1. Introduction

The aim of the present article is to examine the English of non-edited texts from the European Commission and compare it with that of the same texts edited by the DGT Editing Service. The interest of this is to investigate editing as a process of mediation, during which language undergoes a process of rewriting and revision wrought by someone who did not write the text. Lefevere (1992: 9) includes editing in his list of rewriting activities, alongside translation, historiography, anthologizing and criticism, and rewriting is interpreted by Ulrych and Anselmi (2008) as a means of mediation, which is extremely important for texts of all types, since mediated texts are actually the form of texts which most readers encounter. In the present paper, mediated texts are investigated in a comparative light, both against the same text previous to their revision, and against the general reference corpus of the BNC.

2. Languages, multilingualism and translation in the European Union

When the six founder Member States signed the treaties of the European Coal and Steel Community (1951) and the European Economic Community (1957), the predecessors of today’s European Union, they acknowledged four official lan-
languages: German, French, Italian and Dutch. In the EC Treaty, which founded the EEC, Article 248 declares:

The present Treaty, drawn up in a single original in the German, French, Italian and Dutch languages, all four texts being equally authentic, shall be deposited in the archives of the Government of the Italian Republic which shall transmit a certified copy to each of the Governments of the other signatory States.

The EC Treaty (formally known as the Treaty of Rome) 1957, which established the EEC

This article set the precedent for other Treaties of the Union, and indeed for all legislative documents: it is standard practice in the European Union (henceforth EU) for the different language versions of a legislative document to be considered “originals”: there is no source document and subsequent translations.

The language scenario of the EU in 2008 has changed as much as the geographical reach of the EU: there are now 23 official languages for 27 Member States. The current President of the European Commission, José Manuel Barroso, in office since 2004, has shown particular sensitivity to the issue of interlingual communication, and has instituted two new portfolios in this regard. In his first year of office, he created a portfolio of multilingualism and assimilated it to that of the Commissioner for Education, Training and Culture, Ján Figel’. This led to a Communication by the Commission to the other institutions outlining a new framework strategy on multilingualism. Barroso subsequently appointed a special Commissioner for Multilingualism, a post currently held by Leonard Orban.

Multilingualism is seen as the strong point of the European Union, the emblem of its linguistic and cultural diversity, the “key feature of Europe” according to the report of the High Level Group on Multilingualism of the European Commission. Of course, full multilingualism with translation and interpreting into and out of all languages is impossible, because of a shortage of skilled translators in the languages of the most recently arrived Member States. Given the practical issues involved, its desirability is also questionable.

Following Gazzola (2006), a useful distinction can be made between the practices of translation and interpreting towards citizens and Member States, where fully multilingual communication is in place, and translation and interpreting within the institutions themselves, where communication is partially multilingual. The institutions that are representative of European peoples, governments and regional and local authorities, that is, the European Parliament, the Council and the Economic and Social Committee and the Committee of the Regions, are all fully multilingual, and offer full translation and interpreting services. On the other hand, the Commission, the European Central Bank, the Court of Auditors and the Court of Justice use a limited number of ‘working’ languages. Council Regulation 1/1958, recognised by some working within the EU (eg. Wagner et al. 2002) as the ‘EU Language Charter’, does not stipulate which languages shall be the working languages: English, French and German (in that frequency order) are called the working languages simply because they are the most commonly used for internal activities (Gazzola 2006: 397).
Among the European Institutions, the European Commission has the largest translation service, the Directorate-General for Translation (DGT). DGT employs over 2000 translators, located in Brussels and Luxembourg, and also outsources a considerable amount of work. Under its present structure, DGT is organised into departments, one for each language, and the departments in turn are made up of units which are topic specialised. In this way, translators only work on texts from certain domains, and this has proved to be more effective.

Translators nowadays have sophisticated aids, such as the online interinstitutional terminology database, IATE\(^1\) (interactive terminology for Europe), which gives free access to thousands of terms translated into 22 of the 23 official languages (Irish is at present not yet included). EUR-Lex,\(^2\) another online database, gives access to all European Union legal documents in all available languages. Besides these resources, which are available for public access, DGT translators also have translation memory tools, such as TRADOS, DGT-VISTA or EURAMIS. The former, a tool which stores translations in real time as they are written, is available commercially, while the latter two are specific to the EU and only draw on a database of EU translated texts. Using these tools, the translator can call up previous translations of phrases from EU texts in specified language pairs via a multilingual concordancer.

DGT does not translate all documents produced within the Commission: the requirement that EU law must exist in all the official languages means that the core of their work deals with proposals for legislation, green and white papers, Commission communications, and scientific and economic reports required prior to legislation drafting. Speeches by Commissioners and personal correspondence are a marginal part of their work.

While in theory all the official languages of the EU can be considered working languages, in practice, the vast majority of texts are drawn up in English. This was estimated at 72% in 2006 by DGT. The prevalent use of English by those working within the EU institutions is a clear sign that practical concerns have the strongest influence on daily lives and working habits, and that English serves as the \textit{lingua franca} of the 21\textsuperscript{st} century, particularly within Europe. It follows that those drafting the documents are, in most cases, not native speakers. The quality of the English in these documents has been cause for concern in recent years, and has led to the institution of an editing service within DGT. The concern about the language of the documents which are translated into other languages was publicly acknowledged by the General Director of DGT, Karl-Johan Lönnroth, at a conference held by the Center for International Cooperation in E-Business in the following terms:

Another challenge which haunts us is quality. This does not only concern the need to develop and update terminology, but also to watch the quality of the originals. In translation we have introduced editing services for French and English to ensure that the source documents (mostly drafted by non-native speakers) are of high linguistic quality and as translatable as we can make them.\(^3\)
3. The EuroCom corpus and research methodology

EuroCom is a corpus containing documents whose common characteristic is that they originate within the European Commission and have passed through the Editing Unit at DGT. It is a monolingual comparable corpus of specialized texts; the two sub-corpora contain different versions of the same texts, non-edited (EuroCom Non-edited) and edited (EuroCom Edited). It covers a span of 4 years: 2005-2008, and contains 156 documents in their non-edited and edited versions. EuroCom Non-edited corpus totals 1,001,804 tokens, while EuroCom Edited corpus is slightly smaller, standing at 995,451 tokens.

The documents in EuroCom have all been drafted by staff in Directorate-Generals or Services within the Commission. They cover a wide range of domains, coming from DGs as various as Agriculture and Rural Development, Competition, Economic and Financial Affairs, Energy and Transport, Information Society and Media, Justice, Freedom and Security and Taxation and Custom Unit, to name just a few. The text types in the corpus vary, but consist mostly of Communications from the Commission, which aim to open up debate within the institutions prior to proposals for secondary legislation (Regulations, Decisions and Directives), reports for the European Council and the European Parliament on the implementation of Regulations and Directives, documents for internal use only, such as working papers, work programmes, minutes of meetings, staff notices, and brochures, guides, and press releases produced for the general public, as well as texts to be published on the Europa website.

The methodology adopted for the present study fits into the framework of corpus-assisted discourse analysis (Partington 2004). This involves both quantitative and qualitative analyses, using the software Wordsmith Tools 4 (developed by Scott 2006) for the former, and reading and comparing the texts in the two corpora for the latter. The importance and advantages of using a complementary approach have been emphasized by many linguists (Biber, Conrad, and Reppen 1998; Kennedy 1998; McEnery and Wilson 1996; Sinclair 2006 inter alia): a quantitative analysis counts and classifies features, and pays considerable attention to frequency, while a qualitative analysis aims to provide complete and detailed description within a context. The investigation of the features of the texts started from a quantitative analysis, examining wordlists of the two corpora, divided into one, two, three and four-word clusters. Comparisons of frequency of occurrences led to their investigation in the two versions of the texts. The qualitative investigations of the text, on the other hand, involve examining the texts in the form they leave the Editing Unit. This is as a Microsoft Word document, where the revision and tracking functions are activated, allowing the editor’s interventions to be displayed visually either within the text, or in ‘text bubbles’ that appear in the right-hand margin of the document.

4. Guides for writers of EC documents and the Editing Service

All writers of EC documents have access to the online Interinstitutional Style Guide, produced by the European Union Publications Office in all the official lan-
guages. This guide provides the conventions and rules for documents from all the EU institutions, and advises that a consistent style in documents will contribute to cohesion between them. The appropriate way to refer to types of legal acts is listed, together with conventions for referring to the institutions, official titles within the institutions, references to countries, languages, currencies, etc. For English, the Guide explains punctuation rules and British spelling conventions which are adopted within the documents, and also warns against interference from French, with a list of the most common false ‘spelling’ friends, such as correspondance /correspondence, and defense/defence.

More specifically for English, two documents for writers have been written by translators at the European Commission: How to write clearly, which was drafted during the Fight the Fog campaign for writing clear English that started from the English section of DGT in Luxembourg in 1998, and The English Style Guide. The basic message of How to write clearly includes the following points:

a) put the reader first – the general lay reader is the one who will find documents about the European Institutions hardest to understand; avoid gender-specific terms; use plurals instead;
b) use verbs, rather than abstract nouns; avoid nominalisation;
c) use concrete nouns rather than abstract ones;
d) use active verbs, not passives, unless they are deliberately chosen to avoid mentioning responsibility or agency; don’t be afraid to use first person agents for the verb;
e) place old information at the beginning of the sentence and new, important or complex information at the end;
f) KISS – keep it short and simple;
g) avoid false friends between French and English.

The English Style Guide, on the other hand, is a handbook for authors and translators in the European Commission, written by translators working at DGT. It offers the consolidated experience of more than 25 years translation of EU documents – the first edition is dated 1982 – and is updated regularly online (latest version April 2008). In its introduction, it clarifies the sense in which “style” is used: “a set of accepted linguistic conventions; [...] recommended in-house usage”. Its aim is to set a good example by using English that is “as clear, simple, and accessible as possible out of courtesy to our readers and consideration for the image of the Commission”. It clearly reflects the preoccupations of the briefer Fight the Fog booklet, and cites the same reference works (Cutts 1996, Williams 1995).

The English Style Guide is divided into two parts: one deals with linguistic conventions applicable in all contexts – spelling conventions, information on using upper and lower case, geographical names, hyphens and compound nouns, punctuation, numbers, dates and times, abbreviations, and acronyms. There is a section on the agreement of verbs with singular or collective nouns, verb tenses (past simple versus present perfect), notes on the tenses of verbs to be used in minutes of meetings, and on appropriate verbs in legislation, as well as advice on the split infinitive. It has a section on scientific language, footnotes, translating correspondence, personal names and titles, and gender-neutral language. The
second part has a more encyclopaedic function, and deals with the workings of the European Union, and how they are expressed in English, sometimes accounting for them by referring to the original French.

Despite the guides that are available, a few years after the Fight the Fog campaign, there was a general consensus within DGT that translators were spending an excessive amount of time editing texts written in poor English. A pilot service was thus set up to start editing documents written in English and French from two DGs, Economics and Finance, and Environment. Staffed by ex-translators with considerable experience of translating for the EU institutions, the Unit met with immediate success, and this led to the establishment of a dedicated Unit – the Editing Unit - within DGT.

5. Classifying the revisions

Revisions to translated texts have been studied by Rega (1999), Scarpa (2008), Mossop (2001) and Cosmai (2007). Cosmai, who works as a translator/reviser at the Committee of the Regions at the European Union, classifies revisers’ interventions into three categories: subjective, objective and specialized (2007: 102). In the case of editing non-native English, however, the question is perhaps more complex. Some revisions are indeed objective, particularly those referring to the house-style of the European Union institutions (capitalisation, names, dates, numbers, etc.), and grammatical revisions (prepositions, verb tenses, concord of noun and verb). But many revisions can be described simply as ways of improving the text, rendering it more natural, or smooth. Changing a post-modifying phrase into a pre-modifying phrase, or a passive verb into an active one may be ways of making the test more incisive, more concise, but they are not corrections of elements that are incorrect.

Two approaches were taken here to the description of the revisions. Firstly, on a sample of 15 documents, all the revisions were categorized under a series of labels. These labels are:

- **Objective revisions**: including layout (capitalisation, acronyms, etc), grammar, punctuation, syntax and spelling
- **Overall improvements**: including passages that are rewritten for content, wrong lexical choices, and changes in style (e.g. changing passive verbs to active verbs, post-modifying phrases to pre-modifying phrases).

Secondly, following the guides recommended by the EU Institutions (such as the English Style Guide, and How to write clearly booklet, and some Tips for Writing English drawn up by translators), the whole corpus of edited texts was analysed for certain overall improvements. Apart from the features listed in b), elements such as foggy phrasing, Eurospeak, false friends, and bureaucratic turns of phrase were examined.

On the first sample of 15 documents, ‘objective’ revisions accounted for 69% of revisions, whereas those which improve the text overall account for 31%. Excerpt 1 shows a text that has been edited in which both types of revisions are visible. Each revision is commented on briefly.
Visits to DGT in September and October 2006

On 18 September 2006, DGT received a group of students from the Drew University, in the USA. It is worth quoting their reasons for wishing motivation to visit our service: “We do understand the importance of teaching our students about learning to communicate in another language and to understand its culture. The European Union is the perfect example and we believe that this visit to your institution could change how American students view the importance of multiculturalism and multilingualism in our world”. These students were particularly curious to learn more about our working methods and computer-assisted translation.

Excerpt 1. Objective Revisions and Overall Improvements

In Excerpt 1, the objective revisions (which refer to layout elements such as capitalisation and acronyms, grammar, punctuation, syntax and spelling) include the deletion of the definite article the before the name of Drew University, the added comma and the deletion of A from the abbreviation USA, the change to single inverted commas from double, and the replacement of the demonstrative adjective these by the definite article the. The overall improvements are the added prepositional phrase in the before US, and the lexical change of motivation to the noun phrase reasons for wishing. While the former type of revisions are objectively explicable (e.g. it is the policy in EU documents to refer to the US, not the USA), the latter are more complex and subjective interventions carefully weighed. The added prepositional phrase in the fleshes out the text, creating better balance, but the original version cannot be said to be wrong. Similarly, the change from the Latinate nominalization motivation to the more phrasal reasons for wishing marks a change from longer words to shorter words, and from Latinate to German compound phrases, which is intended presumably to make the text more reader-friendly (the text was destined for the DGT website, so for a non-specialized readership). Other types of lexical revisions are classed as objective revisions, if for example they regard typical EU terms. An example is services being changed to departments: while there are some services which have kept their original French name (such as Interservice consultation) or the Legal Service, generally most services are called departments within the EC. Changing the word is thus not a lexical choice, as it were, but a conforming of terms to the objective norms of the Commission.

6. Reflections on the edited-mediated corpus

In this section, selective observations will be made about the results of the quantitative analyses of the two corpora. Further findings are reported in Murphy (2008).

From a corpus-driven perspective (Tognini-Bonelli 2001), one way of investigating comparable corpora is to compare single-word lists and cluster word lists. Clusters of two and three words were examined across the EuroCom Edited and Non-Edited corpora, and findings from each examination will be reported here.
From the comparison of two-word clusters, it emerged that the two-word cluster *of the* is edited out more than 700 times in the EuroCom edited corpus. Occurrences in EuroCom Non-edited stand at 11,996, and in EuroCom Edited at 11,286. Close examination of some of the occurrences reveal that *of the* is often edited out when it is part of a post-modifying phrase. The editors frequently turn the *of the* phrases into pre-modifying phrases. To find examples of this, the three-word wordlist was useful. In this wordlist, it was seen that phrases beginning with *of the + adjective/noun* were fewer than in the edited corpus, examples from the non-edited corpus being *of the European*, *of the EU*, *of the Member*, *of the Council*, *of the Commission*, *of the Directive*. Examples 1 and 2 show how one such phrase *of the* European citizens is changed by the editors from a post-modifying position to a pre-modifying one:

1) It will provide a clear signal of support to the **rights** to information and freedom of expression **of the European citizens** and confirm communication as an EU policy at the service of citizens and democracy.
   
   (EuroCom Non-edited Corpus)

> 

2) It will provide a clear signal of support for **European citizens’ rights** to information and freedom of expression and confirm communication as an EU policy, at the service of citizens and democracy.

   (EuroCom Edited Corpus)

Another finding from the comparison of three-word clusters was the pattern *the + noun + of*, with tokens such as *the implementation of*, *the application of*, *the protection of*, *the creation of*, *the adoption of*, *the definition of*, *the admission of*. These patterns involve a nominalised verb. In EuroCom Non-edited, there are 3012 occurrences of these phrases, and 2651 in EuroCom Edited. This shows a tendency to eliminate nominalisations, although they are still present in the Edited corpus. Example 3 shows an example of a nominalisation in the pattern *the + noun + of* (i.e. the provisions of) that has been edited out:

**Example 3**

Moreover, all staff members are **in particular reminded in particular of**:

- the obligations laid down\* by the [Staff Regulations of officials](https://example.com) and the [Conditions of employment of other servants of the European Communities](https://example.com), and, notably, the provisions of [Article 12 of the Staff Regulations](https://example.com), which states that an official “shall refrain from any action or behaviour which might reflect adversely upon his position”. Therefore, any abuse linked to personal use of the ICT services at the disposal of staff available to [staff member](https://example.com) is prohibited (see Report on the work of the Investigation and Disciplinary Office (IDOC)\* and relevant case law);

**Example 3 the provisions of** (line 4) edited out
The third observation reported here regards the comparison of phraseology in EuroCom Edited corpus and the BNC. By comparing the two versions of EuroCom, it came to light that the three-word clusters *in order to* and *as well as* are among the most frequent in both corpora. Although on some occasions they are added by the editors, the tendency is for editors to edit them out: occurrences in EuroCom Non-edited amount to 183 and 125 respectively, and 139 and 99 in EuroCom Edited. They remain among the 9 most frequent three-word clusters in the edited corpus. A look at the most frequent three-word clusters in the BNC reveals that neither of these phrases occur within the top 1000 3-word clusters in the BNC. This is an interesting finding, in that it would appear to characterize the EuroCom Edited corpus, but not the BNC. It becomes more significant, perhaps, if one reflects that both phrases have filler functions: *in order to* (a complex preposition) makes the purpose of something explicit, it extends and clarifies the meaning of the to-form of the verb. Example 4 illustrates one such case from the corpus, where the complex preposition has been added in by the editors, presumably for the sake of clarity:

4) With a view to further development of a coherent immigration policy, *in order to narrow* the rights gap between EU citizens and third-country nationals legally working and to complementing the existing immigration *acquis*, a set of rights should be laid down, in particular in the form of specifying the policy fields where equal treatment with nationals is provided for third-country workers legally admitted into a Member States but not yet long-term residents. (EuroCom Edited corpus)

The interpretation given here is that the editors are making the purpose of *to narrow the rates gap* clearer, more explicit, by lengthening *to* to *in order to*. Explicitation is also seen as the process behind the phrase *as well as*. In Excerpt 2 it can be seen that one example of *as well as* has been kept (in line 8), whereas in line 14 it has been deleted.

### 4. FRANCE

A new law on equal rights and opportunities, participation and citizenship of persons with disabilities came into force on 12 February 2005 (Loi No 2005-102 of 11 February 2005). This law *puts in place establishes* the principle of non-discrimination on grounds of disability, for both the private and public sectors. It obliges employers to provide reasonable accommodation for disabled workers, as long as the cost is not disproportionate. Special arrangements can be made for disabled candidates to take entrance exams for the public service, and it will be possible to employ disabled civil servants under contract as well as by entrance *competitions examinations*. Disability associations which have been legally established for at least five years will be able to give legal help to victims of discrimination. *Measures of positive action in favour of disabled workers* are *is not considered discriminatory*. Employers will have to negotiate annually on *disabled persons' the conditions of access to employment of disabled persons, of their vocational training and working conditions, as well as and actions to raise awareness of all workers on disability issues*. The Government must produce an annual report on the employment of disabled people in the public service. A national Observatory on training, research and innovation in the field of disability will be created.

Excerpt 2 Examples of *as well as*
This complex conjunction is also a filler, a longer way of saying *and*, and it is eliminated to a greater extent than *in order to*, but it still occurs within the 9 most frequent three-word phrases in the Edited corpus, but not among the 1000 most frequent three-word phrases in the BNC.

7. Conclusions

Research into edited-mediated language in the EuroCom corpus is still in its infancy. However, a few preliminary conclusions may be drawn on the basis of the empirical investigations. Firstly, from the point of view of the editors, by looking at the revisions of texts overall, it appears that approximately two thirds of the alterations made to the text regard objective matters: i.e. matters of house style (capitalisation, layout, abbreviations), grammar or spelling. While grammar remains something of which the individual writer may have a shaky grasp, house style and spelling could be corrected by a computer programme. Intervening in many situations where the accepted version is an objectively established one is perhaps not the best use of editor’s time. Secondly, at a higher level of mediation of the text, the overall improvements wrought to the text are rather personal. Changing post-modifying phrases to pre-modifying ones or eliminating nominalisation is certainly a subjective exercise, which undoubtedly varies from one editor to another. Further research might compare the work of individual editors, and gauge the extent to which such changes are uniform across the service.

From the descriptive point of view, it can be said that two tendencies have emerged through comparing the EuroCom Edited Corpus with the EuroCom Non-edited one: firstly, a move towards conciseness. Moving post-modifying phrases to a pre-modifying position makes for tighter syntax and tauter prose. Conciseness may be a characteristic of mediated text. Contemporaneously, an opposing tendency is noted: that of explicitation. The complex phrases *in order to* and *as well as* lengthen and make links that are already in the text more explicit, and characterise the EuroCom Edited corpus significantly in terms of phraseology, whereas they do not characterize the reference BNC. Further research into such issues within these corpora may explain how two opposite tendencies happily co-exist within mediated text.
NOTES

1 URL: http://iate.europa.eu.
4 I gratefully acknowledge the permission of the Head of the Editing Unit, David Crowther, for the opportunity to study these documents.

REFERENCES


Partington, Alan. 2004. “Corpora and discourse, a most incongruous beast”. In Partington, Alan, Morley, John, Haarman, Louann. (eds.) Cor-


