ISSN 1974-4382

THE EVOLUTIVE NATURE OF TRANSLATION AND TEACHING WORKFLOW: TEXTOMETRIC ANALYSIS OF MULTIPLE REVISION CYCLES IN A SIMULATED WEBSITE TRANSLATION PROJECT

MARIA ZIMINA-POIROT, CHRISTOPHER GLEDHILL
UNIVERSITÉ PARIS CITÉ

maria.zimina-poirot@u-paris.fr, christopher.gledhill@u-paris.fr

Citation: Zimina-Poirot, Maria e Cristopher Gledhill (2023) "The Evolutive Nature of Translation and Teaching Workflow: Textometric Analysis of Multiple Revision Cycles in a Simulated Website Translation Project", in Maria Margherita Mattioda, Alessandra Molino, Lucia Cinato e Ilaria Cennamo (a cura di) *L'intelligenza artificiale per la traduzione: verso una nuova progettazione didattica?*, *mediAzioni* 39: A116-A146, https://doi.org/10.6092/issn.1974-4382/18790, ISSN 1974-4382.

Abstract: In this paper, we explore the practical and theoretical status of revision in the context of a Masters-level website translation project integrating machine translation (MT) for teaching purposes. We use textometric analysis to explore the nature of modifications made at each of the different stages of revision during the lifetime of one particular project. Computation of *characteristic elements* and *Correspondence Analysis (CA)* reveal regular patterns of linguistic intervention, with significant differences observed between the initial stages of the project (MT-EN0 generated by SYSTRAN Pure Neural Machine Translation[1] and post-edited MT-EN1) and subsequent revision cycles. We combine quantitative and qualitative research methods to study characteristic revision patterns in different project cycles (*chronological textual series*). While some of these patterns represent unpredictable "punctual" edits, the majority are "progressive", involving modifications that are predictable or gradually built up throughout the project. More generally, it appears that different types of revision closely mirror the guidelines that are presented to students during the teaching workflow. This observation leads us to posit an "evolutive" view of the translation project: just as each webpage in the project has to undergo several cycles of revision, so each stage of revision needs to be carefully primed during the preparation of the project with the clients as well as the students.

Keywords: Chronological textual series; progressive and punctual revisions; simulated translation project; textometric analysis; translation and revision workflow.

[1] We work with a Pure Neural Server platform deployed at Université Paris Cité: https://plateformes.u-paris.fr/pure-neural-server-clillac-arp-plateforme-de-traduction-automatique-sur-serveur-universite/

1. Introduction

In this paper we use a variety of textometric methods to analyse the revisions made by students during a simulated website translation project. The aim of this contribution is not only to quantify what types of linguistic revisions are carried out during the translation workflow, but also to identify at which stages of the teaching workflow these edits are made. Using both quantitative and qualitative methods to examine the chronology of revisions (or *chronological textual series*, Salem 1988; 1991), we hope to shed light on how the developing skills of our students interact dynamically with the workflow of the project, as well as with the evolving macrostructure of the translation course itself.

The website translation project is part of a Masters-level university course in French-to-English (FR > EN) translation entitled: *Traduction de site web vers l'anglais* (TSA). The project is carried out as part of the International Network of Simulated Translation Bureaus (INSTB)[2]. In previous studies we focused on the impact of Neural Machine Translation (NMT) on various aspects of the TSA course. For example, we examined the difficulties and opportunities involved in integrating NMT into a complex pedagogical translation project (Gledhill and Zimina 2019). We also examined the challenges our students face when learning how to manage raw MT output (Zimina and Gledhill 2021), as well as how to reconcile the results of specialised MT with the requirements of institutional style guides and the recommendations of Plain Language guidelines (Gledhill and Zimina 2022; Froeliger, Gledhill and Zimina in press).

In this contribution, we set out to examine a previously unexplored dimension, namely the evolutive nature of revision within the TSA project. In particular, we set out to establish correlations between the different types of linguistic revisions made during the translation project and the different stages of the teaching workflow. Our aim here is to contribute to the nascent field of revision studies, in particular the on-going discussion of translation/revision competencies (TRC, as discussed in Koponen *et al.* 2021). In this perspective, there has been much research on procedural aspects of translation/post-editing/revision within the project workflow, with revision being seen as a problem-solving exercise which is essentially driven by the need for quality control: "[revision involves] a complete examination and up-dating of a translation" (Hernández Morin 2009: 1 [our translation]). Revision has also been theorised as an autonomous process, in terms of professional, cognitive and linguistic skillsets. For example, there has been much experimental research involving keystroke and eye-tracking analysis, as well as data looking at the different types of resources used by trainee translators, in an attempt to discover whether the same procedures, skills and resources are used across the translation project (Goulet *et al.* 2017; Robert *et al.* 2017). In the following discussion, we attempt to expand this research paradigm by using a new set of experimental tools (textometric analysis) to look at previously unexamined data (cycles of revision over the lifetime of a translation project).

[2] Details available here: https://www.instb.eu.

https://doi.org/10.6092/issn.1974-4382/18790

The following discussion is divided into three parts. In the first section, we set out briefly some of the particular features of the TSA course, including the general macrostructure of the course (the teaching workflow). Then, we look at the central role of technology within the project, setting out how NMT is integrated into the teaching and translation workflows. In the final sections, we use textometric analysis to look in detail at the editorial history of a project running from 2021 to 2022. As we see in the following discussion, there appears to be a correspondence between the different types of edits made and particular stages of the project workflow. In particular, we suggest two different revision types: 1) "progressive" edits, which appear to follow the natural progress of the project and which often involve modifications which themselves entail further revisions down the line, and 2) "punctual" edits, one-off modifications, which appear to react to different task constraints, often corresponding to particular phases of the project. In the following sections, we examine some of these evolutive processes in more detail.

2. Salient Features of the TSA course

The main characteristics of the TSA course have been described elsewhere (Gledhill and Zimina 2019). Here we set out some of main features of the TSA course that are relevant to the management and processing of linguistic revisions within the overall project.

2.1. Creating and managing a translation workflow for teaching

Managing a translation project is a complex task that becomes even harder when it comes to developing a translation project workflow for teaching purposes. Building a translation / revision workflow for a translation project such as TSA is a process in constant evolution. It encompasses nearly all facets of training management: curriculum adjustments, group work and assessment planning, learner feedback, technology updates, lab facilities, etc. Each year, much flexibility is expected from project coordinators in order to fine-tune various practical aspects of teaching.

2.1.1. Project roles

In terms of group work, a characteristic feature of the TSA project is that different project management roles are assigned to each student (with a single role sometimes being shared, or the same person having several roles). For example, here are project roles from the period under study (2021 – 2022):

- Contract Writer(s)
- Communication Manager(s)
- Website Auditor(s)
- Task Manager(s)
- Translation Resources Auditor(s)

https://doi.org/10.6092/issn.1974-4382/18790

- Technological Resources Coordinator(s)
- Translator(s)/Reviser(s): **all students**
- Terminologist(s)
- Quality Assurance Editor(s)
- Project Editor
- Website Administrator

2.2. Data Formats

With respect to data management, the trainee translators involved in the TSA project deal with several types of content and data formats:

- Website pages (html, pdf, etc.)
- Project dictionaries (txt, xls)
- Translation memories (txt, tmx, sdlxliff), editorial history EN0-EN5 (xlsx, pdf)
- Project reporting forms (doc, odt, pdf)

2.3. Tools workflow

We have tried different configurations of CAT technology in relation to our teaching workflow. The sequence in which each platform or tool is implemented into the workflow clearly has an impact on the translation project. For example, RWS Trados Studio is intricately interconnected with SYSTRAN Pure Neural Server using an API key. It would be interesting to study the cognitive effects of these tools on revision edits, and to examine how our students appropriate these tools into their editing practice. However, such questions remain unanswered, and require further study. It is sufficient for our purposes here to consider the different types of tools that are presented to students on the project:

- Machine Translation Engine: SYSTRAN Pure Neural Server (specialised MT with custom resources: project dictionaries, translation memories)
- Translation Management System: RWS Trados
- Moodle, Google Drive, Excel (project management documentation, reporting forms and editorial history).

2.4. Editing and revision within a "Qualitative Translation / Revision Workflow": the concept of the Revision Cycle

The "revision cycle" is a critical component of the TSA project (Gledhill and Zimina 2022). Our trainee translators are expected to deliver quality output after several complex cycles of post-editing and revision. To this end, students are provided with a number of documents (notably the English Style Guide published by the Directorate General for Translation of the EU 2016/2023) and we also encourage them to build their own Quality Assurance Document (including their own style guide, inclusive language guide, etc.). In addition, in future iterations of the course, the MQM standards (https://themqm.org) are planned to be integrated into the teaching workflow, thus giving students a comprehensive set

https://doi.org/10.6092/issn.1974-4382/18790

of quality control guidelines regarding project management, linguistic quality control, etc. In order to characterise our translation project design philosophy, we use the term "Qualitative Translation/Revision Workflow" (QTRW). QTRW involves several revision cycles, and each revision cycle (RC) is given a code (i.e. EN0, EN1, EN2 etc.). Here is a summary of the main cycles in a complete QTRW:

- EN0 (machine translation) > EN1 (post-edited translation). Generation of specialised machine translation output using custom resources (translation memories, terminology, specialised translation engine).
- EN1 (post-edited translation) > EN2 (first round of revision). Post-editing of all the task segments by the Translator.
- EN2 (first round of revision) > EN3 (second round of revision, style guide compliance). Revision of the same segments by the Reviser.
- EN3 (second round of revision) > EN4 (third round of revision in accordance with Editor comments). A further round of revision by students in the first instance (sometimes addressing specific editorial requests, style guide recommendations etc.).
- EN4 (third round of revision) > EN5 (translation delivery, translation quality control). Validation of the same segments by the Editor (i.e. the teacher).
- Delivery (with potentially more revision cycles, according to customer feedback).

2.5. Managing editorial history

Many features of computer-assisted translation tools and software have been conceived to facilitate the revision process. However, generating a permanent editorial "history" of changes made to each segment is not yet a standard feature of translation technology. As a rule, existing translation/revision logs are not designed for linguistic analysis and are not designed to produce quantitative data that can be stored outside the translation system. We suggest that this aspect reflects an underlying tendency in translation technology design, in which detailed linguistic analysis of the translation/revision process is seen as of little immediate relevance to the needs of the translation industry.

This may be part of a more general tendency to downplay the importance of language skills in society at large, as noted by commentators such as Hagège (2012). On this view, skills which are apparently seen as objective and value-free, such competencies in data management or image manipulation, are preferred over skills which involve potentially face-threatening decisions involving language, such as judging whether to re-write an unclear sentence, or the ability to identify key arguments in relatively long texts (Hagège 2012: 72-73). Thus, we have observed personally that some students see editing as merely "error correction", and they usually delete initial (non-revised) versions or the initial machine translation that they have generated. Similarly, some students see revision as a time-consuming obstacle rather than an opportunity for further learning, and uncritically accept the changes made by their classmates or by their teachers. As we have previously mentioned, this goes against the philosophy of the TSA course: we feel that revision should be seen as an opportunity to engage with the text and to add value to the translation project. More generally, the TSA

https://doi.org/10.6092/issn.1974-4382/18790

course should be seen as a collaborative project, with revision being the responsibility of the whole team (Gledhill and Zimina 2019: 63).

3. Textometric Analysis of Translation/Revision Processes

Multilingual textometric analysis is a field of research which brings together knowledge from several related disciplines such as translation theory, natural language processing (NLP) and textometrics. It covers a series of methods that enable the researcher to formally reorganise regularly-occurring textual sequences and to conduct statistical analysis based on the vocabulary of a corpus of texts (Salem 1991; Lebart, Salem and Berry 1998; Zimina 2005).

Abrupt changes that occur in the distribution of a textual unit in different parts of a corpus may raise questions concerning the identification of other related textual units (different realisations of the same lemma, related forms at the semantic level, etc.). Textometric tools (such as *iTrameur*: https://itrameur.clillac-arp.univ-paris-diderot.fr) allow the analyst not only to subdivide the text into tokens, but also to identify other types of textual units on several annotation layers (repeated segments calculated at the part-of-speech level, cooccurrence relations with syntactic dependencies, etc.).

In the following section, we describe a series of experiments that have been carried out in order to identify characteristic revision patterns in the TSA translation/revision data by means of textometric browsing (Zimina 2005). This approach allows the user to navigate among the results produced by different methods of textometric analysis and the original corpus. As we see below, the observation of the frequencies and co-occurrences of the textual units in the translation/revision corpus reveals patterns that are indicative of different types of discourse that are characteristic of the source text (the FFD website, as described below).

3.1. The TSA revision corpus

We collected all the revision data from a project conducted by a group of 2nd Year Master's students (2021-2022) who translated and revised 49 pages from the FFD website (the official website of the French Federation of People with Diabetes: https://www.federationdesdiabetiques.org).

3.1.1. Translating into English and related challenges

We have previously commented on the challenges faced by the students on the TSA course, the foremost of which being the ability (and confidence) to translate website content into English (Zimina and Gledhill 2021: 68, 76). It is now increasingly recognised that an ability to work towards English is a viable core skill for many L2 (non-native) students of translation (Taviano 2018; Garnier 2020). In parallel, it is also increasingly acknowledged that L1 ("native" or "national") varieties of English are less important in distributional terms than international forms of the language, especially since so much online material,

https://doi.org/10.6092/issn.1974-4382/18790

including translation memory, is generated by L2 users. This concept is encapsulated by terms such as Global English or English as a lingua franca (ELF) (Jenkins 2009; House 2013). But although many academics and industry insiders recognise the validity of ELF, such a "plural" view of English has not penetrated public discourse, including institutions such as the Directorate General for Translation. Thus, the DGT's English Style Guide still defines the norms of English in terms of national standards (although it is notable that since the UK's withdrawal from the EU, the DGT has replaced "British English" by "Irish/British English"):

> 1.1. Language usage. The language used in English texts should be understandable to speakers of Irish/British English (defined in the introduction to this guide as the shared standard usage of Ireland and the United Kingdom). As a general rule, Irish/British English should be preferred, and Americanisms that are liable not to be understood by speakers of Irish/British English should be avoided. (EU 2016/2023:7)

Notwithstanding such a "nativist" stance, we generally encourage our students to adopt an international or neutral variety of English. In this respect, our students also face a number of complementary linguistic challenges that are related to the ability to handle the subtleties of English, namely:

1) the students also have to acquire skills in Plain Language (a requirement of most style guides, and the general practice on many Anglophone websites),
2) the students also have to familiarise themselves with a variety of "atypical" discourses and text-types (genres) which they are often unfamiliar with (in the L2 or even their L1), especially since the websites we are dealing with often adopt 'militant' or 'unorthodox' discourse styles.

3.1.2. Editorial history

The editorial history of the project consists of the machine translation output (EN0) and five consecutive revisions (EN1-EN5). Table 1 shows the main quantitative characteristics of the TSA French/English parallel corpus. The translation/revision data (EN0-EN5) are aligned with the source web page segments (FR) at the sentence level.

Table 1. Quantitative characteristics of the TSA corpus (type/token).

Part	**F token**	**F type**
FR	52617	6301
EN0	46319	5468
EN1	46779	5522
EN2	46604	5499
EN3	47027	5492
EN4	47073	5496
EN5	47219	5486

https://doi.org/10.6092/issn.1974-4382/18790

3.1.3. Syntactic tagging

This aligned corpus was syntactically tagged using UDPipe 1 (https://ufal.mff.cuni.cz/udpipe/1). The annotated data in the CoNLL format were then reformatted into an annotated database for *iTrameur* via a perl script.

We thus obtained an aligned textometric database with six annotation layers (form; lemma; category (pos); XPOSTAG for language-specific part-of-speech tag; underscore if not available; FEATS for list of morphological features; underscore if not available; DEPREL(HEAD)) for universal dependency relation, cf. Table 2. The data were partitioned to trace the editorial history (FR, EN0-EN5).

Table 2. The TSA corpus as a textometric database with six annotation layers.

#num	#Type	#forme	#pos	#leme	#XPOSTAG	#FEATS	#DEPREL(HEAD)
118043	forme	French	PROPN	French	NNP	Number = Sing	COMPOUND(118045)
118044	delim		BLANK	BLANK	_	_	_
118045	forme	Ministry	PROPN	Ministry	NNP	Number = Sing	OBL(118039)
118046	delim		BLANK	BLANK	_	_	_
118047	forme	of	ADP	of	IN	_	CASE(118049)
118048	delim		BLANK	BLANK	_	_	_
118049	forme	Health	PROPN	Health	NNP	Number = Sing	NMOD(118045)

3.2 Characteristic elements computation based on POS annotations of individual tokens

To trace important changes in the revision process across different parts of the TSA corpus, we calculated characteristic elements on several annotation layers. Tables 3-4 show a variety of significant variations of text unit distributions across the entire project (from EN0 to EN5) on the layer #XPOSTAG corresponding to language-specific part-of-speech tags. The tags used in annotation come from the Universal Dependencies (UD) project: https://universaldependencies.org. The main trends in Tables 3-4 can be interpreted as follows:

FW: “Foreign Words” are significantly underrepresented by the machine translation (EN0), but become stable during the main revision cycles (EN1-EN4). They are then overrepresented during the final revision cycle (EN5). As we suggest below, this is part of a general tendency to re-introduce key foreign items from the source text. The abrupt increase during the final stages of the revision process probably represents the work of terminologists (who by this stage have established a stable glossary for the project).

https://doi.org/10.6092/issn.1974-4382/18790

Table 3. Results of characteristic elements computation on XPOSTAGs: Part 1.

Item	FQ	EN0 / fq	EN0 / sp	EN1 / fq	EN1 / sp	EN2 / fq	EN2 / sp	EN3 / fq	EN3 / sp	EN4 / fq	EN4 / sp	EN5 / fq	EN5 / sp
FW	687	57	-11	114	0	125	0	125	0	127	0	139	3
VBG	4910	692	-7	825	0	824	0	847	0	864	0	858	0
NNS	22046	3476	-4	3684	0	3676	0	3737	0	3727	0	3746	0
-RSB-	77	5	-3	23	4	13	0	11	0	11	0	14	0
-LSB-	76	5	-3	23	4	13	0	11	0	11	0	13	0
RB	7970	1259	-2	1331	0	1310	0	1343	0	1353	0	1374	0
IN	30429	4999	0	5040	0	5031	0	5103	0	5111	0	5145	0
DT	22152	3718	0	3700	0	3654	0	3698	0	3689	0	3693	0
NNP	20636	3356	0	3341	-2	3432	0	3457	0	3500	0	3550	0
JJ	16690	2767	0	2806	0	2765	0	2781	0	2782	0	2789	0

Table 4. Results of characteristic elements computation on XPOSTAGs: Part 2.

Item	FQ	EN0 / fq	EN0 / sp	EN1 / fq	EN1 / sp	EN2 / fq	EN2 / sp	EN3 / fq	EN3 / sp	EN4 / fq	EN4 / sp	EN5 / fq	EN5 / sp
NNPS	661	148	5	112	0	107	0	101	0	103	0	90	-3
PRP	8762	1550	4	1447	0	1443	0	1447	0	1442	0	1433	0
,	10914	1883	3	1825	0	1802	0	1811	0	1794	0	1799	0
SYM	259	56	3	36	0	47	0	38	0	41	0	41	0
NN	41209	6915	2	6903	0	6858	0	6857	0	6836	0	6840	0
VBP	6390	1107	2	1044	0	1052	0	1063	0	1069	0	1055	0
IN	30429	4999	0	5040	0	5031	0	5103	0	5111	0	5145	0
DT	22152	3718	0	3700	0	3654	0	3698	0	3689	0	3693	0
NNP	20636	3356	0	3341	-2	3432	0	3457	0	3500	0	3550	0
JJ	16690	2767	0	2806	0	2765	0	2781	0	2782	0	2789	0

VBG: "Verb gerund-participles" are also significantly underrepresented at EN0, while their usage stabilises during cycles EN1-EN5. As mentioned below, there are several explanations for this, mostly relating to phraseological preferences in the target language.

NNS: "Plural nouns" are similarly underrepresented at EN0, while they become more prevalent during cycles EN1-5. Below, we suggest that this can be explained by a series of terminological expansions and explanations introduced during the revision process.

LSB/ RSB: "Left / Right Square Brackets" are underrepresented at EN0, but then increase in use during cycle EN1 (the first revision cycle). We suggest below that this modification is part of the standard practice of inserting a "Translator's Note", although the timing here represents changes that are particular to the FFD project.

RB: "Adverbs" are underrepresented at EN0 and then become stable during cycles EN1-EN4. Below, we suggest that this comes from subtle reformulations

https://doi.org/10.6092/issn.1974-4382/18790

involving the expansion of the verbal group in English across a variety of different segments.

In the following sections, we examine each of these special cases in more detail, giving examples and contexts of use.

3.2.1. Foreign Words

As mentioned above, the category **FW** (Foreign Words) is underrepresented at the machine translation stage (EN0, specificity index -11), but then becomes over-represented towards the end of the revision process (EN5, specificity index +3). The hypergeometric model used to calculate the specificity index is described in (Lebart *et al.* 1998; Lebart and Salem 1994). It identifies statistically significant elements that occur more frequently in a selected part of the corpus as compared to the corpus in its entirety. The explanation of the results appears straightforward: the FFD website contains several references to French addresses, names of organisations, titles and so on. If proper nouns and "ergonyms" (organisation-specific terminology) are not recognised by the NMT, they can be incorrectly rendered as English forms. The DGT English Style Guide gives specific recommendations on how to represent source language nomenclature (section 5):

> If an official body or a company does not have an English name, treat this as proper name and leave it in the original form: Bundesbank, CNRS, Médecins Sans Frontières.

There are many instances of "foreign word restoration" in the FFD project. One interesting example involves a postal address, in which the NMT deals correctly with items such as the name of the Federation, the street name and the postcode, but fails to recognise the name for a specific part of the organisation. It is notable that the original error means that the students have had to reintroduce capitalisation twice during the revision process:

> FR *Fédération Française des Diabétiques,* ***Service relations donateurs,*** *88 rue de la Roquette, CS20013, 75544 Paris cedex 11.*
> EN0 French Federation of Diabetics, **donor Relations Department**, 88 rue de la Roquette, CS20013, 75544 Paris cedex 11.
> EN1 French Federation of Diabetics, **Donor Relations Department**, 88 rue de la Roquette, CS20013, 75544 Paris cedex 11.
> EN2 Fédération Française des Diabétiques, **Service relations donateurs**, 88 rue de la Roquette, CS20013, 75544 Paris cedex 11.
> EN3-5 Fédération Française des Diabétiques, **Service Relations Donateurs**, 88 rue de la Roquette, CS20013, 75544 Paris cedex 11.

Other occurrences include names which go beyond the scope of the FFD project. The following segment, for example, is only ever encountered on the website as a link to an event (*États Généraux > General Congress / Assembly*) and thus constitutes an ergonym whose form in French has had to be restored during the revision process:

https://doi.org/10.6092/issn.1974-4382/18790

FR *Les États Généraux du Diabète et des Diabétiques*
EN0 The States General of Diabetes and Diabetics
EN1-5 **États Généraux du Diabète et des Diabétiques**

3.2.2. Verb Gerund Participles

The tag **VBG** refers to gerund-participle constructions involving the form *being*. The category of "gerund-participle" covers a variety of different verbal constructions (Huddleston and Pullum 2002), and tends to be used in a wider set of contexts in English when compared with equivalent verbal structures in French (Guillemin-Flescher 1981; Gómez-Castejón 2012). Thus, some very general principles of comparative grammar may explain why the form is strongly underrepresented at the machine translation stage (EN0) of the project (specificity index: -7).

One source of gerund-participle forms involves the translation of non-clausal structures in French by clausal forms in English, as in the following:

FR *La Fédération Française des Diabétiques est une association régie par la loi du 1er juillet 1901, fondée le 25 mars 1938,* ***reconnue d****'utilité publique le 3 décembre 1976.*
EN0 The French Federation of Diabetics is an association governed by the law of July 1, 1901, founded on March 25, 1938, **recognized of** public utility on December 3, 1976.
EN1-4 The French Federation of Diabetics is an association governed by the law of July 1, 1901, founded on March 25, 1938, **recognized as** of public utility on December 3, 1976.
EN5 The French Federation of People with Diabetes is an association governed by the law of 1 July 1901. It was founded on 25 March 1938 and **recognised as being** of public interest on 3 December 1976.

The explanation for this is phraseological (lexico-grammatical preferences). In both French and English, the verb *reconnu / recognized* can introduce an attributive complement (in French *d'utilité publique*). However, mental process verbs in English such as *consider* and *recognise* do not license prepositional phrases for this function, instead preferring a bound clause headed by *being*. It is notable that the students partially recognised this pattern (as can be seen in EN1-4, where the subordinator as erroneously introduced after *recognised*).

A similar example involves a fronted clause in French, which is initially rendered by the NMT by a very marked form in English (a reduced clause, headed by the ADJ *free*). During the revision process, this was repaired progressively, and then replaced by a less marked structure (expanded passive, subordinate clause) in the final revision:

FR *Ainsi,* ***débarrassés*** *des peurs ou lourdeurs qui nous empêcheraient d'oser et d'agir, nous pouvons construire de meilleures réponses à nos propres besoins et à ceux de nos parties prenantes.*

https://doi.org/10.6092/issn.1974-4382/18790

EN0 So, **free** of the fear or heaviness that would prevent us from daring and acting, we can build better responses to our own needs and to those of our stakeholders.
EN1-3 Thus **free** of the fear and heaviness that would prevent us from daring to act, we are able to forge better responses suited to our own needs and to those of our stakeholders.
EN4-5 Without **being weighed** down by fear that would prevent us from daring to act, we are able to forge better responses suited to our own needs and to those of our stakeholders.

Finally, another potential source of the VBG tag involves the "resultative passive". In this structure, a simple passive can be used to express an on-going process in French, while the equivalent structure in English should be rendered by a progressive form. The difference can be seen in the French ATM message *Votre demande est traitée* > *Your request *is made* which should be rendered as > *Your request is being made*. Here is a similar example from the FFD project, made more complex by a change in perspective (lexical verb) and transitivity (nominalisation of the verb):

FR Déjà 3500 personnes **accompagnées**, pourquoi pas vous ?
EN0 Already 3,500 people **followed**, why not you?
EN1-4 Already 3.500 people **are being guided**, why not you?
EN5 Already, 3 500 people **are receiving guidance**, why not you?

3.2.3. Plural Nouns

The presence of the tag **NNS** (cf. Table 3) indicates that plural nouns are under-represented at the machine translation (EN0) stage of the project, and then introduced gradually during the revision process. There are many potential sources for this. One simple explanation may involve certain very high frequency keywords that happen to be singular in French but are plural (or have a superficial plural form) in English. This is certainly the case for the single most important term on the FFD website *diabète / diabetes,* which in English can be used in contexts which are lexically ambiguous (and thus prone to be mis-tagged), as in the following:

FR *Faites reculer **le diabète**!*
EN0 Get **diabetes** back!
EN1-5 Help us fight **diabetes**!

A more sophisticated example involves FR *ma glycémie* > EN *my glycaemia,* a term that occurs right across the website, including entire pages devoted to the topic. In this case, the student team responsible for terminology had found that the compound term *blood sugar* is preferred to *glycaemia* on a number of Anglophone websites. This corroborates our experience with previous projects, in which technical terms in FR (which often have Graeco-Latin form) are often replaced by "plain language" expansions or explanations in EN. In this case, *my glycaemia*

https://doi.org/10.6092/issn.1974-4382/18790

was at first rendered as a singular form *my blood sugar* at the EN1 (post-editing) stage, but then during subsequent revisions, the translators changed the frame of reference (moving from the referent to a facet of the referent: here referring to a variable *level*), with a subsequent grammatical change (making *blood sugar* the modifier of *level* and introducing a plural form):

> FR *Ma **glycémie***
> EN0 My **blood sugar**
> EN1-5 My **blood sugar levels**

In a related example, the NMT appears to have treated *glycaemia* as a Foreign Word (or perhaps as a potential typo). In this case, the students correctly decided to harmonise their translation in line with the rest of the website:

> FR *Des substances qui font varier la **glycaemia***
> EN0 Substances that make **glycaemia** vary
> EN1-5 Substances that vary **blood sugar levels**

3.2.4. Left / Right Square Brackets

As mentioned above, the tags **LSB / RSB** are generally underused during the post-editing cycle EN1. This result is not surprising, as the use of parentheses and brackets corresponds to standard practice when introducing extra material or explanatory notes (metacomments, etc.), as mentioned in the DGT English Style Guide (section 2):

> When translating, use square brackets to insert translations or explanations after names or titles left in the original language.

An example of this can be seen in this (admittedly unusual) example, in which the student has attempted to repair a poorly-rendered version of the Koran into English (a topic relevant to Diabetes because of the religious practice of fasting):

> FR *La Sourate II verset 183 précise : "Si le jeûne peut altérer de manière significative la santé du jeûneur ou lorsque la personne est malade, l'Islam l'exempte du jeûne".*
> EN0 Surat II verse 184 states: "If fasting can significantly affect the fast's health or when the person is ill, Islam exempts him from fasting."
> EN1 Surat II verse 184: "**[Fast for]** a limited number of days. So whoever among you is ill or on a journey [during them] – then an equal number of days **[are to be made up]**. And upon those who are able **[to fast, but with hardship]** – a ransom **[as substitute]** of feeding a poor person **[each day]**. …"

It is notable that the material included in these reformulations goes beyond the translation task, including reinterpretation and accompanying commentary on the original Koranic text (thus going beyond the source text on the French

https://doi.org/10.6092/issn.1974-4382/18790

website). This type of editing is clearly of a different nature to the other revisions we have seen do far. We argue below that while most other forms are "progressive" (in that they involve changes that are relatively predictable and implemented across various stages of the revision process), this type of "one-off" or occasional intervention is more likely to be encountered at just one stage of the process.

3.2.5. Adverbs

The tag **RB** refers to various adverbial types. Again, the textometric analysis picks up on the fact that these forms are underrepresented during machine translation and then become stable during the remaining revision cycles. One explanation for this may be the tendency for translators to introduce 'modulation' in the English text. By 'modulation' we are not referring to the term familiar to many translators (Vinay et Darbelnet 1977), but rather to a subtype of modality in systemic functional grammar (Matthiessen *et al.* 2014: 145-146) which involves the expression of obligation / inclination, on a scale from positive (commands, declaratives, performatives, etc.) to negative (hedging items, attenuating adverbials, etc.)[3]. The following example shows a rather sophisticated form of this, involving reorientation (change of person, from "I" to "you") as well as the introduction of two modal adverbs ("really" and "perhaps"). We suggest that such changes modify the pragmatic function of the segment, encouraging the addressee to interpret the clause as a suggestion or a possible explanation, rather than as a statement of declarative fact:

> FR *Je ne suis pas intéressé par l'activité physique.*
> EN0 I am not interested in physical activity.
> EN1-4 You are not **really** into sports.
> EN5 **Perhaps** you are not **really** into sports.

It is important to note that the students who came up with this reformulation used the same technique of reorientation (i.e., moving from first-person to second person perspective) for many similar segments. This reformulation therefore corresponds to a deliberate strategy of communication, and forms part of a long list of similar statements that can be found on one page of the FFD website. Thus, it may be that this reduplication of the same form may have had an impact on our characteristic elements results. Although this form of revision may resemble highly specific "punctual" rather than "progressive" editing, we see below that there is also a tendency to find adverbials in co-occurrence in a variety of complex verb group structures, as detailed in the following sections.

[3] This contrasts with 'modalisation', which is concerned with the expression of propositions on a scale of positive (affirmation) to negative (negation) (Matthiessen *et al.* 2014: 144).

https://doi.org/10.6092/issn.1974-4382/18790

3.3 Characteristic elements computation on POS n-grams

In the previous section, we looked at outstanding features which are recognised by characteristic analysis at different stages of the revision process in the form of individual items (represented by one particular tag). In this section, we examine the co-occurrence of two or more tags in sequences, or POS n-grams. In phraseological terms, this is known as "colligation": the co-occurrence of one POS tag with another (e.g., Noun + Verb), or one POS tag with another in a particular sequence (e.g., Noun ^ Verb, etc). Either configuration may be indicative a regularly recurring lexico-grammatical pattern. For this analysis, we use "pos" and "XPOSTAG" annotations (cf. Table 2).

One particular observation is that verbs of all types are generally underrepresented in machine translation (specificity index: -2). In general, the use of verbs then increases through the revision cycles (cf. Figure 1):

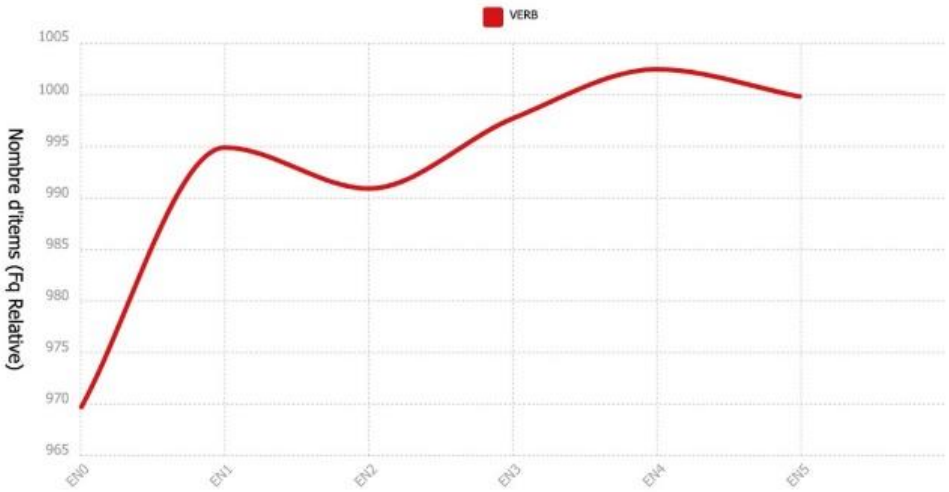

Figure 1. Relative frequencies of VERBS in translation/revision cycles.

The increased use of verbal forms during the revision stages of the project is an expected feature when translating from French to English. There are however several different contexts in which this can occur, and it is interesting to see which of these are prevalent in the revision process of the TSA project.

https://doi.org/10.6092/issn.1974-4382/18790

3.3.1. Pronoun + Verb

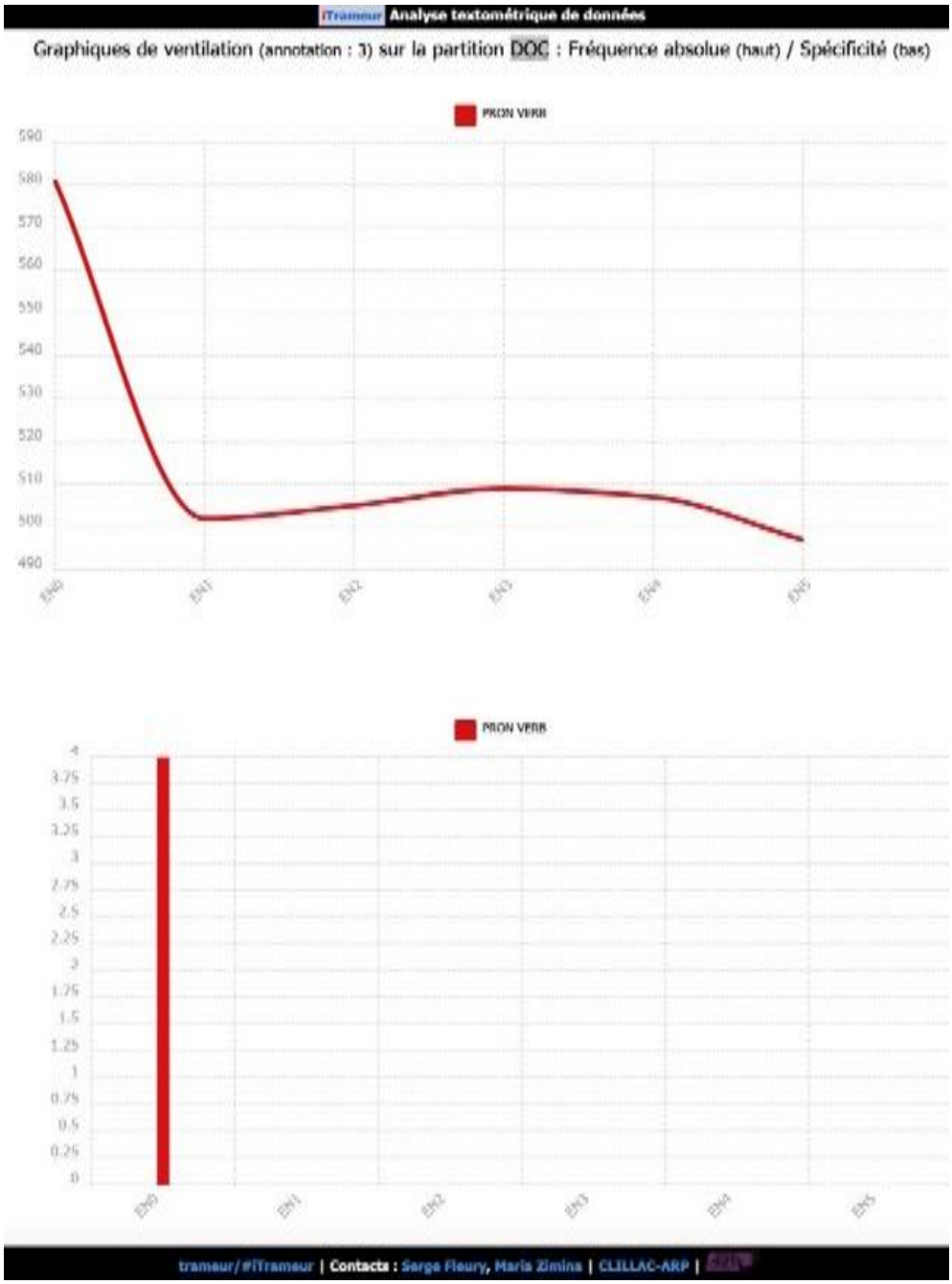

Figure 2. Distributional profile of the POS n-gram PRON + VERB in the translation/revision cycles.

As seen in the previous section, there is a strong tendency for verbs to be introduced in the English text during revision (cf. Figure 1). There is however a parallel tendency, which involves the maintenance of Pronoun + Verb in the original machine translation, followed by the expansion of the verbal group in later revisions (cf. Figure 2). One explanation for this may lie in the transposition of predicative nouns from the French source text into clauses in English: for each new clause in EN, there is a subsequent need to introduce a pronoun subject for

https://doi.org/10.6092/issn.1974-4382/18790

the verb. The following is a typical but also rather complex example, in that the English version involves a different verb:

> FR *La régularité est importante dans la **pratique** d'une activité physique.*
> EN0 Regularity is important in the **practise** of physical activity.
> EN5 Regular exercise is important when **it comes** to physical activity.

Another source for this pattern involves the translation of "performatives", that is to say, instructions expressed not in the form of imperatives, but rather by the first person in French (using "*je* + verb", as though the authors of the website are adopting their readers' voice). Such examples are often rendered by the first person at EN0, but are subsequently re-oriented during the revision process, either towards second-person forms ("you") or the imperative. The following selection gives an idea of the range of forms involved:

> FR ***Je peux** joindre les différents services de la Fédération par email et poser mes questions via le formulaire de contact.*
> EN0 **I can** reach the different services of the Federation by email and ask my questions via the contact form.
> EN4: **You can** reach the different services of the Federation by email and ask your questions via the contact form

> FR ***Je reste** informé, **je m'inscris** à la newsletter*
> EN0 **I stay** informed, **I subscribe** to the newsletter
> EN1-2 **I stay i**nformed, **I subscribe** to the newsletter
> EN3-5 **Stay** informed, **subscribe** to the newsletter

> FR ***J'équilibre** mon alimentation*
> EN0 **Balance** my diet
> EN4 **Eat** a balanced diet

Since the imperative in English removes the need to represent the pronoun in French, the consequent decrease in the relative frequency of PRON tags may have had an impact on the frequency distribution. We have also noted that there are differences in the use of interjections (INTJ) in EN0 and EN5, and this may be associated with the use of *please* in these contexts, as in this example:

> FR ***j'appelle** le 01 40 09 24 25.*
> EN0 **I call** 01 40 09 24 25.
> EN5 **Please call** +33 1 40 09 24 25.

https://doi.org/10.6092/issn.1974-4382/18790

3.3.2. Auxiliary + Adverb + Verb

Another feature of the revised translation is revealed by characteristic changes in the distribution of the POS pattern AUX + ADV + VERB ("should also follow" "are now included"). This pattern is underrepresented (specificity index: -3) in the machine translation output (EN0), but then undergoes a significant increase in use in the remaining revision cycles (cf. Figure 3).

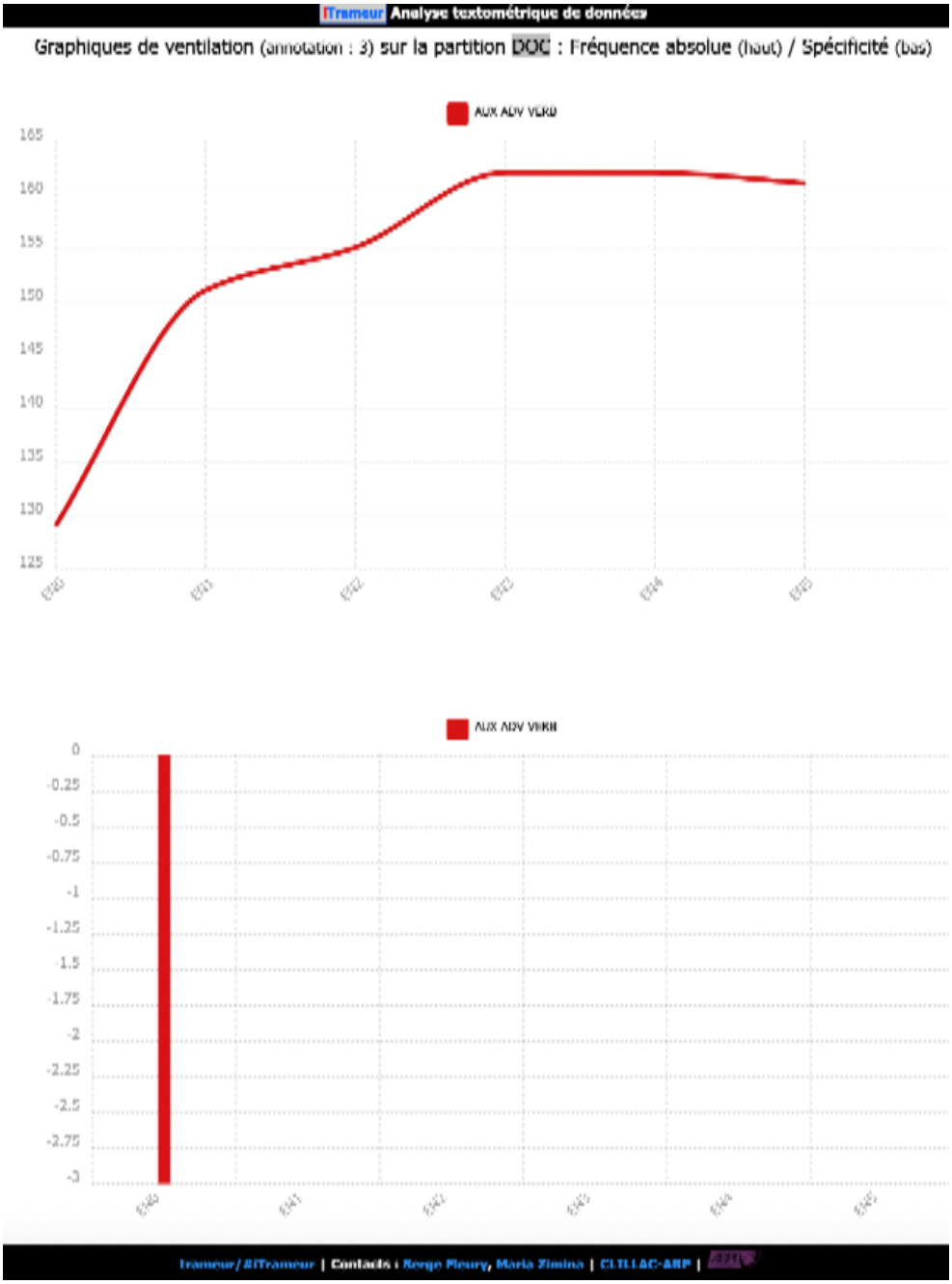

Figure 3. Distributional profile of the POS n-gram AUX + ADV + VERB in the translation/revision cycles.

Given the syntactic preferences of French and English, there are various explanations for this result. Generally, while adverbial and prepositional phrases

can be placed freely after lexical verbs in French, this position is highly marked in English (as shown in Gledhill 2011), a restriction that may account for many examples of material being moved away from the post-verbal position and placed elsewhere, as in the following:

> FR *Si les activités physiques ou sportives **sont recommandées en général** pour les personnes diabétiques, il faut adapter ses pratiques en fonction de…*
> EN0 If physical or sporting activities **in general are recommended** for people with diabetes, you have to adapt your practices according to...
> EN1 Sport and physical activities **are usually recommended** for people with diabetes, but it is necessary to adapt your practices according to your condition.

The sequence AUX + ADV + VERB can also occur when a simple verb in French is replaced by an expanded form in English, as in the case of active-to-passive reformulations:

> FR *Elles **relèvent aujourd'hui** du secteur hospitalier.*
> EN0 They **are now** in the hospital sector.
> EN1-4 They **are now included** in the hospital sector.
> EN5 These **are now included** in the hospital sector.

A more unexpected source for AUX + ADV + VERB can be found in *ad hoc* additions. Although most of these examples are unique, they generally involve the breaking up of the original segment and the addition of new clause. Thus, the following example involves a short re-statement of (nominalised) advice that is only expressed indirectly as a pronoun in the source text:

> FR *[conseils du médecin] d'adapter son alimentation et ses doses d'insuline en fonction de l'activité choisie, etc. **Et celles** de tout exercice physique en général :*
> EN0 [your doctor's advice is] to adapt your diet and your insulin doses according to your chosen activity, etc. **And those** of any physical exercise in general:
> EN5 [you should follow you doctor's advice] adapting your diet and your insulin doses according to your chosen activity, etc. You **should also follow the rules** of any physical activity in general:

In the following example, the translator has decided to explain the difference between two sources of medical insurance in the French system. This new material is expressed as a secondary clause which restates the modal verb first encountered in the primary clause:

> FR *Leurs tarifs diffèrent mais généralement ce sont ceux de médecins conventionnés du **secteur 1 ou 2**.*
> EN0 Their rates differ, but generally they are **sector 1 or 2** physicians.
> EN1-4 Their rates differ, but they are usually **sector 1 or 2** fund doctors.

https://doi.org/10.6092/issn.1974-4382/18790

> EN5 Their charges will differ. They can be imposed by the French Health Insurance ("sector 1" doctors) or **can fully depend** on the practitioner ("sector 2" doctors).

A further example of *ad hoc* addition involves the division of single segment into two separate segments. Here a simple French imperative is expanded into a full clause, with the addition of the subject from the previous text, a modal verb and textual adjunct (*can also*):

> FR *Mais si vous préférez le cadre d'un club ou d'une salle de sport, pensez aux possibilités de paiement en plusieurs fois,* ***voyez*** *si la mairie, une association locale ou le centre social ne propose pas la même activité à moindre coût.*
> EN0 But if you prefer the setting of a club or a gym, think about the possibilities of payment in several times, **see** if the town hall, a local association or the social centre does not offer the same activity at a lower cost.
> EN1-5 However, if you prefer practicing within a club or a gym, think about the option of paying in several times. **Check** if the town hall, a local association or the community centre offers the same activity at a lower cost.
> EN6 (additional revision cycle, student's reporting form data) However, if you would like to do sports in a club or a gym, consider paying over several instalments. You **can also check** if the town hall, a local association or the community canter can offer the same activity at a lower cost.

3.3.3. Noun + Noun

Characteristic elements analysis reveals that the pattern Noun + Noun is underrepresented at the machine translation stage (EN0, specificity -3) but stabilises during the rest of the revision process (EN1-5). We interpret this as evidence for the increased use of pre-modifying nouns and other reformulations of the nominal group during revision.

Figures 5 and 6 reveal there is a tendency to translate simple nominals in French by compound nominals in English in human revision, especially in contexts where non-technical ("plain") language is recommended. As mentioned above in relation to the plural (3.2.3), this statistical trend becomes particularly prevalent in the FFD website translation project, because of very high frequency terms such as *glycémie,* which is regularly translated by *blood sugar (level)* in the revised translation (EN1-5).

https://doi.org/10.6092/issn.1974-4382/18790

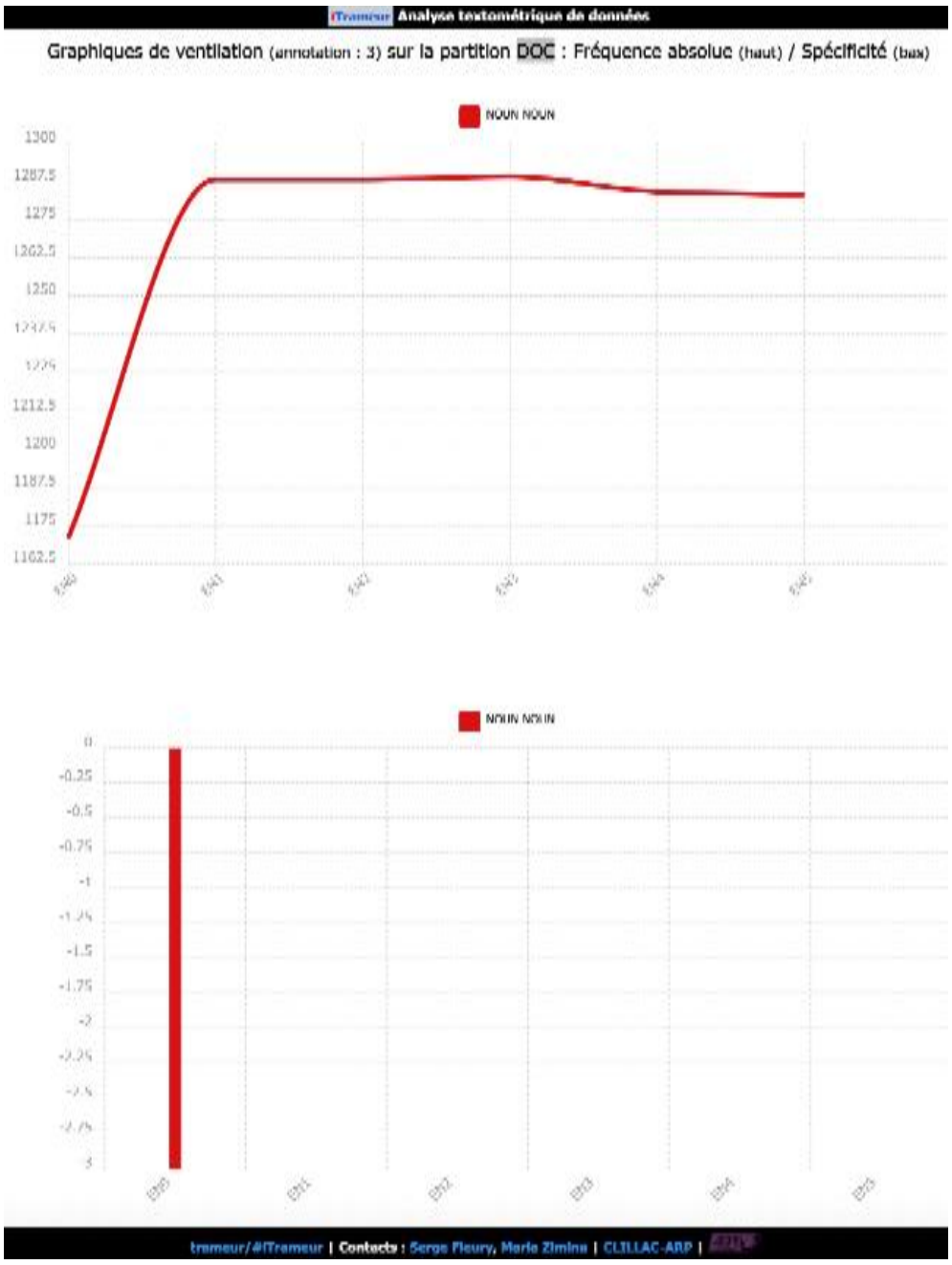

Figure 4. Distributional profile of the POS n-gram NOUN + NOUN in the translation/revision cycles.

https://doi.org/10.6092/issn.1974-4382/18790

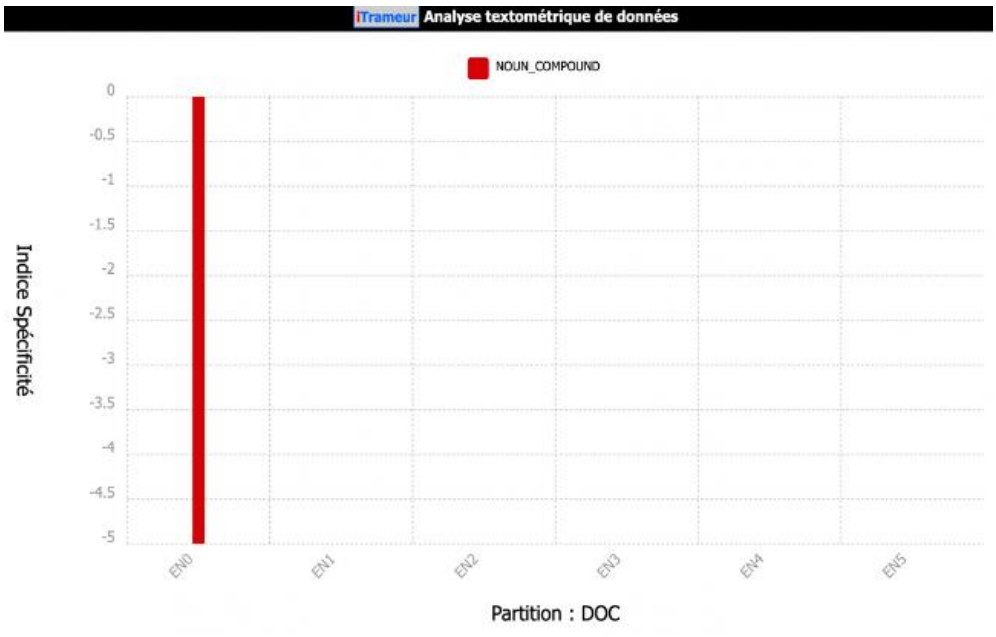

Figure 5. Distributional profile of NOUNS identified by the COMPOUND dependency in the translation/revision cycles.

iTrameur Analyse textométrique de données

Copy CSV Excel PDF Print

Recherche : COMPOUND

Item	FQ	EN0 / fq	EN0 / sp	EN1 / fq	EN1 / sp	EN2 / fq	EN2 / sp	EN3 / fq	EN3 / sp	EN4 / fq	E /
level_COMPOUND_sugar	646	34	-19	117	0	124	2	124	0	125	
federation_COMPOUND_French	778	128	0	133	0	129	0	128	0	129	
sugar_COMPOUND_blood	625	57	-8	113	0	116	0	114	0	114	
reduction_COMPOUND_tax	260	45	0	44	0	43	0	43	0	43	
level_COMPOUND_blood	178	10	-6	31	0	33	0	34	0	35	
Ministry_COMPOUND_French	167	15	-3	23	0	31	0	32	0	33	
return_COMPOUND_tax	181	30	0	31	0	30	0	30	0	30	
diabete_COMPOUND_type	183	32	0	29	0	29	0	31	0	31	
medal_COMPOUND_gold	166	27	0	27	0	28	0	28	0	28	
benefit_COMPOUND_federation	156	20	0	28	0	27	0	27	0	27	

Affichage de 1 à 10 des 631 items (filtrage à partir des 15,790 items)

Figure 6. Results of characteristic elements computation for NOUNS identified by COMPOUND dependency in *iTrameur*.

It is interesting to note that this type of reformulation has an impact not only on the translation of nominals, but also affects forms that are derived from these nominals. For example, the compound *blood level* is also used introduced to translate the adjective *glycémique,* as in:

> FR: *Des études démontrent en effet ses nombreux bienfaits, notamment l'amélioration de l'équilibre* ***glycémique****.*

https://doi.org/10.6092/issn.1974-4382/18790

EN 0: Studies show its many benefits, including improved **glycaemic** equilibrium.
EN 5: Studies show how beneficial it can be, including with improving **blood sugar** balance.

Similarly, compounds such as *blood sugar / blood glucose* can be further pre-modified to render terms such as*glycémie à jeun* > *fasting blood sugar* or used as a pre-modifier to translate *lecteur de glycémie* > *blood glucose meter* (although in this case the multi-word compound has been replaced by a multi-morpheme term at a later stage of revision, as can be seen at EN5):

FR: *Elle se surveille de deux façons : en laboratoire d'analyses médicales : pour mesurer sa* ***glycémie à jeun*** *et tous les 3 mois, son hémoglobine glyquée (HbA1c) avec un* ***lecteur de glycémie****…*
EN0 It monitors itself in two ways: in the laboratory of medical analysis: to measure his **fasting blood glucose** and every 3 months, his blood glucose (HbA1c), with a **blood glucose meter** …
EN5: It can be monitored in two ways: Laboratory tests: measure your **fasting blood sugar** (FBS) and glycated haemoglobin (HbA1c) every three months, with a **glucometer** …

While these examples all involve technical terminology, a further example demonstrates that the sequence N + N can also emerge in a more discursive context, this time replacing an adjective by a pre-modifying N:

FR *Au fil de plus de 80 ans d'existence, notre Fédération, reconnue d'utilité publique, a su se transformer mais également apporter sa contribution à des* ***évolutions décisives****.*
EN0 Over more than 80 years of existence, our Federation, recognized as a public utility, has been able to transform itself but also to make its contribution to **decisive developments**.
EN1 In over 80 years of existence, our Federation, recognised under French law as a public utility, has undergone constant evolution and has also contributed to **landmark developments.**
EN5Throughout its 80 years of existence, our Federation has undergone constant evolution, contributing to **landmark developments**, which has led it to be recognized under French law as public utility.

Although N + N is clearly associated with technical term formation in English, other sources for this pattern can also be observed. For instance, complex nominals in French are regularly formed using a sequence of nouns plus the preposition *de* (*N* + *de* + *N* + *de*). In the machine translation, this can often result in a "cascade of of's". During the subsequent revision process, the students are encouraged to create more balanced nominals, either using prepositions other than "of" or by transforming a post-modifier into a pre-modifier in English. Both techniques can be seen in the following example:

https://doi.org/10.6092/issn.1974-4382/18790

FR *100 ans **de** la découverte **de l'insuline***
EN0 100 years **of** discovery **of insulin**
EN1-5 100 years **since** the discovery **of insulin**
EN6 100-year anniversary **of insulin** discovery

A similar type of revision can be seen at work in the following example, although here the machine translation has proposed a solution using a specific preposition (*of>for*), while the translators prefer a three-element compound:

FR *Concernant les données relatives aux prospects issues d'un fichier loué ou échangé, elles ne sont pas conservées au-delà de la durée nécessaire à la réalisation de **la campagne de collecte de dons** pour lesquelles elles ont été louées ou échangées.*
EN0 For data concerning prospects from a leased or exchanged file, it shall not be kept for longer than is necessary for carrying out **the campaign for the collection of the donations** for which it has been leased or exchanged.
EN5 Data concerning prospects from a leased or exchanged file shall not be kept for longer than the time necessary for the corresponding **donation collection campaign**.

3.4 Editorial history, longitudinal partition, textual time series

3.4.1. Correspondence analysis of the TSA translation / revision

The first quantitative analyses of TSA revision cycles reveal the importance of phenomena related to text progression. We use Correspondence Analysis (CA) to further explore chronological progression of translation/revision cycles (Lebart *et al.*, 1998).

CA is a multivariate statistical technique that graphically displays a contingency table by computing the coordinates representing its rows (textual units in our case) and columns (corpus parts) and their interdependencies. As one can see from the plot of the CA dimension 1and 2 (cf. Figure 7), there is a clear serial structure in the revision data (i.e., the so-called "Gutmann effect": Lebart et Salem, 1994; Salem 1991). It can be seen here that CA is capturing the relative chronological order of the revision cycles. Red lines connecting consecutive corpus parts (revision cycles) have been added to highlight the outcome provided by the CA. With reference to the dimension 1and 2, we can see that the text revision progresses from right to left. Moreover, the third dimension (i.e., the vertical one) is capturing a trend of variation related to human/machine translation, separating MT output (EN0) from human revisions (EN1-5).

https://doi.org/10.6092/issn.1974-4382/18790

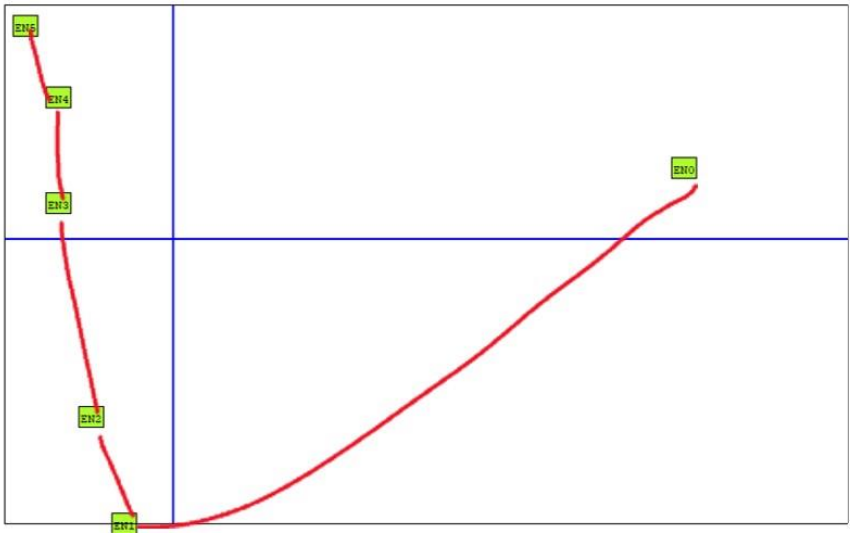

Figure 7. Correspondence analysis of the TSA revision cycles.

3.4.2. Temporal analysis of small contextual shifts in revision cycles

We notice that the chronological order of the partition gives the TSA revision data special properties in terms of the chronological textual series (i.e., homogeneous periodical series constituted by texts produced in similar enunciation situations, Salem 1991). In the case of chronological textual series, the results obtained by textometric methods must be interpreted according to specific rules. Coefficients calculated on the basis of the distribution of textual units through the different periods of the corpus make it possible to compare and explore the revision cycles. We rely on the periodisation metrics provided by temporal variations of barycenter positions (BT) and von Neumann stability analysis (VN) (explained in Salem 1988), to explore the revision data (cf. Figure 8). Thus, CA allows us to highlight the elements that change over time due to the chronological structure of the data. The results calculated by *iTrameur* are informative because they enable us to identify small contextual changes that occur in specific revision cycles. Such successive changes are sometimes difficult to capture without taking into account the chronological dimension of the corpus (lexical time).

In EN4, for example, the form ‘EUR’ replaces the symbol € in compliance with the English Style Guide (cf. Figure 8).

https://doi.org/10.6092/issn.1974-4382/18790

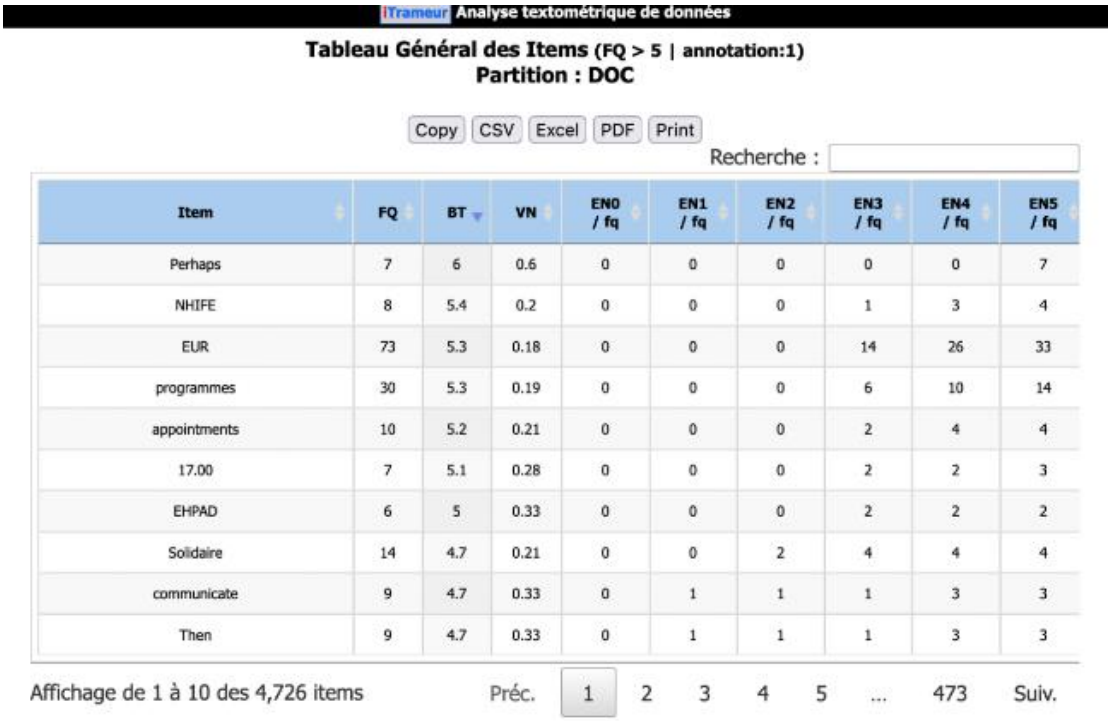

Tableau Général des Items (FQ > 5 | annotation:1)
Partition : DOC

Copy CSV Excel PDF Print

Recherche :

Item	FQ	BT	VN	EN0 / fq	EN1 / fq	EN2 / fq	EN3 / fq	EN4 / fq	EN5 / fq
Perhaps	7	6	0.6	0	0	0	0	0	7
NHIFE	8	5.4	0.2	0	0	0	1	3	4
EUR	73	5.3	0.18	0	0	0	14	26	33
programmes	30	5.3	0.19	0	0	0	6	10	14
appointments	10	5.2	0.21	0	0	0	2	4	4
17.00	7	5.1	0.28	0	0	0	2	2	3
EHPAD	6	5	0.33	0	0	0	2	2	2
Solidaire	14	4.7	0.21	0	0	2	4	4	4
communicate	9	4.7	0.33	0	1	1	1	3	3
Then	9	4.7	0.33	0	1	1	1	3	3

Affichage de 1 à 10 des 4,726 items Préc. 1 2 3 4 5 ... 473 Suiv.

Figure 8. Periodisation metrics provided by the temporal barycentre (TB) / von Neumann coefficient (VN).

Legend: The 'BT' (temporal barycenter, *barycentretemporel* in French) measures the distribution of a given unit: we measure where in the partition we have as many occurrences before as after. The "VN" index measures the differences in frequency of a unit between the different parts: if the differences are stable, the unit evolves continuously within the chronology; if the frequency varies greatly from one part to another, a "coarse" profile with ups and downs is identified. Finally, the combination of these two indications highlights the parts that evolve according to the chronology (we sort by "BT" and identify low "VN" to spot shifting elements that change over time due to the chronological structure of the data, such as "Perhaps" or "EUR").

4. Discussion

The patterns observed in the previous sections can be discussed in terms of two general tendencies:

1. **Progressive revisions:** "on-going" or partial modifications which can be observed right across the project and can involve several further changes as the project progresses (and can thus be seen as potentially stretching across one or more revision cycles).
2. **Punctual revisions:** "one-off" or isolated modifications which are often associated with a very specific part of the project, and can be completed (or reversed) in one revision cycle; these can be unpredictable one-off edits

https://doi.org/10.6092/issn.1974-4382/18790

(reacting to a specific problem) or edits which are more or less predictable (reflecting collective editing decisions, editorial instructions, etc.)

Within both revision types, it is also possible to distinguish between **project-initial, project-central** and **project-final** revisions**.**

Project-initial revisions are implemented at the very beginning of the project workflow, and can be either progressive or punctual. The clearest example of project-initial modifications involves Noun ^ Noun compounds, which as we have seen are introduced in significant numbers very early-on in the project.

In contrast, project-final revisions tend to involve punctual revisions that are often introduced as part of the harmonisation process (in terms of the style guide) or as part of the terminology project (since ergonyms and project terminology often only emerge at the tail end of the project). The clearest examples of this include the restoration of Foreign Words (as discussed in section 3.2.1) and punctual modifications such as the correction of the currency symbol (section 3.4.2).

Finally, it is interesting to note that project-central revisions (that is to say, edits which predominate in cycles EN2-3) are not frequent, but they provide some of the most interesting data. For example, it could be argued that the Pronoun ^ Verb pattern (mentioned above, section 3.3.1) is a project-central progressive revision, in that many examples of this sequence are removed at stage EN0, but then re-introduced during later stages, as students learn to replace nominal predicates with clauses. We also suggest that the pattern AUX ^ ADV ^ VERB is also a good example of a project-central progressive revision, since this pattern emerges gradually over the course of cycles EN1-4, reaching a peak at EN3. However, as discussed above (3.3.2), this pattern involves many different types of structure, and so cannot be seen as a reaction to a particular editorial decision.

Table 3. Summary of revision types according to duration (Progressive / Punctual) and timing (Project-initial / Project-central / Project-final).

Progressive revisions	**Project-initial revisions** NNS Plural Nouns N + N Noun ^ Noun RB Adverbs (Modal Adverbials) VBG Verbal Gerund-Participles
	Project-central revisions AUX + RB + VB Auxiliary ^ Adverb ^ Verb (listed twice) PRON + VERB Pronoun ^ Verb VBG Verbal Gerund-Participles
	Project-final revisions VBG Verbal Gerund-Participles
Punctual revisions	**Project-initial revisions** LSB/RSB Left / Right Square Brackets
	Project-central revisions [No examples found in our dataset].
	Project-final revisions FW Foreign Words

https://doi.org/10.6092/issn.1974-4382/18790

5. Conclusion

What lessons can be gleaned from the data we have set out in this paper? As mentioned in the introduction, we have previously attempted to systematise the TSA course by creating a model for the translation workflow project as a whole (the Quality Translation Revision Workflow) and by devising a teaching workflow built around the core activity of revision. Perhaps the most important message we infer from the above analysis is that revision data are just as important to the analysis of a translation as the initial and final stages of the project. At the very least, this kind of study emphasises just how important it is to conserve the intermediary stages of a translation project; a feature of the project workflow that is often underdeveloped within translation platforms.

From a broader theoretical and empirical perspective, the data set out above open a number of interesting research perspectives. We would claim that the patterns of revision identified here are not randomly distributed, but represent a number of significant regularities of practice. Perhaps the clearest distinction can be seen between project-initial revisions that are conducted immediately after machine translation (EN1 onwards), and project-final revisions (at EN4 or later). Project-initial revisions tend to be "progressive", in that once made the edits also often involve further modifications (as can be seen in the iconic example: *glycaemia* > *blood sugar* > *blood sugar levels*). Examples such as these show very clearly that some kinds of revision are not one-off modifications, but can involve a chain of reformulations, many of which being implemented in reaction to other considerations. For example, it may have been that our students decided to use "blood sugar" collectively, but the addition of "levels" turned out be a local, isolated addition. Such revisions clearly represent very complex phenomena indeed.

It occurs to us therefore that it is important to establish a typology of revisions that takes account of their status and function in the chronology of the project. In this paper, we have made a modest proposal to distinguish in the first instance between progressive and punctual revisions. The recognition that some revisions may be "partial" or "on-going" allows us to see the revision process as truly "evolutive", that is to say as a developing, accumulative process rather than as an exercise in "getting it right first time". Such an approach would certainly be a reassurance for our students, who are sometimes stressed by the need to finish a translation / revision cycle, often forgetting that they will (indeed must!) have another opportunity to improve their text at a later stage. In any case, it remains to be seen whether such a distinction is truly useful, and it would in any case be necessary to conduct further analysis (on a much broader data set, for example). However, the category of "progressive revisions" that we have identified here (including syntactic reformulations such as adverbial displacement, the formation of nominal compounds, the preference for clauses instead of predicative nominals, and so on) may reflect a number of lexico-grammatical patterns that have already been identified as translation universals in the literature (Mossop 2014).

The examples above also show very clearly that certain types of revision can skew the data very strongly. This may be particularly true of punctual and

https://doi.org/10.6092/issn.1974-4382/18790

project-final revisions, which tend to be modifications that emerge as part of the specific editorial problems associated with the project (as we have seen with the example of Left and Right Square Brackets, which may have been introduced as an early-stage, one-off "punctual" revision in the translation of the Koran, section 3.2.7).

In sum, although the data that we have presented in the preceding sections do at times seem to point in some very different directions, we would claim that they show at least one thing: that the revision process is "evolutive", that is to say, the modifications that can be observed are not random, but involve a complex interplay of different editing constraints that can be objectively revealed combining quantitative and qualitative research methods.

REFERENCES

European Union (2016 [2023]) *English Style Guide. A handbook for authors and translators in the European Commission* (8th ed. Updated 27 February 2023). European Commission, Brussels. https://commission.europa.eu/system/files/2023-03/styleguide_english_dgt_en.pdf.

Froeliger, N., C. Gledhill and M. Zimina-Poirot (in press) "Intégrer des plateformes de traduction automatique neuronale dans l'enseignement de la traduction spécialisée", in *Actes du colloque Tralogy3: Human Translation and Natural Language Processing Forging a New Consensus?* Traduction humaine et traitement automatique des langues : vers un nouveau consensus ? (Université Paris 8 Vincennes – Saint Denis, 7-8 April 2022, Paris France).

FFD (Fédération Française des Diabétiques) : https://www.federationdesdiabetiques.org.

Garnier, L. (2020) "Traduire les sciences humaines et sociales du français vers l'anglais", *Traduire* 7. http://journals.openedition.org/bssg/498.

Gledhill, C. (2011) "A lexicogrammar approach to checking quality: Looking at one or two cases of comparative translation", in I. Depraetere (ed) *Perspectives on Translation Quality (Text, Translation, Computational Processing 9)*, Berlin: Mouton de Gruyter, 71-98.

------ and M. Zimina (2019) "The Impact of Machine Translation on a Master's Course in Web Translation: From Disrupted Practice to a Qualitative Translation/Revision Workflow", *Translating and the Computer 41* (London, 21-22 November 2019), Geneva: Editions Tradulex, 60-73. http://www.tradulex.com/varia/TC41-london2019.pdf.

------ and M. Zimina (2022) "Human-Machine Interaction: How to Integrate Plain Language Rules in the Revision Cycles of Neural Machine Translation (NMT) output", *De Europa: European and Global Studies Journal*, Special Issue (2022), 76-88.

Gómez-Castejón, Á. (2012) "Contrastive analysis and translation study from a corpus linguistics perspective", *International Journal of English Studies* 12 (2): 111-132. http://revistas.um.es/ijes.

https://doi.org/10.6092/issn.1974-4382/18790

Goulet, M.-J., M. Simard, C. P. Escartín and S. O'Brien (2017) "La traduction automatique comme outil d'aide à la rédaction scientifique en anglais langue seconde : résultats d'une étude exploratoire sur la qualité linguistique", *Asp* 72: 5-28.

Groom, N. and J. Grieve (2019) "The Evolution of A Legal Genre. Rhetorical Moves in British Patent Specifications, 1711 to 1860", in T. Fanego and P. Rodriguez-Puente (eds) *Corpus-based Research on Variation in English Legal Discourse*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 201-234.

Guillemin-Flescher, J. (1981) *Syntaxe comparée du français et de l'anglais : problèmes de traduction*, Paris : Editions Ophrys.

Hagège, C. (2012) *Contre la pensée unique*, Paris : Odile Jacob.

Hernández Morin, K. (2009) *La révision comme clé de la gestion de la qualité des traductions en contexte professionnel*. Thèse de doctorat, École doctorale Sciences humaines et sociales, Université européenne de Bretagne, Université Rennes 2, Rennes.

House, J. (2013) "English as a Lingua Franca and Translation", *The Interpreter and Translator Trainer* 7(2): 279-298.

Huddleston, R. and G. Pullum (2002) *Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press.

Jenkins, J. (2009) "English as a Lingua Franca: interpretations and attitudes", *World Englishes* 28(2): 200-207.

Koponen, M., B. Mossop, I. S. Robert and G. Scocchera (eds) (2021) *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes*, London: Routledge.

Lebart, L. and A. Salem (1994) *Statistique textuelle*, Paris: Dunod.

Lebart, L., A. Salem and L. Berry (1998) *Exploring textual data,* Text, speech, and language technology, v. 4*,* Dordrecht, Boston: Kluwer Academic Publishers.

Matthiessen, C., K. Teruya and M. Lam (2014) *Key Terms in Systemic Functional Linguistics*, London: Continuum.

Mossop, B. (2014) *Revising and Editing for Translators (3rd edition)*, New York: Routledge.

Robert, I. S., A. Rigouts Terryn, J.J. Ureel and A. Remael (2017) "Conceptualising Translation Revision Competence: A Pilot Study on the 'Tools and Research' Subcompetence", *JoSTrans: the Journal of Specialised Translation* 28: 293-316.

Salem, A. (1988) "Approches du temps lexical", *Mots. Les langages du politique* 17: 105-143. https://www.persee.fr/doc/mots_0243-6450_1988_num_17_1_1401.

------ (1991) "Les séries textuelles chronologiques", *Histoire & Mesure* 6(1/2): 149-175. https://www.persee.fr/doc/hism_0982-1783_1991_num_6_1_1390.

Taviano, L. (2018) "ELF as a Translational Lingua Franca: Reciprocal Influences Between ELF and Translation", *The Translator* 24(2): 1-14.

Vinay, J-P. and J. Darbelnet (1977) *Stylistique comparée du français et de l'anglais,* Paris: Didier.

Zimina, M. (2005) "Bi-text Topography and Quantitative Approaches of Parallel Text Processing", in *Proceedings from the Corpus Linguistics Conference Series* (12-17 July 2005, Birmingham University), vol. 1, n° 1, Centre for Corpus

Research, Birmingham University. https://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx.

------ and C. Gledhill (2021) “L’impact de la traduction automatique sur les pratiques langagières et professionnelles des apprentis-traducteurs : entre apports en efficacité et menaces pour la diversité des discours”, in J-C. Beacco, J. C. Herreras and C. Tremblay (eds), *Traduction automatique et usages sociaux des langues. Quelles conséquences pour la diversité linguistique ?,* Paris : L’Observatoire européen du plurilinguisme (OEP), 63-82.

https://doi.org/10.6092/issn.1974-4382/18790