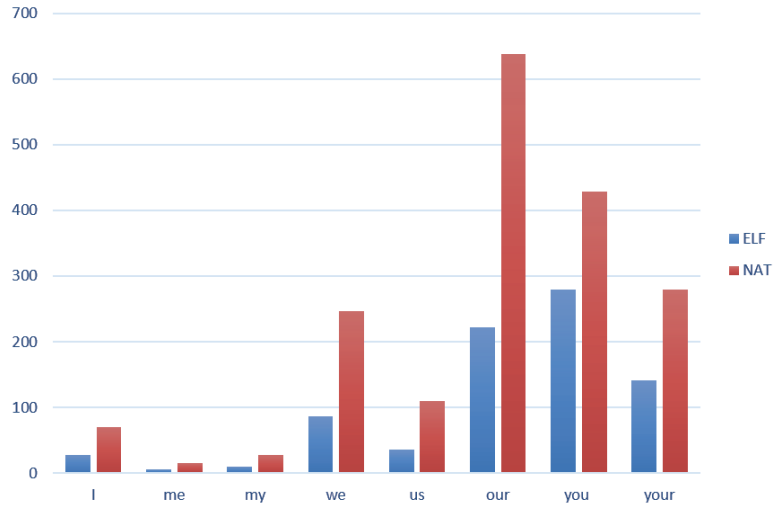


Fig. 6.5 Distribution of personal pronouns in native English and ELF texts (per 100K words).

an inclusive rather than exclusive way. To further investigate the differences among these pronouns, a subsequent qualitative analysis was conducted.

Although thinly spread across both varieties of English, the first-person pronoun *I* can offer insights into how promotional contents are conveyed, since most of its instances can be found in two promotional genres, i.e. testimonials and welcome messages. FAQs can also be written in first person but it is quite unusual to find them in this set of highly promotional texts (Section 6.1). As Figure 6.6 shows, present tense verbs are frequent colligates of the first-person pronoun *I* in both native English and ELF texts. However, it seems that the difference in the use of present and past tenses is particularly marked in ELF documents, where past tenses are scarcely used as compared to native English texts. The analysis of concordance lines may provide a hint on the reasons why this occurs. Past tense verbs tend to describe past students' experiences about any aspects of their academic life, as shown in the examples below:

(1) I chose to go to Edinburgh because I knew people who had been there and loved it [The University of Edinburgh – UK]

(2) I found King's to be a very inclusive university [King's College London – UK]

(3) I learned about Scottish Literature and culture in the most beautiful city in the world [The University of Edinburgh – UK]

(4) I benefited from CV support, job application workshops and a mock interview [The University of Manchester – UK]

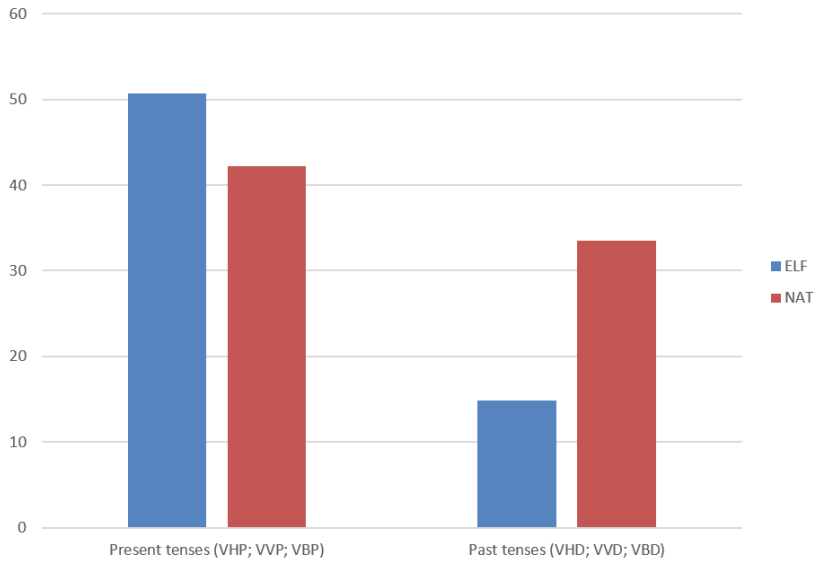


Fig. 6.6 Distribution of colligation candidates (POS-tags) for the "I" pronoun over native English and ELF texts (per 100K words).

(5) I had so much fun and had the chance to travel all over Asia [The University of Warwick – UK]

By contrast, present tenses can either describe a current state of mind or reveal future plans, as exemplified below:

(1) Turku is beautiful and I like it here, maybe I want to stay [University of Turku – Finland]

(2) I consider the most fun part of the programme to be the international student body. [Leiden University – The Netherlands]

(3) I hope you will like the idea of studying dentistry in Szeged and we can welcome you as our student soon [University of Szeged – Hungary]

The above examples suggest that past tense verbs describe an action that really happened in the past (in particular with reference to student personal experience), while present tense verbs are more related to the expression of feelings, personal opinion and future expectations.

As for the remaining first-person singular pronouns, namely *me* and *my*, there are very few occurrences in the corpus, especially in the ELF subcorpus, which only features 5 occurrences of *me* in the first decile of the promotional dimension (per 100K words). The pronoun *me* occurs as the object of causative verbs indicating positive results, such as *helped*

*me to develop*, *allowed me to engage*, *gives me the freedom to interact*, *enabled me to acquire*. Once again, this pronoun seems to be associated with a personal experience in testimonials. The pronoun *my* displays 50 occurrences in the ELF subcorpus (first decile in the promotional dimension). Although there are too few occurrences to detect reliable collocation candidates, *life*, *career* and *studies* are identified as immediate right collocates of *my*. In the native English subcorpus, *research* is the most frequent collocate of *my*, such as is *my research interests*, *my research commitments*, *my research skills*. It is noticeable that there are no instances at all of *my research* in ELF promotional texts, possibly suggesting a higher engagement in research within native English universities or that research is not described using such an involved, personal style.

Turning to first-person plural pronouns, Figure 6.7 displays the top twenty right collocates of *we* in both varieties of English, while Figure 6.8 illustrates the immediate left collocates of *us*. It will be noted from Figure 6.7 that native English pages show a higher frequency of contracted verb forms, which may suggest a more colloquial style. A further feature to note is the higher use of *must* in ELF texts. This is generally associated with university commitment towards one of its stakeholders, as in:

(1) we must offer our students a comprehensive education [Universitat Pompeu Fabra – Spain]

(2) we must create a wide range of incentives [Ruprecht Karls Universitaet Heidelberg – Germany]

(3) we must conduct research of the higher quality [Aarhus University – Denmark]

As for the pronoun *us*, it is noticeable that native English collocates includes some –ing forms in the top twenty collocates and further down in the list as well, while there are no occurrences of *-ing + us* forms in ELF promotional pages.

An examination of the top twenty collocates of *our* in both varieties of English reveals some typical areas in which the academic community is actively involved, such as teaching, research, international cooperation and the business sector (Figure 6.9). Collocates refer to similar institutional-academic areas, although terms vary slightly. For instance, native documents use the word *staff* to a greater extent than ELF texts, which may indicate a higher sense of inclusion towards any academic member other than students (e.g. *The profile of our staff in terms of research and teaching*, *our staff feel they are part of a supportive*, *inclusive and highly motivated community*, *We take the well-being of our staff very seriously*). On the other hand, ELF collocates comprise culture-specific elements, such as the existence of *faculties* and a larger use of the word *society*, which appears at the very bottom of the native

| | Frequency |
|---|---|
| P \| N are | 170 |
| P \| N have | 136 |
| P \| N can | 104 |
| P \| N will | 80 |
| P \| N do | 62 |
| P \| N 're | 62 |
| P \| N offer | 57 |
| P \| N need | 31 |
| P \| N work | 27 |
| P \| N 'll | 24 |
| P \| N aim | 21 |
| P \| N also | 20 |
| P \| N believe | 20 |
| P \| N live | 18 |
| P \| N 've | 18 |
| P \| N were | 15 |
| P \| N support | 14 |
| P \| N provide | 14 |
| P \| N hope | 13 |
| P \| N continue | 12 |

(a) Native English collocates.

| | Frequency |
|---|---|
| P \| N are | 62 |
| P \| N have | 41 |
| P \| N offer | 27 |
| P \| N can | 26 |
| P \| N want | 15 |
| P \| N also | 14 |
| P \| N will | 12 |
| P \| N must | 12 |
| P \| N need | 10 |
| P \| N do | 10 |
| P \| N believe | 8 |
| P \| N support | 7 |
| P \| N live | 6 |
| P \| N would | 6 |
| P \| N aim | 6 |
| P \| N encourage | 6 |
| P \| N provide | 6 |
| P \| N strive | 5 |
| P \| N work | 5 |
| P \| N welcome | 5 |

(b) ELF collocates.

Fig. 6.7 Top twenty collocates of *we* in native English and ELF promotional pages.

| | Frequency |
|---|---|
| P \| N with | 114 |
| P \| N about | 60 |
| P \| N contact | 49 |
| P \| N join | 32 |
| P \| N help | 24 |
| P \| N follow | 21 |
| P \| N tell | 20 |
| P \| N allows | 14 |
| P \| N joining | 13 |
| P \| N of | 12 |
| P \| N let | 12 |
| P \| N allow | 12 |
| P \| N to | 12 |
| P \| N enable | 11 |
| P \| N helping | 10 |
| P \| N for | 10 |
| P \| N gives | 8 |
| P \| N around | 8 |
| P \| N enables | 7 |
| P \| N find | 7 |

(a) Native English collocates.

| | Frequency |
|---|---|
| P \| N with | 29 |
| P \| N contact | 27 |
| P \| N join | 17 |
| P \| N visit | 11 |
| P \| N help | 10 |
| P \| N for | 8 |
| P \| N about | 7 |
| P \| N helps | 5 |
| P \| N allows | 5 |
| P \| N find | 5 |
| P \| N around | 4 |
| P \| N reach | 4 |
| P \| N of | 4 |
| P \| N tell | 4 |
| P \| N enable | 3 |
| P \| N allow | 3 |
| P \| N let | 3 |
| P \| N follow | 3 |
| P \| N meet | 3 |

(b) ELF collocates.

Fig. 6.8 Top twenty collocates of *us* in native English and ELF promotional pages.

| | Frequency |
|---|---|
| P \| N students | 289 |
| P \| N research | 270 |
| P \| N staff | 74 |
| P \| N work | 57 |
| P \| N graduates | 55 |
| P \| N alumni | 48 |
| P \| N academic | 46 |
| P \| N international | 45 |
| P \| N teaching | 42 |
| P \| N undergraduate | 37 |
| P \| N global | 37 |
| P \| N researchers | 29 |
| P \| N postgraduate | 28 |
| P \| N current | 27 |
| P \| N programmes | 26 |
| P \| N campus | 26 |
| P \| N commitment | 24 |
| P \| N academics | 23 |
| P \| N own | 23 |
| P \| N website | 23 |

(a) Native English collocates.

| | Frequency |
|---|---|
| P \| N students | 89 |
| P \| N university | 70 |
| P \| N research | 45 |
| P \| N international | 36 |
| P \| N faculty | 28 |
| P \| N researchers | 21 |
| P \| N graduates | 20 |
| P \| N alumni | 20 |
| P \| N teaching | 17 |
| P \| N phd | 16 |
| P \| N society | 16 |
| P \| N partner | 16 |
| P \| N website | 15 |
| P \| N academic | 15 |
| P \| N partners | 15 |
| P \| N master | 14 |
| P \| N programmes | 14 |
| P \| N time | 12 |
| P \| N world | 11 |
| P \| N current | 10 |

(b) ELF collocates.

Fig. 6.9 Immediate right collocates of *our* in native English and ELF promotional pages (top twenty).

English list of collocates. The word *society* is related to challenges, values and it may also refer to the local community.

Second-person pronouns *you* and *your* are used to address specific members of university audience directly. Due to the high number of occurrences of these features, the analysis of collocations helps us identify specific patterns. Figure 6.10 shows the immediate right and left collocates of *you*, together with their raw frequencies[15]. From a semantic point of view, left collocates in both English varieties are associated with some offers of goods and services. Actional verbs, such as *help*, *give*, *provide*, *invite*, *offer* imply some non-linguistic action as their objective. In this sense, they constitute a response given to an implicit request for information. From a grammatical point of view, it can be noted that *would* and *do* occur as left collocates of *you* in ELF texts, indicating a higher frequency of questions as compared to native texts (e.g. *Would you like to study*, *Would you like to know more*, *Do you wish to study*, *Do you need assistance*). A more detailed analysis of questions will be provided further down in Section 6.2.1.3. Finally, as already observed above, native English right collocates for *you* hint at a more colloquial style through the use of contracted forms (*you'll* and *you're*).

The last personal pronoun feature concerns patterns of *your*. An initial examination of right collocates for *your* in native English texts reveals some typical university addressees: 1) students, in phrases like *your career*, *your studies*, *your learning*; 2) researchers (e.g. *discuss your research proposal*), and 3) business companies or organizations (e.g. *Kickstart your business with our incubation*). The third element represents key stakeholders in universities'

[15]Raw frequencies are shown in the figures because the size of the subcorpora are comparable.

| P \| N | Frequency |
|---|---|
| if | 257 |
| , | 207 |
| help | 153 |
| that | 98 |
| for | 77 |
| give | 61 |
| and | 59 |
| whether | 53 |
| what | 50 |
| to | 44 |
| when | 33 |
| how | 32 |
| provide | 31 |
| with | 30 |
| where | 28 |
| offer | 23 |
| hope | 23 |
| gives | 23 |
| do | 23 |
| as | 22 |

(a) NAT left collocates

| P \| N | Frequency |
|---|---|
| , | 249 |
| if | 141 |
| are | 69 |
| help | 55 |
| do | 54 |
| for | 47 |
| that | 38 |
| offers | 31 |
| would | 28 |
| give | 27 |
| to | 24 |
| provides | 23 |
| where | 23 |
| whether | 21 |
| gives | 21 |
| and | 20 |
| when | 20 |
| invite | 19 |
| here | 18 |
| provide | 15 |

(b) ELF left collocates

| P \| N | Frequency |
|---|---|
| to | 252 |
| will | 216 |
| can | 194 |
| are | 167 |
| have | 106 |
| 'll | 91 |
| 're | 75 |
| . | 71 |
| need | 58 |
| the | 57 |
| with | 55 |
| want | 51 |
| would | 46 |
| in | 35 |
| a | 34 |
| for | 32 |
| may | 30 |
| and | 25 |
| develop | 23 |
| find | 19 |

(c) NAT right collocates

| P \| N | Frequency |
|---|---|
| will | 207 |
| to | 137 |
| can | 131 |
| are | 99 |
| with | 61 |
| have | 58 |
| need | 45 |
| want | 42 |
| . | 33 |
| a | 32 |
| like | 30 |
| would | 29 |
| the | 27 |
| 'll | 22 |
| in | 19 |
| for | 18 |
| find | 16 |
| may | 15 |
| an | 14 |
| get | 13 |

(d) ELF right collocates

Fig. 6.10 Immediate left and right collocates of *you* in native English and ELF promotional pages (top twenty).

management since companies, enterprises and organisations can provide economic funding and jobs. As for ELF texts, it can be noted from Figure 6.11 that almost all of the immediate right collocates of *your* are associated with some education-related terms and that direct communication with the business world is indeed missing.

While pronouns work as a persuasive tool through the engagement of readers into the discourse, other strategies, such as the use of imperatives, persuasively push readers into action.

We now turn to the feature of imperatives, that can be considered key constructios in advertising material since they work as a call to action and are a way to trigger an immediate reaction from the reader.

#### 6.2.1.2 Imperatives

As shown in Figure 6.4, imperatives occur with a higher relative frequency in native texts as compared to ELF texts. Frequency lists and concordance searches were carried out in order to look at the use of imperatives in the surrounding context. Three classes of verbs can be identified in Figure 6.12. First, "navigation verbs", which serve to connect various website sections (e.g. *find, read, see, learn, visit, go*). Second, verbs that we could refer to as "action verbs", because they prompt a response in the real world, such as *join* and *apply*. Third, "verbs of help", used to ask for economic funding (*help* and *support*).

The first class represents most of the imperatives in both varieties of English, used in phrases like *find out more*, *read more*, *learn more about*, *visit the…page/portal/online prospectus/Department*, *take a look at*, *use the browse function*. *Explore* and *Discover* seem

| | Frequency |
|---|---|
| P \| N career | 61 |
| P \| N own | 55 |
| P \| N time | 54 |
| P \| N business | 42 |
| P \| N research | 40 |
| P \| N degree | 37 |
| P \| N studies | 37 |
| P \| N organisation | 36 |
| P \| N chosen | 31 |
| P \| N skills | 30 |
| P \| N academic | 26 |
| P \| N knowledge | 23 |
| P \| N experience | 22 |
| P \| N future | 22 |
| P \| N life | 18 |
| P \| N interests | 18 |
| P \| N subject | 17 |
| P \| N learning | 17 |
| P \| N personal | 17 |
| P \| N course | 17 |

(a) Native English collocates.

| | Frequency |
|---|---|
| P \| N studies | 57 |
| P \| N own | 55 |
| P \| N career | 26 |
| P \| N home | 19 |
| P \| N research | 17 |
| P \| N degree | 14 |
| P \| N master | 13 |
| P \| N application | 12 |
| P \| N life | 12 |
| P \| N stay | 12 |
| P \| N academic | 12 |
| P \| N professional | 12 |
| P \| N knowledge | 11 |
| P \| N field | 11 |
| P \| N questions | 10 |
| P \| N bachelor | 10 |
| P \| N personal | 9 |
| P \| N fellow | 8 |
| P \| N future | 8 |
| P \| N university | 8 |

(b) ELF collocates.

Fig. 6.11 Top twenty collocates of *your* in native English and ELF promotional pages.

to be more typical of the native English and ELF environment respectively, although they convey almost the same meaning, as in *Explore research courses* (University of Oxford – UK) and *Discover the programme* (Science Po – FR). The second group of verbs is aimed at inviting them to take action outside of the website, in phrases like *contact us*, *join us*, *apply*. *Get* is used rather differently in the two contexts of English; while in the native context it clearly belongs to the second group (as part of verbal patterns like *get involved* and *get in touch*), in the ELF context it is used in constructions used to suggest that information should be obtained, falling under the first category, as in *get informed* or *get an overview*. The third semantic class comprises verbs aimed at asking for economic support (*support us*, *make a gift* and *help us* further down in the list of verbs). These verbs do not occur in the top twenty imperatives in the ELF subcorpus, or they do not appear at all, as is the case of *help*. This confirms previous observations whereby ELF texts miss a direct attempt to communicate with the entrepreneurial world in highly promotional pages.

Moving on to our next topic, together with personal pronouns and imperatives, question marks provide potential surface evidence of reader engagement since they construe a relationship with readers, imitating a real exchange of information.

#### 6.2.1.3 Questions

Although Figure 6.4 shows that direct questions are used more frequently in ELF texts than in native English texts, a closer analysis through concordance lines reveals a different use in

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | . find | 341 |
| P \| N | . read | 236 |
| P \| N | . see | 42 |
| P \| N | . visit | 33 |
| P \| N | . learn | 31 |
| P \| N | . explore | 30 |
| P \| N | . view | 15 |
| P \| N | . contact | 15 |
| P \| N | . support | 11 |
| P \| N | . make | 11 |
| P \| N | . get | 11 |
| P \| N | . apply | 11 |
| P \| N | . join | 10 |
| P \| N | . develop | 10 |
| P \| N | . use | 9 |
| P \| N | . meet | 9 |
| P \| N | ? find | 8 |
| P \| N | . watch | 8 |
| P \| N | . go | 7 |
| P \| N | . build | 7 |

(a) List of native English imperatives.

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | . read | 197 |
| P \| N | . find | 62 |
| P \| N | . see | 30 |
| P \| N | . learn | 30 |
| P \| N | . discover | 25 |
| P \| N | . watch | 13 |
| P \| N | . get | 13 |
| P \| N | ! read | 11 |
| P \| N | . join | 10 |
| P \| N | . go | 10 |
| P \| N | . visit | 8 |
| P \| N | . contact | 8 |
| P \| N | ? find | 7 |
| P \| N | . take | 7 |
| P \| N | . follow | 7 |
| P \| N | . profile | 6 |
| P \| N | . apply | 6 |
| P \| N | ? read | 5 |
| P \| N | . meet | 5 |
| P \| N | . let | 5 |

(b) List of ELF imperatives.

Fig. 6.12 Top twenty imperative forms in native English and ELF promotional pages.

context. Native English texts display various instances of *how* questions, as is evident from the most frequent tri-grams listed in Figure 6.13.[16]

Questions introduced by *How can we*, *How are we* and *How do you* tend to be used as a sort of FAQ (e.g. how can we help?) or are more frequently used to promote university activities in research, teaching and third mission. Example (1) refers to a research project dealing with aging at Trinity College Dublin. Examples (2) and (3) are a way to introduce core issues in the relation between universities and society, specifically addressing the African community and the best students. Example (4) instead refers to a UCL fund-raising campaign for research projects. This last example goes back to previous discussions in Section 6.2.1.2 about the tendency in native English to ask for economic support through promotional texts.

(1) How can we keep people very healthy right to the very end of life? [Trinity College Dublin – Ireland]

(2) How are we preparing a new generation of African leaders? [King's College London – UK]

(3) How can we support the brightest students? [King's College London – UK]

(4) How do you want your gift to benefit UCL? [University College London – UK]

By contrast, ELF questions favour *what* and *why* forms, as can be noted from Figure 6.13. Although there are many occurrences of *why* questions coming from Scienco Po (France),

[16]Tri-grams were extracted from the native English and ELF subsets of sentences returned by a search for questions in the respective subcorpora.

| | word | Frequency |
|---|---|---|
| P \| N | How can we | 12 |
| P \| N | What is the | 10 |
| P \| N | How are we | 7 |
| P \| N | and a Language | 6 |
| P \| N | Why Imperial ? | 6 |
| P \| N | Who is the | 6 |
| P \| N | What will you | 6 |
| P \| N | Studies and a | 6 |
| P \| N | Business Studies and | 6 |
| P \| N | ? How do | 6 |
| P \| N | ? How can | 6 |
| P \| N | should attend ? | 5 |
| P \| N | Who should attend | 5 |
| P \| N | How do you | 5 |
| P \| N | you study ? | 4 |
| P \| N | with us ? | 4 |
| P \| N | will you study | 4 |
| P \| N | the scheme ? | 4 |
| P \| N | the programme for | 4 |
| P \| N | the UK ? | 4 |
| P \| N | programme for ? | 4 |
| P \| N | look like ? | 4 |
| P \| N | is the programme | 4 |
| P \| N | has been the | 4 |

(a) List of native English trigrams.

| | word | Frequency |
|---|---|---|
| P \| N | Sciences Po ? | 14 |
| P \| N | choose Sciences Po | 9 |
| P \| N | What is a | 9 |
| P \| N | they choose Sciences | 8 |
| P \| N | did they choose | 8 |
| P \| N | Why did they | 8 |
| P \| N | ? What do | 7 |
| P \| N | is a Research | 6 |
| P \| N | a Research Master | 6 |
| P \| N | Research Master 's | 6 |
| P \| N | Master 's programme | 6 |
| P \| N | 's programme ? | 6 |
| P \| N | like to be | 5 |
| P \| N | What is the | 5 |
| P \| N | What are the | 5 |
| P \| N | the UvA ? | 4 |
| P \| N | the Best Choice | 4 |
| P \| N | is Sciences Po | 4 |
| P \| N | for You ? | 4 |
| P \| N | after graduation ? | 4 |
| P \| N | What is Sciences | 4 |
| P \| N | What do you | 4 |
| P \| N | What Makes Our | 4 |
| P \| N | School the Best | 4 |

(b) List of ELF trigrams.

Fig. 6.13 Top twenty trigrams extracted from concordance lines for questions in both varieties of English.

it is still remarkable that how-questions are used twice as many times in the native English promotional subset as in ELF promotional texts (10 vs. 5 occurrences per 100K words).[17]

(1) What is a Research Master's programme? [University of Amsterdam – The Netherlands]

(2) Why did they choose SciencePo? [Science Po Paris – FR]

(3) What is the application procedure? [Humboldt University – Germany]

(4) What is a liberal arts education? [Saint Petersburg State University – Russia]

(5) What is the European Campus in the Upper Rhine region? [Albert Ludwigs Universitaet Freiburg – Germany]

(6) What are the Grand Challenges for the coming decennia? [Delft University of Technology – The Netherlands]

Why-questions are often related to the genre of Why choose us pages, providing a list of reasons for choosing a university or – as is the case of example (2) – sharing stories from university ambassadors. What-questions are extremely general and can be adapted to various needs. From a functional point of view, what-questions can give definitions, explanations or provide a context to some issues. On the other hand, how-questions enable recipients to

[17]The frequency of occurrences was obtained by counting the number of "how" items in the full list of native English and ELF trigrams. Tri-grams were extracted from a native English and ELF subcorpus composed of concordance searches for questions exclusively.

get control over a specific situation, eventually leading to better understanding. This can be clearly exemplified by the term *know-how* in the field of technology transfer, where it refers to all sort of competences and skills that tend to be shared within a confidential circle only, due to the invaluable contribution they provide. In this sense, how-questions can be interpreted as an indication of true engagement with university stakeholders.

Another group of features are modal and semi-modal verbs, which have been suggested to be "the most common grammatical device to mark stance in academic registers" (Biber, 2006a) and which will be analysed in the next section.

#### 6.2.1.4 Modals

As was done for the other features, searches for modals were carried out using the SketchEngine tool and results were normalised per 100K words. Findings are displayed in Figure 6.14, showing the overall distribution of modal and semi-modal verbs. As was found out in a previous case study conducted on a different corpus of ELF and NAT uni websites (Ferraresi and Bernardini, 2013), modals are used to a greater extent in the native subcorpus than the ELF one. Although this analysis differs from Ferraresi and Bernardini (2013) in that it focuses on pages with a promotional function only, modal verbs seem to behave in a rather similar way. *Will* and *can* are the most common modal verbs in both varieties of English. Furthermore, volition/prediction and permission/possibility modals are used more frequently in the native subcorpus, while modals of necessity/obligation display a higher or equal number of occurrences in the ELF component. Modals of necessity can appear in specific "administrative" sections within more promotional pages, for instance course descriptions:

(1) Students must be keen to work in groups and be sensitive to ethical and cultural issues. [KULeuven]

Yet, necessity modals are persuasively used as a self-constraint in which the university commits to reaching key goals, as in the examples below:

(1) An excellent university must be able to offer excellent conditions to both researchers and students [Ruprecht Karls Universitaet Heidelberg – DE]

(2) We must identify ways to increase the size of our research base, so as to secure and sustain our place as a globally leading institution [Warwick University – UK]

The analysis of modal verbs is particularly fruithful since it confirms and complements findings from Ferraresi and Bernardini (2013). While the authors could not link their final

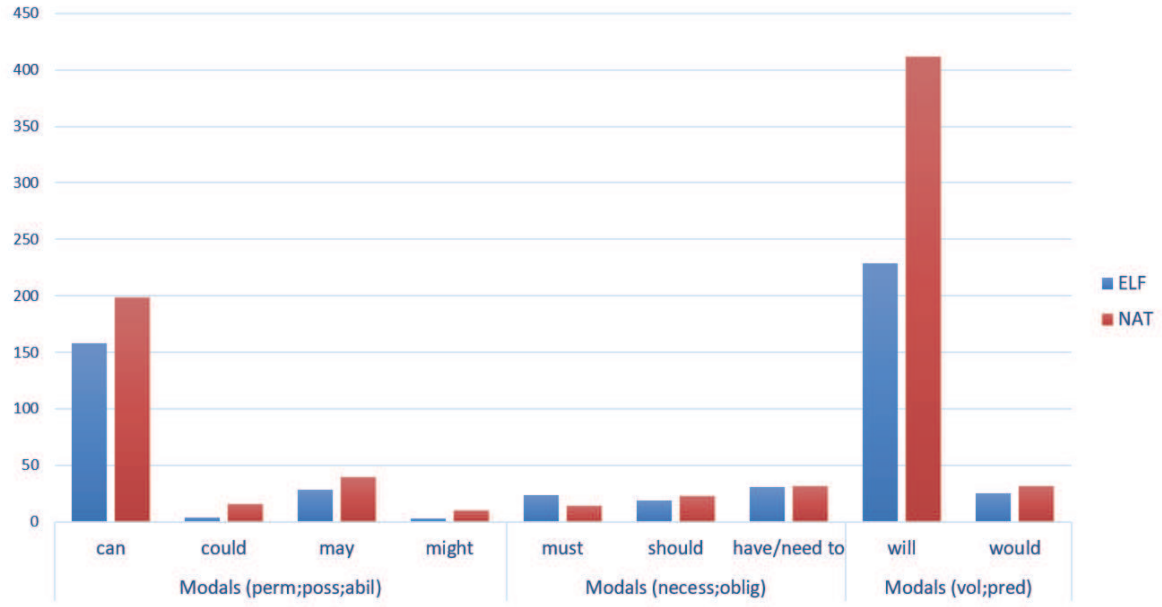


Fig. 6.14 Distribution of (semi-)modal verbs over native English and ELF texts.

observations to comparable functions in the two varieties of English, in this PhD thesis, corpora are controlled by function. Therefore, it can be confirmed that the different use of modal verbs in promotional texts is related to different native English and ELF strategies, rather than different text functions.

### 6.2.2 Lexico-grammatical forms

This section examines lexico-grammatical markers of stance, namely stance adverbials or short prepositional phrases and stance complement clauses (cf. Section 2.3 for a theoretical overview and Section 5.2 for examples and search queries).

As can be noted from Figure 6.15, amplifiers or intensifiers such as *very*, *highly* and *fully* are among the most common promotional strategies in this set of texts (together with certainty markers and stance verbs followed by a to-clause). Biber et al. (1999:565) find that amplifiers are generally more frequent in conversation than in academic prose, which confirms that university promotional strategies tend to rely on a spoken-like register. On the whole, all these lexico-grammatical patterns are used slightly more frequently in the native subcorpus, with two exceptions, i.e. markers of perspectives (e.g. *generally, personally, according to*) and stance nouns followed by that-clauses (e.g. *fact that, view that*).

In terms of the use of individual adverbials, the most common markers of stance in the corpus include *really*, *likely*, *clearly*, *sure*, *certainly*, *definitely*, *of course*, *indeed* and *in fact*. Some of them seem to be more common in the ELF subcorpus and others in the

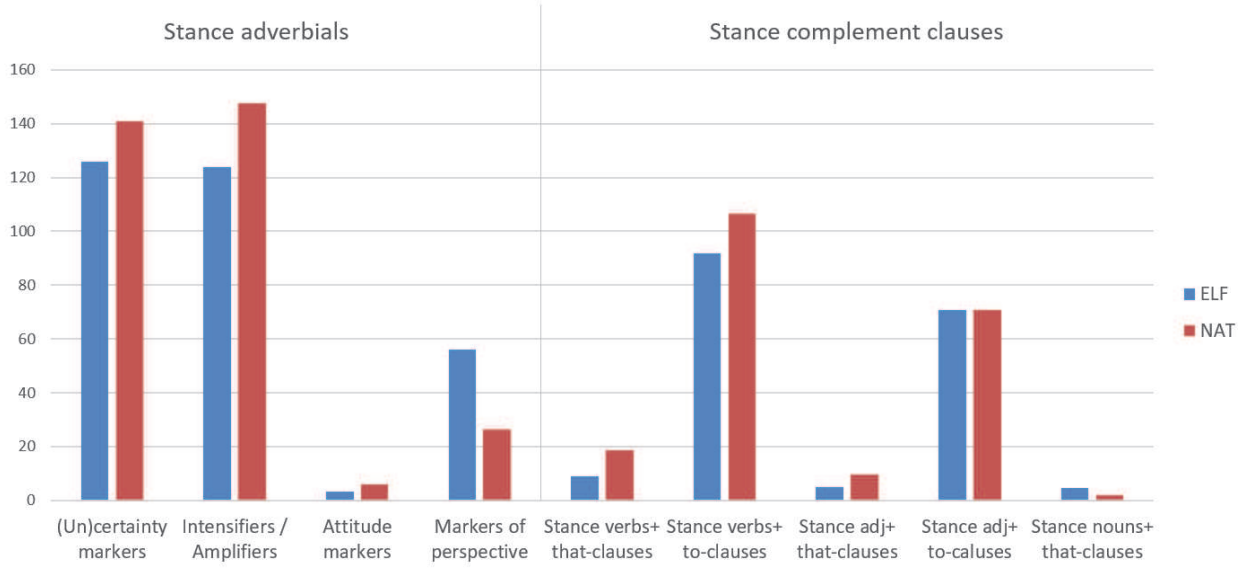


Fig. 6.15 Lexico-grammatical markers of stance in native English and ELF promotional texts (per 100K words).

native English one. In particular, *really*, *definitely* and *indeed* are slightly more common in native English texts, while ELF texts favour *clearly* and *in fact*. Even if the preference for an expression or another does not affect textual intelligibility, these (apparently) interchangeable options may have some implications for style and communicative strategies. For instance, the analysis of concordance lines show that *really* is used rather colloquially and is often adopted in student testimonials or other informal texts and the same is true for the few instances of *really* in ELF texts (Example 3):

(1) I really enjoyed my studies, and my undergraduate experience was underpinned by my sports and society activities [Warwick University – UK]

(2) Smaller class sizes mean you will really get to know your classmates and tutors [King's College London]

(3) That pretty much sums the core of successful career planning. You have to think about what really motivates you [University of Helsinki – Finland]

As for amplifiers and intensifiers, *very* and *highly* account for most of the normalized frequencies shown in Figure 6.15. Furthermore, *very* is used slightly more frequently in native countries, while *highly* tends to be more common in ELF settings, a finding which seems to challenge previous work portraying ELF as an elementary or simplified version of native English. Yet, it is worth mentioning that some differences can be observed through

| | Frequency |
|---|---|
| P \| N best | 32 |
| P \| N active | 9 |
| P \| N high | 7 |
| P \| N wide | 7 |
| P \| N well | 7 |
| P \| N much | 7 |
| P \| N important | 6 |
| P \| N good | 6 |
| P \| N successful | 6 |
| P \| N positive | 5 |
| P \| N own | 5 |
| P \| N latest | 5 |
| P \| N proud | 5 |
| P \| N first | 4 |
| P \| N strong | 4 |
| P \| N different | 4 |
| P \| N diverse | 4 |
| P \| N highest | 4 |

(a) Native English collocates.

| | Frequency |
|---|---|
| P \| N good | 10 |
| P \| N high | 8 |
| P \| N much | 8 |
| P \| N best | 7 |
| P \| N well | 6 |
| P \| N close | 5 |
| P \| N attractive | 5 |
| P \| N wide | 5 |
| P \| N important | 5 |
| P \| N active | 4 |
| P \| N successful | 4 |
| P \| N first | 4 |
| P \| N few | 4 |
| P \| N young | 3 |
| P \| N affordable | 3 |
| P \| N useful | 3 |
| P \| N helpful | 3 |
| P \| N rewarding | 3 |

(b) ELF collocates.

Fig. 6.16 List of immediate right-hand collocates of "very" in native English and ELF texts.

the analysis of concordance lines. *Best* and *good* are the top right-hand collocates of "very" in native English and ELF texts respectively, while they both show up further down in the opposed list of collocates (Figure 6.16).

This suggests that there are indeed some typical native English and ELF patterns in the use of *very*. *Very best* is widely used among native texts in each of the three main university sectors (teaching, research, third mission). In the examples provided, universities take a reflexive stance, with no attempt to provide further justification to their evaluative claims:

(1) Campus life includes access to some of the very best teaching, research and leisure facilities [Imperial College London – UK]

(2) We're committed to providing a world class education and access to the very best teachers, resources and support. [University of Cambridge – UK]

(3) We do our very best to make sure that the transition to Ireland and to University is as smooth as possible [University College Cork – IE]

On the other hand, *very good* appears more frequently in the ELF subcorpus where it is used to report an explicit opinion or evaluation by a third party, thus providing concrete evidence of university claims:

(1) The University of Warsaw holds very good positions in comparative evaluations [Warsaw University – RU]

(2) Human resources representatives gave a very good evaluation for Aachen graduates [RWTH Aachen University – DE]

(3) After completing their PhD they eventually work as professors or researchers, and our placement records are very good [KU Leuven – BE]

Markers of perspective lead to complementary observations. An examination into these specific markers indicates that the higher normalized frequency in the ELF subcorpus is the result of one marker, namely *according to*, which serves to convey evaluative meaning through an external source of knowledge:

(1) According to a recent investigation by the European Commission, Leiden University ranks alongside Oxford and Cambridge as the academic top of the European Union. [Leiden University – NL]

(2) Since 2012, SPbU business school has been the best education centre in Eastern Europe, according to EdUniversal [Saint Petersburg State University – RU]

(3) UAB is the second university in Spain, the 49th in Europe and 119 in the world according to the 2015-2016 edition of the URAP ranking [Universitat Autonoma de Barcelona – ES]

Turning to the last features among stance adverbials, attitude markers (including adverbials such as *hopefully* and *wisely*) are the least frequent in both varieties of English, suggesting that they are less typical of promotional language in academic settings, as compared to grammatical devices. Still, attitude adverbials are more common in native English universities which rely on a wider range of linguistic resources compared to ELF ones, a finding that is consistent with a large body of work on native vs. non-native language (cf. Granger 1998). For instance, while *importantly* and *hopefully* are shared by both varieties of English, other adverbs such as *astonishingly*, *surprisingly* and *wisely* are found exclusively in the native English subcorpus, suggesting that native English universities choose more extreme forms of boasting.

As for stance complement clauses (Figure 6.15), the normalized frequency of stance verbs followed by a that-clause is relatively low in both subcorpora, although native English texts make larger use of this promotional strategy as compared to ELF texts. As stated in Biber et al. (1999:675), *that-clauses* controlled by verbs such as *think* and *know* are a typical feature of interpersonal communication (*I know that you, I think it is*), while academic prose favours extraposed that-clauses controlled by adjectives (e.g. *it is likely that*, *it is possible that*). Furthermore, this gap betweeen native English and ELF texts in the use of *stance that-clauses* is even more evident if we consider less formal usage of that-clauses, such as the occurrences in which *that* is omitted (e.g. *we believe you will*). A quick search for stance

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | believe that | 27 |
| P \| N | hope that | 26 |
| P \| N | know that | 13 |
| P \| N | recognise that | 9 |
| P \| N | say that | 7 |
| P \| N | feel that | 7 |
| P \| N | think that | 4 |
| P \| N | recognize that | 3 |
| P \| N | expect that | 2 |
| P \| N | conclude that | 2 |
| P \| N | determine that | 1 |
| P \| N | claim that | 1 |
| P \| N | argue that | 1 |

(a) Native English stance verbs.

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | believe that | 15 |
| P \| N | say that | 10 |
| P \| N | think that | 7 |
| P \| N | hope that | 7 |
| P \| N | know that | 5 |
| P \| N | expect that | 3 |
| P \| N | feel that | 2 |
| P \| N | claim that | 2 |
| P \| N | worry that | 1 |

(b) ELF stance verbs.

Fig. 6.17 List of native English and ELF stance verbs followed by that-clauses in highly promotional texts.

verbs + that-clauses (without *that*)[18] shows that this pattern is used twice as many times in native English texts as in ELF texts. In addition, the search for this feature produced consistent results with those obtained from the analysis of attitude markers. While some recurrent verbs like *believe*, *hope*, *say*, *think*, and *know* are top-ranked in both the native English and ELF lists of stance verbs (Figure 6.17), a number of other evaluative verbs tend to be exclusive of native English texts. This is the case of *argue*, *conclude*, *determine* and *recognise*. In other words, it looks like native English authors draw from a larger pool of linguistic resources, while ELF authors make use of the most frequent verbs only, although they might well be familiar with many other linguistic resources.

A second point emerges from this set of data. Quite surprisingly, three out of four of the verbs mentioned above that are exclusive to the native English component (namely, *conclude*, *determine* and *recognise*) fall under the "epistemic certainty" category in Biber's list of stance features (Biber, 2006a). This could be interpreted as either a form of lexical simplification in which ELF users rely on high-frequency verbs exclusively (i.e. *know*) or a cultural shift in which ELF universities tend to use different strategies to evaluate themselves as compared to native English universities. Although these two hypotheses are not mutually exclusive, the latter may also explain why the verb *say* (Example 1 below) appears to be the second most frequent verb in this specific pattern of stance within the ELF subcorpus (Figure 6.17).

Once again, while ELF texts make large use of communication verbs to provide explicit evidence of their excellence (Example 1 below), native English texts favour verbs like *know* and *hope* (Examples 2 and 3 below), which express either certainty or future beliefs:

[18]The query used to check this feature was: `[lemma="conclude|determine|know|recognise|recognize|assume|feel|argue|claim|report|say|believe|doubt|guess|think|expect|hope|worry" &tag="V.*"][tag="DET|N.*|PP.*"][tag="V.*"]within<textcompuff="1"/>`

(1) Östling says that KTH has in recent years performed very well in this dimension [KTH – Royal Institute of Technology – SE]

(2) We also know that international students have made the right choice in coming to UCC, and that not only will they receive an excellent medical education, but they will make friends for life [University College Cork – IE]

(3) We hope that you will enjoy the challenge and inspiration of taking forward your education at King's [King's College London – UK]

Therefore, although this lexico-grammatical pattern shows a relatively low frequency in both varieties of English, subtle differences exist that might reveal specific native English and ELF discourse patterns, whether deliberate or not. In particular, ELF texts show a lower degree of variety in the choice of verbs and tend to exploit the high-frequency verb *say* in order to report an external opinion or factual evidence.

The analysis of stance verbs followed by to-clauses (e.g. *wish to enrich your knowledge*) led to different but arguably complementary results. It is interesting to note that not only are the top lemmatized verbs introducing a to-clause identical in native English and ELF promotional texts, but they also show the same rank ordering (Figure 6.18). Yet, there is an exception to that, namely the pattern *allow to*, which is ranked sixth in the ELF subcorpus, while it does not appear even lower down in the native English list of stance verbs. The qualitative analysis of *allow to* shows that, despite the regulatory nature of this verb, ELF texts use it to express an opportunity rather than an authorization:

(1) Participation in the 5th and 6th RTD Framework Programs [...] has allowed to develop Centers of Excellence [University of Latvia – LV]

(2) We wish to be an inspiring, progressive and gender-aware institution, attracting the world's best scientists and most gifted students in the knowledge that their talents will be allowed to develop fully [Delft University of Technology – NL]

On the one hand, this particular usage conveys a bureaucratic or regulatory nuance which increases complexity, and on the other hand, it shifts the centre of attention from the opportunity itself to its initiator. This switch in perspective is likely to enhance the status of whoever creates the opportunity (i.e. the university), as compared to a student-centred perspective where the main initiator is hidden, and the focus is placed on either the addressee or the opportunity itself.

Therefore, two conclusions can be drawn from the analysis of this feature of stance, i.e. verbs + to-clauses (e.g. *wish to pursue*). First, native English and ELF texts show many similarities, suggesting that there is a common set of features widely used in both contexts

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | want to | 150 |
| P \| N | need to | 123 |
| P \| N | wish to | 89 |
| P \| N | help to | 66 |
| P \| N | intend to | 30 |
| P \| N | try to | 15 |
| P \| N | hope to | 14 |
| P \| N | tend to | 6 |
| P \| N | love to | 5 |
| P \| N | happen to | 4 |
| P \| N | consider to | 3 |
| P \| N | attempt to | 3 |
| P \| N | seem to | 2 |
| P \| N | believe to | 2 |

(a) Native English stance verbs.

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | want to | 143 |
| P \| N | need to | 107 |
| P \| N | wish to | 82 |
| P \| N | help to | 53 |
| P \| N | intend to | 39 |
| P \| N | allow to | 16 |
| P \| N | try to | 11 |
| P \| N | hope to | 10 |
| P \| N | consider to | 9 |
| P \| N | tend to | 7 |
| P \| N | love to | 4 |
| P \| N | seem to | 2 |
| P \| N | attempt to | 1 |

(b) ELF stance verbs.

Fig. 6.18 List of native English and ELF stance verbs followed by to-clauses in highly promotional texts.

and that university websites' authors are experienced practitioners of the English language. Second, and to a lesser degree, some typical features of ELF emerge which might be due to differences in the cultural footprint of ELF authors, as a sort of conscious or unconscious mental map.

Turning to adjectives followed by that- and to-clauses (e.g. *sure that* and *impossible to*), findings indicate some trends that were already discussed above. First, in terms of frequencies, native English and ELF texts behave in a rather similar way. It will be remembered from Figure 6.15 that that-clauses are generally less common than to-clauses in both English subcorpora. Second, native English texts show again a higher variety of lexis, particularly with regard to adjectives followed by to-clauses, such as *able to*, *proud to* and *necessary to* (Figure 6.20). Furthermore, if we look at the list of stance adjectives introducing that-clauses (Figure 6.19), native English texts strategically adopt very positive adjectives such as *sure*, *confident*, *clear*, *important*, *proud*. On the other hand, the most frequent adjective among ELF texts is *aware*, which has rather negative connotations, usually referring to some risks or issues that need to be dealt with in everyday language (e.g. *smokers are well aware of the dangers to their own health; She was acutely aware of the noise of the city*).[19] This choice is made either to explicitly acknowledge one's inadequate services (Example 1 below) or used as a synonym of "know" (Example 2 below). Besides contributing to developing a much more complex structure, as in Example 2, the latter strategy may lead to a misinterpretation since the reader may well doubt that, e.g., the relation with international research communities is a common practice at the university.

(1) Considering the points of development, we are aware that we have to improve our internationalisation [University of Turku – FI]

[19] https://www.collinsdictionary.com/dictionary/english/aware [last consulted on 15 December 2017].

| | word | Frequency |
|---|---|---|
| P \| N | sure that | 10 |
| P \| N | confident that | 7 |
| P \| N | clear that | 6 |
| P \| N | important that | 5 |
| P \| N | proud that | 4 |
| P \| N | likely that | 4 |
| P \| N | aware that | 4 |
| P \| N | vital that | 3 |
| P \| N | noteworthy that | 2 |
| P \| N | essential that | 2 |
| P \| N | surprising that | 1 |
| P \| N | imperative that | 1 |
| P \| N | great that | 1 |
| P \| N | fortunate that | 1 |
| P \| N | critical that | 1 |
| P \| N | amazing that | 1 |

(a) Native English stance adjectives.

| | word | Frequency |
|---|---|---|
| P \| N | aware that | 5 |
| P \| N | sure that | 4 |
| P \| N | important that | 4 |
| P \| N | clear that | 3 |
| P \| N | glad that | 2 |
| P \| N | essential that | 2 |
| P \| N | crucial that | 2 |
| P \| N | true that | 1 |
| P \| N | surprising that | 1 |
| P \| N | proud that | 1 |
| P \| N | possible that | 1 |
| P \| N | likely that | 1 |
| P \| N | imperative that | 1 |
| P \| N | great that | 1 |

(b) ELF stance adjectives.

Fig. 6.19 List of native English and ELF stance adjectives followed by that-clauses in highly promotional texts.

(2) As an institute of higher learning and research, the ETH Zurich cultivates an international standing. It is aware that its scientific contribution has to be confirmed by the international research community [ETH Zurich – CH]

Another aspect to note in this list of adjectives + that-clauses is that *confident* is the second most frequent adjective in promotional native English texts, where it is used with a persuasive function to encourage students:

(1) We want you to be confident that this is the sort of environment, wherever you come from, where you can be yourselves [University of Oxford – UK]

Conversely, it does not appear at all in the list of ELF adjectives followed by that-clauses, as Figure 6.19 illustrates. In a context where trust between university and its stakeholders is becoming such a key university value, it is questionable whether a native English high-frequency pattern such as *confident that* is lacking in the ELF promotional subcorpus because of a strategic choice.[20]

The last feature to be analysed in this set of stance complement clauses is nouns followed by that-clauses (e.g. *view that*, *conclusion that*).[21] Although this feature is quite rare in both varieties of English, in the ELF subcorpus it is twice as frequent than the native subcorpus,

[20] A quick search for *confident that* in the whole ELF subcorpus with no functional restriction shows that this expression scores highly on the A7 dimension (academic writing) and on the A8 dimension (hardnews) as well, since it is used to present research outputs, such as *We're confident that this method can improve the reconstruction of events at a crime scene* (University of Amsterdam – NL) or *Arenz is confident that his discovery can show the way for economically viable fuel cell production* (University of Copenhagen – DK).

[21] Nouns followed by to-clauses (e.g. *failure to*) are not considered here because there are no occurrences in this set of promotional texts.

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | able to | 140 |
| P \| N | proud to | 33 |
| P \| N | keen to | 28 |
| P \| N | easy to | 22 |
| P \| N | necessary to | 21 |
| P \| N | likely to | 18 |
| P \| N | eligible to | 17 |
| P \| N | happy to | 16 |
| P \| N | possible to | 13 |
| P \| N | important to | 13 |
| P \| N | free to | 12 |
| P \| N | willing to | 9 |
| P \| N | difficult to | 7 |
| P \| N | hard to | 6 |
| P \| N | eager to | 6 |
| P \| N | sure to | 5 |
| P \| N | essential to | 5 |
| P \| N | due to | 4 |
| P \| N | lucky to | 2 |
| P \| N | impossible to | 2 |
| P \| N | grateful to | 2 |
| P \| N | glad to | 2 |
| P \| N | unlikely to | 1 |
| P \| N | unable to | 1 |
| P \| N | nice to | 1 |
| P \| N | fit to | 1 |
| P \| N | certain to | 1 |
| P \| N | afraid to | 1 |

(a) Native English stance adjectives.

| | lemma (lowercase) | Frequency |
|---|---|---|
| P \| N | able to | 236 |
| P \| N | proud to | 19 |
| P \| N | necessary to | 17 |
| P \| N | possible to | 16 |
| P \| N | free to | 16 |
| P \| N | important to | 15 |
| P \| N | happy to | 14 |
| P \| N | eager to | 13 |
| P \| N | willing to | 11 |
| P \| N | eligible to | 9 |
| P \| N | difficult to | 8 |
| P \| N | essential to | 7 |
| P \| N | easy to | 7 |
| P \| N | sure to | 6 |
| P \| N | keen to | 6 |
| P \| N | likely to | 4 |
| P \| N | unable to | 3 |
| P \| N | glad to | 3 |
| P \| N | apt to | 2 |
| P \| N | hard to | 1 |
| P \| N | due to | 1 |
| P \| N | certain to | 1 |
| P \| N | afraid to | 1 |

(b) ELF stance adjectives.

Fig. 6.20 List of native English and ELF stance adjectives followed by to-clauses in highly promotional texts.

suggesting that ELF texts are more likely to adopt complex noun phrases. This is particularly evident in the expression *fact that*:

(1) Teacher Awards are further proof of the fact that the didactic talent of the lecturers at Utrecht is highly valued [Utrecht University – NL]

(2) We are proud of the fact that our graduates go on to assume important positions in our society [University of Ljubljana – SI]

As can be noted from Examples 1 and 2, not only does this expression lead to a rather complex structure, but it also tends to be superfluous, since both sentences can convey the same meaning even if *fact that* is omitted (*further proof that*; *we are proud that*). Since the omission of *fact that* would create a simpler structure, facilitating reading and understanding, it can be argued that this is a further example of ELF complexity, formality and over-explicitness.

### 6.2.3 Attitudinal lexis

The semantic system of appraisal can serve as a tool to describe explicit evaluative lexis (referred to as "attitude" in Martin and White (2005), as will be remembered from Section 2.3 of this thesis). Linguistic resources for grading lexis, i.e. *graduation*, and resources

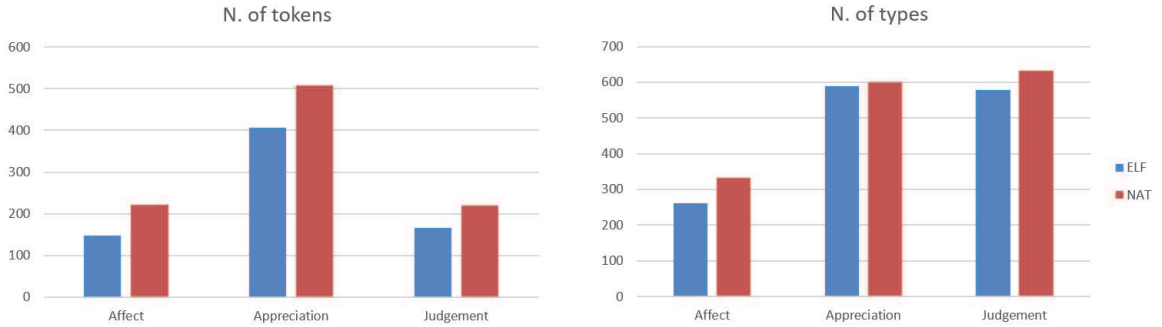


Fig. 6.21 Distribution of attitudinal lexis in native English and ELF texts.

| | NAT | ELF |
|---|---|---|
| N. of types | 156 | 142 |
| N. of tokens | 949 | 719 |

Table 6.2 Different types (per 10K) and tokens (per 100K) of attitudinal lexis.

of *engagement* that also compose the system of appraisal in Martin and White (2005), have already been analysed within other lexico-grammatical patterns above. The analysis of attitudinal lexis expands Sections 6.2.1 and 6.2.2 of this Chapter by presenting lexical resources of evaluation systematically and using a different approach (cf. Section 2.3 on the different approaches to evaluation).

By looking at Figure 6.21 which displays the distribution of attitudinal lexis in native English and ELF texts, one aspect becomes immediately clear: in terms of both types and tokens, native English texts make a larger use of explicit evaluative lexis in any category of attitude (see also Table 6.2). Furthemore, the percentage of coverage (i.e. the number of attitudinal lexical items found in the two subcorpora over the total number of attitudinal lexical items included in the categories of Attitude, Appreciation and Judgement) is higher in the Appreciation category. This might not be surprising considering that the Appreciation category includes resources to construct the *value* of things (Martin and White, 2005:36). Conversely, the category of Affect (which deals with resources for constructing emotional reactions) and the cateogry of Judgment (which deals with attitudes towards behaviours) play a less central role in both subcorpora.[22]

[22]In order to avoid skewed results, some lexemes that were consistently used in a different meaning were not counted and were removed from the search query. For instance, the adjective *natural* is frequent among ELF texts in expressions like *natural and social sciences* and was therefore discarded altogether.

| | Types in category | Types in NAT | Types in ELF |
|---|---|---|---|
| Affect | 92 (100%) | 31 (34%) | 24 (26%) |
| Appreciation | 114 (100%) | 56 (49%) | 54 (47%) |
| Judgement | 156 (100%) | 59 (38%) | 53 (34%) |

Table 6.3 Percentage of coverage in each category of attitude.

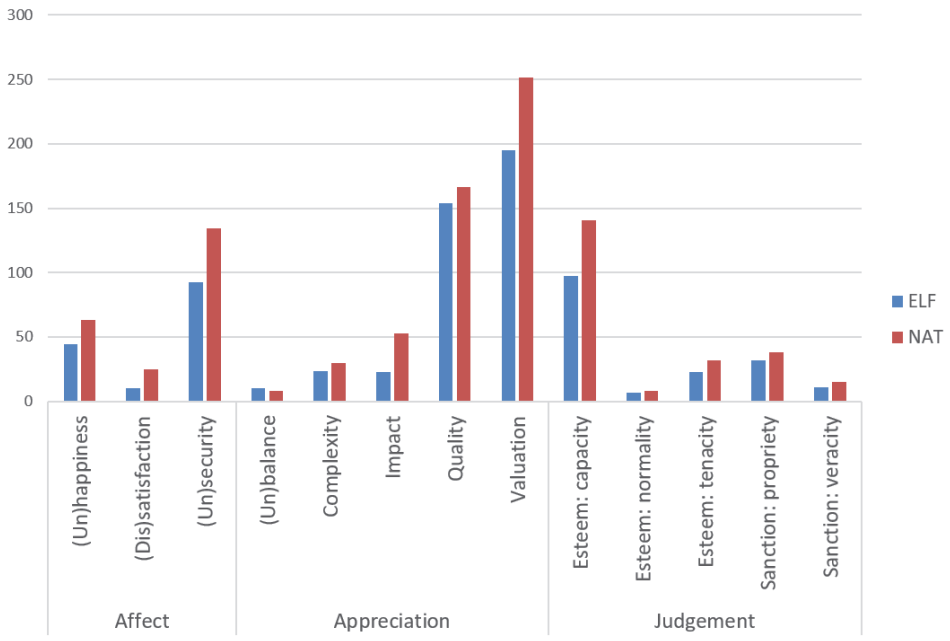


Fig. 6.22 Normalised frequency of attitude subcategories in native English and ELF websites.

| | | |
|---|---|---|
| P \| N | satisfaction | 38 |
| P \| N | reward | 30 |
| P \| N | pleasure | 12 |
| P \| N | busy | 9 |
| P \| N | involved | 7 |
| P \| N | compliment | 4 |
| P \| N | attentive | 2 |
| P \| N | satisfied | 1 |
| P \| N | pleased | 1 |
| P \| N | industrious | 1 |
| P \| N | bored | 1 |

(a) Native English lemmas.

| | | |
|---|---|---|
| P \| N | reward | 17 |
| P \| N | attentive | 6 |
| P \| N | pleasure | 5 |
| P \| N | satisfaction | 4 |
| P \| N | busy | 4 |
| P \| N | involved | 3 |
| P \| N | satisfied | 1 |

(b) ELF lemmas.

Fig. 6.23 List of native English and ELF lemmas expressing Satisfaction.

The Affect category includes values of Happiness, Satisfaction and Security, which all refer to positive emotions and feelings (Figure 6.22).[23] While the value of Happiness is delivered through very similar lexical items in both ELF and native texts (*like, love and happy* are the most frequent lemmas), (Dis)satisfaction points to some interesting results.

The value of Satisfaction is at the very top of the native English list of lemmas, while ELF texts refer predominantly to Reward (Figure 6.23). Satisfaction and Reward are two sides of the same coin: they both increase university value, but they do it from very different perspectives. Satisfaction is related to a student perspective since it refers to students' degree of success in their study and/or future work (Example 1), while Reward focuses on university ability to satisfy its students (Example 2) or to gain public recognition (Example 3).

(1) We will measure success through: [....] Increase in student satisfaction scores as measured by undergraduate and postgraduate student surveys [Warwick University – UK]

(2) We do reward students for merit and provide a strong basis for future career success [Polito]

(3) the UvA has been rewarded the HR Excellence Log [University of Amsterdam]

It will be remembered from Section 6.2.2 that other lexico-grammatical features of stance are consistent with these findings, which may indicate an underlying strategy or a cultural specificity.

As for the value of (In)security, some very frequent lemmas include *commit*, *security* and *confidence*, related to *confident* a few lines further down in the list (Figure 6.24). Similarly to what was noted in Section 6.2.2 with reference to the phrase *confident that*, the words *confidence* and *trust* both show high frequencies among native English texts, while ELF

[23] Note that both varieties of English adopt positively-loaded lexis, while there are no or very few negatively-loaded words. Yet, since this section includes explicit lexical items, no effort was made to account for grammatical negation, as in "not happy".

| | | |
|---|---|---|
| P \| N | together | 364 |
| P \| N | commit | 180 |
| P \| N | security | 69 |
| P \| N | confidence | 49 |
| P \| N | trust | 43 |
| P \| N | confident | 18 |
| P \| N | comfortable | 11 |
| P \| N | restless | 2 |
| P \| N | anxious | 2 |
| P \| N | uneasy | 1 |
| P \| N | entrust | 1 |

(a) Native English lemmas.

| | | |
|---|---|---|
| P \| N | together | 363 |
| P \| N | commit | 112 |
| P \| N | security | 37 |
| P \| N | trust | 14 |
| P \| N | comfortable | 9 |
| P \| N | confidence | 5 |
| P \| N | entrust | 3 |
| P \| N | confident | 1 |

(b) ELF lemmas.

Fig. 6.24 List of native English and ELF lemmas expressing Security.

texts mainly rely on the word *trust*. While il seems that ELF texts consistently favour the word *trust* rather than *confidence*, it might be more difficult to find out the reasons why this happens. It could derive from institutional guidelines referring to one term instead of the other, it could be due to an influence from their first languages, or it might be just another example of the overall reduction of English vocabulary among ELF speakers that was already noted in previous studies (Mauranen 2012).

Turning to the Appreciation category, the value of Balance is one of the two cases in which ELF texts show a higher frequency of certain lemmas (*balance* in particular) as compared to native English texts. The analysis of concordance lines points to a number of different contexts in which this lemma is used: the balance between time spent in class and individual study, work-life balance, gender balance, balance between teaching and research. On the other hand, ELF universities seem to prioritise work-life balance in this set of promotional texts, where this sense accounts for almost one third of the total number of occurrences for *balance*. Examples include *work-life balance*, *balance between career and private life*, *balancing family and career*, *balance between work/study and family life*. Fewer occurrences among ELF texts refer to the balance between theory and research or gender balance. A search by university name shows that most of these occurrences come from Northern European countries (Germany, Denmark and the Netherlands).

As for the values of Complexity and Impact, top-ranked lemmas are rather similar in both varieties of English and include *rich*, *clear*, *detailed* in the Complexity category, and *exciting*, *lively* and *engaging* in the Impact category. Lexis from both of these categories is used to a greater extent in the native subcorpus, where it also includes a higher number of low-frequency types, specifically *unclear*, *plain* and *arcane*. While high-frequency words convey positive meaning, low-frequency ones may include negatively-loaded words that work as strategies for conflict resolution and show university support system:

(1) If these [rules] are unclear, first contact the International Student Team [University of Oxford – UK]

The Quality and Valuation subcategories seem to be the most standardised since top-ranked lemmas are very similar in both types and frequencies. *Good*, *welcome* and *beautiful* are high-frequency lexemes referring to the value of Quality, while *innovative*, *unique*, *creative*, *effective* and *common* refer to the Valuation subcategory. It is somehow paradoxical that these texts are declaring their uniqueness and innovative aspects through the use of such standardised lexis. This tension between uniqueness and commonplaceness also emerges from the list of evaluative words. This is one of the few cases in which two opposite lexemes are both regarded as positive, namely *common* and *unique*. They are indeed used to deliver promotional messages even within the same university:

(1) Åbo Akademi University is an international, inspiring and unique university. [University of Turku – FI]

(2) The strategy is based on the common values of the University: ethicality, criticality, creativity, openness and communality [University of Turku – FI]

Turning to the last category, corresponding to the value of Judgement, most of the lexis positively assesses the university according to various principles, especially teachers' experience and university ethical issues. There are no particular points to be made here since occurrences are rather few. Yet, it is remarkable that the value of Tenacity is slightly more frequent among ELF texts, against what was observed in the other categories. The analysis of concordance lines shows that this higher frequency is due to the adjective *thorough*, which does not appear at all in the list of native English attitudinal lexis:

(1) You'll have thorough knowledge of current developmental, learning, and teaching theories as well as the neurobiological basis of education and learning [Utrecht University]

(2) Relations programme provides students with a thorough understanding of the politics and power of Europe [University of Amsterdam – DK]

As Examples 1 and 2 suggest, *thorough* is used very positively in the ELF subcorpus to assess university performance in delivering detailed, precise and meticulous courses.

## 6.3 Discussion

This last section discusses the main findings with regard to the initial research questions, further characterising the similarities and differences between native English and ELF academic websites.

The first part of the analysis (Section 6.1) showed that the texts with a promotional function in the native English and ELF promotional subsets belong to a similar range of genres, although they are distributed in different proportions. There are very few cases in which a genre occurs in one of the varieties only. Specifically, staff profiles were found in the native English subset exclusively, while contact pages were found among ELF texts. In addition, the genres showing a large difference in proportion between native and ELF texts can be associated with an alternative genre in the other variety of English since they both perform the same function. For example, the ELF component has a larger number of *About us* pages, while the native English component has more *Why choose us* pages. Although these two genres have different names, in fact they contribute to achieving the same purpose, namely to advertise the university either explicitly or implicitly. Moreover, considering that *Why choose us* pages can be found as subsections of *About us* pages, there are good reasons to state that these two genres fall within the same functional macro-category. In this sense, the promotional genres of native English and ELF academic websites are sufficiently similar to justify the linguistic comparison carried out in this thesis. In other words, native and ELF university websites have been found to resort to similar genres to advertise themselves, such that the micro-differences observed in Sections 6.1 and 6.2 can be safely attributed to different promotional strategies rather than different genres. Since the analysis in Chapter 6 focuses on a selection of texts among those automatically classified as promotional, no general conclusions can however be drawn about the totality of the institutional academic genres, beyond the promotional ones.

In addition to any difference that may characterise each of the two genres from a structural point of view (cf. section 2.4), their very name is one of the elements signalling that *Why choose us* pages are explicitly promotional, while *About us* pages hide their promotional purpose behind their stated informative function. In other words, while *About us* pages (prominent among promotional ELF pages) explicitly provide information and implicitly convey promotional messages, *Why choose us* pages (more numerous in the native subcorpus) work the other way round, giving priority to the promotional aspect. The same is true for the genre of course descriptions. As will be remembered from the analysis in Section 6.1, while genres such as *Why choose us* are typical of corporate websites, and are being imported into academic communication, course descriptions are "native" academic genres. It is especially revealing that these informative academic genres that are now in the process of being infiltrated with promotional features are more typical of ELF websites than native ones. The genre of research descriptions is a further example of ELF academic genres turning into promotional ones.

A final caveat is in order at this point. Most research description pages were found in universities from Austria, Finland, Germany, Switzerland and the Netherlands. In this thesis no attempt was made to distinguish different trends belonging to specific ELF countries or groups of countries. The careful process of corpus construction and the qualitative analysis carried out in Chapter 6 were meant to ensure that the generalizations made apply to ELF universities in general. While typical trends of specific (groups of) universities were ignored here, they might make the object of further, more focused studies aiming to highlight how a university lingua-cultural background spills over into its English-language communication strategies.

Summing up, the comparison of native and ELF promotional websites has pointed to a substantial overlap, but also to several differences. The latter can be summarised thus: openly promotional genres coming from a business-related context (such as *Why choose us* pages and staff profiles) are more common in the native English subcorpus. Conversely, most of ELF texts belong to typical academic genres that were enriched with promotional features. Besides confirming Fairclough's theory according to which new hybrid promotional genres are arising (Fairclough, 1993:145), this may be a first indicator that English universities are pushing their corporate-based strategies even further then ELF ones.

As seen in the first part of the analysis, native English and ELF subcorpora are comparable in terms of genres, which provided the necessary common ground to carry out a comparison of linguistic features expressing evaluation within European academic websites. While the relevant literature provides a full set of evaluative indicators that can be used across any language register or text type (spoken, written, disciplinary academic English and so on), research findings in Section 6.2 highlighted some recurrent strategies for the specific case of institutional academic English. Findings show that, from a grammatical perspective, some commonalities can be identified. The verbs *will* and *can* are the most common verbs to express modality in both sets of promotional texts, which are indeed associated with possibilities/opportunities and future prospects. On the other hand, modals of necessity and obligation (e.g. *must*, *should*), as well as modals of volition (*would*) tend to be scarse overall. The same is true for the grammatical categories of imperatives and questions, which show relatively low frequencies in both subcorpora. Two semantic categories of imperatives are shared between native English and ELF subcopora, namely verbs that connect website sections (e.g. *find*, *read*, *see*, *learn*, *visit*) or verbs that prompt a response in the real world, such as *join* and *apply*.

Within lexico-grammatical forms, markers of certainty (e.g. *certainly*, *definitely*) and amplifiers or intensifiers (*very*, *highly* and *fully*) are among the most common promotional strategies in both varieties of English. Some common markers of certainty include *really*,

*clearly*, *sure*, *certainly*, *definitely* and *of course*. As for amplifiers and intensifiers, *very* and *highly* account for most of the normalized frequencies shown in Figure 6.15. Another common feature of evaluation with regard to stance complement clauses are stance verbs or adjectives controlling to-clauses. The top lemmatized stance verbs are *want*, *need*, *wish*, *help*, *intend* and *try*. These stance verbs showed the same rank ordering in the native English and ELF subcorpora, suggesting that they are key within promotional texts. As for adjectives + to-clauses, native English and ELF texts behave in a similar way in terms of frequencies since *able to*, *proud to* and *necessary to* are all ranked high. As compared to to-clauses, that-clauses controlled by stance verbs or adjectives (*we believe that*, *it is possible that*) are definitely less common, while stance nouns followed by that-clauses (e.g. *idea that*, *view that*) are the least frequent. Turning to low-frequency stance adverbials within lexico-grammatical forms, attitude markers (including adverbials such as *hopefully* and *wisely*) are equally rare.

From a purely lexical perspective, most of the evaluative items from Martin and White's list belong to the Appreciation category, which is supposed to turn feelings into resources for constructing the value of things (Martin and White, 2005:36). Among the range of meanings associated with Appreciation, the most common ones include Quality, which answers to the question "did I like it?" and Valuation itself, which answers to the question "is it worthwhile?" (Martin and White, 2005:56). The Quality and Valuation subcategories seem to be the most standardised since top-ranked lemmas are very similar in both types and frequencies. *Good*, *welcome* and *beautiful* are high-frequency lexemes referring to the value of Quality, while *innovative*, *unique*, *creative* and *effective* refer to the Valuation subcategory. Conversely, the category of Affect – which deals with resources for constructing emotional reactions (e.g. *like*, *love*, *happy*) – and the category of Judgment – which deals with attitudes towards behaviours (e.g. *expert*, *successful*, *generous*, *ethical*) play a minor role in both subcorpora.

The picture that emerges is one in which web-based strategies for promoting universities from a semantic point of view focus on the delivery of the most traditional academic values and qualities, through the category of Appreciation, which seems to be rather standardised across native English and ELF texts. Universities describe themselves as being innovative, creative and effective institutions, with specific elements that make them unique as compared to other universities in Europe or in the world. A tension between uniqueness and commonplaceness emerged from the list of frequent evaluative words, in which two opposite lexemes, namely *unique* and *common*, are both associated with positive values. On the other hand, the category of Judgment is mainly aimed at praising the capacities of university staff, whereas evoking affective feelings – such as happiness and positive experiences – through the system of Affect is not a typical strategy.

A comparable picture emerges from lexico-grammatical indicators of evaluation in which attitudinal markers expressing the writer/author feelings are definitely rare. Finally, and from a lexico-grammatical perspective again, the general preference for to-clauses can be interpreted as an attempt to simulate interpersonal communication, considering that to-clauses are more common in the spoken registers than the written registers (Biber, 2006a:108).

Although several similarities were observed across the two subcorpora, differences also emerged. For instance, the strategic use of patterns that are typical of spoken registers is more noticeable in the native English subset where most of the evaluative features are consistently higher as compared to ELF texts, with few exceptions that will be mentioned below. From a grammatical perspective, the largest differences in proportions between native English and ELF texts were found in the use of first and second person pronouns, especially *our*, *you*, *your* and *we*, which make texts similar to face-to-face communication. In particular, it is remarkable that native English texts make substantial use of the pronoun *our* in an inclusive, rather than exclusive way. Besides engaging readers through personal pronouns, a further point involves the greater use of contracted verb forms as immediate right-hand collocates of *you* and *we*, within native English texts only. The ability of native English writers to take advantage of evaluative grammatical features to successfully engage readers is also evident from the analysis of questions. Although from a quantitative point of view, questions seem to be more typical of the ELF context, a further investigation into the *type* of questions showed that native English universities display various instances of more engaging how-questions, as compared to general-purpose questions within ELF texts. In contrast with this general prominence of evaluative features in native English texts, the modal verb *must* was found to be slightly more frequent in ELF texts. Looking at concordance lines in the ELF promotional subcorpus, it was noted how *must* is used persuasively as a self-constraint in which the university commits itself to reach strategic goals.

Turning to lexico-grammatical markers of evaluation, frequencies are again consistently higher in the native English subset with two exceptions, i.e. markers of perspectives (e.g. *generally, personally, according to*) and stance nouns followed by that-clauses (e.g. *fact that, view that*). An examination of the markers of perspective indicated that the higher normalized frequency in the ELF subcorpus is the result of one marker (i.e. *according to*), which serves to convey evaluative meaning through an external source of knowledge. In other words, ELF universities are less likely to boast about their own success, and resort to a third-party opinion to buttress their claims. A further point consistent with this observation was made through a closer look at other features of evaluation. For instance, two comparable expressions were found to be frequently used in the two varieties of English, namely *very best* in native English texts and *very good* in ELF texts. Besides expressing different degrees

of promotion, these two expressions were also used differently. In native English university websites, no attempt was made to justify such a controversial – although quite common – expression, whereas ELF universities were more likely to provide contextual evidence of their claims.

A second exception going against the general trend of native English texts using more evaluative features than ELF ones is represented by stance nouns controlling that-clauses. Although this feature was found to be quite rare in both varieties of English, in the ELF subcorpus it was twice as frequent than in the native English one, due to the higher number of occurrences of *fact that*. In Section 6.2 it was observed that the omission of this expression would not undermine communication at all; rather, it would create a simpler structure and facilitate reading and understanding. Besides reducing complexity and/or formality, the same structure can be regarded as an example of linguistic over-explicitness characterising ELF university websites.

Finally, other minor observations were made regarding the list of stance adjectives controlling *that-* or *to-* clauses. While native English texts strategically adopt very positive adjectives (e.g. *sure*, *confident*, *proud* and the like), *aware* was found to be the most frequent stance adjective followed by that-clauses among ELF texts. Besides lowering the degree of promotion as compared to native English texts, this adjective was revealed to have rather negative connotations, since it refers to some issues that need to be dealt with or something that is missing. This point can be somehow related to previous observations about ELF universities making a larger use of the modal verb *must* to commit themselves to reach an ideal standard. Both of these features create a consistent picture where ELF universities seem to build their strategic communication around self-commitment rather than self-assertion.

Summing up, this Chapter has offered a global view on the promotional strategies used within academic websites, which makes it possible to abstract results and provide an overall interpretation of promotional strategies between ELF and native English universities, both on the level of genre and on the level of language. If taken together, all these observations can point to two parallel trends associated with promotional strategies in academic websites. On the one hand, university promotional pages in native English and ELF show a substantial overlap. They have been found to resort to similar genres to advertise themselves while, from a linguistic point of view, they seem to pursue an increasingly spoken-like type of discourse. Concurrently, a number of differences were found which can be summarised through at least three main points. First, native English texts adopt openly promotional genres, while ELF universities rely on existing academic genres to which they add linguistic promotional strategies. In other words, ELF universities show some reluctance to resort to explicit advertising material, such as openly promotional genres coming from corporate websites,

which lead them to increase the degree of promotional features within more typical academic genres. Second, native English texts display an informal and colloquial style leading to a higher degree of engagement, whereas ELF texts show a higher degree of complexity.[24] Third, native English websites make use of implicit promotional strategies for self-assertion, while ELF ones favour self-commitment and an "ethical" approach to promotional strategies. In contrast with the native English trend to adopt "boastful" strategies, ELF universities rely more on external sources of knowledge to support their self-evaluating statements or they commit themselves to do their best (rather than presenting themselves as "the best and only" option). In other words, native English texts would seem to adopt corporate strategies both in their promotional genres and in their promotional language, while ELF universities may still prefer to build their promotional strategies on institutional credibility and accountability.

The next Chapter will provide a brief summary of the full thesis, including its main outputs, limitations and suggestions for further research.

[24]One may argue that ELF complexity contrasts with the higher variety of lexico-grammatical features found in native English texts as compared to ELF ones. Yet, findings have shown that ELF texts do not necessarily resort to high-frequency simple structures. For instance, the most frequent adjective controlling that-clauses in the ELF subcorpus is *aware*, while in the native English subset, *aware* appears lower down in the list of adjectives (Figure 6.19). As shown in the examples provided in Section 6.2.2, *aware* followed by that clauses is used in sentences like *we are aware that* as a less straightforward alternative of *we know that*. Another important point here is that *know that* is less frequent in ELF texts as compared to native English ones (Figure 6.17), which may confirm the ELF preference for *be aware that* as a more complex alternative of *know that*. Finally, a further point to note is that one lexico-grammatical feature only was found to be twice as frequent in the ELF subcorpus as compared to the native English one, namely nouns controlling that clauses. This feature is not typical of colloquial style since it is generally restricted to academic registers like textbooks (Biber, 2006a:109).

# Chapter 7

# Conclusion, limits and future work

This Chapter summarises the study, presents its limitations and offers suggestions for future work.

Chapter 1 introduced the motivations for this research and listed the key research questions, objectives and final goals. This research aimed to investigate the promotional function as it is expressed in university websites in native English and ELF. Focusing on genres and language features, it seeked to identify shared strategies to promote European universities on the web, as well as any differences that may characterise either variety of English.

Chapter 2 reviewed the most relevant studies in institutional academic settings (Section 2.1), it discussed the main approaches to English as a Lingua Franca (Section 2.2) and evaluation language (Section 2.3). Section 2.4 addressed the controversies related to the concept of genre, text type and register, attempting to provide a working definition for each of these terms, while Section 2.5 provided a brief account of web-as-corpus techniques and the different approaches to text classification. It also illustrated the potential and limitations of the "web-as-corpus" methodology for building specialised corpora from the web.

Chapter 3 described the procedure for corpus construction and its internal composition, while Chapter 4 focused on automatic classification based on Functional Text Dimensions (FTDs). Besides offering an account of how automatic classification was performed, Chapter 4 includes a manual assessment of classified pages, focusing on the promotional dimension. Although very much dependent on preliminary human judgement, the good performance of the method was proved by an extensive number of examples provided in Section 4.5, which deals with a post-hoc evaluation of text classification, and in Chapter 6, particularly in the qualitative examination of concordance lines. Section 6.1 also confirms that the method works since most of the webpages were found to belong to externally-defined promotional genres (e.g. About us, Why choose us). Yet, a limitation of this method is that it may not assign a promotional function to ELF texts that are written with a promotional intent, but do

not display any *native-like* characteristics of promotional texts. A third step in the analysis could help to solve this issue, for instance by selecting a subset of promotional genres through external criteria (URL or page title) and ascertain whether and to what extent these genres adopt atypical evaluation features, i.e. words and structures not typically associated with the promotional function in standard (native) English. On the other hand, one may argue that even if texts are selected using 'external' criteria – such as the very name associated with a genre – these are still largely dependent on language choices, and there might be many other similar genres that simply adopt a different webpage title (cf. Section 2.4.2). In other words, this turns into a philosophical problem about the role of language in structuring our experience of the world, that is well beyond the scope of this thesis.

Moving now to Chapter 5, the features chosen for conducting a corpus-based analysis of evaluation were presented, together with a brief discussion of their adequacy as markers of evaluation. Features were selected from a wide range of relevant research studies and more theoretical works, in particular Biber (2006a), Hyland (2008) and Martin and White (2005). Since the list of evaluative features taken from the relevant literature is not meant to be applied mechanically (Martin and White, 2005:52), it was used as an instrument to look at the context surrounding the features, especially through concordance lines. Furthermore, features were manually checked and some of them were removed altogether. Semantic tagging might be used as an alternative, but this would not avoid the mechanical procedure problem, since an automatic or "mechanical" procedure would be carried out in the first place.

Finally, Chapter 6 presented the study findings: in order to characterise the promotional strategies of academic websites linguistically, features of evaluation were identified, quantified and analysed within the native English and ELF subcorpora. In the analysis, examples were selected from various country/linguacultural backgrounds to show how promotional strategies apply across Europe, although there are inevitable differences in the distribution of occurrences across countries and universities. This might be an issue to explore in future studies, that could help understand the effect of the different language backgrounds or academic traditions.

This thesis contributed to current research in at least three ways. First, it provided evidence that text classification based on FTDs (Sharoff, 2018) is an effective method for identifying promotional pages automatically. Although this method was assessed on the promotional dimension only, there is no reason to believe that it would not apply equally well to other scenarios. Second, it provided a corpus that is made available through an online platform[1] and that can be used to support promotional writing in academic websites or other

[1] https://corpora.dipintra.it [last consulted on 15 December 2017].

text types (e.g. instructional texts and news). Furthermore, it offered a full set of strategies that can be used to express the promotional function of texts within academic websites in English. Third, it identified some distinctive characteristics of either native English or ELF promotional texts within university websites. In future work, it would be interesting to understand whether these observations regarding native English and ELF promotional strategies can be generalised to other non-academic settings. The list of promotional strategies obtained from a thorough scrutiny of the relevant literature and the set of queries produced for this study (which are also made available to the corpus linguistics research community) can work as a toolkit that might favour comparable studies within other contexts, for instance the business sector or non-profit organisations/social businesses, thus contributing to mapping the ever-expanding territory of evaluation in English(es).

# References


Askehave, I. (2007). The impact of marketization on higher education genres: The international student prospectus as a case in point. *Discourse Studies 9*, 723–742.

Baroni, M. and S. Bernardini (Eds.) (2006). *Wacky! Working papers on the Web as Corpus*. Bologna: GEDIT.

Bernardini, S. (2016). Discovery learning in the language-for-translation classroom: corpora as learning aids. *CADERNOS DE TRADUÇÃO* (36), 14 – 35.

Bernardini, S., A. Ferraresi, and F. Gaspari (2010). Institutional academic English in the European context: a web-as-corpus approach to comparing native and non-native language. In *Professional English in the European context: The EHEA challenge*, pp. 27 – 53. BERNA: Peter Lang.

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

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

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistics Computing 8*(4), 243–257.

Biber, D. (2006a). Stance in spoken and written university registers. *Journal of English for Academic Purposes 5*(2), 97–116.

Biber, D. (2006b). *University Language. A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., U. Connor, and T. A. Upton (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins.

Biber, D. and S. Conrad (2009). *Register, Genre, and Style*. New York, NY: Cambridge University Press.

Biber, D., S. M. Conrad, R. Reppen, P. Byrd, M. Helt, V. Clark, V. Cortes, E. Csomay, and A. Urzua (2004). *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. ETS Research Memorandum.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan (1999). *Longman grammar of spoken and written English*. London: Longman.

Caiazzo, L. (2011). Hybridization in institutional language: Exploring we in the 'about us' page of university websites. In S. Sarangi, V. Polese, and G. Caliendo (Eds.), *Genre(s) on the move. Hybridization and discourse change in specialized communication*, pp. 243–260. Napoli: Edizioni Scientifiche Italiane.

Campbell, K. and K. Jamieson (1978). Form and genre in rhetorical criticism: An introduction. In K. Campbell and K. Jamieson (Eds.), *Form and Genre: Shaping Rhetorical Action*, pp. 9–32. Falls Church, VA: Speech and Communication Association.

Casañ Pitarch, R. (2015, 04). The genre 'about us': A case study of banks' corporate webpages. *9*, 69–96.

Cogo, A. and J. Jenkins (2010). English as a lingua franca in europe: A mismatch between policy and practice. *European Journal of Language Policy 2*(2), 271–294.

Conrad, S. and D. Biber (2000). Adverbial marking of stance in speech and writing. In S. Hunston and G. Thompson (Eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford University Press.

Costales, A. F. (2012). The internationalization of institutional websites: The case of universities in the European Union. In A. Pym and D. Orrego-Carmona (Eds.), *Translation Reasearch Projects 4*. Terragona: Intercultural Studies Group.

Crystal, D. (2012). *English as a Global Language* (second ed.). Cambridge University Press.

Davies, A. (2004). The native speaker in applied linguistics. In A. Davies and C. Elder (Eds.), *The Handbook of Applied Linguistics*, pp. 431–450. Oxford: Blackwell Publishers Inc.

Dröschel, Y. (2011). *Lingua Franca English: The role of simplification and transfer*. Bern: Peter Lang.

Egbert, J., D. Biber, and M. Davies (2015). Developing a bottom-up, user-based method of web register classification. *Journal of the Association for Information Science & Technology 66*(9), 1817–1831.

Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. London: Pinter Publishers.

European Commission (2012). The development of European identity/identities: Unfinished business: A policy review.

European Language Council (2013). The role of languages in the european higher education area.

European University Association (2013). Internationalisation in European higher education: European policies, institutional strategies and EUA support.

Evert, S. T. C. D. T. (2016). The ims open corpus workbench (cwb) cqp query language tutorial. cwb version 3.4.

Fairclough, N. (1993). Critical discourse analysis and the marketization of public discourse: The universities. *Discourse & Society 4*(2), 133–168.

Fairclough, N. (1995). *Critical Discourse Analysis: The critical study of language*. London: Longman.

Fairclough, N. (2001). *Language and Power* (second ed.). Harlow: Pearson Education.

Ferraresi, A. and S. Bernardini (2013). The academic Web-as-Corpus. In *Proceedings of the 8th Web as Corpus Workshop*, Lancaster.

Ferraresi, A., S. Bernardini, G. Picci, and M. Baroni (2010). Web corpora for bilingual lexicography: A pilot study of english/french collocation extraction and translation. In R. Xiao (Ed.), *Using Corpora in Contrastive and Translation Studies*. Newcastle: Cambridge Scholars Publishing.

Fletcher, W. H. (2004). Making the web more useful as a source for linguistic corpora. In *Corpus Linguistics in North America*, pp. 191–205. Rodopi.

Forsyth, R. and S. Sharoff (2014). Document dissimilarity within and across languages: a benchmarking study. *Literary and Linguistic Computing 29*(1), 6–22.

Foskett, N. and J. Lumby (1999). *Managing External Relations in Schools and Colleges: International Dimensions*. Sage Publications.

Freedman, A. and P. Medway (1994). Locating genre studies: Antecedents and prospects. In A. Freedman and P. Medway (Eds.), *Genre and the New Rhetoric*. London: Taylor & Francis.

Fuoli, M. (2012, February). Assessing social responsibility: A quantitative analysis of appraisal in BP's and IKEA's social reports. *Discourse & Communication 6*(1), 55–81.

Garzone, G. (2007). Introduction. In C. Garzone, Giuliana; Ilie (Ed.), *The use of English in institutional and business settings: an intercultural perspective*, pp. 9–27. Bern: Peter Lang.

Gesuato, S. (2011). Course descriptions: Communicative practices of an institutional genre. In S. Sarangi, V. Polese, and G. Caliendo (Eds.), *Genre(s) on the move. Hybridization and discourse change in specialized communication*, pp. 221–241. Napoli: Edizioni Scientifiche Italiane.

Halliday, M. A. K. (1978). *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Hasan, R. (1989). *Linguistics, Language and Verbal Art*. Oxford University Press.

Hoang, T. V. Y. and I. Rojas-Lizana (2015). Promotional discourse in the websites of two australian universities: A discourse analytic approach. *Cogent Education 2*(1).

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations* (second ed.). Thousand Oaks CA: Sage Publications.

House, J. (2003). English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics 7*(4), 556–578.

Hundt, M., N. Nesselhauf, C. Biewer, I. Wissenschaftsforum, F. Thyssen-Stiftung, U. Heidelberg, and S. U. Heidelberg (2007). *Corpus Linguistics and the Web*. Corpus Linguistics and the Web. Rodopi.

Hunston, S. (1989). *Evaluation in experimental research articles*. Ph. D. thesis, Department of English, Faculty of Arts, University of Birmingham.

Hunston, S. (2010). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. Routledge.

Hunston, S. and G. Thompson (2000). Evaluation: an introduction. In S. Hunston and G. Thompson (Eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford University Press.

Hyland, K. (2008). Persuasion, interaction and the construction of knowledge: representing self and others in research writing. *International Journal of English Studies 8*(2), 1–23.

Hyland, K. and P. Tse (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics 25*(2), 156.

Jenkins, J. (2000). *The phonology of English as an international language: new models, new norms, new goals*. Oxford University Press.

Jenkins, J. (2004). Elf at the gate: The position of english as a lingua franca. *The European English Messenger 13*(2), 63–39.

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford University Press.

Jenkins, J. (2012). English as a lingua franca from the classroom to the classroom. *ELT Journal 66*(4), 486–494.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The english language in the outer circle. In R. Quirk and H. Widdowson (Eds.), *English in the World: Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press.

Kanaris, I. and E. Stamatatos (2007). Webpage genre identification using variable-length character n-grams. In *ICTAI (2)*, pp. 3–10. IEEE Computer Society.

Kehoe, A. and A. Renouf (2002). Proceedings of the www 2002 conference. In *WebCorp: Applying the Web to Linguistics and Linguistics to the Web*, Honolulu, Hawaii.

Kilgarriff, A. and G. Grefenstette (2003, September). Introduction to the special issue on the web as corpus. *Comput. Linguist. 29*(3), 333–347.

Lee, D. Y. W. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning & Technology 5*(3), 37–72.

Leech, G. (2007). New resources or just better old ones? the holy grail of representativeness. In M. Hundt, N. Nesselhauf, and C. Biewer (Eds.), *Corpus linguistics and the web*, pp. 133–150. Amsterdam and New York: Rodopi.

Machin, D. and A. Mayr (2012). *How to Do Critical Discourse Analysis. A Multimodal Introduction*. London: Sage Publications.

Martin, J. (1984). Children writing: A reader. In F. Christie (Ed.), *Language, Register and Genre*. Geelong, Vic.: Deakin University Press.

Martin, J. and P. White (2005). *The Language of Evaluation. Appraisal in English*. Palgrave Macmillan UK.

Mautner, G. (2005). For-profit discourse in the nonprofit and public sectors. In G. Erreygers and G. Jacobs (Eds.), *Language, communication and the economy*, pp. 25–44. Amsterdam: John Benjamins.

Mayr, A. (2008). *Language and Power: An Introduction to Institutional Discourse.* London: Continuum.

Mazzei, A. (2002). *La comunicazione per il marketing dell'Università*. Milano: FrancoAngeli.

McEnery, T. and A. Hardie (2012). *Corpus Linguistics: Method, theory and practice.* Cambridge: Cambridge University Press.

Mehler, A., S. Sharoff, and M. Santini (2010). *Genres on the Web: Computational Models and Empirical Studies* (1st ed.). Springer Publishing Company, Incorporated.

Meierkord, C. and K. Knapp (2002). Approaching lingua franca communication. In C. Meierkord and K. Knapp (Eds.), *Lingua Franca Communication*. Peter Lang.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech 70*(2), 151–167.

Mokhtarnia, S. and R. Ghafar-Samar (2015). American vs iranian university websites: A critical discourse analysis of about pages. *JOURNAL OF SOCIAL SCIENCES RESEARCH 8*(3), 1628–1640.

Mollin, S. (2006). English as a lingua franca: A new variety in the new expanding circle? *Nordic Journal of English Studies 5*(1), 41–57.

Morrish, L. and H. Sauntson (2013). Business-facing motors for economic development: an appraisal analysis of visions and values in the marketised uk university. *Critical Discourse Studies 10*(1), 61–80.

Nickerson, C. (2007). English as a lingua franca in business contexts: Strategy or hegemony? In C. Garzone, Giuliana; Ilie (Ed.), *The use of English in institutional and business settings: an intercultural perspective*, pp. 351–363. Peter Lang.

Palumbo, G. (2013). Divided loyalties? Some notes on translating university websites into English. *CULTUS 6*, 95–109.

Penny, U. (2010). English as a lingua franca: A teacher's perspective. *Cadernos De Letras (UFRJ)* (27).

Poncini, G. (2007). Communicating within and across professional worlds in an intercultural setting. In G. Garzone and I. Cornelia (Eds.), *The Role of English in Institutional and Business Settings: An Intercultural Perspective*, pp. 283–311. Bern: Peter Lang.

Reid, T. B. (1956). Linguistics, structuralism, and philology. *Archivum Linguisticum 8*, 28–37.

Reithofer, K. (2010). English as a lingua franca vs. interpreting: battleground or peaceful coexistence? *The Interpreters' Newsletter 15*, 143–157.

Saichaie, K. (2011). *Representation on college and university websites: An approach using critical discourse analysis*. Ph. D. thesis, University of Iowa.

Sanigar, M. A. (2013). Selling an education. universities as commercial entities: a corpus-based study of university websites as self-promotion. *LWPLP 18*.

Santini, M. (2007). Characterizing genres of web pages: Genre hybridism and individualization. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, Washington, DC, USA, pp. 71 – 81. IEEE Computer Society.

Schäfer, R. and F. Bildhauer (2013). *Web Corpus Construction*. Morgan and Claypool.

Scott, J. (2001). *Power*. Cambridge:Polity Press.

Seidlhofer, B. (2010). A concept of international English and related issues: from 'real English' to 'realistic English'? Council of Europe. http://www.coe.int/T/DG4/Linguistic/Source/SeidlhoferEN.pdf.

Sharoff, S. (2018). Functional text dimensions for annotation of web corpora. *Corpora 31*(1).

Smola, A. J. and B. Schölkopf (2004, August). A tutorial on support vector regression. *Statistics and Computing 14*(3), 199–222.

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

Thornborrow, J. (2002). *Power Talk: Language and Interaction in Institutional Discourse*. London: Longman.

Tipping, M. E. (2001, September). Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res. 1*, 211–244.

Van Dijk, T. A. (1981). Discourse studies and education. *Applied Linguistics II*(1), 1–26.

Wodak, R. (1997). Critical discourse analysis and the study of doctor-patient interaction. In *The Construction of Professional Discourse*, pp. 173–200. London: Longman.

Wodak, R. and N. Fairclough (2010). Recontextualizing european higher education policies: the cases of austria and romania. *Critical Discourse Studies 7*(1), 19–40.

Xiong, T. (2012). Discourse and marketization of higher education in china: The genre of advertisements for academic posts. *Discourse & Society 23*(3), 318–337.