Alma Mater Studiorum · Università di Bologna

DIPARTIMENTO DI INTERPRETAZIONE E TRADUZIONE

Corso di Laurea magistrale in Specialized Translation (classe LM-94)

TESI DI LAUREA
in CORPUS LINGUISTICS

Round-Trip Translation: A Method for Estimating Revision and Editing Difficulty of English as a Lingua Franca Academic Texts

Candidata:
Novella Tedesco

Relatrice:
Prof. Silvia Bernardini

Correlatori:
Prof. Alberto Barrón-Cedeño
Prof. Federico Garcea

Anno Accademico 2020/2021
III Appello

Acknowledgements

When I enrolled in the first year of the Master's Degree Programme in Specialized Translation at the Department of Interpretation and Translation of the University of Bologna (DIT), my knowledge of the professional role of specialized translators and revisors was very limited. My previous studies and work experience had not yet brought me into contact with either the practice of translation or the use of technology in linguistics studies.

After the first month of classes at DIT, my perspectives on what I wanted to do in the future had already radically changed. In particular, Professor Bernardini and Professor Partington made me discover corpus linguistics; during their course I developed a deep interest in the subject. Professor Bernardini has been a guide and reference point for me throughout my studies and I want to thank her sincerely for her continual support.

I would also like to thank Professor Garcea for sharing with me his intuitions about round-trip translation and Professor Barrón-Cedeño for his constant engagement with my work and his essential help in the technical aspects related to the present dissertation. Professor Ferraresi was also present in the weekly meetings that we had to monitor the management of the experimental part, and I would like to thank him for his insights and support in data analysis.

When the DOT project at Bologna University Press (BUP) began its first cycle of traineeships, I took the opportunity to start putting into practice the knowledge I had acquired during my first year of the Master's. I had the chance to experience translation and revision in a real working environment, and I had the possibility of using technological tools to complete translation and revision tasks. Moreover, my first

traineeship at BUP has shown me a glimpse into the world of academic research and publication, inspiring this work, which has been carried out within the framework of my second traineeship at BUP. I would therefore like to thank BUP and, in particular, Ilaria Laurenza, who supervised me during this period.

Last but not least, thanks to my best friends, to Kenny, my fellow companion in this adventure, and thanks to my whole family, my siblings, my grandmother, my mother and my dear aunt Picca: without your constant support and caring attention none of this would have been possible.

Abstract

The aim of this work is to examine the feasibility of automatically evaluating English as a Lingua Franca (ELF) academic writing for revision purposes in the context of the publishing industry, thus optimizing the performance of revisors and minimizing the costs and time needed to plan the revision workflow. The study also represents a starting point for achieving fairer and more standardized revisor rates.

The dissertation has two main objectives: on the one hand, shedding light on some theoretical issues regarding English linguistics and revision, which provide the background for the practical developments of this work. In particular, the concept of English as a lingua franca is framed in the context of international academic communication in the era of post-globalization. Machine translation and technological tools for revision are also analyzed from a theoretical perspective.

On the other hand, this study encompasses the design, implementation and testing of an automatic method for *a priori* evaluation of revision difficulty, based on the similarity between a source text and its round-trip translation (RTT). Accordingly, the experiments and the work carried out at Bologna University Press (BUP) have had the aim of developing specialized methods for revision difficulty evaluation. Nonetheless, the variety of data used in the experiments demonstrates the applicability of the method beyond the cases under study.

The results show that the RTT-based method, in conjunction with the BLEU score as a text similarity measure, can be successfully used to estimate revision difficulty, although further tests are needed before this method can be implemented in real working settings. These reflections are included in the discussion of each set of experiments,

leading to the final remarks and further developments of the study.

The possible application of RTT for language learning is also briefly considered as a way of generalizing about the potentials of this innovative method.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Framework and Specific Objectives

> Che cosa vuol dire tradurre? La prima e consolante risposta vorrebbe essere: dire la stessa cosa in un'altra lingua. Se non fosse che, in primo luogo, noi abbiamo molti problemi a stabilire che cosa significhi "dire la *stessa* cosa" [...]. In secondo luogo, perché, davanti a un testo da tradurre, non sappiamo quale sia *la cosa*. Infine, in certi casi, è persino dubbio che cosa voglia dire '*dire*'. (Eco, 2003: 9)

These are the first famous sentences of Umberto Eco's important work on translation, entitled 'Dire quasi la stessa cosa'. Paolucci, Professor in Interpretative Semiotics at the Alma Mater Studiorum University of Bologna and one of Eco's former students, drawing inspiration from the semiotic theories developed by his mentor, considers the practice of translation – conceived, in a broad sense, as the interpretation of signs through other signs – as playing a fundamental role in the human processes of meaning construction (Paolucci, 2010, 2020).

Taking inspiration from these philosophical reflections, the present dissertation aims to employ an unconventional approach to address a practical problem in the world of academic publishing: applying (automatic) translation as a method of estimating editing and revising difficulty of a text written by non-native speakers of English, used

as a lingua franca (Mossop, 2001; Hall, 2013[1]).

Revision – that can be generally defined as the act of modifying a text in order to improve its linguistic quality (Haar, 2016) – is a practice that has been particularly investigated in the field of translation studies, for several reasons. Firstly, because specialized translators – who are experts in the use of languages in editorial and scientific contexts – often find themselves acting as revisors at different levels and degrees of professionalism. Secondly, language revision is mostly practiced on works written by non-native speakers or translations, which, by the way, have been found to share many common linguistic features (Mossop, 2001). Scholars who dealt with revision from a theoretical point of view have been seeking to identify, in a more or less systematic manner, possible criteria for the detection, evaluation and correction of errors, which may be, in turn, due to a variety of causes (inadequacy to the textual and contextual standards of the target language and culture, linguistic interference, structural or semantic ambiguity, etc.). It should be noted that, much like translation and, in fact, all the other metalinguistic disciplines, revision, both theoretically and practically, is not exempt from being conditioned by specific conceptions of language. Therefore, the theoretical background, which guides studies on revision as well as its actual practice, should always be taken into consideration.

The problem that this thesis seeks to address is of a twofold nature. On the one hand, it attempts to find a method for assessing the difficulty of revision, conceived according to a pragmatic and functional vision of language, abandoning the constraints and prejudices deriving from prescriptivist ideologies that can lead, as various authors in the field of sociolinguistics have pointed out (e.g., Knowles, 1997; Canagarajah, 1999, 2002, 2012; Seidlhofer, 2009, 2011), to linguistic imperialism. On the other hand, the assessment of the quality of a manuscript in terms of revision difficulty is practically necessary to estimate costs and time of a revision: hence, the possibility of predicting this estimate automatically constitutes a new advantageous perspective for

[1] See the second section of the Introduction (1.2) for more information about the main references on this work.

revisors, publishers and clients. The problem of pre-assessing manuscripts has been recognized as critical by various studies; in this respect, Mossop (2001) notices that evaluating manuscripts can take a lot of time and effort, especially because evaluations need to be made in light of the publishing and costumer requirements. In particular, I will refer to the final work of a former student at the Department of Interpretation and Translation based in Forlì, Xiaoli Wang (2021). Her dissertation has, indeed, highlighted the necessity for an automatic tool that can determine revision difficulty *a priori*.

BUP, the Bologna University Press, is the context within which this work has been conducted, providing the link between theoretical speculation, theory of practice and actual revision practices. Indeed, the datasets of the experimental research, which, in truth, represents the central focus of this dissertation, have been published and revised at BUP. Furthermore, the objective of this experimental work is to examine the possibility of using the method of round-trip translation (RTT) to obtain an indication of the revision difficulty of manuscripts written in English as a lingua franca, with the idea that such a parameter can then become an estimate of costs and time of revision. Round-trip translation is the process of translating a text from a source language into a target language (forward translation, FT), then translating the result back into the source language (back translation, BT), using machine translation (MT) software (Moon et al., 2020). In this work, RTT is used for the production of similarity references to which the manuscripts under analyses are compared employing text similarity metrics, with the aim of obtaining an indication of their revision difficulty. In this way, both the problem related to finding objective criteria of correctness and the necessity of *a priori* evaluation for revision purposes can be overcome.

Circling back to Eco's work, cited at the beginning of this introduction, I would like to recall another important aspect of translation intended as a theory for meaning interpretation: according to Eco, translation is not an exact science; its regulating principle is that of 'almost'. The same, I believe, can be said for revision. But the author adds something more: this fact — that translation is not a one-to-one match, in which the translation B always and only corresponds to source text A — does not

prevent it from being an essential process through which, according to Eco, human knowledge is constructed. The 'almost' principle, therefore, far from being limiting, opens up many possibilities.

In this sense, I intend to explore the possibility of approaching the problem of revision from the point of view of translation – that is, investigate the possibility of understanding and evaluating the adequacy of a text by translating it. And, in an era of digital transformation, this work particularly aims to solve such a problem exploiting some of the most advanced technologies for automatic text processing.

1.2 Thesis Structure

This work is essentially interdisciplinary, positioning itself at a crossroads between English sociolinguistics[2], data science, corpus and computational linguistics, and translation studies.

In Chapter 2, the latest research trends in English as a lingua franca, digital humanities and corpus linguistics are examined, particularly as concerns academic writing, automatic text evaluation and revision techniques. I will refer to the works and theories by Baird et al. (2014), Canagarajah (1999, 2002, 2012), Hyland (2006, 2011, 2015), Hyland & Shaw (2016), McEnery & Hardie (2012), Seidlhofer (2011), Mossop (2001), Mitkov (2003). Then, Round-Trip Translation (RTT) is introduced as a method to automatically generate a similarity reference to be compared with the source text, and BLEU and BERTScore, the text similarity metrics used in the experiments, are presented and comprehensively described.

Chapter 3 begins with the description of the work I have done as a trainee at BUP, experimenting with revision from the 'inside' of a publishing house, outlining the issues that may arise during work as a revisor and keeping note of a publisher's needs (Section 3.2). The third section includes details on the compilation of the database, used in the experiments which constitute the main focus of Chapter 3. The basic idea of this

[2] 'Producing language is above all a social activity' (Mitkov, 2003: 285).

dissertation is to calculate the similarity between the original ELF English text and its round-trip translation, through the application of appropriate methods, obtaining with this procedure information on the needs for revision. The hypothesis underlying this study is that, when a manuscript needs significant revision, the similarity score between the revised text and its RTT is higher than the score between the manuscript and its RTT. A number of supporting hypotheses are consequently tested in the experimental part of this work. To test their validity, a series of experiments have been carried out which are described in Chapter 3.

Specifically, in Section 3.4.1, first the preliminary experiments, which were conducted on texts extracted from a volume on the COVID-19 pandemic (Bellettini and Goldstein, 2020) in order to refine the method and the starting hypotheses, are described. Then, a second set of experiments is presented. This concerns texts related to the XXXV Congress of the International Committee of the History of Arts (CIHA), published in the volume 'Motion: Transformation' (Faietti and Wolf, 2021), all of which were revised by the same revisor, who provided her evaluation before the actual RTT experiments were performed. This second set of experiments, specifically designed to test the validity of the method, are described in detail in Subsections 3.4.1, 3.4.2, 3.4.3, 3.4.4 and 3.4.6.

As a way of concluding the experimental work, in Chapter 4, RTT is performed on some unrevised data provided by BUP. In particular, a test of the method carried out using a set of three texts is presented and the encouraging results obtained from it are discussed. This last chapter also offers some insights on the limits and possible further developments and applications of the RTT method.

In Chapter 5, the final remarks of this work will be illustrated and conclusions will be drawn on the possibility of applying the RTT method and the BLEU score for the evaluation of revision difficulty of ELF manuscripts.

Chapter 2

Research Background

2.1 Introduction

This chapter provides an overview of the theoretical background underlying this work. The considerations here presented are the starting point from which the hypothesis could be formulated; they guided the management and interpretation of the experiments described in Chapter 3. Moreover, in the light of these theories, the motivation for this study can be better understood.

Because of the interdisciplinary nature of the problem I am addressing, it is impossible to aspire covering a complete review of literature while maintaining the scientific focus with which this dissertation was born. Therefore, I selected specific and up-to-date reflections that are relevant for the practical aim of this work – i.e., *a priori* automatic evaluation of revision difficulty – and that can help us to reframe and approach the problem from the still unexplored point of view of round-trip translation.

For this reason, the present chapter has macro-sections, referring to the different disciplines which contribute to the theoretical background for this study, and subsections where the most relevant topics are treated in a detailed manner.

First, in Section 2.2, the concept of English as a lingua franca (ELF) is addressed, with a particular focus on the context of English for Academic Purposes. Adopting

the ELF perspective seems fundamental for the aims of the present research, because it sets innovative criteria for quality evaluation of Academic Writing in English – which has to be measured considering communicative purposes more than an ideal native-speaker's standard. Indeed, the concept of *nativeness* itself carries with it meaningful reflections on social dynamics of power and colonialism.

The second theoretical field which needs to be explored is that of revision (Section 2.3). In this respect, the perspectives of both writing and translation studies are acknowledged with the purpose of reframing the issues related to revision practices in the specific context of the academic publishing industry. This section includes a review of studies on text quality evaluation, which – adopting a functionalist perspective – will be related to revision difficulty. Therefore, criteria for revision and some relevant revision practices will be analyzed.

Since this dissertation is based on experimental activity, Section 2.4 of this chapter is dedicated to the presentation of the theoretical background for the method employed in this study, which involves the implementation of NLP text similarity metrics. In order to address these topics properly, some elements of corpus linguistics, computational linguistics and artificial intelligence are presented.

Section 2.5 begins with a brief overview on the state of the art of machine translation. Finally, in Subsection 2.5.2 the method of round-trip translation is introduced as a feasible way to leave the 'native standard' behind in the evaluation of text quality for revision purposes, making it possible to overcome problems related to the definition of objective criteria of evaluation while allowing for a time-saving automatic pre-assessment of the manuscripts.

2.2 Towards a Definition of a Style: ELF and Academic Writing Perspectives

2.2.1 English as a Lingua Franca

In the era of post-colonialism and globalization *linguae francae* have developed as fundamental means for intercultural communication and knowledge sharing (Hall, 2013). Within this context, the massive spread of the English language has raised many questions concerning its usage in all the different domains of human interaction. For this reason, contemporary research in linguistics has variously explored the history and use of *linguae francae*; in this respect, the studies on English as a lingua franca (ELF) have always had a leading importance (Jenkins, 2015).

In her milestone work, 'Understanding English as a Lingua Franca', Seidlhofer (2011) dedicated much attention to the description of what she defined 'real vs. realistic English' and to the analysis of linguistic aspects specific to international English, referring to the principle of adaptation, the idiom principle, the compilation of an ELF corpus, and theories on World Englishes. In this work, ELF is defined through a functional perspective (Hülmbauer et al., 2008) as 'any use of English among speakers of different first languages for whom English is the communicative medium of choice and often the only option' (Seidlhofer, 2011).

In order to map the usage of ELF, Jenkins (2013) refers to the Three Circles model introduced by the linguist Braj Kachru (1992a, 1992b), who organized the territories influenced by the use of English into three main areas: Inner Circle, Outer Circle and Expanding Circle. As Schneider (2012) underlines, this model is generalized and only indicative of the actual linguistic situation. However, it can be a good starting point to describe in a sketchy manner the several different contexts where ELF has spread. Jenkins (2013) importantly includes the Inner Circle speakers communicating with speakers from the Outer or the Expanding Circle among the ELF users, providing a definition of ELF that combines an idea of deterritorialization with intercultural nego-

tiation processes (Ventola, 1991) of communication and knowledge creation, eventually defining ELF as a 'contact language'.

It has been claimed that 'the study of English as a Lingua Franca (ELF) is essentially the study of adaptive variation', a variation that is deeply invested with sociocultural values (Widdowson, 2017: 101). This idea is largely explored by Seidlhofer (2011) in the second chapter of her 'Understanding English as a lingua franca'. In particular, the two sections 'Anglo-Saxon Attitudes' and 'The Assumption of Native-Speaker Authority' focus on language ideologies. In this context, the description of English as a monolith, a fixed and homogenous linguistic system is presented as functional for the perpetration of linguistic imperialism, a notion that will be soon analyzed in more detail.

Following a similar school of thought, authors like Larsen-Freeman (2018) and Baird et al. (2014) have contextualized the study of ELF within the framework of complexity theories, unveiling the problematic nature of using terms with an intrinsically contrastive meaning, such as 'variation' and 'deviation', to characterize ELF. In particular, the article 'The Complexity of English as a lingua franca' problematizes the monolithic concept of 'ideology': the authors believe that a more complex approach, which takes into account the many factors affecting communication processes, could provide more powerful arguments to overcome the dichotomies which traditionally characterize the discourse on language, such as *nativeness/non-nativeness* and *norms/deviation* (Baird et al., 2014). Also Anna Mauranen (2020) follows in the footsteps of complexity theories (Baird et al., 2014) and analyzes the macro, meso and micro levels from which the concept of English as a lingua franca can be approached. This operation highlights that adopting a sociological perspective in the study of linguistic varieties and standards always involves the consideration of different social levels.

It is important to underline that, from the ELF point of view, language standards are continuously set and reset by users in the process of affirming their role as active speakers in the frame of sociopolitical communicative contexts (McKinley and Rose, 2018). The latter element is of crucial importance in the framework of this study, as it highlights the connection between ELF studies and translation, editing, and revision

practices for the publishing industry.

Furthermore, it should be noticed that ELF studies, by their own nature, have particularly benefited from the application of corpus linguistics approaches and methodologies. In this respect, two corpus-informed studies have proposed a description of English as a lingua franca: Prodromou's 'English as a lingua franca: a Corpus Based Analysis' (2009) and 'Analysing English as lingua franca: a Corpus Driven Investigation' by Cogo and Dewey (2012). Both works particularly focus on the point of view of language learners, providing useful reflections on competence evaluation, that have many points in common with the study of ELF within the context of academic communication. For example, Prodromou describes the difficulties of learners with respect to idiomatic phraseology. Although, idioms tend to be excluded from the academic register, phraseology is a big issue in revision. Cogo and Dewey's work, instead, covers ELF speaking practices in particular and is focused on elements of pragmatics, that normally occupy too little a space in the reflections concerning ELF use. This perspective is in line with the functional and pragmatical approaches that will be explored in this section, as well as in 2.3. Thus, descriptive studies of this kind play the decisive role of building empirical resources for both theoretical and practical purposes.

Chew's (2010) volume 'Emergent Lingua Francas and World Orders', which features the explicative subtitle 'The Politics and Place of English as a World Language', collocates ELF in the broader discussion on *linguae francae*, linking historical research to a detailed analysis of the relationship among current varieties of World Englishes. This volume is at the crossroads of ELF, linguistic imperialism and World Englishes theories. Such theories have first and foremost problematized the concept of *nativeness* applied to a language so steeped in colonialist history and dynamics of power negotiation as the English language is.

Taking into consideration the Foucauldian idea of power, the transdisciplinary concept of linguistic imperialism has been theorized in the last decades in the context of postcolonial studies by scholars such as Canagarajah (1999, 2012), Knowles (1997), Ashcroft (among his many contributions to this subject, of a particular relevance is 'The Em-

pire Writes Back: Theory and Practice in Post-Colonial Literatures', Ashcroft et al., 2002). By the expression 'linguistic imperialism', or 'language imperialism', Knowles, as quoted by Feng (2002), means 'the transfer of a dominant language to other people. [This] transfer – he explains – is essentially a demonstration of power' (ibid.: 98). As will be seen later, this perspective is particularly relevant for the identification of the social aspects that affect academic communication (Belcher, 1997; Canagarajah, 2002). Moreover, the post-colonial point of view is crucial when referring to language quality evaluation. In this sense, it certainly does not come as a surprise that British English has been, until recently, the standard to which every foreign English speaker and writer aimed. This is very clearly stated, among many others, by James and Lesley Milroy. In their preface to the volume 'Authority in Language', they write: 'it is well known that in British and American society judgments are made about "correct" and "incorrect" use of English and that in some countries, such as France and Italy, academies exist which prescribe "correct" use of the language concerned' (2012: vii). The standards of American English and the so-called Oxford English have gained importance in recent years, earning the role of main standards for publication when it comes to styles of academic writing (Bennett, 2009).

In the above-cited volume by Milroy and Milroy, the dynamics which have characterized the process of standardization of the English language are described in detail, as well as the mind frames that construct an imagery on language as a – sometimes discriminatory – indicator of one's social and cultural background. The discussion on standardization, as shown by the authors, involves a deep reflection on general linguistics and the way the study of language should be approached in the first place:

> although it is understandable that linguists should have to place clear limitations on their field of inquiry (especially if they are to make progress in *formal* linguistics, following Chomsky (1965) etc.), we are unlikely to make great progress in understanding the nature of language if we entirely ignore its *social* functions and characteristics. Amongst these are phenomena such as language standardization, the nature of literacy, notions of prestige in language and popular attitudes to usage. (2002: 7)

At this point, the concept of 'attitude', described by Sarnoff (1970) as a 'disposition to react favourably or unfavourably to a class of objects' can be used to understand how deeply relations of power permeate the use and study of languages. In recent times, the investigations of scholars such as Sapir and Whorf, have acquired a renewed value in the field of psycholinguistics (Liuzza et al., 2010), where the latest theories on language bias (see, for example, Pesciarelli et al., 2019 on the implicit processing of grammatical and stereotypical gender violations) have shown how languages are products of cultures just as much as they contribute to the attribution of meaning and interpretation of life (Prinz, 2012). Such reflections clarify that communicative systems are of a complex nature, and investigating human language necessarily involves the examination of social and political relations of power. This obviously applies to the study of academic writing and revision too, because in the practice of revision an actual negotiation process concerning both knowledge and language takes place (Hyland, 2006).

It should be noted that also feminist theories have played an essential role in this discussion, promoting key concepts – such as, those of *agency*, *situatedness*, *negotiation*, *transcreation* – for the socio-political debate on language (Baccolini, 2005). For instance, the concept of agency has been presented by many authors in different branches of gender studies. In the interdisciplinary field of feminist translation studies, scholars such as von Flotow, Tymozcko, Spivak and Simon have highlighted the need for a renewed transcultural sociopolitical consciousness in the practices of translation and revision, envisaging a stronger agency of the translator/revisor (Simon, 1996).

At this point, I'd like to cite Jennifer Jenkins (2018:2), who writes: 'ELF promotes radical change in the way we think about English as well as language more broadly'. Thus, the ELF lesson is crucial for any studies on English applied linguistics because it overcomes the traditional dichotomy between native and non-native speaker varieties. In the same contribution, the author (2018:3) explains that 'it is also evident that speakers of different first (and other) languages are influenced, albeit to a greater or lesser extent, by their language backgrounds'. Indeed, while one cannot pretend that the native language of authors writing in English as a lingua franca does not affect

their language production – and this aspect will be explored in relation to the case-studies under analysis (see Section 4.2) – my research is based on the common-ground constituted by all the aspects shared across ELF productions.

Furthermore, there is a special relationship between the establishment of English as a lingua franca as a standard and of academic writing as a text type that I am going to explore in more detail in Subsections 2.2.3 and 2.2.4.

2.2.2 Linguistic Interference

The interference of the authors' mother-tongue in their productions in English as a Lingua Franca should also be acknowledged. In linguistics, this phenomenon is known as 'language transfer', which can be defined as 'the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired' (Odlin, 1989: 27).

The general topic of language contact in a multilingual perspective has been explored by Zabrodskaja (2012) in her study on crosslinguistic influence on language and cognition. In fact, the expression 'language contact' is itself problematizd by the author, who wishes to underline the agency of users as well as the fact that languages are systems which are deeply rooted in culture. This study particularly focuses on the difference in cognition between monolingual and bilingual subjects, noticing that linguistic transfer – which is a notable feature in language productions of bilingual subjects and foreign speakers – does influence cognitive and social behaviors.

From the point of view of linguistics, this topic has been largely treated both in research on language learning and in translation studies. In the second language acquisition (SLA) field, for example, in 1992 Gass and Selinker published a book dedicated to this topic and the many approaches which have been proposed by different theories during the years – from contrastive studies, which focus on the contrastive analysis of language pairs, to general linguistics and Universal Grammar, which integrates language transfer in the conceptualization of language.

From the standpoint of translation studies, it has been noticed that translations share some features with learners' interlanguages and are, in general, different from the language used in original non-translated texts. This phenomen is recognized as 'translationese' (Zauberga, 2001; Schäffner & Adab, 2001; Volansky et al., 2013). Within this framework, Baroni and Bernardini (2006) and Avner et al. (2016) have addressed this issue from a computational linguistics point of view. In particular, Baroni and Bernardini have examined the possibility of exploiting machine learning techniques for differentiating between original texts and translations using a corpus of journal articles in Italian.

For what concerns the present dissertation, it has not been possible to understand whether this factor influences the outcomes of the method here presented, but this is a problem that should definitely be addressed in future research.

2.2.3 Academic Writing as a Text Type

The aim of this subsection is to understand the main aspects of academic writing from the perspective of text linguistics and pragmatics. This is an important operation that allows one to shed some light on theoretical issues related to text classification in the communicative context of academia while providing the possibility for an analysis of the main features of academic writing in a systematic way (Ferrari, 2014). In general, text linguistics offers fundamental support to studies on translation, revision and editing, providing the language specialist with helpful means and methodologies for meaning interpretation (Mazzoleni, 1996).

The fundamental bases of contemporary text linguistics are unquestionably to be found in the work of the Austrian linguists de Beaugrande and Dressler 'Introduction to Text Linguistics', first published in 1972 and translated in English in 1981. Of an essential relevance for this research are the chapters VI, VIII and X, on 'Intentionality and Acceptability', 'Situationality', and 'Research and Schooling'. From Grice's conversational maxims to anthropological approaches, the authors clearly establish the importance of pragmatics in linguistics.

The issue of text classification is not of a simple, clear-cut nature. For this reason, first it can be useful to explicate definitions and distinctions among basic concepts of text linguistics, such as text typologies and genres. On this specific matter, Mazzoleni (2002: 151) explains:

> le due categorie "genere" e "tipo" sono da distinguere accuratamente poiché si riferiscono a concetti decisamente diversi anche se interagenti: il genere viene elaborato empiricamente in maniera induttiva partendo dalle caratteristiche superficiali dei testi [...], mentre il tipo è un costrutto di portata teorica, più astratto, che non permette di classificare direttamente i testi reali ma che identifica in modo deduttivo le caratteristiche essenziali di diversi procedimenti e modalità di comunicazione in base agli scopi del mittente e ai rapporti instaurati con i destinatari.[3]

Albeit diametrically opposed to the traditional conceptualization of genres and text types in the English literature (Swales, 1990; Bhatia, 1993), such a position has been here adopted because it allows for the functional definition of the characteristics shared by the many different textual realizations of the academic discourse (such as terminological density and communicative purpose). The quotation above summarizes three essential concepts: (a) text types do not represent texts, as much as they refer to the context, procedures and conventions; (b) the purpose of communication is crucial for the interpretation of meaning; (c) communication contributes in and is shaped by social relationships. These elements are all present in the analyses provided by Charles et al. (2009: 1) in the volume 'Academic Writing. At the Interface of Corpus and Discourse', where the authors 'explore the interaction between two traditions of investigating written academic prose' – discourse analysis and corpus linguistics. Moreover, this book

[3] The two categories 'genre' and 'type' are to be carefully distinguished because they refer to clearly different – albeit interacting – concepts: genres are empirically elaborated in an inductive way according to the superficial characteristics of texts [...]. On the other hand, a text type is an abstract construct of theoretical scope, which does not directly enable the classification of real texts, but deductively identifies the essential features of different procedures and modes of communication related to the sender's purposes and the relationships established with the receivers [my translation].

is a valuable example of the possibilities offered by corpus linguistics methodologies – which are in almost all cases of a functionalist inspiration (McEnery & Hardie, 2012) – when related to the study of the use of language in academic settings. More details on this matter will be given in the following subsection.

My suggestion is that academic writing should be treated as a text type, although every production can take the form of a different genre – or genres. Indeed, a survey conducted by Karen Bennett between 2004 and 2007 on English Academic Styles Manuals (2009: 43),

> revealed a remarkable consensus as regards general principles, methods of textual construction, and the kinds of grammatical and lexical features to be used. This suggests the existence of a common framework underlying all EAD [English Academic Discourse], thereby supporting the claim made by Systemic Functional linguists that there is an "essential continuity between humanities and science as far as interpreting the world is concerned". (Martin, 1993)

It is beyond the scope of this study to reflect on genres of academic writing, whilst categorizing it as a text type allows for some needed clarifications on the specific features that all written linguistic productions of an academic nature share. In this sense, Troia et al. (2015: 291) notice how 'typical writing instruction generally does not reflect evidence-based practices (EBPs)', calling for the need to undertake corpus linguistics practices related to language description in teaching and evaluation of linguistic productions.

In this respect, it is important to keep in mind that such forms of categorization are abstract and fictive: they serve the purpose of highlighting structures and patterns of meaning at multiple levels, and should never be intended as limitations or prescriptions, also when it comes to revision. Referring to the theories on genre exposed by Derrida and Ronell in "The Laws of Genre" (1980), Shaw (2016) explains that 'that the relationship of a genre to a text [...] is not simply that of a class to one of its members (e.g Hymes 1974; Frow 2005)' (Shaw, 2016: 244). He represents the complex relationship between texts and genres as follows:

> The relation between texts and genres differs from that between species and individual or class and instantiation in at least three ways. One is that a text affects its genre. Innovations in a text change the definition of the genre, and as innovations accumulate the genre changes. Another is that texts do not have to draw on a single genre; genres can be mixed and texts can have features of several genres. [...] Third, a text can be more or less prototypical of its genre (Swales, 1990; Paltridge, 1997). Texts perform or draw on genres rather than instantiating them, so many will have the most frequent characteristics, but some will not. (Shaw, 2016: 244)

These reflections are particularly relevant because they emphasize the active participation of writers, revisors, publishing companies and academic communities to the consolidation of structures and features which are then regarded as typical of the academic writing.

From a post-colonial perspective, this matter has been investigated by Athelstan Suresh Canagarajah. At the beginning of his volume named 'A Geopolitics of Academic Writing' (2002), the scholar affirms something that may sound obvious, but that one should always remember when approaching the theme of academic writing: 'the representation of any research or academic inquiry is considerably mediated by the rhetorical processes of writing and publishing. Research findings cannot stand unaffected by textual forms of knowledge dissemination' (ibid.: 20). In this sense, the point of view of text linguistics on the central role of the communicative context in textual analysis become even more relevant.

Such approaches confirm and amplify those psycholinguistics theories that present language as a tool, such as the study by Borghi et al. (2013), 'The embodied mind extended: using words as social tools'. The authors write in the abstract to their article – that 'words, also due to their social and public character, can be conceived as quasi-external devices that extend our cognition' (Borghi et al., 2013: 1). This view perfectly fits with the conception of academic writing discussed here.

Going back to Canagarajah, he brings up a further problem related to the fact that writing is not used as the main means for knowledge communication in many non-

Western cultures, that is: before language comes into play, non-Western scholars willing to communicate within the international academic community must adapt to a foreign communication tool. As an exemple, the author refers to a work written in Tamil by Jaffna on the role of 'the *pundit* and indigenous scholarship' in the Indian culture (2002: 50), a paper which had not however been welcomed by the Indian readers, who found it offensive due to its ironic tone. In other words, the unequal relationship between center and periphery, which characterizes non-native English speakers (NNES) writing, results in the difficulty of confronting with a foreign means of expression.

But Canagarajah doesn't stop there. He understands that this fact has effects also on the projection of authorial identity. On this matter, it can be useful to refer to Hyland's article 'Authority and invisibility: authorial identity in academic writing' (2002), where he compares two corpora, one composed of Hong Kong undergraduate theses, and another one representing research articles by English native-speaker students. The features shown by Hyland's study (such as the lack of first person pronouns in NNES writing) represent important factors that should influence the work of revisors confronting with non-Western writings. These linguistic productions might not follow logical and structural patterns that are easy-to-interpret from a Western perspective. At the same time, as the author notices, 'if publishing conventions have such gatekeeping potential for knowledge construction, they raise concerns about the hidden interests they may harbor' (2002: 34).

The third chapter of the volume opens with a quotation from 'The Archeology of Knowledge and Discourse on Language' by Michel Foucault (1972), that says: 'in every society the production of discourse is at once controlled, selected, organized and redistributed according to a certain number of procedures, whose role is to avert its powers and its dangers, to cope with chance events, to evade its ponderous, awesome materiality' (Canagarajah, 2002: 77). This becomes evident in textuality every time that what could be defined as a *conventions-breakdown* takes place, just in the way described by Canagarajah.

Some lines above, referring to his book, I have mentioned the dichotomy center/periphery.

The whole work seems to concentrate on this topic and, in particular, on academic writing 'from the periphery', thus shedding light on its *situatedness* (Haraway, 1988). Besides the Sri Lankan author, Ken Hyland (2016) also offers a geopolitical view on the world academic publishing industry. Particularly, in the chapter 'Language: visibility and inequality' he describes the results of globalization processes in terms of 'inequality in the production and distribution of knowledge' (2016: 95). In this sense – and for what concerns the aims of my work – it could be interesting to consider some factors specific to the European context because, although not all the texts in the database used for this study were written by European writers, it is in Italy (precisely at the Bologna University Press) that the writings were revised.

With a focus on university environment and students written productions, Kruse's study on academic writing in European higher education (2013: 38) elucidates on the multilinguistic landscape of the continent. Indeed, he writes:

> different basic writing cultures [...] were established in Europe in the past and [...], to a certain extent, still form the base for academic writing. The linguistic and cultural diversity of the European continent is both an enriching and a dividing factor [...]. Writing routines [...] are strongly rooted in national cultures but are difficult to compare and difficult to evaluate.

This is clearly witnessed by the many academic style manuals that try to standardize writing habits and conventions, in particular when it comes to English productions (Bennett, 2009). A book that I would like to mention for its innovative and inclusive approach is 'Writing Spaces 1: Readings and Writing' (Lowe & Zemliansky, 2010). At the beginning of the volume, some 'myths' about writing are debunked: the authors downsize the role 'good grammar' plays in quality writing, while openly accepting some conventions typical of ELF such as the notable use of first-person pronouns compared to traditional British English academic writing. With respect to these changes in standards, it has to be acknowledged that recent studies on the spread of English in the European context have focused on the definition of a precise variety – the European English, namely the language of the European institutions (Murphy, 2008).

I introduced the discourse on text linguistics citing De Beaugrande and Dressler. In the early seventies they had already stated that 'all the levels of language are to be described in terms of their utilization' (De Beaugrande, Dressler, 1981: 31), according to what was defined as 'procedural approach' in a chapter that prophetically closes with some insights on automatic text processing, possible future developments of such technologies and their implications for textuality standards and text quality evaluation[4]. This, I believe, is a lesson that must never be forgotten by revisors.

The functionalist perspective has been intrinsically linked to the study of English for academic purposes (EAP) since the very beginning of this discipline, that emerged from the area of studies known as English for specific purposes (ESP), which in turn focuses on 'specific, purposeful uses of language' (Hyland & Shaw, 2016). This standpoint has been brought up, among the others, by Susan Hood (2016). In her chapter contained in 'The Routledge Handbook of English for Academic Purposes' (2016: 193) the author precisely refers to the systemic functional theory, which is defined as 'a social semiotic theory of language'.

Moving from elements of structural linguistics and following Halliday's model of communication (1961, 1985), Hood describes the communicative context as field, tenor and mode, where field refers to the institutional background as well as the particular goals of academic communication, tenor 'constitutes complex relations of power and solidarity that are played out in patterns of interpersonal meaning in discourse', and mode 'has to do with the ways in which interactions are mediated, impacting on potential feedback and enabling the relative distancing of discourse from material reality' (Hood, 2016: 194). In this way, the author demonstrates the complex influence that context has on academic productions, which revisors should always take into account.

These reflections highlight that communication is a complex process, thus explaining why automatic correctors such as Grammarly (to which I will refer in Section 2.4 and

[4] The influence of functional and procedural approaches to linguistics can be noticed in computational linguistics applications, in particular, in the task of natural language generation, as highlighted by Mitkov (2003).

in Chapter 3), despite being undoubtedly helpful tools for writers and learners, can be no more than a starting point for practices of self-editing. A lot still remains unexplored when it comes to the revision process, characterized by judgements hard to objectivize and, therefore, hard to translate into computational processing. Translation, as presented in the introduction, can be a means through which a given text is interpreted and procedurally re-elaborated (Eco, 2003; Paolucci, 2010; Mazzoleni & Menin, 2011). For this reason, the idea of using outputs from machine translation systems could overcome problems related to biased evaluation.

2.2.4 ELF, ESP and EAP: The Common Ground

In the introduction to the 2020 edition of the Routledge Handbook of English for Academic Purposes, Hyland and Shaw write: 'The term English for academic purposes (EAP) covers language research and instruction that focuses on the communicative needs and practices of individuals working in academic contexts' (2020: 1). As it can be read shortly after this definition, studies on EAP formally began when Tim Jones first used the term in 1974; since then, it has progressively become one of the major research topics in the study of English for specific purposes (ESP).

As it has been already shown earlier in this section, the study of English for Academic Purposes must take into account two main levels: (a) the contextual level, which covers socio-political relations, reflections on writers' *nativeness/non-nativeness* considering both the implications that arise from the use of English as the lingua franca for academic communication and knowledge construction, and the influence of political views and cultural frames of mind on the concrete linguistic acts; (b) the descriptive level, that has massively been informed by corpus linguistics methodologies. In this subsection I will briefly try to shed some light on both these aspects, which are inherently connected.

As earlier explained from a broader perspective (see 2.2.1, a geopolitical view has been adopted by many scholars in the attempt of analyzing academic writing in context. Bennett (2013), among others, concentrates on the 'overwhelming dominance of English

as a lingua franca in the academic domain' and brings up some instances of EAP used in the Portuguese context, highlighting the ambiguities related to the fact that academic writings in Portuguese and English do not always share the same logical paths and structures. The author sharply condemns what she defines 'epistemicide' – namely, the process through which hegemonic discourse takes over authors' epistemology – and proposes some possible counter-practices, with particular attention to translation.

The geopolitical approaches pave the way for an ethnographic analysis on EAP, which, as Paltridge et al. explain, seeks to 'explore the socially situated nature of the use of language in academic settings' (2016: 218). Among many others, the work by Lillis and Curry (2010) exemplifies some of the typical reflections in this field. Furthermore, some conclusive chapters of their volume contain radical calls toward open-accessibility and 'advocate the idea of "knowledge as a gift economy"' (Meẑek, 2011: 188).

In her volume dedicated to the use of English as a lingua franca in the university, Jenkins (2013: 2) thoroughly differentiate between the two concepts of globalization and internationalization. While the first is an economic phenomenon, the latter is defined as 'a key strategy by which universities have responded to the influence of globalization and mean[s] – here she cites Maringe and Foskett (2010) – "the integration of an international or intercultural dimension into the tripartite mission of teaching, research and service functions of Higher Education"'. In order to investigate the university context, Jenkins exploits Held et al.'s (1999) reflections on the three possible perspectives on globalization linking them to specific attitudes towards English. First, she defines the hyperglobalizers as believing that English is 'the property of its native speakers' and that it should be 'distributed in its native form' (2013: 8). Secondly, there are the skeptics, according to whom 'the findings of ELF research are trivial, that there is nothing particularly new about them, and/or that there is no need to move away from the teaching of English according to standard (i.e. native English) norms' (ibid.). Finally, she compares these two types of attitudes to the trasformationalist position, which relates to the need of addressing 'the considerable reshaping that movements in the socio-political world order have produced' (ibid.).

> This – Jenkins comments – is where ELF enters the frame. According to the ELF school of thought and supported by extensive empirical ELF research, innovative English features are emerging in intercultural communication not because speakers are unable to 'master' the forms of native English, but as a result of their desire (whether conscious or unconscious) to promote effective communication in interactions that are characterized by a far greater degree of diversity among English users than has been the case until recent times. (ibid.: 9)

Hence, the above-mentioned studies all support the same claim that *nativeness* is a rather problematic concept in the study of language. Ute Römer (2009: 89) provides some valuable evidence in favour of this statement. In the article titled: 'English in Academia: Does Nativeness Matters?', the author argues that EAP is a rather specific case of non-native linguistic production. As he clarifies,

> we are dealing with high-proficiency English [...]. We do not normally find non-standard features like missing articles or third-person-singular '-s' in academic English as represented in research articles, academic lectures [...] – and yet, it is a fact that a large and growing number of these (and other) types of text in academic contexts are produced by non-native speakers of English'.

Consequently, the author presents a corpus-driven analysis which aims at comparing productions by native and non-native learners of academic English with published official articles written by experienced authors, who are, in most cases, native-speakers. The author uses a tool named kfNgram, 'a free stand-alone Windows program for linguistic research' which – at the time when the study was published – was an innovative tool for n-Grams extraction, because it also provides a list of 'p-frames', collostructions which are recurrent in n-Grams[5]. The results of her analysis show that 'native speakers also have to learn the language (and phraseology) of academic writing. The native academic writer does not seem to exist' (2009: 99). The authors conclude their article shedding light on the possible implications of this finding on language teaching.

[5] https://www.kwicfinder.com/kfNgram/

Here I would like to suggest that similar implications can be encountered also for what concerns the work of revision. Moreover, this analysis is particularly relevant for the aim of the present thesis: by unveiling the fundamental centrality of phraseology, which in the linguistics landscape is known for Sinclair's formulation of the idiom principle (1991), they provide inspiration regarding the way this principle can be a) taken into consideration when revising b) exploited by automatic text productors (i.e. machine translation systems, as will be seen in Sections 2.3 and 2.5).

As Hilary Nesi explains in her chapter form The Routledge Handbook of English for Academic Purposes,

> corpora are now common in EAP research and practice, both to provide quantitative information about discourse, and to corroborate insights derived from more qualitative studies. They also play an increasingly important role in EAP pedagogy, providing syllabus items, examples to illustrate accepted usage, and opportunities for data-driven learning. (2016: 206)

A corpus-driven approach – here I refer to the difference between corpus-based and corpus-driven approaches as explained by McEnery & Hardie (2012) (see Subsection 2.4 for a detailed summary of corpus linguistics approaches) – is the one that, in my opinion, best fits the study of EAP because it allows for a description entirely 'driven' by the data at hand. For instance, investigating the corpus of academic writings and revisions that I created for BUP can provide some important insights on revision techniques specifically used at that publisher, as well as the company's preferences for register and style[6]. By way of a mere example, I conducted a simple analysis on key nGrams (three to five words) of the sub-corpus Manuscripts, using the sub-corpus Revisions

[6] The parallel corpus that I have compiled at the beginning of my second cycle of traineeship at BUP, the English Monolingual Parallel Corpus of Academic Writings and Revisions (see 3.2, could be used for analysis similar to this one, although the primary aim of that work was to create a sort of *revision memory* (see 3.2.1). Using the latest version of the web concordancer SketchEngine (developed by Lexical Computing in 2003) this resource can easily be explored, while data privacy is protected.

as the reference corpus. In this way, items which characterize the language of unrevised manuscripts are highlighted against the corrections of the revisor(s). Centuries spelled in words are easy-to-understand cases – the revisor has probably opted for ordinal or Roman numbers; predictably, the frequency of the verb 'make' is halved in the Revisions corpus, as well as that of the verb 'assume'.

When talking about corpora and Academic Writing the ELFA project, among the others, deserves to be mentioned. The project, based at the university of Helsinki, is described in an article by Mauranen et al., (2010): it consists of two corpora, ELFA, grouping materials 'with a broad range of language backgrounds, genres and disciplines' (2010: 186), and SELF, which features texts within the context of university.

Nowadays many corpora represent EAP in an extensive way – for example, the BAWE, the PICAE and the OCAE corpora. Indeed, the latest resources for learners are corpus-informed, such as the textbook named 'Academic Writing with Corpora. A Resource Book for Data-Driven Learning' (Karpenko-Seccombe, 2020). Thus, the descriptive approach in EAP teaching moved from these first steps of corpus-informed research in academic writing that I am here – without any claim to completeness – summarizing.

The literature on the typical features of EAP that exploits corpus methodologies is at this point vast. Among many others, Thompson and Diani (2015) describe the phraseology of academic English in detail. Chapter 4 of their book, for example, focuses on the language of abstracts and the main thematizing strategies, divided into the groups of 'observation', 'interpretation', and 'argumentation'.

Another feature which authors have repeatedly investigated in EAP is the specificity of texts, which include studies on terminological density (such as Ferraresi, 2019). Hyland (Hyland & Shaw, 2016: 21) highlights the connections between research on specificity and constructionist theories which 'stresses that disciplines are largely created and maintained through the distinctive ways in which members jointly construct a view of the world through their discourses'. This view of discipline-specific variation is a feature that characterizes EAP from any other type of texts and has to be therefore addressed even in studies – like this one – which do not focus on the *specialized side* of academic

writing productions, as much as they focus on discourse structure, correctness, register, style and all those expressions coming from general English that assume a sort of institutional fixity when employed in academic contexts.

In 1999, John W. T. Smith published the article 'The deconstructed journal – a new model for academic publishing'. Although it dates back to more than twenty years ago, the article is still remarkable for its radical and innovative tone, foregrounding research on global Englishes and inclusiveness in the publishing industry, such as those by Rose, Galloway and McKinley (Galloway & Rose, 2015; McKinley and Rose, 2018). In the article, not only linguistic matters but also general issues related to the publishing workchain were addressed, in a way that foresaw accessibility policies of open access, which the academic publishing world is experiencing nowadays (on this topic, see Section2.3).

The description of EAP naturally plays an essential role in the quest for balanced and impartial standards. Going back to more recent works, in 2019, an interesting debate took place around the publication of an article on journal submission guidelines by McKinley and Rose (2018). In order to respond to the observations made by Stapleton (2019), the authors clarify the implications of assuming an ELF perspective in the publishing industry, specifically for what concerns standards. They explain that, while there is a need for errors to be reconceptualized because 'editors and reviewers are often seen as the "custodians" and gatekeepers of research publishing, [...] this should not be interpreted as an 'anything goes' ideology' (2019: 116). Nevertheless, the intense spread of English in academia and in the academic publishing industry practically means that the majority of papers are now written in English by L2 users, for an L2 public. 'Author guidelines, therefore, must strive to be inclusive and reflective of the 21^{st} century reality of English used for global research and publication purposes' (ibid.).

In short, why is the relationship between ELF and Academic Writing so important? First, because Academic Writing is one of the fields where ELF developed the most. In his article 'English as a lingua franca: Facts, Benefits and Costs', Jacques Melitz

describes the fields of human interaction where English is used as a lingua franca noticing that the cultural and academic world is 'the *single* one where the extraordinary progress of English threatens to go too far' (2018: 1750). In this section, I have analyzed the contributions of scholars who propose approaches and practices to transform English from being a *colonial language* to a landscape for possibilities of agency and renewed, inclusive interchange among cultures. Secondly, some considerations on ELF productions are necessary because academic writing, for its own nature, needs to be comprehensible and appropriate to the meaning it is meant to convey, thereby some standards assuring clarity, unambiguity and readability must be respected. In other words, it appears as if, from the ELF binocular, the main characteristics of academic writing can be better understood and contextualized. Conversely, EAP and, in particular, the corpus linguistics approaches to its description, serve as a vast field of investigation for the studies on ELF in the general framework of the studies on ESP. A complex connection is established between these topics, to the extent that one could argue that both are essential for the definition of each other.

2.3 Editing and Revision for the Publishing Industry: Definitions and Background

2.3.1 The Academic Publishing Industry

According to the 2021 Global 50, the 'World Ranking of the Publishing Industry', Europe and UK are still territories where the majority of publications are produced despite the descending trend of the last years[7]. This is confirmed by the fact that among the many book fairs around the world, one the most prestigious has certainly been, since its first edition in 1949, the Frankfurter Buchmesse (see Section 3.2).

In an online article in the British newspaper 'The Guardian', a few years ago Stephen Buranyi (2017) eloquently wrote: 'the scientific article has essentially become the only

[7] www.wischenbart.com/ranking.com

way science is systematically represented in the world. [...] If you control access to the scientific literature, it is, to all intents and purposes, like controlling science.'[8] In Sections 2.2.1 and 2.2.3 the essential role of the institutional context in academic writing was examined from a macro geopolitical perspective. It is now time to reflect on the characteristics of the specific context which pertains to the subject in question – the academic publishing industry, and on the actors that participate in the creation of academic writings as publishing products. Within the European landscape, the major turnover is made by companies based in the United Kingdom, followed by eight publishing houses based in Germany and six in France. A comparison with previous editions of the Global 50 also confirms the strategic role of the academic sector: almost 60% of world turnover is achieved by publishing groups active in this sector. The strong concentration of revenues in the first part of the ranking is also confirmed. To be more precise, the top ten groups in the ranking make more than the 50% of the turnover and the first twenty reach over 70% of the total turnover (Global 50, 2021). These facts point to the need for some reflections on the world of academic publishing; indeed, 'the approach of analyzing the industry through a business perspective is important so that a clearer understanding of the industry landscape can be drawn' (McGuigan and Russel, 2008: 10).

The Guardian's article to which I referred earlier follows in the footsteps of previous research on this subject, including the article by McGuigan and Russel (2008), named 'The Business of Academic Publishing: A Strategic Analysis of the Academic Journal Publishing Industry and its Impact on the Future of Scholarly Publishing'. The starting point of this paper is the economic crisis that had, in the last decade, affected libraries and that is only partially being solved by the introduction of open accessibility policies, as will be explained later. The academic publishing industry is responsible for the 'creation, review, packaging and distribution of knowledge in multiple formats for use mainly by academic and scientific consumers' (2008: 5). In other words, academic writing is produced by the scientific community – meaning that its contents are devel-

[8] https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

oped mainly thanks to public or anyway institutional funds – and is sold back to the academic institutions in the form of publishers' products. As Buranyi (2017) points out:

> The way to make money from a scientific article looks very similar [to that of books, in general], except that scientific publishers manage to duck most of the actual costs. Scientists create work under their own direction – funded largely by governments – and give it to publishers for free; the publisher pays scientific editors who judge whether the work is worth publishing and check its grammar, but the bulk of the editorial burden – checking the scientific validity and evaluating the experiments, a process known as peer review – is done by working scientists on a volunteer basis. The publishers then sell the product back to government-funded institutional and university libraries, to be read by scientists – who, in a collective sense, created the product in the first place.

The fact that most of the world's biggest publishing companies are based either in the U.K. or in U.S. (Global 50, 2021) – six out of the first ten – is a fact to consider also in the light of the reflections about ELF and linguistic imperialism contained in the previous section, especially because an idea of *prestige* is linked to these publishing companies (Fyfe et al., 2017). In this respect, book fairs also serve as the stage for these publishing companies to hegemonize the publishing landscape. In some meetings with the trainees at BUP[9], the fairs were presented as the possibility for the big publishing companies to consolidate their power, even if they also give space to smaller publishers to become known and, hopefully, to create useful connections for their businesses. Buranyi (2017) quotes some considerations made by Randy Schekman, the Novel prize winner American biologist, on the constantly growing importance that is given to *where* an academic work is published: 'at the start of my career, nobody took much notice of where you published, and then everything changed [...] Suddenly, *where* you published became immensely important.'

[9] BUP provides trainees with some theoretical lectures, as well about the publishing scenario.

In this regard, it may be interesting to consider another classification, this time particularly focused on academic publishing companies: the SENSE ranking[10]. This resource is co-produced by the Research School for Socio-Economic and Natural Sciences of the Environment and the Royal Netherlands Academy of Arts and Sciences. The ranking is estimated using the evaluations of scholars and does not take into consideration the number of publications, or the turnover, as much as it accounts for the publisher's 'prestige'. A number of similar documents can be found online, which presents rankings by different universities around the world. Confronting several different reports (for example, the ones produced by the University of Utrecht[11] and the Australian Political Studies Association[12]) the same few names recurrently appear in the first positions: Cambridge University Press, Routledge (Francis & Taylor Group), Springer, Oxford University Press, Elsevier Science Ltd, Chicago University Press, and some others. On this matter, McGuigan and Russel (2008: 8) write:

> because of the oligopolistic structure of the industry, rivalry between publishers is low [...]. Rivalry is further attenuated because there is little direct competition between the individual journals produced by each publisher. This is due to the specialized character of academic journals which are targeted to specific academic disciplines thus each journal has its own distinct target audience. This is a form of product differentiation. Moreover, the publishers that own prestigious journals are able to take advantage of another form of differentiation since faculty and libraries will always seek out the most influential journal within any given discipline.

In this sense, the publishing houses are represented as 'gatekeepers' of scientific knowledge (Hyland & Shaw, 2016), on the one hand, and of the academic publishing business on the other, as explained by McGuigan and Russel: 'Entry into the traditional publishing industry by potential competitors is also difficult due to cost advantages due to the economies of scale, the learning curve effect, and established market share' (2008: 8). So, as can be deduced, the power of publishing companies exponentially grows

[10] https://sense.nl/

[11] http://ceres.fss.uu.nl/component/option,com_wrapper/Itemid,7/

[12] https://www.eduhk.hk

with their prestige and – of course – turnover. In this respect, a project that deserves to be mentioned is DORA[13], the Declaration on Research Assessment, which aims at providing impartial assessment of research through blogs, community interviews, presentations, conferences, as well as structured research with universities and institutions around the world.

Not much space is left in this landscape, however, for smaller publishing companies, which strive to gain financial security and acknowledgement, while the research they publish is often in need of more funds. Moreover, in Buranyi's analysis, very little consideration is given to the internal work organization of the publishing houses. Revision, as Mossop (2001) has argued, is very often a process to which not much attention nor a large budget are dedicated. In fact, in most cases, revisions are made by external experts, whose compensations are estimated not only on the basis of the amount and difficulty of the work, but also – and mostly – based on the limited budget available for such practices (ibid. 2001). The 2021 Istat report shows that about 15% of the Italian publishing houses rely on external professionals for revision services, as illustrated by Rambelli (2021).

A further detail to be added to this picture, which reflects the circular nature of the academic publishing business, is that 'faculty scholars [often] provide editorial services' (McGuigan & Russel, 2008: 2) both professionally and pro bono. Further evidence of the peripheral role assigned to revisors in publishing companies is given by Joshi, is his article 'Native and non-native speakers of English as copy-editors of research papers' (2011). He differentiates between two types of costumers for revision services: authors and publishers. Regarding the publishers as costumers, revision is presented as something outside the internal processes of editing, whilst, for what concerns authors, they are very often researchers whose work need language revision in order to be accepted for publication (see Subsections2.3.3 and 2.3.4).

In a chapter from the volume Cyberspace Divide (Loader, 1998), Trevor Haywood highlights the contradictions hidden under the myth of a new possible equality, which

[13] https://sfdora.org

were generated by the establishment of the so-called *information society* and the global networks. After more than twenty years from the publication of this book, many of these contradictions still characterize our society, which from a sociological perspective, has now become the *knowledge society* (Guerra, 2010). But it must be said that technology has repeatedly proved to be an ally of the weak- at least, for what concerns the publishing sector, where open access resources as well as social networks have provided possibilities of participation to research as well as working environments that were hardly opened to cultures and societies, characterized by a weak publishing scenario. Of course, this does not mean that technology in itself prevents inequalities; in fact, access to information and knowledge is still massively influenced by economic and political power games (Hyland, 2016). In 2008, McGuigan and Russel wrote: 'The potential exists to radically transform the academic publishing industry through the adoption of new electronic publishing technologies using the internet as a medium for transmission' (ibid: 10). Indeed, these seem to be the intentions behind the latest policies put in place by the European Union to enlarge accessibility and the amount of open access (OA) resources, as testified by the European Accessibility Act published in June (Audrain et al., 2021). After all, one of the major side-effects of the Covid-19 pandemics has been the acceleration of technological development, particularly in the fields of education and academic communication. Then again, in 2015 Pienfield had already noticed that 'higher education institutions are now at the centre of making OA work in practice' (ibid: 620).

OA is rapidly developing:

> The use of metrics as an integral part of the scholarly communication infrastructure (in this case quality evaluation) is a prominent example of an increasingly important wider issue: the prospect of a "network-enabled" OA literature. [...] Whilst still in its infancy, this vision is now becoming a more realistic prospect. Features of such an infrastructure are likely to include interoperable text and data. [At this point], the main challenge associated with scholarly communication is no longer whether OA should be at the centre of the system but how. (Pinfield, 2015: 620)

On the other hand, as can be read in the Wellcome Trust report named 'Economic analysis of scientific research publishing':

> the existence of the means to create significant change does not mean that change will occur. The fact that electronic media exist has implications for the market. It is up to the players in the market to decide how they will use the means at their disposal. The dominance of the commercial publishers will be challenged only if the other players use the opportunities available to them. (2003: v)

To sum up, this complicated institutional context throws light on the complexity of the discourse on revision and point to the need for more structured linguistics services in the academic publishing industry. As studies on academic materials produced by non-native English speakers, such as the one by Vasconcelos et al. (2007), have shown, publication rates are higher for authors who are proficient users of English. Therefore, editing and revision should be provided by publishing houses to guarantee that research has equal importance regardless of the native language of the authors. Revision can serve as a means of empowerment and legitimation for authors who are not proficient English speakers. This is also the reason why revision practices need to be clearly planned, with the final purpose, of course, to stay loyal to the aims and vision of the authors.

In this sense, I would conclude by drawing a parallel between hegemonic economic systems of publishing and hegemonic use of English as the language of academia: OA envisages a change towards more equal practices of knowledge construction, a similar change of direction should happen – and is, indeed, happening – also for what concerns the use, evaluation and revision of texts.

2.3.2 Editing or Revision?

From the point of view of revision, Mossop has dedicated a section of his volume, which concentrates on the editing and revision of translations, to the characteristics of what he defines as 'non-native English'. In a sentence that sounds almost like a warning for

revisors, he notices: 'the biggest problems seen in non-native English are not micro-errors [...], but failures in English composition: since the writers were not educated in English, they may never have learned how to organize sentences in the English manner' (2001: 37). Although some of the terms used by the author, such as 'non-native English' undoubtedly collide with the points of view of the ELF scholars that I have mainly considered in this study (see Section 2.2), what Mossop is saying is not too far from the considerations by Canagarajah (2002) on Western-centred academic publications.

Approaching the issue of revision and editing of academic writing productions by non-native speakers for the publishing industry, the first problem with which one is confronted is that of defining the subject under analysis. Translation and writing studies are the research sectors where the problem of defining and formalizing editing and revision practices has been dealt with in more detail. This leads to a first important consideration: although it might be hard to believe, it appears as if very little has been said about revision and editing from the point of view of publishing issues. Later on in this section, I will present the way in which revision and editing processes at BUP are assigned to different figures inside (and outside) the publishing company in the next chapter. However, the whole bibliography I have analyzed, despite recognizing the difference between practices of revision and editing in the way intended in the present study, often use these two terms, sometimes together with 'copy-editing', with very blurred definition boundaries. For this reason, before deepening the analysis of such studies, I would like to refer to some websites that provide a clear distinction between the practice of revision and that of editing.

Extracting definitions about editing and writing from the many web guides available on the internet about writing and revising can be a useful way to have some updated documentation about the actual understanding of the two terms and their use in the context of the writing studies. For what concerns translation studies, as it will be explained below, these words, in truth, have not always been used with the same meaning. According to my web survey, it seems that 'editing' is used mainly to refer to changes related to grammar and spelling mistakes, or changes concerning style (including, for

example, referencing)[14]. For example, on the Scribbr website (for information about Scribbr see Section2.4) a clear and concise definition can be found:

> Revising, proofreading, and editing are different stages of the writing process.
>
> - Revising is making structural and logical changes to your text reformulating arguments and reordering information.
> - Editing refers to making more local changes to things like sentence structure and phrasing to make sure your meaning is conveyed clearly and concisely.
> - Proofreading involves looking at the text closely, line by line, to spot any typos and issues with consistency and correct them.

Thus, revision includes – to a wider or smaller extent – the practice of re-writing: modifications are made to the text structure, terminology and phraseology are checked, characterizing this intervention with little or greater subjectivism. Consequently, it can be deduced that a) while editing is simple to be decolonized and unbiased, revision can be seen as a more subjective practice as it involves deeper meaningful levels of language and communication; b) this is not a reason, however, to give up on revision: revision is necessary to maintain certain standards of quality in writing; c) it is therefore necessary to find a way to understand from an objective and depersonalized point of view which are the changes that actually make the text better, because it is true that the problems of over-revision and subjective revision exist (Joshi, 2011).

[14] http://info.francis.edu/writing-center-resources-for-students/
https://slc.berkeley.edu/writing-worksheets-and-other-writing-resources/editing-vs-revision
https://www.middlebury.edu/system/files/media/Revision,%20Editing%20and%20Proof-reading.pdf
https://www.smekenseducation.com/understanding-the-difference-be/
https://thinkwritten.com/difference-between-revising-and-editing/
https://www.scribbr.com/frequently-asked-questions/difference-between-revising-proofreading-and-editing/
https://www.scribbr.com/frequently-asked-questions/difference-between-revising-proofreading-and-editing/

Among other practices, some of which are relatable to the revisor's techniques and workflows, a good starting point to guide the human evaluation of revision difficulty could be that of automatizing the process of quality assessment – as will be done in the experimental part of this research.

For what concerns BUP on this terminological matter, in the traineeship presentation named 'Il lavoro del redattore' (2021) everything that is inherent to formatting, whether 'mechanical' or 'logical', i.e. applying the editorial standards of the publishing house, is called editing. Linguistic revision is kept separate, for various reasons including the optimization of revisors' services. This distinction is evident in the characterization of the roles of the chief-editor versus side-editors, among whom linguistic revisors are included, who can be external to the company.

Analyzing the definition of revision in writing studies, the term mostly refers to self-revision. In the book 'The Work of Revision' by Sullivan (2013) revision is referred to as self-editing. Even if, self-editing and self-revision practices are not the main topic of this dissertation, some considerations in her work can be related to the practice of revision in general. She assigns to revision an idea of transformation. 'Modernist writers [...] used revision, an action that implies retrospection, not for stylistic tidying-up but to make it new through large-scale transformations of length, structure, perspective and genre'. Some lines later she writes 'the problem of textual variation is as old as writing and copying. [...] Despite this, revision has not historically been of great concern to textual critics' (ibid.: 16), highlighting the serious lack of theories concerning this practice. Instead, the use of the term revision with the meaning here intended is consistently attested in translation studies (Mossop, 2001; Koponen et al. 2020). According to Wang (2021), in writing studies the term 'revision' – as mentioned above – mainly refers to self-revision, whereas,'in translation studies *editing* and *revision* are less clearly distinguished and are sometimes treated as synonyms with the former being the preferred term in American English and the latter in British English (Scocchera, 2015; Delisle, 1988)' (ibid.: 7).

In this scenario, the article 'Revision for Quality' by Chakhachiro (2005) examines the

parameters for the revision process from a contrastive perspective, analyzing translations from English into Arabic. But there are also authors who have a different approach to revision focused more on the target text than on the source text. From these studies many useful ideas can be adapted to analyze the practice of revising texts written by non-native speakers. For instance, it is clear that 'interference' plays a crucial role in non-native speakers' productions. If we compare contact linguistics studies (Jarvis & Pavlenko, 2008) to those analyzing the so-called *translationese* (Avner et al., 2016), we might find that the two have many points in common (see Subsection 2.2.2).

A volume that has been particularly influent in the teaching of revision within translation studies is 'Editing and Revising for Translators' by Brian Mossop, first published in 2001. In this volume, the use of terminology is very accurate; the author examines different types of editing and revision practices. Both editing and revision are fundamental tasks in the creation of a piece of writing, that might require improvements in typography, use of language (i.e., collocations, vocabulary, register), anaphoric structure, genre and domain specific styles and terminology, coherence in meaning, consistency and cohesion (Mossop, 2001). As outlined by this author, the tasks of the editor are very similar to the ones that BUP trainers associate to the role of chief-editor. For what concerns the 'types of amending work', Mossop differentiates between copyediting, 'correcting a manuscript to bring it into conformance with pre-set rules', stylistic editing, 'work done to improve rather than correct text', structural editing, that is, 'reorganizing the text to achieve a better order of presentation' (ibid.: 30,31), and content editing, which includes suggestions on the topics to be covered. This distinction will be useful in Chapter 3 and Chapter 4 to understand to what extent RTT can work as a method for evaluating text quality for editing and revision purposes.

Although he does not make a clear formalized distinction between editing and revision, Mossop dedicates the second part of his volume to revision, characterizing it, first of all, as 'a reading task'. A section of his work is dedicated to the characteristics of what the author defines as 'non-native English'. In a sentence that sounds almost like a warning for revisors he notices: 'the biggest problems seen in non-native English are not micro-errors [...], but failures in English composition: since the writers were not

educated in English, they may never have learned how to organize sentences in the English manner' (2001: 37). In this sense, he suggests that writing in English as a lingua franca generally requires revision, according to the definition of revision that I am trying to outline.

While the various tasks and roles of editors and revisors are analyzed with unparalleled precision, the two terms are only vaguely differentiated. But, then again, the very first sentence of the book 'Editing and Revising Text' states: 'editing means different things to different people' (Billingham, 2002: 1). When Billingham describes editing, he does not seem to make any difference between editing and revision. Indeed, he simply writes that these processes are aimed 'to make the text as good as [...] possible' (ibid: 6). He makes, on the other hand, a distinction between editing and rewriting (revision as intended in the present work could be positioned in the middle between these two extremes).

To sum up, authors have variously examined revision practices as well as the revisor's status as a professional (see Subsection 2.3.3) using revision and editing as concepts which sometimes are the same, sometimes overlap and sometimes are completely different. However – and despite the relatively low attention paid to this subject – the need for linguistic revision when it comes to tailoring writing products is unquestionable, for what concerns both native and non-native speakers' productions (Joshi, 2011). Therefore, it is essential to distinguish good revision practices from over-revision and poor-quality revisions (see Subsection 2.3.5), keeping in mind that revision, like editing, is actually a much-needed practice. Indeed, functional-systemic approaches have brought attention to the essential role of revision as academic writing skills in English are seldom taught in educational contexts in countries where English is not the official language (Gosden, 1995). As a matter of fact, there is a correlation between written English proficiency and research productivity, as explained by Vasconcelos et al. (2007).

The need for revision has been identified also in the field of computational linguistics and text processing. In an article by Ito et al. (2019) revision is included together with

editing and proofreading, in the main stages of the writing process. Here a difference is outlined between surface-level issues, which, according to the authors, are related to editing, and 'sentence-level' revision, a practice that include major changes in the way the text is structured.

In conclusion, after a general review of the literature on this matter, it seems that the majority of studies on the topic of linguistic revision do not even try to solve this dilemma; they examine essential themes of editing and revision without a clear field of investigation.

For the purposes of this work, the difference between editing and revision, as I have tried to define in the present subsection, should be considered in terms of two practices that can be placed at the extremes of a continuum. On the one hand, editing tout-court refers to all those changes that guarantee grammatical cohesion, textual coherence and respect for conventions – either general of a language or specific of a publishing house; on the other hand, revision includes all those changes that refine and improve the style and linguistic rendering of a text, which can be more or less necessary according to the communicative contexts (Bell, 2012; Sullivan, 2013, Wang, 2021). While this distinction is fundamental for an appropriate discussion on this matter, the present work covers both topics. Defining the difference between the two practices is a way of clarifying the boundaries within which the two processes take place – with all the implications that this brings for the revisor's status and workflow (see the following subsection). For example, the considerations about linguistic ideologies (Subsection 2.2.1) imply that the choice of editing standards is biased by social constraints whereas different approaches to revision can unveil precise political conceptualizations of the English to be revised. However, the method that will be presented and applied in Chapter 3 addresses the issue of evaluating both revision and editing difficulty.

2.3.3 The Revisor: Profile and Workflow

As we have already outlined in the previous subsection, revisors can be both internal and external to publishing companies, entertaining with them different types of working relations (Rambelli, 2021). This trend is confirmed at BUP, where a team of in-house and free-lance translators and revisors work to guarantee the best linguistic quality for their products, given the budget available for every project.

For what concerns the revisor's profile and professionality, an interesting article named "'Convenience editors" as legitimate participants in the practice of scientific editing: An interview study' (Willey & Tanimoto, 2013) provides an insight on the working situation of those English teachers that find themselves in the role of unexperienced editors and revisors, offering more convenient rates compared to those of professional revisors and translators. Through the examination of these cases, Willey and Tanimoto's study sheds light on two important aspects related to the revisor's professionality. On the one hand, the necessity for a specialized training becomes evident; on the other hand, it addresses the issue of conceiving *nativeness* as a parameter for evaluating a revisor's proficiency at work. The notion that one should be able to edit scientific manuscripts simply because one is a native English speaker is itself problematic.

The authors examine the possibility of relating Lave & Wenger's theory about *situated learning* (1991), which is based on three fundamental notions – legitimate peripheral participation, communities of practice, apprenticeship – to convenience editors: they find it problematic to apply such a theory to the specific situations they studied, principally because of the editors' 'ambivalent attitudes' (Willey & Tanimoto, 2013: 31) towards the task, noticing, at the same time, that some skills are acquired by these 'budding revisors', as the situated learning theory would predict. According to the authors, such results imply that scientific editing can be a difficult task 'for even highly experienced English-teaching editors' (ibid.).

Considerations of this kind are the main focus of the afore-mentioned 2011 article by Yateendra Joshi (see Subsection 2.3.2), where the author investigates the relationship

between proficiency in a language, *nativeness* and the ability to revise – or, adopting the author's terminology, 'copy-editing', arguing that it 'is not a straightforward issue of quality versus costs' (ibid.: 38). First, the author reflects on the many aspects of copy-editing that are unrelated to language, such as style, referencing and terminology. Then, according to Joshi, what has to be taken into account are the receivers, that is, the readers who are going to consult the academic products. Not all of them are native speakers – actually, the majority of them are not, hence they will not – in theory, or however they should not – pay much attention to features that might be specific of a certain 'non-standard' variety of English, also because they are aware that the author is not a native-speaker (ibid).

Although the subject under analysis is slightly incongruous with the one of this dissertation, the study by Stevenson et al. (2006) seems to confirm the trend according to which proficiency in editing and revision tasks is not directly linked to *nativeness*. The authors conduct observations on self-revision practices by students combining the use of think-aloud and keystroke-logging techniques, showing that 'although revisions made to linguistic elements were more frequent in FL, there was little evidence that [...] revision processes were inhibited in FL writing' (2003:201). While this study favors the hypothesis that expertise in revision can be achieved by both native and non-native speakers, in contrast with Joshi's assertations on the attitudes toward variation by non-native speakers, it also confirms a trend that was noted by others, such as Ehrensberger-Dow (2006), for non-native speakers revisors – and readers, in general – to be less accommodating towards linguistic errors compared to native-speakers.

Moreover, what Willey and Tanimoto's study (2013) also suggest is that the working environment, including the communication with the manuscript's author, are of primary importance for the development of editing skills: from this point of view, the concept of situated learning can thus provide new perspectives for the enhancement of editing working contexts. Indeed, the study spotlights the main issues with which non-professional editors need to confront. The following two problems are the ones that could be generalized beyond the borders of Willey and Tanimoto's research: difficulty of dealing with technical terminology and disciplinary styles, and 'the authors

involvement in the editing process' (2013: 31). With respect to the latter issue, it has to be noticed that the workflow of the revisor is naturally influenced by the working circumstance and by the type of document to be revised. For example, when working on team documents interaction among authors and revisors are always continuative and profitable. Conversely, working alone and on somebody else's piece of writing can imply different degree of involvement by the author (Billingham, 2002).

Revision typically follows steps which go from a more general reading to scanning and then back to viewing the document as a whole (ibid.). In this regard, it is important to notice that the first reading is the only time the document will be read 'in the way the intended readers will read it' (ibid.: 15).

Xiaoli Wang (2021) in her dissertation on revision techniques focuses on the workflow and techniques put in place by revisors working at BUP, highlighting the difficulty arisen from the lack of any type of guide from the publishing house. Indeed, one of the results of her research has been the development of documents – into which I personally took part – that lead the revisor through the process of revising, also summarizing the techniques that should be adopted and the resources that the company makes available for them.

Besides the style guides with specific information on editing standards and formatting, a questionnaire for the client and a document including the suggested workflow and key points to heed have been ideated. The questionnaire for the client covers some important issues which should be clarified before the revision begins: not only the degree of revision, but also time and the author's involvement are elements that appear in the document.

For what concerns the workflow guide, it is primally meant for trainees and team revisions (such as Wang's case-study); it divides the revision process into the following main stages: orientation, reading, correction, presentational adjustment, and language check. Such documents have an essential value that goes beyond practical implications: they represent an attempt of professionalizing the work of revisors in the publishing industry.

A fundamental piece still missing to this picture is the pre-evaluation of manuscripts, which – if automatized – could save precious time in the workflow of the revisor, lower costs of revision, thus enabling revision for project with budget constraints (see the case of New Medit in Chapter 4 as an example), and could instruct more purposeful revision practices.

2.3.4 Text Quality Evaluation for Revision Purposes

In the previous section it has been seen that the English used in academic writing has all features and functions to be considered a particular – but, at the same time, very prototypical – use of English as a lingua franca. The consequent problem that arises refers to the possibility of evaluating productions in such a language. This difficulty is clearly expressed by Jenkins, who writes:

> linguistic resources are deployed so dynamically in ELF settings that nativization as such does not have time to take hold: there is simply not the longer-term stability required for sedimentation to take place, with the effect that language forms remain more or less continually in a state of suspension. (2013: 36)

At this point, it can be useful to consider once more the important consequences of what has been defined by Seidlhofer (2011) as 'Standard English Ideology'. The author sheds light on an important issue related to the problematization of the notion of StE.

> The calls for resistance to "central norms of language" (Pennycook), or "native-speakerism" (Halliday), and for "relocating English" (Saraceni, 2009) and "reclaiming the local" (Canagarajah, 2005), are timely, much-needed challenges to conventional ways of thinking and acting, and they make for inspiring reading for applied linguistics. But they do not address the conceptualization of "English" itself, and so the central question remains what form this resistance should take, what is actually involved in relocating English or reclaiming the local. (Jenkins, 2013: 36)

While the particular topic of revision and academic writing are not expressively touched by the author, some ways in which English can be reconceptualized taking into consid-

eration processes of appropriation and adaptation are shown. The central importance of a 'community of practice', based on 'mutual engagement', 'jointly negotiated enterprise' and 'use of members' shared repertoire', in order to rethink the concept of competence is highlighted. It is important to keep in mind these considerations when approaching the theme of ELF productions quality evaluation – of course, with all the necessary precautions that relates to the specific case of academic writing and revision.

Indeed, it should be noted that

> the past two decades have accordingly witnessed a burgeoning of articles, dissertations, conferences, corpora, and a dedicated journal devoted to the topic; yet in many circles the phenomenon has still remained unnoticed or unacknowledged, acquired many misunderstandings, is raising mixed opinions or encountering strong resistance. (Paradowski, 2013: 312)

This is probably true also of studies on revision practices. In this regard, it is relevant to cite two of the practical proposals that Paradowski makes and that could also be taken into consideration to improve revision techniques in the light of ELF studies: 'regular forms might be preferred over irregular ones; structurally more explicit constructions should be chosen over less transparent ones' (2013). This is a mere example of the many possible reflections that can arise adopting this perspective in studies on revision, and in revision practices.

In this light, I would like to consider the contribution by Tribble (2006), eloquently named 'Written In, Written Out: Who Sets the Standards for Academic Writing'. Tribble reinterprets the theme of linguistic imperialism as it was conceived in the 1980s and 1990s (see Section 2.2) endowing it with new characteristics that reflect the academic world at the beginning of the open-access revolution. The author, who focuses especially on the criteria for identifying standards for the teaching and learning of academic writing skills, advocates for the adoption of models free from the native-speaker constraints, echoing the arguments about the use of English as a lingua franca for scientific communication, which have been discussed in detail in Section 2.2. He suggests that the principle of *nativeness* should be replaced with the concept of expertise. But, how

can linguistic expertise be assessed?

According to the author in question, language quality is closely linked to terminological precision. While not denying the intrinsic relationship between the content of a paper and its linguistic rendering[15], this reflection has important consequences both as regards the evaluation of scientific texts, where clarity and accuracy assume greater importance than concepts such as 'readability' (as will be later presented), and as regards the profile of the revisor, who more and more needs to be a competent figure prepared to use terminological resources, glossaries and reference corpora. Besides, the author shows how sectorial linguistic expertise also concerns general aspects of language, such as syntax and phraseology, which are used in a certain way in a certain genre to confer a certain meaning. To this concept, the author gives the name of 'discourse expertise', which appears as an objective criterion unhinged by political or personal implications. Apart from two examples concerning theme-rheme structure and noun phrases, however, the author does not reveal much about possible parameters for measuring the degree of authors discourse expertise.

In the Routledge Handbook of English for Academic Purposes (Cushing Weigle and Malone, 2016), a chapter is dedicated to the assessment of EAP. Its focus is predictably on testing language proficiency in learners academic writing, meaning that the evaluation described by the authors does not share the same purpose with the one that is being discussed here. For this reason, I will not concentrate much on their considerations, although important testing procedures such as IELTS and TOEFL should at least be mentioned in this context for comprehensiveness purposes.

Moreover, Cushing Weigle and Malone refer to White's (1985) holistic approach to testing, which provides the excuse to define the limits of comparing the evaluation of students' writing productions with that of manuscripts meant for publication. Indeed,

[15] On this issue, from a computational linguistics perspective, the article by Jafaritazehjani et al. (2020) have experimented on the possibility of automatically modifying style preserving content and fluency and found that 'style cannot be usefully separated from content' in the systems they developed. This highlights the constitutional relationship between language and knowledge, thus showing the limits within which studies on automatic revision need to move.

while in the first case a holistic approach to assessment could be taken into account, the lack of well-defined criteria in the evaluation of academic writing is exactly the problem the present dissertation is trying to address in the first place. However, this chapter provides a good overview on the existing methods of assessment – including the automatic ones, which will be treated in more detail in Section 2.4.2.

For reasons that can be deduced from the above discussion, in the present study a functional-system perspective has been adopted; the research on evaluation parameters is mainly guided by the consideration of the intended readers and the intended aims of the document under revision. As such, its central objectives are a) overcoming subjectivity (i.e., excluding the possibility of 'holistic', thus undefined, assessment), and b) overcoming prescriptivism in evaluation. In this respect, it is useful to refer to Mossop's work (2001), since the author draws a clear connection between these two problems overshadowing linguistic quality assessment.

On the subjectivity of evaluation and, therefore, revision, Mossop traces an important difference between problems which can be seen as objectives, because they relate to 'rules which are inherent in the spoken language [...], such as the position in a sentence of an adverb like frequently' (2001: 53), and problems which relate to 'something variously called "correct usage", "good grammar", "correct English" or "proper English". This is one of the central problems that this study wishes to address: as seen earlier in Section 2.2, it is a problem which carries and mirrors a number of controversial sociopolitical implications. In order to overcome this problem, the author problematizes prescriptivist approaches relating them to the concept of communicative success.

Using a similar terminology, Gosden evaluates the quality of research papers in terms of 'success', as 'judged from the processes of peer review, negotiation, revision, and eventual acceptance for publication' (1995:37). The author analyzes the writing practices of NNS researchers in the light of 'the regulating mechanisms of [the academic] discourse community' (ibid: 38). Following the path outlined by Couture (1985), Gosden applies Halliday's model of tripartite meta-functional organization of language. He suggests that 'success in scientific writing can be evaluated by analyzing revisions in

relation to a network of Ideational, Interpersonal and Textual functions as determined by the rhetorical purposes and structure of the scientific RA' (ibid.: 41). In this view, revision assumes the role of reframing the text according to the standards of the receptive community, at the same time working as faithfully as possible to the author's conceptualizations, his way of presentation and the intended purposes of his work.

In his volume, Mossop (2001) specifically addresses the problem of quality assessment in revision. He describes this process with the following words: 'identifying problems in one or more randomly selected passages of a text in order to determine the degree to which it meets professional standards and the standards of the [client] organization', thus addressing the problem in a way that he defines 'business-oriented' (ibid.: 128, 129). When he refers to the process of evaluation, he introduces the concept of 'revision parameters' (ibid: 134), grouping the problems that publishers may encounter when confronting with a text into four main categories: problems caused by transfer procedures, problems related to content, problems related to language and style, problems related to visual aspects. Apart from this latter, which is in most cases an issue with which graphics experts or the chief-editor deal, the other three groups of problems all pertain to the revisor; therefore, they influence the estimation of time needed for the revision as well as the difficulty of revision.

These criteria, however, need to be confronted by the intended degree of revision (see Subsection 2.3.3) – which reflects not only the number of changes and type of changes (grammar vs. structural changes) but also how many times the revisor will go through the text (ibid.). In other words, the degree of revision is estimated according to the needs of the client and the quality of the writing.

Mossop makes use of one-word descriptors to refer to the writing quality summarizing their meaning with the table reported below (2001: 155) (Table 1)[16].

Collada and Alvarez (2018) quote Brian Mossop on the possible errors that revisors may encounter in translations: transfer, content, language, or presentation errors. Then,

[16] The column 'accuracy' has been removed, because it refers to the contrastive analysis between translation and source-text.

Table 1: A summary of writing quality standards adapted from Mossop (2001)

Commissioner's Purpose	**Writing Quality**	**One-word Descriptor**
For speedy, basic understanding	Minimally readable and clear	Intelligible
For information	Fairly readable and clear	Informative
For publication	Very readable and clear	Publishable
For image	Finely crafted wording and very clear	Polished

they link the changes deriving from the identification and resolution of this type of errors to different degree of revision, marking the difference among full revision, general revision and partial revision – the criteria for determining which kind of revision is most suitable depend on the client's requirements as well as the initial quality of the text. The guides produced at BUP, which have been mentioned earlier in this Chapter, resemble this kind of considerations.

In recent years, the concept of readability has colonized the debate on both human and automatic editing, becoming one of the main criteria to establish the quality of a text. It seems, however, that in most cases 'readability' is a rather vague and undefined idea of something that is easy to read and no clear-cut criteria are defined for the evaluation of such a feature. In a more precise way, this problem is addressed by Mossop when he describes 'smoothing' techniques, among which he lists: 'parallel ideas [must] expressed through parallel forms, the antecedents of pronouns [must be] immediately clear, connector words [must not be] misleading' (2001: 70-72). He then differentiates between clarity and readability, in the sense that clarity, unlike readability, 'is a feature of the meaning of the text, rather than its wording' (ibid.: 72), thus delineating the limits within which language revisors can apply changes to the text. But, because clarity and readability are – at least, up to a certain extent – related to each other, the communication between authors and revisors facilitates the revision process, while assuring accuracy in meaning and high-quality standards in language (Billingham, 2002; Bermann and Porter, 2014).

The classifications here outlined, however, still seem to be, in a way, text-oriented,

because they refer to the evaluation of the piece of writing to be revised itself. While this is surely necessary, a passage remains unspoken – that is, the revisors point of view in terms of their workflow, time and engagement. In other words, are we sure that text quality and revision difficulty completely overlap? With the aim of overcoming this hypothetical problem, and because the evaluation was made retrospectively, the criteria according to which the dataset of this study has been evaluated, while largely taking inspiration from Mossop's considerations, could be described as *revisor-oriented*, as will be seen in Chapter 3.

Earlier in this section, I have referred to Mossop as one of the few authors who treat revision specifically – as a separated practice from translation which needs its own teaching and reflections. Another volume which provides accurate and specific insight in the practice of revision is 'Enhancing Translation Quality: Ways, Means, Methods' (Forstner et al., 2008). In particular, a chapter written by Gyde Hansen is dedicated to the identification and classification of errors according to the CBS (Copenhagen Business School) longitudinal model. Some of the problems are, of course, related to translation errors. More interestingly for the purposes of this investigation, other errors concern the target text as it is – its language, structures and the overall acceptability in the target language/culture. The classification proposed by Hansen features a first type of errors that are performed 'in relation to the affected units and levels of linguistic and stylistic description'. While the first sub-category relates to misinterpretation, the second is labeled 'text-linguistic errors' and includes 'violation of the semantic, logical or stylistic coherence', which is inherent to the type of revision I am addressing here. Moreover, as I was recalling in the previous subsection, errors caused by the interference of the translator's – or, in our case, the author's – mother-tongue are a feature that any text written in a language other than the writer's native one share.

The author also reflects on the difficulty of grading positive quality, which is an issue that partially persists in the method proposed by the present study (see Chapter 3).

2.3.5 Revision Practices

As Hartse and Kubota (2014) explain, the possibility of bringing inclusive practices of pluralizing English into revision is still little explored. Approaches of this kind are gaining more attention in the scenario of language teaching and learning; here I want to confirm the significance of such perspectives also for what concerns revision practices, because it is important to differentiate between changes that are needed and subjective changes (Murphy, 2012). They note that native-speaker intuitions massively orientate revision practices, making them 'idiosyncratic'. 'This idiosyncrasy further poses skepticism about the applicability of both error-oriented approaches and pluralistic theories about L2 writing to copyediting in high-stakes academic publishing' (ibid.: 71). Hence, the authors conclude that lexis and grammar are linguistic levels where pluralizing practices are still hardly experimented.

In this framework, they distinguish between error-based and variation-based approaches to revision, the latter including world Englishes approaches, which advocate the acceptance of multiple varieties of English, translingual approaches, 'which view texts as hybrid constructions influenced by rhetorical factors', and 'the Written English as a lingua franca approaches' (ibid.: 73), which relates to features of the use of English specific of the ELF discourse. The authors also note that there are very few academic journals which explicitly accept articles written in English varieties that deviate from standard English. In fact, the authors write: 'many journals give explicit guidelines to writers regarding the issue of deviation from what are assumed to be agreed-upon standards of good writing' (ibid: 76). As an example, they quote a clarification on language support and acceptability from the Springer website, where it can be read:

> Manuscripts that are accepted for publication will be checked by our copyeditors for spelling and formal style. This may not be sufficient if English is not your native language and substantial editing would be required. In that case, you may want to have your manuscript edited by a native speaker prior to submission. A clear and concise language will help editors and reviewers concentrate on the scientific content of your paper and thus smooth the peer review process. (ibid.: 76)

Commenting on the examples brought up in their paper, Hartse and Kubota notice, again, that changes are applied based on native-speaker intuitions, which are sometimes contradictory. They conclude that the 'NES copyeditor is the final authority on appropriate language use'; for this reason, 'scholars who are committed to pluralizing English can and should work toward more progressive policies in the publishing endeavors with which they are involved' (ibid.: 79,81).

Following a similar trend, Englander (2006) specifically addresses academic journal editors when referring to the problem of quality assessment and linguistic gate-keeping, while highlighting the essentiality of revision. She writes: 'manuscripts submitted by nonnative-English-speaking scientists are sometimes criticized for their language usage and they require revision' (ibid.: 129).

Consequently, the problem of over-revision needs to be considered (Billingham, 2002), because this issue is strictly connected to subjectivity in evaluation. As Hartse and Kubota (2014) suggest, revisors should always ask themselves the question 'why do I want this change to be made?'. 'The answer – they explain – may have to do with intelligibility, acceptability, grammaticality, preferred variations, or any number of factors' (ibid.: 81). Nevertheless, posing such a question before applying any change would probably result in 'a greater variety in expression, more equity for NNES writers, more tolerance for difference, and ultimately, published texts of a higher linguistic and ethical standard' (ibid.: 81).

In this light, some studies have reflected upon the cognitive aspects related to the process of revision. Among the others, Willey and Tanimoto have studied peripherical participation of editors to the final product. At this point, a brief insight in the most widespread techniques of revision is in place. In the above-cited article 'Success in Research Article Writing and Revision: A Social-Constructionist Perspective', Gosden examines the revisions of research articles written by a group of biochemists. From this analysis, three main strategies of text modification stood out: 'the deletion', 'the reshuffling of original statements', 'changes in the modality of certain assertions' (1995: 42). These resemble – at least, partially – the strategies identified by Juan Li (2000),

namely adding, deleting, substituting, rearranging. Gosden connects the three strategies outlined by him to the communal aim of rendering the text more appropriate to the communicative context and the genre framework of reference. Ventola (1991), referring to experiences of expert writers, assigns only a small amount of the whole finalization process of writing products to editing, highlighting the importance of negotiation processes – which I will analyze shortly, when dealing with the author's involvement – and the essential role of proper revising and rewriting practices for cases where texts are written by authors who have little proficiency in English. In this respect, some practical examples are provided by the author, related to thematic patterns and the use of connectors. On this latter subject, it can be read:

> as far as connectors are concerned, it seems that revisers do not pay systematic attention to their use. Incorrectly used connectors will naturally be changed or placed appropriately. But usually no suggestions are made on using explicit connectors to improve the propositional or global organization of texts, although native readers expect such markers and also find them in texts written by native writers. (ibid.: 468)

I would now like to consider the practice of involving the author in the revision process, which was also identified by Ventola as an important step of the whole revision process. Such a practice allows for a more horizontal relationship between authors and revisors, and can save precious time, in particular when it comes to accuracy and clarity. In my personal experience as a revisor of scientific articles on the topic of vulcanology and seismology – of which I am by no means an expert – my workflow as well as the final product of my revision largely profited from the exchanges I could have with the authors. This practice can also spare the involvement of other terminology experts, which, while being quite time- and budget-consuming, is a widespread costume in revision circumstances where authors cannot be contacted (Billingham, 2002). In the above-cited article on 'convenience editors' a similar view is expressed. Willey and Tanimoto (2013) suggest that 'involving authors in the editing process' can help language professionals who 'can never hope to have a sufficient scientific knowledge'

(Willey & Tanimoto, 2013: 31).

I would like to conclude this section with a quotation from the above-cited work by Hartse and Kubota, who discuss on the contradictions deriving both from error-based approaches, which do not accept any variation of lexicogrammatical forms, and pluralist approaches, which have to confront with standards demanded by the publishing industry. The authors write:

> We do not endorse guilt and anxiety about copyediting and acknowledge that errors, of course, do exist. But each person involved in the writing, editing, and publishing process is responsible to uphold standards of intelligibility as well as ethical treatment of NNES scholars. Editing decisions reflect a complex interplay among native-speaker intuition, desire to avoid stigmatization, and conformity to standard varieties of English. (2014: 80)

2.4 Computational Thinking and Linguistics

2.4.1 Data from Texts: Corpus Linguistics and Computational Linguistics

In linguistics, a fixed collection of texts in computer-readable format, representative for a specific linguistic feature or language variety, is called a corpus. Thereby, corpus linguistics can be defined as the study of language data on a large scale through computer-aided analysis (McEnery & Hardie, 2012). However, as McEnery and Hardie explain in the preface to their textbook, 'corpus techniques tend no longer to be the preserve of a clearly delimited field of specialists, but rather have become a critical resource across linguistics as a whole. Thus, [...] the future of the field is in "corpus methods in linguistics"' (2012: XIV).

The history of the discipline is to be found within the studies on the English language 'from the 1960s onwards' (ibid.: 71). On the one hand, English Corpus Linguistics' (ECL) role has been of central importance for the development of the broader discipline;

on the other hand, though, some authors, such as Léon (2005), as reported by McEnery and Hardie (2012), have highlighted the Anglo-centric bias in this field. However, this issue has been addressed by many studies based 'on a large number of languages other than English' (ibid.: 71).

Moreover, it has to be noted that ECL is not homogenic. From its very beginning, schools such as the one at the University of Birmingham differentiated themselves from the ones that particularly focused on the annotation and computing side, adopting corpus-driven perspectives for the study of grammar and highlighting the importance of collocations for the study of both linguistic production and meaning interpretation. Earlier in Subsection 2.2.1, I used the term 'corpus-driven' referring to studies which use data in the corpus as the 'sole source of hypotheses' (ibid.: 6). These kind of studies – as opposed to 'corpus-based' methodologies where the corpus is used in order to test a pre-conceived theory or hypothesis – follow in the footsteps of the famous linguist Sinclair, whose work was developed at the University of Birmingham.

In sum, it can be said that the history of corpus linguistics has seen two broad phases. The first, up to the end of the 1980s, was characterized by the contrapposition between the aforementioned ECL schools. The second phase of CL – as McEnery and Hardie outline – has seen 'the shift in the nature of corpus linguistics, [... which] has become an indispensable component of the methodological toolbox throughout linguistics' (ibid. 226).

In general, it can be said without any doubt that the introduction of corpus methodologies in the study of language has been revolutionary for more than a reason. First and foremost, corpora 'cover subsets of all the research questions that a linguist might ask.' (ibid.: 27). Moreover, the adoption of corpus-informed methodologies resulted in the radical shift in general linguistics from formalist theories (i.e., Chomsky's school in the 1960s) and prescriptive approaches to the descriptive study of languages (ibid.).

While discussing the problematic nature of perscriptivisim in linguistics, Milroy and Milroy, similarly to McEnery and Hardie, notice that descriptivism in linguistics has been the main approach for a long time before this trend became accepted from the

'general public', referring, for example, to users of dictionaries (Milroy & Milroy, 1985: 4). The two authors also relate prescriptivism to evaluation practices (see Subsection2.3.5): indeed, it is important to underline that descriptivism and corpus-informed linguistics have had an essential role toward the establishment of less biased practices in language teaching and evaluation (ibid.). Of course, as McEnery and Hardie clearly state, the process of building, annotating and consulting corpora is never totally objective: even the same results, notice the authors, can be interpreted differently from different corpora users. Nevertheless, since corpus linguistics has developed linguistics as become a field to be investigated with scientific methodologies in a way that could never have been possible before.

In Chapter 8 of their textbook, the authors analyze the 'convergence of corpus linguistics, psycholinguistics and functionalist linguistics', which can be easily explained in the light of what has been said in Subsection 2.4.1. As the authors summarize at the end of the chapter, 'these areas of convergence include the inseparability of grammar from lexis [and] grammar as a phenomenon that emerges from patterns in actual usage' (ibid.: 221). McEnery and Hardie also highlight the importance of corpus linguistics for – and its intrinsic connection with – another discipline, which developed from the 1960s onwards: computational linguistics.

A broad definition of computational linguistics is provided in the conclusion of McEnery and Hardie's volume, according to whom it is 'the field of computer science that looks at how computer systems can be created that work with language in some way'. (ibid.: 228). The authors notice that the two fields of corpus and computational linguistics have converged from the 1980s on, 'as methods based on corpus data became an essential part of computational linguistics' (ibid.: 228), in particular after generative grammar theories were surpassed by probabilistic approaches and the use of stochastic models – but I will come back to this in a few lines (Abney, 2011).

From the computational linguistics side, the definition given by McEnery and Hardie could sound a little reductive, compared to the great revolutionary changes that computational linguistics has brought into the field of linguistics. Such an idea is supported,

for example, by Steven Abney who arguments that computational linguistics has 'its own philosophy and methodology of language study', thus properly constituting 'an alternative linguistics: a linguistics characterized by systematic data collection and rigorous, experimental testing of predictions' (2011: 1). And he clarifies: 'its subject matter is language, not technology' (ibid.). The author also quotes Kay (2005), for whom 'computational linguistics is trying to do what linguists do in a computational manner' (Abney, 2011: 1). Hence, Abney decides to call this type of linguistics 'data-intensive experimental linguistics', considered as a 'genuine linguistics' which 'enables fundamental advancements' in the field (ibid.).

The author briefly outlines the history of computational linguistics, collocating its birth at the beginning of the 1960s, signposting the ALPAC reports (1966) as the place where the term, coined by Hayes four years earlier, publicly appeared for the first time. The report was a negative evaluation of the research in machine translation. To avoid a complete stop in research on the subject, the scholar founded the Association for Machine Translation and Computational Linguistics, which in 1968 became the Association for Computational Linguistics. At the time when the subject was founded, another name was rejected, namely 'natural language processing' (NLP) – a term that is nowadays used interchangeably with computational linguistics, or specifically in reference to NLP in studies related to artificial intelligence (AI) (ibid.)[17].

Abney condemns the lack of interaction between general linguistics and computational linguistics. Indeed, he notices that, until recently, the only linguistic theory which systematically relied on computational linguistics had been the one for which the study of competence and performance should be kept separated – that is, Chomsky's formalism

[17] On the name and definition of the discipline, Jurafski and Martin (2020) introduce their volume as follows:

> This book is about a new interdisciplinary field variously called computer speech and language processing or human language technology or natural language pro- cessing or computational linguistics. The goal of this new field is to get computers to perform useful tasks involving human language, tasks like enabling human-machine communication, improving human-human communication, or simply doing useful pro- cessing of text or speech (2006:1).

and generative grammar.

The revolutionary shift in the field of computational linguistics, which also caused the convergence of corpus and computational linguistics, arrived with the introduction of statistics and, particularly, the application of Hidden Markov Models to some big unsolved problems of computational linguistics. In this respect, Abney cites Church (1988) and DeRose (1988): 'Probabilistic methods were soon being applied to nearly every problem of natural language processing, and within a few years had reshaped the field' (Abney, 2011: 7). The 'statistical revolution', as defined by Abney, brought changes in the very way problems were approached, and experiments were conducted[18].

The closure of his article summarizes his main thesis on computational linguistics as being a new philosophy of linguistics. He writes:

> emerging from computational linguistics is a new approach to linguistic research that is predicated on systematicity and experimentation enabled by large-scale data collection. [...]
>
> The new approach reflects a deeper understanding of the scientific method, and places linguistic inquiry firmly within the paradigm of data-intensive research that has come to characterize modern sciences. (ibid.: 26)

Jurafsky and Martin point to the second half of the 1990s as the period when the dichotomies that had characterized the field for the previous three decades resolved in the incorporation of probabilistic theories and empiricism in all the main research issues. Computational linguistics made rapid progress in recent years favored by the technological advancement of computers, the expansion of the Internet network and by the progressive constitution of large textual dataset – so-called *big data* (Mitkov, 2003; Jurafsky & Martin, 2006). In other words, corpora and big data serve nowadays as the backbone for computational linguistics research.

From the beginning of the new century, machine learning techniques – both supervised and unsupervised – became to spread, investing all applications of NLP. In this sense,

[18] For an extensive review of the applications of stochastic theories to NLP problems, see Krenn & Samuelsson (1997).

connections with neuropsychology became increasingly more relevant, as attested by Aslin (2017) who compared statistical learning to human learning techniques. I will give some more information on the topics of machine learning, deep learning and neural networks in the next subsection.

In order to understand the main trends of research in nowadays computational linguistics the proceedings of the latest editions of the International Conference on Computational Linguistics can be consulted. Scanning the Proceedings of the 2020 Conference, for example, the large presence of studies that employ the BERT algorithm – which I used in the experiments (see Chapter 3) and that will be presented in Subsection 2.4.3 can be noticed. Moreover, the topics of linguistic productions quality evaluation, text similarity, academic writing as well as editing and revision are present (as means of mere example, Anthonio & Roth, 2020; Lepori & McCoy, 2020; Mordido & Meinel, 2020; Gotou et al., 2020; Muangkammuen et al., 2020; Saberi et al, 2020; Jafaritazehjani et al., 2020; van der Lee et al., 2018; Madnani & Cahill, 2018; Paetzold & Specia, 2016; Rubino et al., 2016; Todirascu et al., 2016; Somasundaran et al., 2016; Pal et al., 2016; Sperber et al., 2016).

2.4.2 Automatic Text Processing and Evaluation: History and Latest Development

The expression text processing generally refers to the creation and modification of texts in electronic format. This is possible thanks to specific software that can be used with a computer or other electronic devices (Jurafsky & Martin, 2006). For instance, text processing makes standard operations such as texts writing and texts manipulation through commands which allow you to search and replace words, to delete, copy and paste parts of texts, to format texts, etc. possible.

These features, which now seem simple and obvious, are largely based on the concept of regular expression. A regular expression consists of a sequence of characters that form a search pattern within a text. These first simple operations of searching, recognizing and eventually replacing patterns in a text have laid the foundations of text processing

which is currently projected towards more complex functions such as the automatic generation of summaries, the prediction and suggestion of words and phrases, the automatic writing assistance, including automatic corrections and evaluation of linguistic quality and adequacy of a text (Mitkov, 2003).

Text processing is inspired by the human ability to use words to communicate and therefore to process, or produce and decode, natural language (ibid.). For this reason, the most recent developments in text processing methods are based on Artificial Intelligence (AI) techniques (Léon, 2021). AI is a scientific discipline based on computer science, which studies the theoretical foundations and techniques that allow for the creation of automatic hardware and software systems, capable of processing data and information implying learning techniques and processing procedures that resemble human intelligence (Boden, 2018).

The discipline has largely drawn from information theory, just as NLP has (Léon, 2021). In this respect, Léon notices similarities in the origins of the two subjects.

> The history of artifcial intelligence – she explains – has several aspects in common with the history of natural language processing. Turing's universal machine, conceived in 1936 as a thinking machine that could manipulate discrete symbols and use rules to operate calculations, was also the first finite-state automaton. It was anchored in the frst mathematisation of language. Cybernetics and information theory, which had appeared immediately after World War II, were also at the junction of both fields: IA and MT. (ibid.: 80)

In fact, machine translation (MT) and IA had already been related before the first MT system was born in 1949: two years earlier Weaver shared his first thoughts about MT in a letter to Wiener[19] – who in 1948 presented his information theory in the publication 'A Mathematical Theory of Communication' (Shannon, 1948). However, as the author specifies, 'Weaver's hypotheses were much less ambitious than those advanced by Alan Turing (1912-1954) in his article on the imitation game (Turing's test), published in

[19] https://www.historyofinformation.com/detail.php?id=2990

1950. For Shannon, the machine could not go beyond the objectives for which it was devised (ibid.: 80,81)' (Léon, 2021: 81).

AI systems can be defined on the basis of their internal 'processes of reasoning' (Wagman, 1991: 1) or of their external behavior. The performances of such systems are evaluated in comparison either with human behavior or with an ideal behavior, so-called rational. The goal of these systems is to act in a similar way to humans, so that the result of the operations performed by the intelligent system is little distinguishable from the result of human action. When these systems are applied to cognitive sciences, the process that leads the intelligent system to solve a problem is similar to the human one. Alternatively, the process that leads the intelligent system to solve a problem is a formal procedure that refers to logic, according to a rational action, that is, pursuing the best results given the available information[20].

Artificial intelligence is deeply rooted in the statistical and mathematical sciences and can be applied to a broad range of interdisciplinary fields. Machine learning, computer vision, natural language processing, data science, generally applied to optimization problems and decision support, cognitive neuroscience applied to automatic decisions are some of fundamental procedures on which the functioning of Artificial Intelligence is based (Wagman, 1991; Boden, 2018). As can be deduced, artificial intelligence is a discipline that also implies ethical as well as theoretical and practical aspects.

In the field of computational linguistics, the most recent systems, particularly oriented to the global world and to the Internet user community, train their algorithms for writing assistance, editing and revision by using English webpages from Wikipedia (e.g., BERT developed by Google, that will be described in Subsection 2.4.3), or texts produced by large communities of users as their corpus – for instance, Grammarly (which will be presented in Section 2.4.4). These systems, given their international – if not global – nature, may be focused on the use of ELF (see Section 2.2). They are able to learn from user contributions and from the feedback they receive to their correction

[20] https://www.javatpoint.com/reasoning-in-artificial-intelligence
https://www.professional-ai.com/reasoning-in-artificial-intelligence.html

suggestions, which can be ignored or accepted by users. This makes them constantly updated tools, which reflect the evolution of languages. In addition, these technologies also have the potential to be adapted to the more specialized needs in the EAP area (Léon, 2021).

Because the present research deals with state-of-the-art machine translation systems as well as text similarity metrics which make use of deep learning techniques, i.e. the BERT algorithm, I will now briefly introduce artificial neural networks. These systems are the basis of the most modern technologies for Natural Language Processing.

Many of the modern technologies of text processing exploit the learning abilities of Artificial Neural Networks (ANN), widely applied to artificial intelligence systems (AI). ANNs are able to learn from examples and, once trained, to recognize and predict patterns. As it was previously mentioned, these analysis techniques were developed starting in the early 1940s and were improved and refined in the following decades. However, only in the last two decades they have been intensively applied in many fields of analysis, thanks to the technological progress: nowadays great computing power permits to carry out complex operations in a short time. It should also be noted that, hand in hand with the technological progress, the need to automatically manage, classify and analyze large amounts of data (big data) and large amounts of textual content disseminated on the World Wide Web (WWW) has also grown (Mitkov, 2003).

An ANN is a computational model inspired by the biological neural networks that constitute the human brain. ANNs are based on a collection of processing units called artificial neurons (or nodes). The artificial neurons are connected to other neurons through links called synapses. Each synapse, like in a biological brain, can transmit a "signal" (represented by a real number) from one neuron to another. Connections between neurons are associated with weights that adjusts as the learning activity proceeds. The weight modulates (increases or decreases) the strength of the signal in a connection. Neurons may have a threshold such that a signal is transmitted only if its "intensity" crosses that threshold. Typically, neurons are aggregated into layers. The

ANN receives external signals (e.g,. an input vector) on a layer of input nodes, each of which is connected with numerous internal nodes, generally organized in several layers. Each node processes the received signals and transmits the result to neighboring nodes. At the end, the output layer provides the final result of the process (Goldberg, 2017; Eckman, 2021).

In most cases, an artificial neural network is an adaptive system that changes its structure based on the information that passes through the network itself during the learning phase. The learning process can be supervised, i.e., based on labeled data (pre-classified by human analysts) to help predict outcomes, or unsupervised, i.e., based on algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns in data without any need for human pre-classification. In all cases, the neural network learns to perform tasks such as recognition, discrimination, classification, clustering and comparison of patterns using specialized criteria chosen by system developers. In this sense, it is important to underline that there is a design choice underlying the training of an automatic AI system, developed, for example, to distinguish a correct English text from an incorrect text.

2.4.3 Text Similarity Metrics

In Natural Language Processing, measuring text similarity is an important task that has extensive implications for the study of linguistics, psychology and information theory.

Most traditional automatic methods for measuring text similarity treat texts as collections of words, analyze the number of times each word appears in a given text, then use the information about word frequencies to represent texts as vectors (Huang et al., 2011). Thus, once two or more texts have been transformed into vectors, they can be compared with various methods that allows for text similarity estimation (Salton, 1971). Of course, these methods ignore the meaning of terms in the text, the existence of grammatical rules, text organizational structure, etc.

In contrast to these first automatic techniques proposed for measuring similarity be-

tween texts, the more modern methods are based on semantics. They expand semantically similar terms in the traditional word frequency vector and further increase the size of the text representation vector. Semantic-based text similarity measurement methods examine the similarity between texts accounting for semantic relationships such as synonyms, redundancy and implications, but may not well reflect the strict similarity between two texts, when it comes to word choice and small changes in grammatical structures.

Text similarity measurement methods have a wide range of applications: in the field of text classification (Ko et al., 2004), in the automatic generation of text summaries (Erkan and Radev, 2004), in the detection of text repetition (Theobald et al., 2008), to retrieve the most relevant document corresponding to web user's query Pradhan et al. (2015) and for Quality Estimation (QE) of machine translation (MT) (Papineni et al., 2002; Moon et al., 2020).

An in-depth discussion of the various methods proposed for measuring the similarity of texts is beyond the scope of this dissertation. What follows is a description of the two methods used in the present work – namely, BiLingual Evaluation Understudy (BLEU), as an example of a well-established traditional method, and Bidirectional Encoder Representations from Transformers (BERT), as an example of a semantic embedding method.

The BLEU method was proposed by Papineni et al. (2002) to evaluate Machine Translation (MT) systems (see Subsection 2.5.1). Developers of MT systems need to frequently monitor the results of changes to their systems, in order to understand which ones produce improvements and which generate deteriorations in the quality of the translation. For this reason, the progress of MT systems depends on efficient and rapid automatic evaluation, which correlates with human evaluation. MT human evaluation takes into account different aspects of translation including adequacy, fidelity, and fluency of the translation and adopts different approaches (Reeder 2001). However, human evaluations are generally expensive and time-consuming. In order to address this problem, Papineni et al. (2002) proposed the BLEU method, which is based on the idea that

the closer the machine translation is to a professional human translation, the higher its quality. For this reason, the authors develop their method around two main elements: a) a numerical "translation closeness" metric and b) a corpus of good quality human reference translations.

Scores are calculated on individual segments, usually sentences, by comparing them to a set of good quality reference translations. In the segment comparison method, individual words are used as the base unit for the comparison between the candidate and the references. The scores obtained for the individual segments are then averaged over the entire text to obtain an overall estimate of the translation quality. Moreover, different types of score calculation are made available. Besides sentence score calculation, it is possible to use the corpus score function. 'Different than averaging BLEU scores of each sentence, it calculates the score by "summing the numerators and denominators for each hypothesis-reference(s) pairs before the division"[21]. In the present study, both sentence score average and corpus score were calculated, which give very similar results.

The output of the BLEU method is a number between 0 and 1 which represents the similarity between the candidate and the reference. A value close to 1 indicates high similarity while a value close to 0 indicates low similarity between the candidate and the reference.

Although BLEU is one of the first metrics that were used for MT systems quality evaluation, it is still valid and represents a benchmark for testing and comparing the performance of any new evaluation metric method. BLEU has applicability limitations when dealing with languages lacking word boundaries; however, these are not relevant for the present study (Callison-Burch et al., 2006).

BERT is a machine learning technique for natural language processing (see Subsection 2.4.1) developed by Google in 2018 (Devlin et al., 2018). Since 2019, Google has been using BERT in its search engine, specifically for English-language queries. BERT

[21] https://colab.research.google.com/github/gcunhase/NLPMetrics/blob/master/notebooks/-bleu.ipynb

uses two main types of models for the English language, Base and Large. $\text{BERT}_{\text{BASE}}$ models have 12 encoders with 12 bidirectional self-attention heads. $\text{BERT}_{\text{LARGE}}$ models are instead based on 24 encoders with 16 bidirectional self-attention heads. Both model types are pre-trained from unlabeled data. $\text{BERT}_{\text{BASE}}$ is pre-trained using a Books Corpus with 800M words (Zhu et al., 2015), while the $\text{BERT}_{\text{LARGE}}$ model is pre-trained using English Wikipedia, with a corpus of 2,500M words[22]. Over time more models have been added, among which RoBERTA-large that was used in this study.

Traditional methods have the limit of offering a single vector for each word, regardless of the context to which the word belongs (context-free). These models cannot handle polysemy. On the contrary, BERT is able to dynamically generate vectors considering the context of the words, through the use of a fully bidirectional model trained without supervision with a simple textual corpus. This approach, based on the concept of transformer language model – i.e., a deep learning model that adopts the mechanism of self-attention, weighting the significance of each part of the input data (Vaswani et al., 2017) – gives BERT its ability to contextualize. Thus, BERT is an important tool for automatic semantic analysis. It is currently implemented, for example, for a better interpretation of query strings on the Google search engine, as it helps to better understand the search focus. Furthermore, BERT's ability to manage polysemy and represent the meaning of words based on context makes it suitable for new applications such as word prediction and suggestion – for example, when writing an email or during a web search.

BERTScore is 'a language generation evaluation metric based on pretrained BERT contextual embeddings (Devlin et al., 2018). BERTScore computes the similarity of two sentences as a sum of cosine similarities between their tokens' embeddings' (Zhang et al., 2020: 1). Since it identifies cosine similarity, BERTScore is expressed with a number ranging between -1 and 1, with values close to 1 which indicate high similarity

[22] links to BERT models are available at https://github.com/google-research/bert https://github.com/Tiiiger/bert_score/blob/master/README.md

between candidate and reference.

Rescale functions are available for both scores. For example, with BERTScore the rescale_with_baseline function, which was used in part of the experiments presented in Chapter 3, allows for a rescale of the score within a natural range (for example, between 0 and 1)[23].

2.4.4 Automatic Text Quality Evaluation and Editing

In the context of automatic text processing, many studies have focused on quality evaluation of linguistic productions. Apart from machine translation quality evaluation, which includes, in truth, text quality evaluation, but goes far beyond that (see Subsection 2.5.1), probably the first approach to this topic as a stand-alone issue was related to the possibility of automatically assessing English learners' productions.

In a technical report produced at the University of Cambridge in 2013, automated assessment was already treated as a machine learning problem. Broadly, it is defined as the task of 'automatically analysing and assessing someone's competence' and the first steps in this research 'can be traced back to the early 1960s' when it 'emerged as a means to overcome issues arising with standardised assessment' (Yannakoudakis, 2013). In the framework of the teaching and assessment of English as a Foreign Language, the corpus which is mainly used as the database for developing automatic evaluation systems is the Cambridge Learners Corpus[24](ibid.).

In 2015, Hamp-Lyons and Lockwood described the automatic evaluation debate that was taking place – and still is – in the U.S., where the government has imposed 'writing requirements at all levels of schooling' and has demanded 'that students receive feedback on their writing' (ibid.: 1). Anyway, when it comes to language learners' productions assessment, many issues – both of practical and ethical nature – arise. Madnani and Cahill (2018), for example, have noticed that automatic scoring is a

[23] https://github.com/Tiiiger/bert_score/blob/master/journal/rescale_baseline.md

[24] https://www.sketchengine.eu/cambridge-learner-corpus/

https://www.cambridge.org/elt/catalogue/catalogue.asp?cid=339

problem that goes beyond NLP. In this respect, they write:

> If NLP researchers initiate the request to add automated scoring to an assessment, they should have already connected with the subject-matter experts to ensure that they have built a system that adequately measures the correct construct. [...] For any automated scoring system that is proposed, NLP researchers need to take into account ethical considerations regarding fairness and validity and evaluate the system on dimensions other than just the agreement with human scores. (ibid.: 1104)

And this is precisely the reason why, despite the practical focus of this work, it is only in the present subsection, after a review of the main theoretical elements that can inform both human and automatic evaluation, that I am starting to address this issue itself.

In general, learners' productions and academic writing in English as a lingua franca share many features, especially when learners are advanced and are confronted with writing tasks such as short essays or narratives. As repeatedly highlighted during this work, the main difference between the evaluation of scientific writing and learners' production is the purpose of the evaluation; in other words, systems developed for automatic scoring in learning contexts might not be appropriate tools that can directly be adapted by publishing companies with fix standards of publication and with little interest in understanding the stage of the candidate's English acquisition process or in providing some kind of feedback for the authors' language learning path. This is, instead, the framework within which the DELTA (the Diagnostic English Language Tracking Assessment) project was developed.

The DELTA project aims at the automatic assessment of undergraduate students' essays, reducing time and costs of evaluation procedures. In the project, three automatic essay scoring applications (EAS) have been compared: Intelligent Essay Assessor developed by Pearson Knowledge Technologies[25], which also provide a feedback version

[25] The information sheet for this system is available at the following link: https://pmark.pearsoncmg.com/templates/assets/upload/IEA-FactSheet.pdf

for students and teachers, Write to Learn; *e-rater*®, developed by ETS, which can be trained according to the institutional needs and use holistic scores[26]; Intellimetric by Vantage Learning, specifically designed for classroom use[27].

Similarly to the latter two systems, in 2009 IADE (the Intelligent Academic Discourse Evaluator) was specifically designed to meet the most diverse needs of learners in the framework of research in intelligent computer assisted language learning (ICALL). Despite its name, though, IADE is not so much an evaluator as it is a tool dedicated to international students who need to improve their writing skills at Iowa State University, providing help for self-improvement of first drafts of papers (Cotos, 2009, 2010). The system 'automatically analyzes learner drafts of research article introductions in terms of discourse development features' (Cotos, 2009: 106). It is meant to be used by students on their own, since the system engages 'learners in an iterative revision process and, at the same time, enhancing the formative assessment aspect of the instructional process' (ibid.). For this reason, while the final purposes of this technology might not completely overlap with the ones of this study, the focus on evaluation for revision – albeit self-revision – is of particular interest for the present research. As Cotos explains, the principles of Evidence Centered Design (ECD) constituted the starting point for the development of IADE. In this sense, the system employs two of the four ECD constituting processes, namely the presentation process, which relates to the web interface, and the response processing. In order to make these processes work a database named Evidence Compository Library has been integrated in the system. The model adopted by the system is based on encoding and decoding modules which allow for texts processing and numerical feedback. The latter part is performed as a classification task, based on the information contained in the database (Cotos, 2009). This last element unveils the problem related to the possibility of applying many automatic evaluators for revision difficulty evaluation, since they need to be trained on very specific data with predefined rules and criteria, thus departing from the features that characterize the problem I am trying to address.

[26] A description of the ETS system can be found here: https://www.ets.org/erater/about

[27] More information can be found at this link: https://www.intellimetric.com/direct/

In 2020, Gotou et al. have investigated the possibility of evaluating the difficulty of correcting errors – a task that seems quite alike the one presented in this study. At a closer view, however, these authors focus on the success rate of automatic correctors, that is the capacity of correctly modifying a word or sentence which present some errors. In this sense, the application of this evaluation procedure is meant to be embedded in automatic correctors – which will be soon treated in more detail – in order to improve their performances.

Text clarity scoring is the problem addressed by Muangkammuen et al. (2020), who explore local coherence relations between two phrases or sentences to evaluate general clarity in writing. The method presented in this paper is pre-trained on the BERT language model. This is implemented with a text clarity model that exploits local coherence relations in order to assess the structure of a text.

The main trend underlying these studies seems to be that of addressing very specific problems, which surely play a role in estimating text quality but are either only a small piece of the main picture or they are linked to very specific purposes – different from ours, so that their application is extremely narrowed. In this respect, an exception that is surely worth mentioning is Coh-Metrix, defined by its creator as 'an automated tool for theoretical and applied natural language processing' (McNamara and Graesser, 2012). The many perks of this tool include the innovative theoretical background according to which it was designed – for example, the concept of text readability (see Subsection 2.3.4) is substituted by several factors which account for a major focus on discourse issues, and the adaptability of this tool to a great variety of NLP problems (ibid.). However, while Coh-Metrix could be implemented for text evaluation purposes in terms of revision difficulty, such an application has not yet been explored – and dealing with such a complex system to investigate its possible adaptations to specific NLP problems would require a deep knowledge of the model and functioning of the tool.

On the other hand, I would now like to take into consideration some systems that are specifically meant for revision. Despite believing that automatic revision for publishing

purposes is still far from being a reality, some systems have been developed with the aim of improving self-revision, revision and editing practices. For reasons related to available time limitations and conciseness, technologies related to aided writing on mobile phones have not been taken into consideration in the present work[28].

Before delving into the topic of 'intelligent' automatic editors, however, text editor software, in general, need to be introduced. Indeed, text editors such as Libre Office[29] and Microsoft Word[30] are some of the most common software employed by revisors. For this reason, these programs feature 'Review' options and automatic spelling check (see Subsection 2.4.2) (Billingham, 2002; Mossop, 2001). Nowadays there are also plenty of text editors that can be used online, such as Google Documents[31]. In this respect, a project that seems particularly interesting, as it moves from the premises of cooperative learning and collaborative knowledge, is CEPT, the Collaborative Editing Tool for Non-Native Authors (Zhu et al., 2015). Although such an editing practice does not guarantee high linguistic standards and has little to do with evaluation, it implements cross-version sentence mapping, summarization of edits from multiple co-authors, and a collaborative editing interface. To some extent, CEPT can be seen as an *improved* collaborative version of Microsoft Word, or Google Documents.

In recent years, some applications have been created for automatic proof-reading. ScribeMedia an online professional publishing service, has published a web article focused on these innovative tools, which 'apply artificial intelligence (AI) to proof-reading'[32]. As explained on the webpage, unlike Microsoft Word's spell checker which works with a dictionary, editing apps are true grammar and style checkers, evaluating possible mistakes in their context of usage[33]. According to ScribeMedia, Grammarly

[28] As an example of the developments of such tools and the main topics addressed in this field of research, see Li et al. (2020).

[29] https://it.libreoffice.org/

[30] https://www.microsoft.com/en-us/microsoft-365/word

[31] https://www.google.com/docs/about/

[32] https://scribemedia.com/proofreading-editing-software/

[33] A review of some automatic editors can be found here: https://mention.com/en/blog/content-editing-tools/

can be considered as the best-performing of such programs.

Grammarly is a proprietary software, developed by Grammarly Inc, that also distributes a free basic version. It is a cloud-based writing assistant designed for the automatic correction and revision of texts written in English. Grammarly is based on Artificial Intelligence (AI) techniques and applies different grammar rules. The system learns the rules and hidden patterns of good writing by analyzing millions of sentences from a large collection of texts that has been organized and labeled for research and development purposes – i.e., a very large training corpus. Grammarly exploits the computational power offered by the cloud. It learns the usefulness of its individual suggestions by analyzing users' feedback, who can ignore or accept corrections. In this way, the system gradually improves its revision skills. Grammarly also provides an evaluation of the analyzed text through a performance score. The score indicates how accurate the analyzed document is compared to documents written by other Grammarly users, who set the same goals as the analyzed document. To calculate the score, Grammarly computes the accuracy level of a document based on the total word count and the number and types of writing issues detected. Then, Grammarly compares the accuracy level of the document to the accuracy levels of other documents. A score of 90, for example, means that writing in the document is more accurate than writing in 90 percent of other documents with similar goals. The text score ranges from 1 to 100 and depends on how many different types of suggestions appear in the document, as well as how the text compares to other texts with similar goals. The fewer the suggestions Grammarly makes, the higher the score of the document[34].

In order to discuss the outcomes of the experiments described in the following chapter, I am going to use the automatic score calculated by Grammarly in comparison with the scores I obtained through the method here investigated (see Section 3.6).

This review of automatic language evaluation applications, far from being an exhaustive summary of the diverse landscape which characterizes language assessment technology, was meant to highlight the great attention given to this problem in recent linguistics

[34] More details about this software can be found at the link: https://www.grammarly.com/blog/

research.

One last aspect which needs to be considered relates to the ethical issues concerning automatic evaluation, as I had previously mentioned. In the general field of language analysis, in recent decades automatic analysis methods based on computer systems (stand-alone software or web-based platforms) have become increasingly widespread and established. Such a phenomenon is accompanied by many side-effects and endeavors reflections on the changes that the fast technological progress of the latest years has caused in human life and in the global society. In the world of post-globalization, where the contradictory nature of progress has highlighted the many inequalities that the process of globalization brought along – including the so-called 'digital divide' (Guerra, 2010), technology enhancement can be a valid support to favor and strengthen strategies for overcoming barriers, even those linked to language diversity, unequal learning possibility and access to knowledge. Conversely, they involve risks regarding the standardization of language which can have negative repercussions on creativity, freedom, and the right of expression in human communications (Mayne, 2021)[35]. This ethical issue has been addressed from many points of view (see, for example, Loader, 1998; Dumouchel & Damiano, 2019; Jahankani et al., 2020; Mayne, 2021). However, if it's true that human nature is intrinsically social and cultural, as Prinz argues in his 'Beyond Human Nature' (2012), then the new reality of augmented humanity should not be ostracized from the outset; instead, the consequent debate around technology should be restated in terms of users' awareness and disparities narrowing.

[35] https://dept.writing.wisc.edu/blog/revisiting-grammarly/

2.5 An Innovative Method: Round-Trip Translation

2.5.1 Automatic Machine Translation: the State of the Art

Before deepening into the description of round-trip translation (RTT) as a method and its application in translation and language studies, some considerations about automatic machine translation (MT) are in place.

'The term machine translation refers to computerized systems responsible for the production of translations with or without any human assistance' (Mitkov, 2003: 503). Most of these systems are nowadays recurrent neural networks, which make intensive use of deep learning techniques with bi-directional layers using end-to-end processes for decoding – first used in speech recognition tasks (Mikolov, 2010; Liu et al., 2014) – and encoding. The state-of-the-art MT systems normally work with very large, updated collections of data and are able to learn from new examples, thus improving their performances (see Subsection 2.5.1). Users of automatic translation can be international companies and organizations, business professionals, learners, university scholars and really anybody needing a translation for the most diverse reasons.

For example, the European Union has had for years now some of the most advanced institutions in translation and, in general, linguistics research. In fact, EU had already had its own machine translation system (MT@EC – initially based on statistical models) (Foti, 2016) for several years when neural networks were implemented, and E-Translation – a system specifically meant to be integrated in public administration – was officially launched in 2017. I will report some extracts from the description of E-Translation available on the EU website to give an idea of the kind of services this tool was born to supply:

> the main purpose of CEF eTranslation is to make all Digital Service Infrastructures (DSIs) multilingual. While CEF eTranslation is mainly intended to be integrated into other digital services, it also offers useful stand-alone services for translating

> documents or snippets of text.
>
> Unlike general-purpose web translators, CEF eTranslation will be adapted to specific terminology and text types that are typical for the usage context (e.g. tender documents, legal texts, medical terminology). It enables multilingual operation of digital services and can be used to reduce the time and cost of translating documents[36].

If one thinks about it, we really live in an era in which texts written in a foreign language are regarded as comprehensible by the general public; the most popular social networks and applications not only are highly internationalized (Heimgärtner, 2014), but they also embed – or even develop – translation systems that allow for immediate translation of new content. Meta Platforms Inc. has lately developed its own MT system based on a multilingual approach. It is open sourced on GitHub and has probably been trained on the immense amount of text data users make available on the social networks and communication software owned by the company – including Facebook, Instagram and WhatsApp[37].

The wide-spread use of MT systems is probably linked to the fact that most of them are free – or a free version is available along with a purchasable one – and have very user-friendly interfaces. Therefore, the majority of the common users of MT systems have hardly any idea of how they work, although fast progress in the field is commonly perceived (Boden, 2018). Considerations about the influence of AI in everyday life, awareness in users and ethical problems that might be related to an improper or excessive use of these types of technology can be found in Subsection 2.4.4, in Chapters 3 and 4.

Undoubtedly, MT systems have developed by leaps and bounds since they were introduced for the first time. The evolution of machine translation has seen three main phases. In the 1960s, the problem of machine translation was addressed using rule-based models. As Rivera-Trigueros (2021) notes, these systems can be very expensive

[36] https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-etranslation_en

[37] https://github.com/pytorch/fairseq/tree/main/examples/m2m_100

because they need regular update and are not able to deal with language ambiguity. Then, since the 1980s – and until very recently – statistical MT systems became the predominant model in use. Statistical MT systems (SMT) have the advantages of generally being accurate in vocabulary and training is mainly automatic. On the other hand, they can produce grammatical and structural errors. It should be noticed that rule-based and statistical MT systems can be integrated. In the second decade of this century, neural machine translators have been developed. These systems have rapidly conquered machine translation landscape, because they allow for good quality translation with much lower processing and memory costs (Rivera-Trigueros, 2021).

An in-depth discussion of the characteristics of MT systems is beyond the scope of this work. However, a broad understanding of the mechanisms behind machine translation is necessary since this study moves from the assumption that MT performances are so refined that they can decode meaning beyond language correctness and that their productions are qualitatively high. In order to confirm this and for practical reasons, related to the choice of the most suitable MT system for the applications of the present research – RTT of academic texts – I have carried an examination of the latest performances evaluation. The Encoder-Decoder (seq2seq) model was first used by Google in 2014; to date, it is of one the best performing neural network models for machine translation (Shah, 2020)[38]. In practice, thanks to this model, it is possible to produce outputs that are also very distant from the inputs (e.g., in terms of length) – a characteristic which is relevant to the use of MT proposed in the present dissertation.

When it comes to MT quality evaluation (QE) two main approaches can be used: human evaluation, which is the most accurate but also the most expensive and time-consuming, and automatic evaluation. The article by Moon et al. (2020), for instance, explores the possibility of using RTT with reference-based evaluation measures for QE, as will be described in the next subsection. On the other hand, Sperber et al. (2016) have proposed a method for QE which they named 'slightly supervised' because

[38] https://towardsdatascience.com/machine-translation-with-the-seq2seq-model-different-approaches-f078081aaa37

quality scores for segments or files in a corpus are automatically predicted on the basis of manual evaluation done only on a small subset of data. However, it should be noticed that there is little consensus in the scientific community about QE. An updated overview of the quality assessment procedures can be found in Rivera-Trigueros (2021).

Their progress is so fast that it is unlikely that descriptions of the latest MT architectures will be found in translation or computational linguistics textbooks; in this sense, web articles and, more importantly, annual reports produced by companies, associations and conferences are certainly the most updated documents[39]. For the purposes of the present research, I have taken into account the Proceedings of the Machine Translation Summit 2021[40] and the Intento report 2021[41].

While probably constituting the most prominent assessment of the state of the art in machine translation research, the MT Summit particularly focuses on two main topics which are little related to this dissertation (see Chapter 3): improving performances with low-data language pairs and comparisons of the most updated techniques applied to a variety of general tasks.

In the framework of the Summit, a study which is particularly interesting in relation to machine translators language generation is 'On Nature and Causes of Observed MT Errors' by Popovic' (2021). Her study analyzes from a both a qualitative and a quantitative point of view the most common errors produced by machine translators and their distribution over 11 MT systems and three language pairs (EN-SR, HR and EN-DEU). For what concerns the two parameters of adequacy and comprehensibility – where adequacy means that the meaning of the source text is preserved, whilst comprehensibility defines an output which is easy to read and understand but could mistranslate the source text. The author adopts human evaluation and notices that overall 33% of all adequacy errors is comprehensible and more than 40% of all incomprehensible words are adequate translations. This confirms the previous findings

[39] An exstensive bibliography on machine translation evaluation can be found at this link: https://www.issco.unige.ch/en/research/projects/isle/femti/refs.html

[40] https://mtsummit2021.amtaweb.org

[41] https://inten.to

that good comprehensibility often "masks" adequacy errors, but also shows a tendency in the opposite direction, namely "forgiving" incomprehensible errors after seeing the source text (ibid. 171).

In general, this means that errors in meaning can be masked by a fluent rendering and errors in structure or lexical choices might not imply that the conveyed meaning is inaccurate. The fact that these errors seem to occur unpredictably should always be taken into account by users.

Moreover, Popovic' lists the most common errors generally occurring in machine translation. Among the most common errors, the ones which should be taken particularly in consideration for the use I am making of machine translation are: rephrasing, ambiguity, mistranslation and untranslated words – which are, indeed, regarded as major adequacy errors by the human evaluators.

Aside from errors analysis, however, in order to run the experiments included in this work (see Chapter 3), I needed some information about performances of specific MT systems in specific domains and language pairs. In this sense, the Independent Multi-Domain Evaluation of Machine Translation Engine (Intento, 2021) has been very helpful, as it provides a summary of the best MT Engines per Industry Sector and Language Pair, obtained using BERTScore. According to the report:

> 5 MT engines cover the best scores for all 13 languages: DeepL, Systran, Yandex, ModernMT, and Baidu or Tencent. [...] Many engines perform best with English to Spanish, Russian, and Chinese. Legal, Financial, and Healthcare require a careful choice of MT vendor, as ew perform at the top level. [...] Despite of having several comparable MT engines per language pair, Education shows relatively low scores, which may indicate the importance of customization in this domain. (ibid.: 22, 23)

Actually, from the figures in the document it can be deduced that DeepL and Google are among the best MT engines to deal with financial texts, whereas ModernMT shows good performances with the language pair EN-ES.

2.5.2 Round-Trip Translation: Description and Previous Applications

The possibility of using MT systems for academic writing purposes has begun a reality since the implementation of neural networks in machine translators, which have been developing to produce outputs increasingly target-oriented. In 2015, such considerations were made by Groves and Mundt, in a study that implied Google Translate[42] – which still featured, at that time, a statistical architecture. Because the massive shift from statistical to neural network machine translators has completely changed the characteristics of MT outputs[43], allowing a modification of the text based on previous and subsequent textual context (Kalchbrenner & Blunsom, 2013), analyzing the results of Groves and Mundt's study would be of little use. Nevertheless, their article not only proved that even in 2014 Google Translator was already able to produce outputs which would meet – though only barely – the standards of acceptance of many international universities, it also pioneered a new field of application of machine translation, namely language generation for EAP. The present dissertation, thus, collocates within this perspective. Here, however, I will not discuss the possible use of automatic machine translation for the production of academic writing as much as I will be exploiting several MT systems using a method defined as round-trip translation in order to evaluate academic writing revision difficulty.

Round-trip translation (RTT) is the process of translating a text from a source language into a target language (forward translation, FT), then translating the result back into the source language (back translation, BT), using machine translation (MT) software. The present dissertation explores the possibility of using RTT for ELF – or, more precisely, EAP -texts quality estimation, taking inspiration from the article by Moon et al. (2020), who propose this method for the evaluation of MT quality.

In the article, more than one sentence similarity metrics were used. A difference is

[42] https://translate.google.com/intl/en/about/

[43] https://www.argotrans.com/blog/accurate-google-translate-2019

outlined between surface-level metrics, such as BLEU and CHRF, and semantic-level metrics, i.e. SBert and BERTScore. This is the reason why, as will be seen in Chapter 3, the experiments conducted for this dissertation employed both BLEU and BERTScore. Starting from this application of RTT, we thought that a similar method could be suitable for quality estimation of written productions, assuming – as said at the beginning of this section – that an MT system with high performance could be used. For considerations on the actual suitability of this method for the issue under analysis see Chapter 3 and Chapter 4.

The method has been applied also to other NLP tasks: for example, Ahmadnia and Dorr (2019) investigate the possibility of exploiting RTT to train MT systems for low-data language pairs.

2.6 Conclusion

The theories and concepts presented in this chapter allow for a deeper understanding of the status of English as a lingua franca within the academic world, and the effects that the massive spread of English as the main language for academic communication has on the work of revisors.

In sum, what is important to notice is that, while editing practices can follow different styles that largely relate – to a smaller or greater extent – to the requirements of the publishing context, it is during the process of revision that changes are made with a specific, though not necessarily entirely conscious, 'standard' use of English in mind: in most cases, collocations and structures typical of the British variety (Phillipson, 2006a). Thus, the present research tries to replace the *nativeness* standard following the path of automatic approaches to quality evaluation through the method of round-trip translation. It must be said that for the purposes of revision this is, in truth, a very difficult operation because of the complex context where revision practices take place.

Ultimately, this chapter has been conceived with two main objectives in mind: first

and foremost, I wished to identify and characterize the specific problem of estimating revision difficulty from a broader perspective; secondly, I wanted to clarify the theoretical framework which inspired and guided the development of the experimental method that will be described in Chapter 3.

Chapter 3

The Experiments and the Work at BUP

3.1 Introduction

The general aim of the experiments on which this chapter focuses is to investigate the possibility of using round-trip machine translation (RTT) (see Section 2.5) for estimating the revision difficulty of texts written in English as a lingua franca (see Section 2.2) by non-native writers within the context of the academic writing publishing industry. In particular, I would like to investigate the effectiveness of RTT employed as a method to evaluate the difficulty of revision of ELF manuscript through the automatic creation of a similarity reference.

The theoretical background as well as a summary of the main revision practices and techniques have already been presented in Chapter 2. Instead, Section 3.2 contains specific information about BUP, including the work of revisors and the traineeship I did there.

Section 3.3 is dedicated to the description of the two dataset that were compiled; one was used for preliminary experiments to refine the overall structure, whilst the second dataset featured also the evaluation and amount of revisions done by the revisor at

BUP.

The results of the experiments described in Section 3.4 and detailed graphical representations can be found in the repository named 'dot-rtt' on the Internet hosting platform GitHub[44].

In order to verify the possibility of evaluating the quality of a text before it is subjected to linguistic revision through a comparison between the text itself and its round-trip translation, the experiments have been designed and carried out using software available in the public domain. For the experiments to be performed, data were collected through the creation of a corpus of aligned comparable texts, as will be described in Subsection 3.2.1 and Section 3.3. Then, a Machine Translator (MT) that would guarantee adequate performances was chosen. RTT were produced both for the manuscripts and their revisions, so that two datasets were compiled. The first one, namely IE (see Subsection 3.3.1), was used for preliminary experiments to refine the overall structure of the experimental procedure. This stage highlighted the necessity of having an accurate pre-evaluation of the manuscript in order to understand the results of the experiments. For this reason, the second dataset (see Subsection 3.3.2) features also a general assessment and the number of revisions done by the revisor at BUP.

The starting hypotheses, that my experiments can prove or disprove, are:

1. the higher the similarity score between a raw text and its revision is, the higher the quality of the raw text is;

2. the score indicating the similarity between the revised version (Rev) and its roundtrip translation (RevRTT) should be higher than the score between the manuscript (Raw) and its roundtrip translation (RawRTT). Hence, I expect the difference between the two scores to be of a higher value the more a text needs to be revised;

3. the similarity between RawRTT and Rev should be higher than the similarity between Raw and Rev. This would mean that the RTT produces an improved

[44] https://github.com/TinfFoil/dot-rtt

version of the source text. If this is true, the output of the roundtrip translation could serve as a good starting point for revision.

These hypotheses are behind the design of my experiments.

In the following sections the dataset and the experiments are described. For what concerns the RTT method and the two texts similarity metrics – namely, BLEU and BERTScore –, they have already been presented in Chapter 2. This work moves from the assumption that automatic machine translation is of a quality that permits such an application of these systems (see Subsection 2.5.1).

Some preliminary considerations need to be made before attempting to solve the questions described above. First of all, evaluating a written production is never an easy – or completely objective – task, as it has been discussed in Subsection 2.3.4. Then, different styles of revision need to be acknowledged (see Section 2.3). Moreover, it should be noted that while automatic detection of grammatical errors can be solved by existing NLP systems (see Subsection 2.4.2), it is still hard to quantify the need for revision and this is partially due to the fact that the time and degree of expertise of the revisor are influential factors. Nevertheless, the time a revisor normally dedicates to the pre-evaluation of the manuscript at hand, before even starting the revision process, is long and sometimes remains unpaid, while being this process essential to identify the time needed and the right compensation for the revision work (see Section 2.3).

3.2 Revising at BUP

Bologna University Press, formerly named Bononia University Press, was founded in 1998 as the publishing brand of the University of Bologna. The brand has increasingly grown since its creation, and this year it has become a foundation counting 11 full-time employees, plus a network of freelance translators and collaborators. BUP currently publishes around 90 titles a year and has a turnover of around one and a half million euros. The new and functional headquarters in Via Saragozza, 10 are equipped with

the latest information technology, as can be read on their website[45]. The entire BUP catalogue (more than 1200 titles) is available on the foundation's website and on Amazon, Ibs, Unilibro, Libro Co and other online bookshops. In addition, BUP books can also be purchased in digital version on Apple and Amazon platforms and read directly from devices such as iPad or Kindle.

The long-standing relationship with international retailers specialized in supplying libraries and research institutions has enabled BUP to establish strong business relations with the world's leading libraries, which are updated and supplied with the latest products on a monthly basis. The international outreach of the business and its rapid growth made BUP participate in the 2019 Frankfurter Buchmesse, one of the biggest and most prestigious book fairs in the world[46].

While the high standards of research quality are guaranteed by the scientific committee, composed of many outstanding professors of the University of Bologna, the linguistic standards, for what concerns English publications, are kept high also thanks to the close relationship BUP has with the Department of Interpretation and Translation (DIT) of the University of Bologna[47]. Having done two traineeships at the publishing house, I have had the possibility of training both in translation and revision. Moreover, theoretical training was provided during the traineeship: trainees were given the chance to meet with the chief-editors and graphic experts working at BUP, getting some precious insights on the publishing world.

During the last two years that I have personally known BUP, their approach to revision practices has radically changed. Indeed, Xiaoli Wang, one of the trainees working at BUP in the academic year 2020/ 2021, wrote her dissertation on the topic of revision, reporting the techniques implied by three different revisors (with different degrees of expertise). The text she analyzed was precisely 'The Italian Economy After Covid-19' (Bellettini and Goldstein, 2020), from which the first dataset used in the experiments

[45] buponline.com

[46] https://magazine.unibo.it/archivio/2019/10/14/bononia-university-press-vola-alla-fiera-del-libro-di-francoforte

[47] https://dit.unibo.it/it

on which this dissertation is focused were extracted (see Subsection 3.3.1). The study by Wang (2021) had highlighted some issues which have been addressed by BUP – as witnessed by the revision guides and documents made available for revisors and clients as well as by the constant interest in research on this matter.

3.2.1 The Database: Building the Monolingual English Parallel Corpus of Academic Writings and Revisions

The compilation of the Monolingual English Parallel Corpus of Academic Writings and Revisions (EMPCAWR) has been inspired by the procedures and methods illustrated provided by the volume Corpus Linguistics for Contrastive Studies (Mikhailov & Cooper, 2016), as well as by the materials made available by the professors at DIT in the last two years, particularly in the courses 'Terminology and Information Mining' and 'Corpus Linguistics' (a.a. 2019/2020).

The necessity of building a corpus of parallel manuscripts and final versions of the same texts arose for two main reasons. First, in the frame of the DOT project at BUP this corpus can be a meaningful resource for the work of revisors and trainees constituting a sort of 'revision memory'. In this sense, the corpus can be used both for analyzing the methods behind revision practices so to improve revisors' performances and for speeding up the process of revision exploiting previous solutions to recurring problematic patterns and errors. The existence of recurring errors is assumed, particularly in relation to errors caused by the interference of the authors' native language (see Subsection 2.2.2). Secondly, this corpus has been created for research purposes in the field of automatic evaluation of academic writing – namely as the starting point for compiling the two datasets used in the experiments that I will describe shortly.

All the texts included in the corpus were collected directly from BUP archives. I intentionally decided to exclude translations and revisions of translations in the need of defining the limits of my study. Moreover, revising a text firstly conceived in English as a lingua franca implies practices that might differ – from a small to a greater extent – from those involved when revising a translation (see Section 2.3).

At the beginning, I wanted to consider only texts written by Italian writers. The influence of a writer's native language on every level of linguistic production, from the choice of vocabulary to prosody – not to mention the recurrency of certain errors depending on the native language of the writer – is proved by numerous studies (among others, Yang, 2001; Jarvis & Pavlenko, 2008; Galvao, 2009; Lorés Sanz, 2020; see Subsection 2.2.2). However, I soon had to face the problem of having too little data at hand. For this reason, I decided to include documents written by non-native English speakers of different nationalities. A possible solution to this problem, which has not been applied yet, could be to encode the information about the author's nationality of each text as metadata.

I uploaded the corpus on SketchEngine, after I checked that privacy measures where appropriate. SketchEngine[48] is an online concordancer and a text-analysis software developed in 2003 by Lexical Computing Limited. It is particularly useful to quickly get information about a corpus. Plus, SketchEngine provides automatic PoS-tagging thanks to the English 3.3 Treetagger Pipeline V2 and the possibility of annotating metadata about what are defined on the software as 'text typologies'. For example, in this corpus I inserted the tag 'raw version' and 'final version'. In this way, I had the possibility to filter the results of searches according to this criterion.

Moreover, I also have commented revisions and other .docx files that attest the revision procedure for each document. These files are not included in the corpus because they could not be transformed into .txt. They were used in the second round of experiments, in order to count the changes made by the revisors and use them as a parameter for the pre-categorization of the dataset, as will be seen in Subsection 3.3.2.

Unfortunately, the corpus cannot be made available to the public, as it contains data protected by copyright and some texts that had not been even published at the time the corpus was compiled. Information about the corpus is shown in Table 2. The names in italics refer to subsets of the precedent sub-corpus.

During the compilation of the corpus, I kept a report of the construction procedure

[48] Sketchengine.eu

Table 2: Information about the Monolingual English Parallel Corpus of Academic Writings and Revisions

Corpus and sub-corpora	**Number of Texts**	**Number of Tokens**
EMPCAWR	332	802,070
ArchAlp_N6	32	6,942
ArchAlp-raw	16	3,407
ArchAlp-revisions	16	3,535
CIHA	238	441,233
CIHA-*raw*	119	218,832
CIHA-*revisions*	119	222,401
EU Disaster Response Law	2	149,050
EU-*raw*	1	74,776
EU-*revisions*	1	74,274
The Italian Economy after COVID-19	60	204,845
Italian-*raw*	30	102,402
Italian-*revisions*	30	102,443

which can be found in the 'readme.txt' file together with some more information about genre and authors of the texts (see Appendix A). Here I will only briefly summarize the main steps of the procedure:

- conversion in UTF-8 character encoding and .txt files;
- cleaning from anything that could cause noise, such as numeric values contained in tables and information about the authors;
- manual alignment using regular expressions.

It should be noted that no .csv or any other format of unified file is available at the moment because it would have been of no use for the present study, but it can be easily compiled using the already aligned files.

From this corpus of texts, the files to build the two datasets used in the experiments were extracted. These files were all double-checked for mistakes that could have been made during cleaning and aligning procedures.

3.3 The Datasets

I set up two different datasets for the experiments. The first, which I named IE (Italian Economy), is made up of texts on socio-economic matters; it was used for preliminary experiments aimed at identifying the most suitable MT system for RTT, choosing the best algorithm between the two considered (BLEU and BERTscore) and for tuning the method (Bellettini and Goldstein, 2020). The second one, named CIHA (Congress of the International Committee of the History of Arts), features texts on the topic of history of art (Faietti & Wolf, 2019). It is accompanied by detailed information on the revision, therefore it was used for experiments, which also involved a comparison with revision difficulty indexes obtained from the evaluations and corrections of the human revisor. Furthermore, being on a clearly different topic than that of dataset 1, the second dataset was used to verify that the results obtained with the preliminary experiments (for example, those aimed at choosing the best MT and the best similarity measurement algorithm) are generalizable and not dependent on the dataset.

3.3.1 Dataset 1 – IE

The dataset 1 includes texts from the volume 'The Italian Economy after COVID-19: Short-term Costs and Long-term Adjustment' (Bellettini and Goldstein, 2020), which was published by BUP in 2020 after being revised by a group of revisors and trainees. This volume was the focus of the dissertation by Xiaoli Wang (2021), which focused on revision practices and highlighted the need for a faster and more objective evaluation of the manuscripts. It was initially divided into three sets, each containing texts revised by the same revisor (A, B and C, respectively). For every set of texts, five folders were created: 'Raw', the manuscripts written in English by non-native writers, 'Raw_RTT', revised, that is, the final versions ready for publication after revision, 'Revised_RTT' and 'Revisions with comments, where the .docx files with all the revisions and comments are stored.

The procedures described in what follows were all carried out using Sublime Text[49]. Texts were converted in Unicode Transformation Format, 8 bit (UTF-8). Then, they were divided into segments. Each of the files has been organized in segments of maximum 500 words delimited by the end-of-line character '\n', they were cleaned from any possible noise (e.g., tables, mathematical formulae). In order to be compared, the data also needed to be aligned. For this task, I manually aligned the texts using regular expressions and the function 'replace'. '\.[ ]' was replaced with '\.\n'; '\n\n' was replaced with '\n' because BLEU and BERTScore would provide a score of 0 if candidate and reference segments are empty; occurrences of '[0-9]+\.\n[0-9]+' (i.e. numeric values) were found and '\n' was deleted. Then, the length of segments was checked because BLEU and BERTScore both work only with segments of maximum 500 characters. Every of this passage has been double-checked. Finally, the RTT were produced. As machine translators I tested DeepL (En(British)-De), ModernMT (En-Es) and the MT system provided by Google (En-Es), which are largely used and can be considered as the state-of-the-art MT systems (see Section 2.5).

Each text has an identification number. The files in 'raw' are named with the identification number followed by an underscore '_' and a word taken from the name of the chapter. The files in 'raw_RTT' are named with the identification number followed by 'RTT_' and the same word in the title of the respective raw. The files in 'revisions' are named with the identification number followed by '.' and the same word in the title of the respective raw. The files in 'revisions_RTT' are named with the identification number followed by 'RTT.' and and the same word in the title of the respective raw.

Table 3 shows the main features of Dataset 1: the root of the file names, the number of segments for each file and the number of words. The table does not show the parameters of the raw RTT and revised RTT texts, in which the number of segments equals the number of segments of the source file, whereas the number of words may differ from that of the source file. Table 4 shows one example from the corpus. More specifically, the first column shows the Raw text; the second column shows the RTT

[49] https://www.sublimetext.com/

of the Raw text; the Rev text is in the third column while its RTT is reported in the fourth column. Some more examples of randomly selected parallel segments from the corpus can be found in Appendix B.

Table 3: Files contained in IE

Text ID	**Number of segments**	**Number of words - Raw**	**Number of words - Revised**
0 introduction	83	2,468	2,470
1 pandemic	204	5,529	5,509
1a figures	13	263	237
2 demography	160	4,138	4,168
10 banking	156	3,939	3,946
13 trade	163	4,539	4,596
17 why	270	5,998	6,087
117 preface	47	965	965

Table 4: IE Dataset – Example of one segment

Text ID	**Raw**	**Raw RTT**	**Revision**	**Revision RTT**
0 introduction	In mid-March, it was almost a cliché to criticize the Italian approach as alarmist and excessive;	By mid-March, it was almost a cliché to criticize the Italian action as alarmist and excessive;	In mid-March, it was almost a cliché to criticize the Italian approach as alarmist and excessive;	By mid-March, it was almost a cliché to criticize the Italian action as alarmist and excessive;

3.3.2 Dataset 2 – CIHA

The second dataset was compiled in a similar way as the first one. Dataset 2 consists of texts taken from the volume 'Motion: Transformation' (Faietti & Wolf, 2019) which includes the proceedings of XXXV Congress of the International Committee of the History of Arts. The dataset features 14 manuscripts (Raw) organized in segments of maximum 500 words, and 14 revised texts (Rev). In order to allow a correct comparison between the texts and their RTT the whole dataset was manually aligned, following the same procedures employed for the preparation of IE dataset (see Subsection 3.3.1). In this case, I decided to collect information that would allow for a pre-categorization of the texts according to the number of changes that the manuscripts have undergone and a general retrospectiv evaluation provided by the BUP revisor. The revisor was

asked to differentiate the texts into three groups according to the difficulty experienced during the revision process following these criteria:

- group 1 includes texts which only needed proofreading and light post-editing, such as checking grammar, spelling, and stylistic consistency. Manuscripts assigned to this group were judged as being overall well-structured, featuring an appropriate use of language;
- group 2 is composed of texts for which the revision required a fairly long time, but no particular effort for the revisor because, although there were many grammatical errors, the syntactic structure of the text was overall clear. General use of language and vocabulary needed some revision;
- group 3 features texts for which the revision required a long time and a high mental effort, because the manuscripts were grammatically incorrect, syntactically ambiguous, they presented evident interferences of the author's mother-tongue, inadequate use of linguistic structures and vocabulary.

Table 5 reports the main characteristics of the CIHA dataset. For each text, the following information is reported: number of segments, number of words, number of revisions (both raw and normalized according to the number of words and segments), the evaluation given by the revisor (on a scale from 1 to 3, where 1 means that a text is very easy to revise, 2 means that a text needs some revision, 3 refers to texts that are very hard to revise).

I referred to the categorization 1-3 as a subjective qualitative revision difficulty index (IRev_S), i.e. estimated by the human revisor based on their own experience. Furthermore, using the statistics of the tracked changes in the revised files, I obtained the average of the corrections per segment and the average of the corrections per number of words for each text. I normalized these values to 3, to obtain a scale consistent with the 3 groups identified by the human revisor. Then, I calculated the average between the normalized corrections per segment and the normalized corrections per number of words, obtaining a parameter that I called subjective quantitative revision difficulty

Table 5: Files contained in CIHA

Text ID	**Number of segments**	**Number of words**	**Number of revisions**	**Revisions per segment**	**Revisions per word**	**Evaluation**
118 emigrants	93	2017	623	6.6	0.308	3
119 kakezukuri	70	1887	391	5.5	0.207	3
84 orient	95	2418	502	5.2	0.207	3
88 symbolism	55	1039	180	3.2	0.173	3
65 small	77	2669	396	5.1	0.148	3
22 introsess1	103	3169	441	4.2	0.139	2
23 painted	163	4368	887	5.4	0.203	3
41 between	103	2481	404	3.9	0.162	2
56 written	121	2547	221	1.8	0.086	2
51 visual	102	2201	27	3.7	0.012	1
47 parer	102	2748	140	1.3	0.050	1
26 trinity	76	1788	129	1.6	0.072	1
38 colony	82	2012	147	1.8	0.073	1
44 archiving	85	3012	42	2	0.013	1

index (IRev_Q). Finally, I calculated the average between IRev_S and IRev_Q for each text, obtaining a human combined revision difficulty index (IRev_H), which represents a reference parameter to evaluate the effectiveness of the method tested in the experiments, as shown in Table 6 (the qualitative index of revision difficulty is reported in the second column, the normalized human quantitative index of revision difficulty is displayed in the third column and the average of the two indices is shown in the last column).

In Figure 1, a representation of the texts ordered according to the human evaluation of revision difficulty is displayed. This classification is compared with the one resulting from the experiments with round-trip translation in Section 3.5.

Table 6: Files contained in CIHA with their respective human evaluation indices

Text ID	**IRev_S (Revisor's qualitative index)**	**IRev_Q (Revisor's quantitative index)**	**IRev_H (Human index of revision difficulty)**
44 archiving	1	0.518	0.759
47 parer	1	0.539	0.769
26 trinity	1	0.714	0.857
38 colony	1	0.762	0.881
51 visual	1	0.899	0.950
56 written	2	0.828	1.414
22 introsess1	2	1.632	1.816
41 between	2	1.675	1.838
88 symbolism	3	1.570	2.285
65 small	3	1.880	2.440
84 orient	3	2.190	2.595
23 painted	3	2.216	2.608
119 kakezukuri	3	2.258	2.629
118 emigrants	3	3.000	3.000

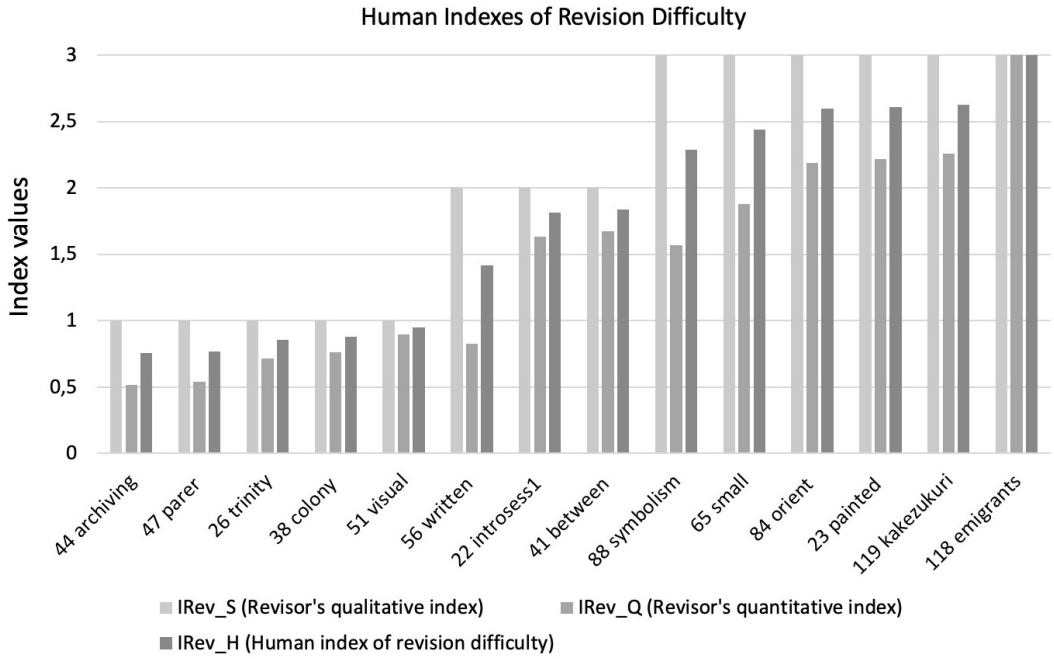

Figure 1: Comparison between revision difficulty indices constructed on the basis of human evaluation.

3.4 The Experiments

3.4.1 Preliminary Experiments

In the experiments with Dataset 1 I compared the round-trip translation offered by three different automatic translators, DeepL, ModernMT and Google translator, to perform the round-trip translation, and two different algorithms to measure text similarity, i.e. BLEU and BERT (see Subsection 2.4.3).

In this phase, I generated the RTTs of the Raw texts with the three aforementioned automatic translators, which offer user-friendly interfaces and are some of the most advanced and best-performing MT systems available to the public (see Section 2.5).

Then, I calculated the similarity, or score, between Raw texts and their RTT using the BLEU and BERTscore algorithms. I adapted the codes I downloaded from the GitHub platform to the specific needs of the experiments using Google's Colab (see 'dot-rtt' on GitHub)[50].

In any case, at this stage I was able to verify that DeepL, ModernMT and Google are all suitable MT systems for the purposes of this study and that the best method to measure the similarity between texts (with reference to the analysis proposed in this thesis) is BLEU, both because it is less sensitive to the context, thus highlighting even small differences between the original files and their RTTs, and because it is time-saving and convenient in terms of computational power. This first set of experiments also made me realize that I needed more detailed information on human evaluation of the revision difficulty.

Indeed, while Wang's study provided precious theoretical insights and references to the techniques put in place by the revisors, one of the problems she outlined was precisely the lack of an objective evaluation about the quality of the manuscripts. So, I built a

[50] I downloaded the code for the BLEU algorithm from the following URL: https://colab.research.google.com/github/gcunhase/NLPMetrics/blob/master/notebooks/-bleu.ipynb, and the code of BERTscore algorithm from https://github.com/Tiiiger/bert_score.git

second dataset, namely CIHA, on which I could have a direct confrontation with the revisor (who is, unlike in the case of IE, an experienced in-house revisor at BUP). The experiments that I will describe in the next subsections were therefore designed on the CIHA dataset in order to verify the hypotheses described in the introduction to this chapter. For the sake of conciseness, I omit here the detailed description of the first cycle of experiments conducted with Dataset 1; however, the results are available on GitHub.

3.4.2 Experiment 1: Comparison between the BLEU and the BERT Scores

After the set of preliminary experiments conducted using Dataset 1, described in the previous subsection, I designed a series of specific experiments that I carried ou using the CIHA Dataset. In these experiments, first of all I repeated the comparison between BLEU and BERTScore by applying both methods to the pairs of Raw and Raw-RTT texts. Even if this comparison had already been executed in the preliminary experiments, as I have explained in 3.4.1, I needed to be sure that the results were not data-dependent.

In order to perform Experiment 1, I produced round-trip translations of the Raw files. The MT system used for this task is Modern MT; as language pair I used EN-ES (English and Spanish). These decisions were made considering the reports about the state of the art in machine translation (see Subsection 2.5.1). In particular, the Intento Report (2021) shows that ModernMT is among the ones which outperform competitors in the field of humanities, education and in the general domain, with respect to the chosen language pair. The obtained scores are shown in Table 7 and Figure 2.

In order to make the two scores more comparable I rescaled BERTScore values on a 0 to 1 baseline. It should be noted that the BERTScore values were always higher and less differentiated than the BLEU one. This happens thanks to the aptitude of the BERT algorithm to capture the context of the text, since it works at the semantic level.

Table 7: BLEU and BERT scores between the raw texts and their Round-Trip Translation

Text ID	**Raw-RTT_BLEU**	**Raw-RTT_Bert**
44 archiving	0.662	0.839
47 parer	0.777	0.915
26 trinity	0.649	0.871
38 colony	0.756	0.914
51 visual	0.617	0.859
56 written	0.668	0.896
22 introsess1	0.648	0.856
41 between	0.668	0.882
88 symbolism	0.646	0.877
65 small	0.592	0.829
84 orient	0.585	0.860
23 painted	0.685	0.868
119 kakezukuri	0.638	0.885
118 emigrants	0.588	0.814

The comparison between the results of these experiments confirms that BLEU is more suitable for the application here discussed, compared to BERTScore, because it is more sensitive to differences. This result should be regarded as positive in the perspective of future practical applications of the method proposed in the present work. Indeed, the implementation of BERTScore requires about 5 minutes for each text, whereas BLEU immediately produces the scores, thus eliminating waiting times. From this point on I have used BLEU only.

3.4.3 Experiment 2: Calculating the BLEU Similarity Score between Raw and Rev

To investigate hypothesis 1, i.e. that the higher the similarity score between raw text and its revision, the easier the revision, I calculated the BLEU score between Raw texts and their respective revisions (Rev), as present in the CIHA dataset. This score has been inverted (using the formula 1 – BLEU score) and normalized on a scale from 0 to 3 in order to be comparable with the indices deriving from the human evaluation; the resulting values are called Automatic Index of Revision Difficulty (IRev_A). The results

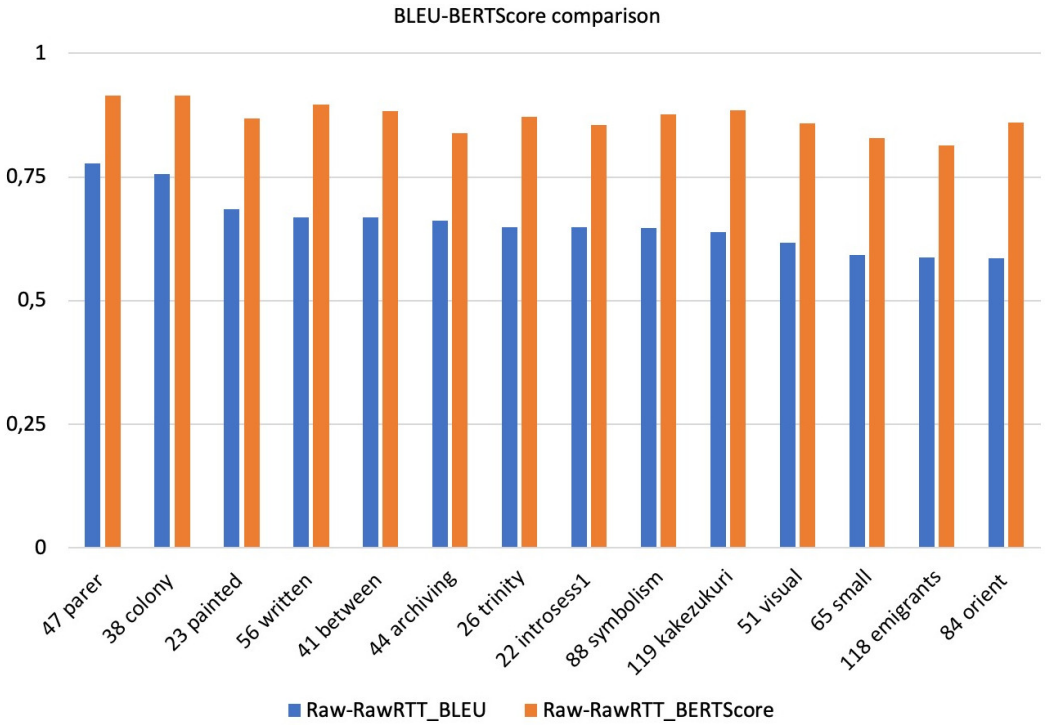

Figure 2: Comparison between the scores obtained using BLEU and BERTScore.

of this experiment confirm the starting hypothesis. Table 8 shows the scores obtained for each text and the automatic index of revision difficulty IRev_A (last column).

Table 8: BLEU score between Raw texts and their Revision. In the last column, values have been inverted and normalized on a scale between 0 to 3

Text ID	**Raw-Rev_BLEU**	**IRev_A**
51 visual	0.980	0.176
44 archiving	0.872	0.344
56 written	0.961	0.600
47 parer	0.932	0.935
38 colony	0.888	0.988
26 trinity	0.894	1.129
41 between	0.742	1.588
88 symbolism	0.685	2.038
84 orient	0.769	2.038
65 small	0.820	2.276
23 painted	0.728	2.400
119 kakezukuri	0.702	2.629
22 introsess1	0.756	2.152
118 emigrants	0.660	3.000

3.4.4 Experiment 3: Comparison between the BLEU Similarity Scores of Raw-RawRTT and Rev-RevRTT

The experiment presented in this subsection is focused on verifying hypothesis 2, namely whether the similarity between a raw text written in English by a non-native writer and its round-trip translation is indicative of the difficulty – and, therefore, costs and time – of its revision. This would imply that the score between the revised version (Rev) and its roundtrip translation (Rev-RTT) should be higher than the score between the raw text (Raw) and its roundtrip translation (Raw-RTT).

For this purpose, I generated the round-trip translations of the revised versions of the texts included in CIHA and I calculated the BLEU scores between the revised texts and their RTTs. Then, I compared the scores obtained for these pairs of texts with the already calculated scores between the Raw texts and their RTTs. In order to make the scores comparable across the dataset, I have calculated the difference between Rev-RevRTT and Raw-RawRTT. This difference has been also inverted and rescaled on a range 0 to 3 to allow comparison with human evaluation (see Section 3.5). In this sense, the hypothesis here tested would be rephrased as follows: the higher this value is, the more a text is in need for revision. Table 9 shows this comparison. In Table 9 the fourth column shows the difference between the two scores displayed in the second and third columns; the normalized values referring to this difference are displayed in the last column. I have named this parameter, which represents the distance between the Raw-RawRTT and Rev-RevRTT score, as Discrepancy Index (Idx_Dis).

The results of my experiments show that the Idx_Dis correlates with the difficulty of revision.

Indeed, this experiment confirms that the similarity between the revised texts and their RTT is greater than the similarity between the raw texts and their RTT. This behavior of the score is especially evident for the raw texts 118 and 84, which were among the hardest to revise (see Table 6 and Figure 1) – a further discussion on the comparison between human evaluation and the values obtained in this experiment will be provided

Table 9: BLEU scores of Raw-RawRTT and Rev-RevRTT

Text ID	**Raw-RawRTT**	**Rev-RevRTT**	**Rev_RevRTT - Raw_RawRTT**	**Normalized Rev-RevRTT - Raw-RawRTT**
47 parer	0.777	0.733	-0.044	0.140
56 written	0.668	0.665	-0.003	1.101
51 visual	0.617	0.615	-0.002	1.125
44 archiving	0.662	0.661	-0.001	1.148
26 trinity	0.649	0.65	0.001	1.195
88 symbolism	0.646	0.658	0.012	1.453
38 colony	0.756	0.771	0.015	1.523
23 painted	0.685	0.705	0.02	1.640
41 between	0.668	0.689	0.021	1.664
65 small	0.592	0.617	0.025	1.757
119 kakezukuri	0.638	0.674	0.036	2.015
22 introsess1	0.648	0.687	0.039	2.085
84 orient	0.585	0.626	0.041	2.132
118 emigrants	0.588	0.666	0.078	3.000

in Section 3.5.

When the raw text is of a good linguistic quality, the difference between the scores calculated for the pairs Raw-RawRTT and Rev-RevRTT is very small or even negative, as in the case of the text '47 parer'. The difference between the score of the revised texts with their RTTs and the score of the raw texts with their RTTs shows that, for manuscripts which were evaluated positively from the revisor, our hypothesis is not always verified. In order to address this aspect, that is, to understand if the method was actually working and its responsiveness, I have designed an ad hoc experiment.

3.4.5 Ad Hoc Experiment on Text 47

First, I generated a sample text by taking the first 5000 characters of the "47_parer" text, which I named 47_0. Then, I created another sample text, named 47_1, deteriorating the text 47_0 by randomly deleting some words (24 words). In this process, I made sure that the number of segments did not change. I repeated this operation so as to have four more sample texts (from 47_2 to 47_5). These texts were gradually deteriorated from the linguistic point of view through the random deletion of words

(see Figure 3). Finally, I generated the RTTs of the 6 sample texts and I calculated the BLEU score between them, and the sample texts themselves. I obtained the results shown in Figure 3. As can be seen, the similarity between Raw texts and their RTTs systematically decreases according to text quality, thus confirming that the method and the measures employed in the experiments are valid.

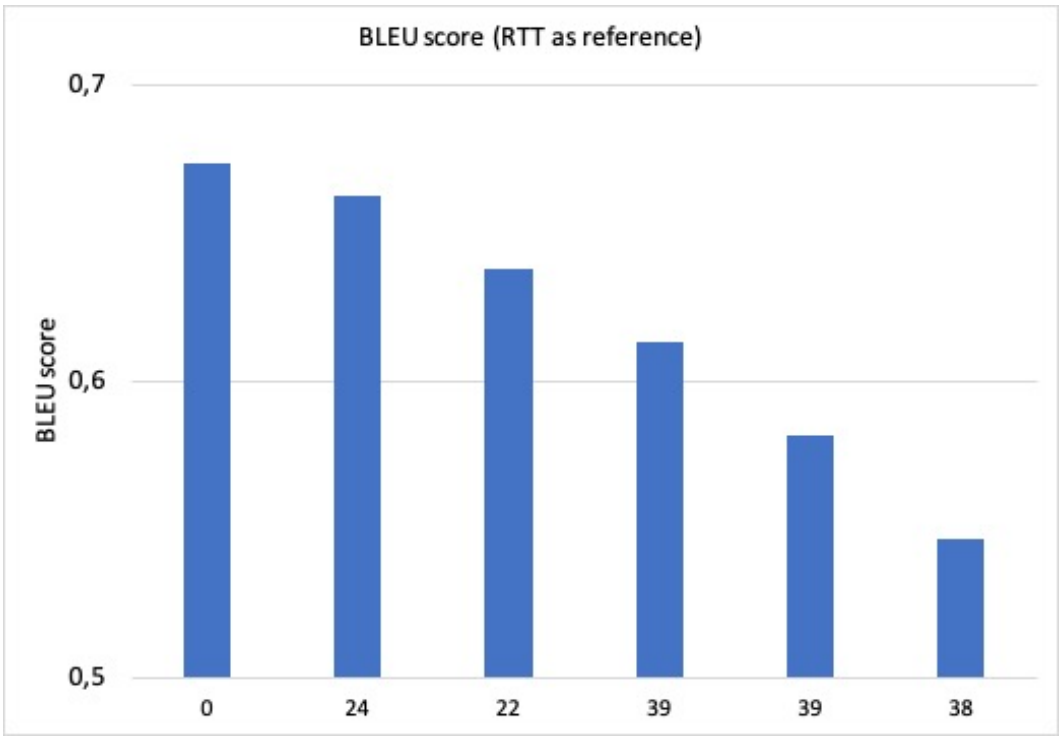

Figure 3: The BLEU scores between the sample texts with decreasing quality. The number of randomly deleted words is reported in the abscissa.

3.4.6 Experiment 4: Comparison between BLEU Similarity Scores of Raw-Rev and RawRTT-Rev

The fourth experiment that I carried out using the CIHA Dataset is aimed at testing hypothesis 3, which concerns the possibility producing an improved version of the source text through RTT. To verify this hypothesis, I calculated the similarity between RawRTT and Rev and compared it with the similarity between Raw and Rev, already calculated for the previous experiment. If the similarity between RawRTT and Rev is higher than the similarity between Raw and Rev, hypothesis 3 is confirmed, meaning

that the output of the roundtrip translation process can be considered a good starting point for revision. Table 10 reports BLEU scores between RawRTT and Revision, while Figure 4 shows the results of the above-described comparison.

Table 10: BLEU scores between RawRT and Revision for each text in CIHA

Text ID	**RawRTT-Rev_BLEU**
47 parer	0.777
56 written	0.670
51 visual	0.575
44 archiving	0.616
26 trinity	0.627
88 symbolism	0.505
38 colony	0.728
23 painted	0.592
41 between	0.567
65 small	0.538
119 kakezukuri	0.510
22 introsess1	0.566
84 orient	0.551
118 emigrants	0.512

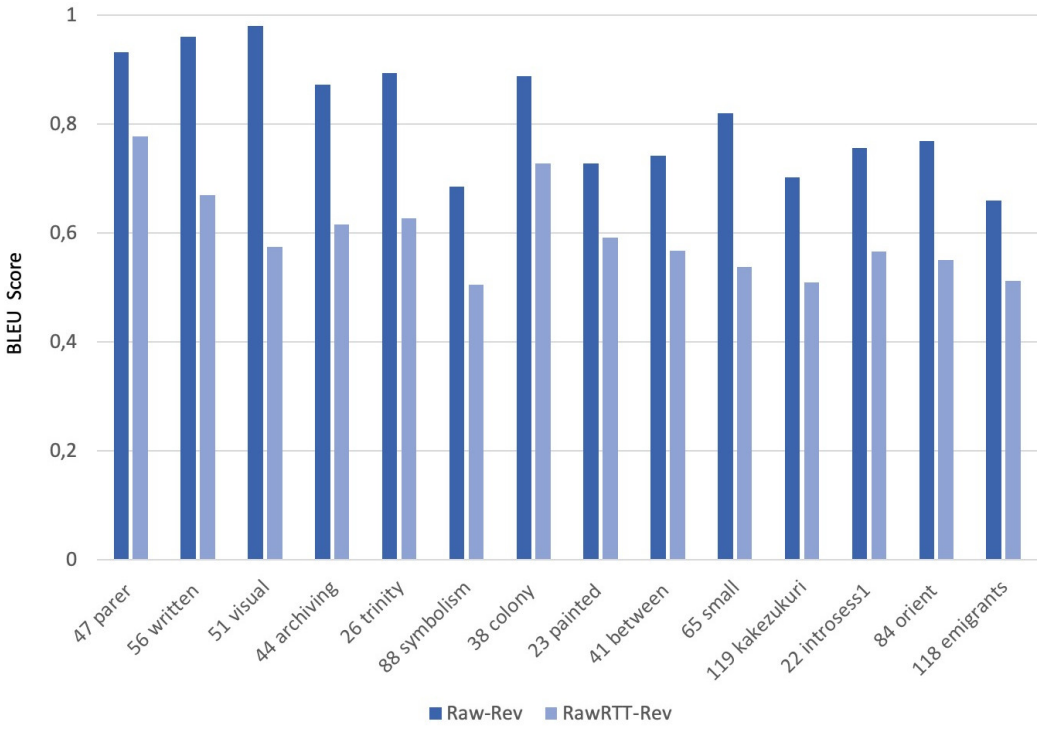

Figure 4: Comparison between Raw-Rev score (see Table 8) and RawRTT-Rev score.

In this case the experiments did not confirm hypothesis 3. The scores between the manuscripts (Raw) and their corresponding revisions (Rev) are systematically higher than the scores between the round-trip translations of the raw texts (RawRTT) and the revisions of the same texts (REV).

3.5 Discussion

The set of experiments performed with the CIHA Dataset highlights that the hypothesis behind this work are partially confirmed.

Experiment 1 confirmed that BLEU is the most suitable text similarity metrics for the aims of this study, as it had already been shown by the preliminary experiments. I decided to repeat the test with the second Dataset to be sure that metric performances did not depend on specific features of the IE Dataset.

Experiment 2 worked as a further demonstration that BLEU can measure the distance between two texts – the candidate, which in this case is the manuscript, and the reference, i.e. its revision. The hypothesis according to which the similarity between Raw and Revision increases with manuscripts which only need little revision is partially confirmed, as I will now explain.

Inverting and normalizing BLEU score values obtained in this experiment, I came up with the automatic indices of revision difficulty (IRev_A), which can be compared with the human evaluation. The two indexes are reported in Table 8 and Figure 1, respectively. In Figure 5, a comparison between the two indexes is displayed. The average between the two values is calculated in order to obtain the final index of revision difficulty (IRev_F). Figure 6 shows the comparison between the different revision difficulty indices (IRev_H, IRev_A and IRev_F) and the discrepancy index (Idx_Dis).

As can be noticed in both figures, the two categorizations do not perfectly overlap. In this respect, a number of elements must be taken into account. First of all, as much as the human evaluation was based on the judgements and revisions of an experienced BUP revisor, only one person was involved in the evaluation process, whereas studies

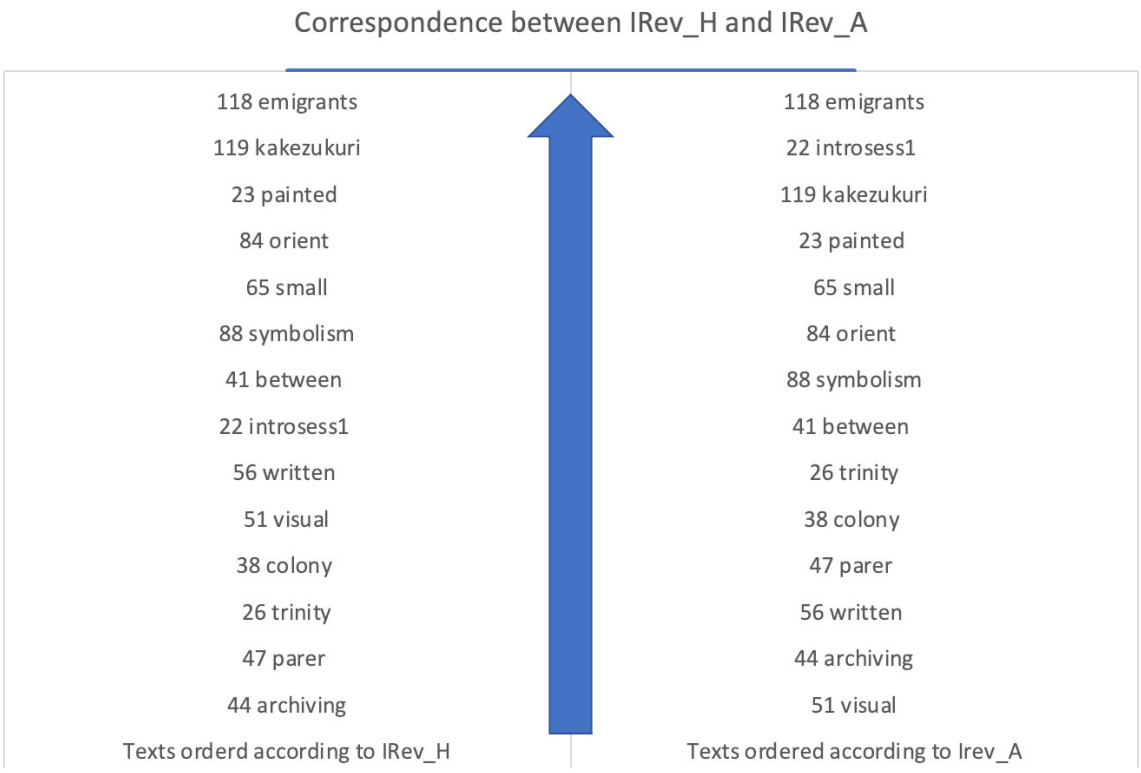

Figure 5: Comparison between texts ordered according to IRev_H and texts ordered according to IRev_A.

adopting human evaluation are normally run with more evaluators. Unfortunately, this was the only available revisor who could provide not only a judgement, but also the actual revision made on Microsoft Word with the option 'track changes' enabled. Moreover, in personal communication the revisor herself has reviewed the task of evaluating texts according to a holistic judgement as very difficult. In particular, she noticed that, while texts which needed major revision interventions could be easily separated from texts which features an elevated linguistic quality, it was particularly hard to assign texts to the intermediate group.

This leads to a second important observation, that is: the outcomes of the automatic comparison are more accurate with clear-cut cases, which include both texts that are very difficult to revise and texts where revision is hardly necessary.

Experiment 3 was designed to test the second hypothesis, as presented in the introduction to this chapter. It can be said that the hypothesis, according to which the

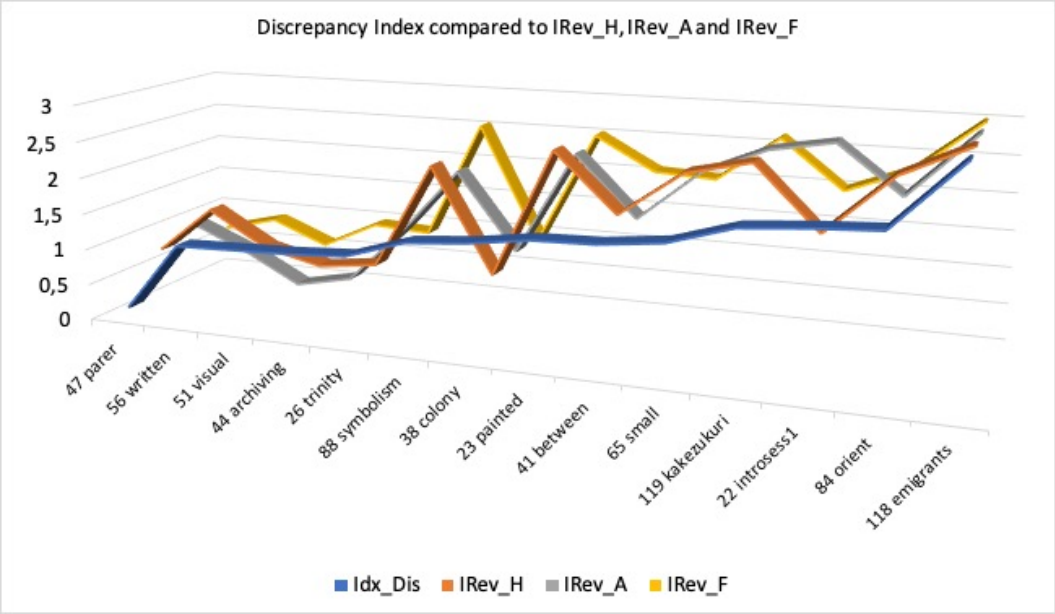

Figure 6: Comparison of discrepancy index (Idx_Dis), human revision difficulty index (IRev_H), automatic revision difficulty index (IRev_A) and final revision difficulty index (IRev_F).

difference between similarity scores of Rev-RevRTT minus Raw-RawRTT is higher when the manuscript is harder to revise, seems to be overall confirmed. Again, the main results of the experiment – the values representing the difference between Rev-RevRTT and Raw-RawRTT – were normalized on a scale from 0 to 3 (Idx_Dis), in order to be compared with the various indexes of revision difficulty.

For a better understanding of the results provided by the second experiment, it can be useful to compare them with the human index of revision difficulty – namely, the evaluation provided by the revisor and the number of revisions – as well as with the automatic index of revision difficulty – that is, the BLEU similarity between manuscripts and their revisions. Moreover, some additional considerations can be deduced from the comparison between the discrepancy index and the final index of revision difficulty (IRev_F), as displayed in Figures 6, 7 and 8. In this way, the accuracy of the results produced by the second experiment can be evaluated. It can be noticed that – with the only exception of text 88 and 23, the evaluation originated from the employment of round-trip translation (Idx_Dis) is overall consistent with IRev_F. This is consistent with the revisor's difficulty of evaluating texts with medium revision difficulty. Figure 7

shows the two values one against another. In theory, if the Idx_Dis would perfectly mirror the IRev_F, the data point should cluster along a straight line.

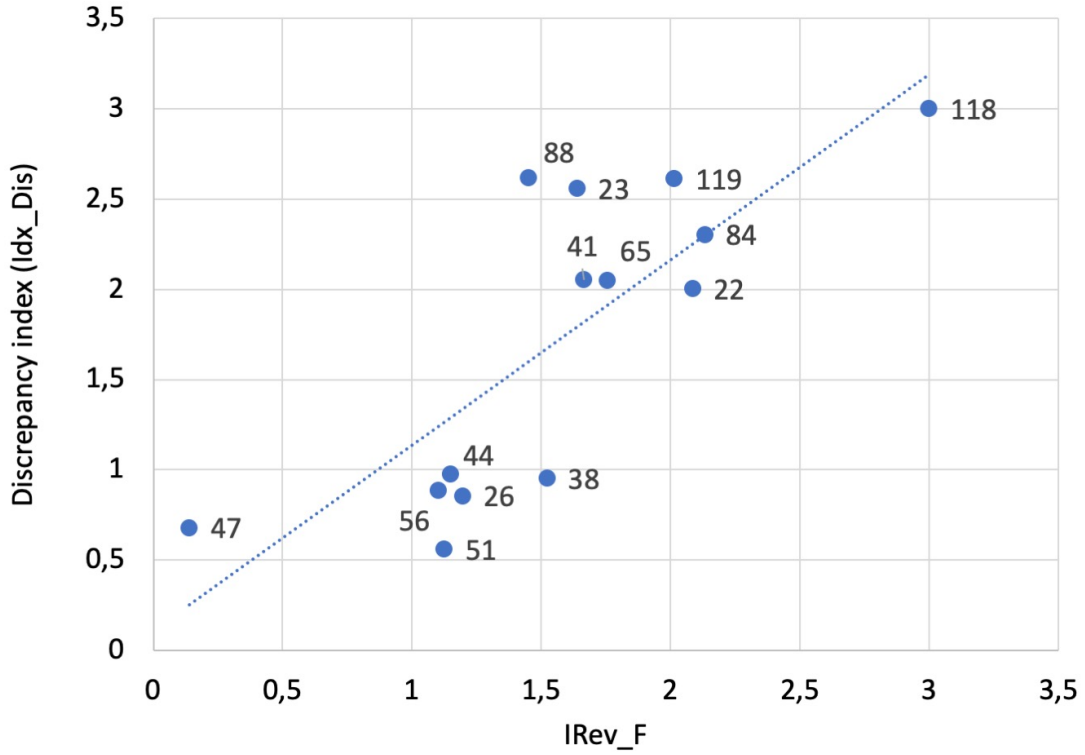

Figure 7: Discrepancy index against IRev F on the two Cartesian axes. The horizontal axis features IRev F, whilst the vertical axis features the discrepancy index.

Although this does not happen, the main difference occurs for texts that fall within the middle of the range, thus further demonstrating the idea that this method can eventually be applied for differentiating between texts which need a very difficult, time-consuming revision and texts which only need minor editing changes.

On the other hand, plotting the linear trends of the two values, IRev_F and Idx_Dis, it can be observed that the two lines are very similar, as shown in Figure 8. This indicates the consistency between the automatic method and the human evaluation.

For what concerns Experiment 4, a noteworthy aspect in Experiment 4 is that, comparing the scores between the Raw texts and their Revisions with the scores of the Raw texts with their RTTs (see Figure 9), it can be noted that the former are systematically

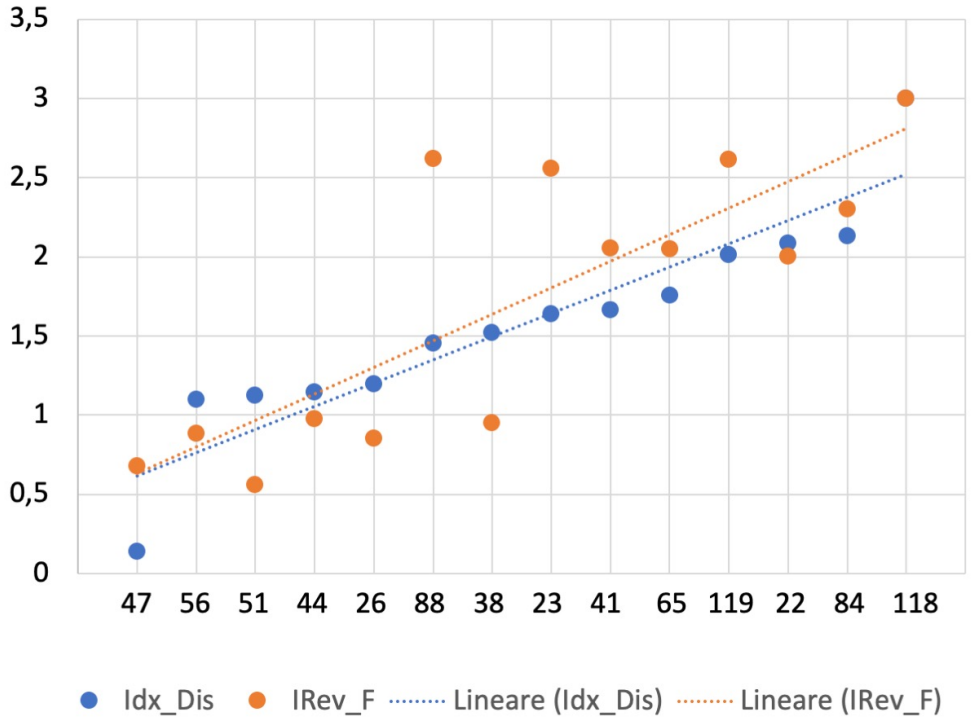

Figure 8: Comparison between the trend of the IRev F (blue) and the trend of the discrepancy index (orange).

higher than the latter.

I interpret this as a high quality of human revision compared to the quality of the RTT version of the Raw text. Thus, these results indicate that the revisions carried out by the expert revisor are closely related to the source texts, meaning that the authors' structural and lexical choices are respected as much as possible.

Basically, RTT cannot be used as a starting point for revision because there is no way to separate random changes (e.g. choice of synonyms) from those that actually improve the text. The human revision performed by an expert revisor achieves the effect of optimally enhancing the raw text while still remaining close to it. At present, this result still confirms a human primacy over automatic systems. For this reason, the scores between manuscripts and their revisions are always high even when revisions include major changes to the initial version.

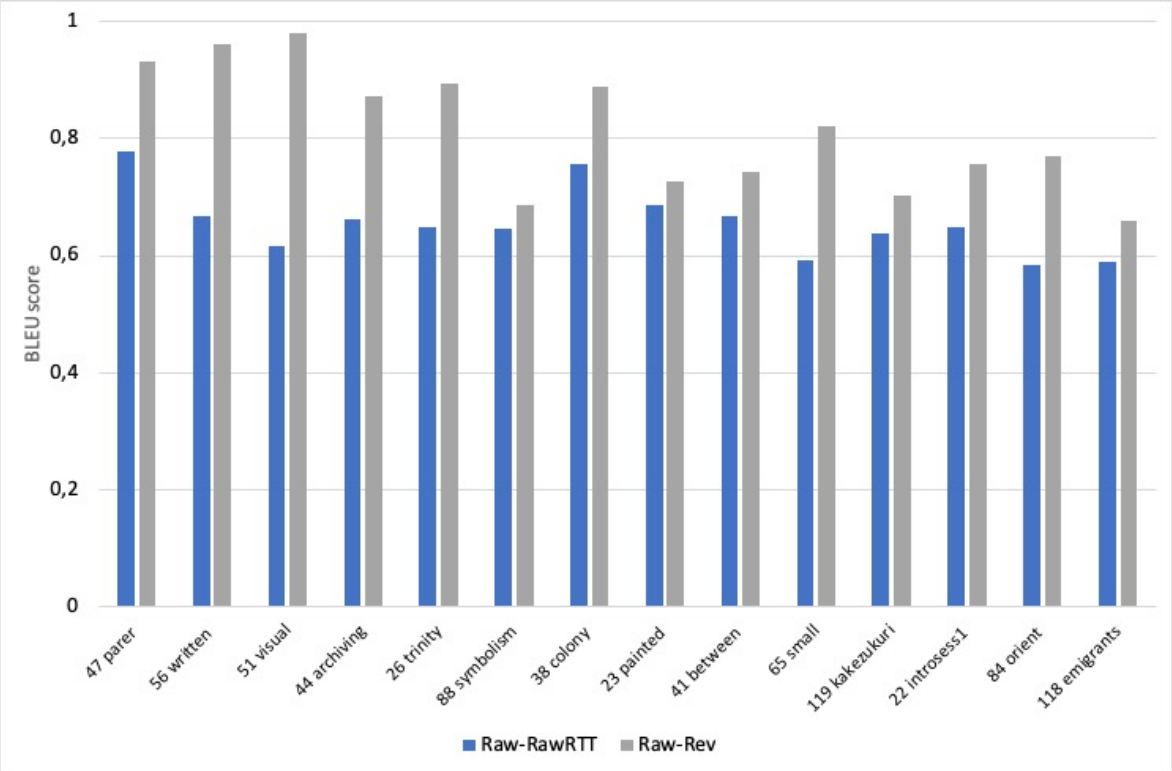

Figure 9: Comparison between BLEU scores of Raw-RawRTT (blue) and Raw-Rev (gray).

There is another relevant observation in this respect. Although the similarity scores between RawRTT and RevRTT have not been calculated, because they were regarded as purposeless for the aims of this work and the initial hypotheses, from a qualitative comparison between the outputs of the RTT process, it can be noticed that changes were applied to both raw texts and revisions regardless of the correctness of the source text. Table 11 provides an example of this phenomenon in text 38.

Table 11: Sample segments from text 38 colony

Raw 4, 18, 24	**RawRTT**	**Rev**	**Rev RTT**
The study starts with the Oath to Fernando VII, King of Spain, in 1809, and finishes examining some cultural events promoted by President Juan Rafael Mora Porras' government.	The study begins with the oath to Ferdinand VII, King of Spain, in 1809, and ends by examining some cultural events promoted by the government of President Juan Rafael Mora Porras.	The study starts with the Oath to Fernando VII, King of Spain, in 1809, and finishes by examining some cultural events promoted by President Juan Rafael Mora Porras' government.	The study begins with the oath to Ferdinand VII, King of Spain, in 1809, and ends by examining some cultural events promoted by the government of President Juan Rafael Mora Porras.

Continued on next page

Table 11 – *Continued from previous page*

Raw 4, 18, 24	**RawRTT**	**Rev**	**Rev RTT**
Then, they rang the church bells, many fireworks flew up and everyone shouted: ¡Long live the King don Fernando VII!	Then, the church bells ring, many fireworks fly and everyone shouts: Long live King Don Fernando VII!	Then, they rang the church bells, many fireworks flew up into the sky and everyone shouted: "Long live the King don Fernando VII!".	Then the church bells rang, many fireworks flew into the sky and everyone shouted, "Long live King Don Fernando VII!"
In this way, Costa Rican elites used popular cultural practices to convey to the people the official position of Spanish colonial rulers about contemporary European conflicts.	In this way, the Costa Rican elites used popular cultural practices to convey to the people the official position of the Spanish colonial rulers on contemporary European conflicts.	In this way, Costa Rican elites used popular cultural practices to convey to the people the official position of Spanish colonial rulers on contemporary European conflicts.	In this way, the Costa Rican elites used popular cultural practices to convey to the people the official position of the Spanish colonial rulers on contemporary European conflicts.

In other words, there is no way to ensure that changes deriving from the RTT method are motivated. On the other hand, this observation is evidence of a particular phenomen, which has been defined 'machine translationese' by Vanmassenhove et al. (2021). The authors 'hypothesize that the "algorithmic bias", i.e. an exacerbation of frequently observed patterns in combination with a loss of less frequent ones, not only exacerbates societal biases present in current datasets but could also lead to an artificially impoverished language' (ibid.: 1).

As a final summary of the outcomes of all the experiments, Figure 10a shows a complete picture of the results of the experiments. To better illustrate the comparison of the results obtained from the specific experiments, I have selected three sample texts: "38 colony", "22 introsess1" and "84 orient", belonging to the groups G1 (low revision difficuty), G2 (medium or unpredictable difficulty of revision) and G3 (high revision difficulty) of revisor subjective evaluation, respectively. These texts were chosen because they represent cases for which the hypotheses were confirmed. Figure 10b shows the details of the experimental parameters of these texts.

Figure 10b summarizes the trends of the different scores that I have calculated and shows their consistency with the fundamental hypothesis that inspired this work, namely that the score between a text and its RTT can be an indicator of the diffi-

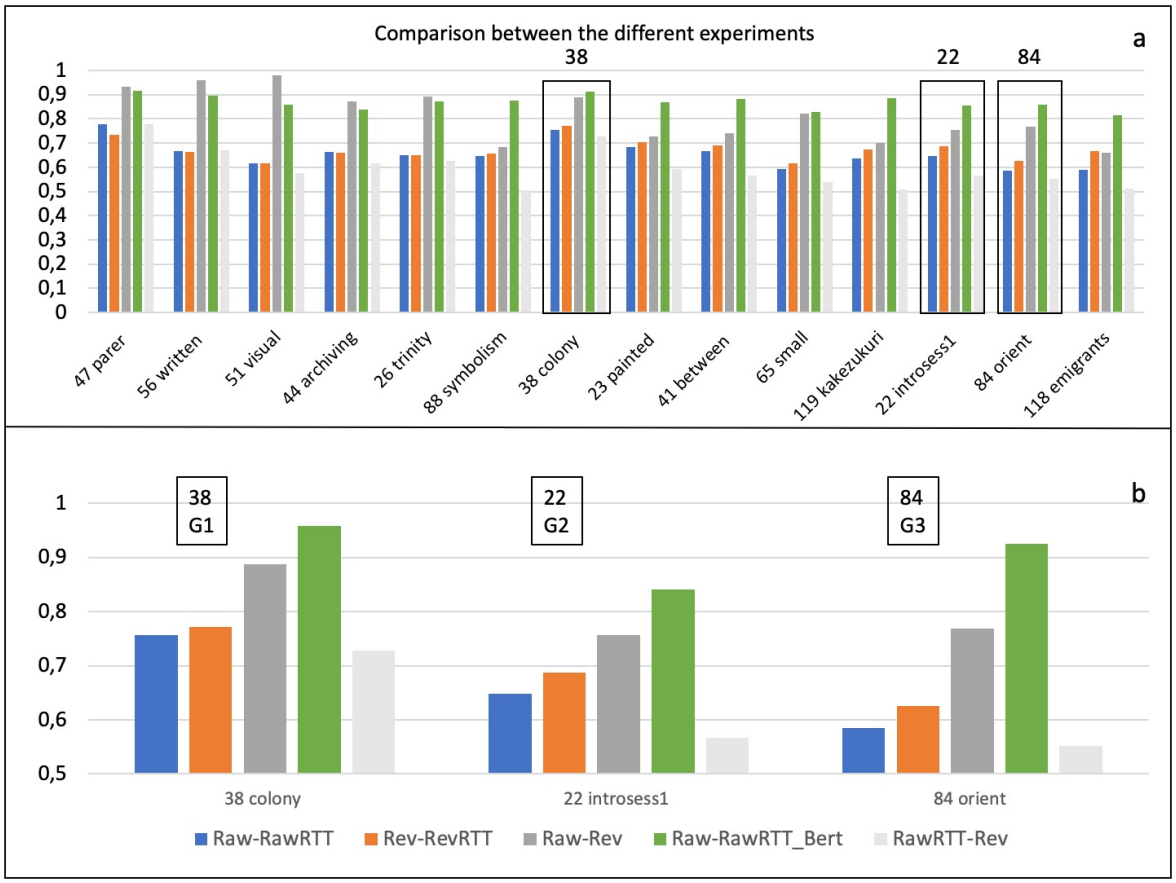

Figure 10: a) Comparison of the results of the different experiments. b) Detail of the results of the experiments on three sample texts with IRev_S 1, 2, and 3 (G1, G1 and G3).

culty of revision. It also highlights some interesting relationships between the source text (Raw), its RTT and its revised version. Consistently with the starting hypothesis, Figure 10b shows that Raw-RawRTT score (blue bars) regularly decreases passing from text 38, belonging to the first group, to text 84, belonging to the third group. On the other hand, considering the scores calculated with BERTscore (green bars), it can be noticed that these are always greater than BLEU scores and do not always correlate with the revision difficulty represented by the groups (G1, G2 and G3).

Furthermore, some considerations can be made on the general trends of scores for the different pairs of candidate/reference. As already mentioned, the Rev-RevRTT score (orange bars) is always higher than the corresponding Raw-RawRTT score (blue bars). Moreover, it also decreases slightly passing from the text of the first group (38) to the

text of the third group (84). This behavior could depend on some kind of influence of the linguistic adequacy level of the source text on the revision.

Finally, it should be noted that the scores between the RawRTT and Rev texts, which in the CIHA experiments are always lower than those between Raw and Rev are always the lowest, are sometimens higher than those between Raw and Rev in the preliminary experiments (see repository 'dot-rtt' on Github). This is also the reason why I decided to test hypothesis 3, which has been, in any case, disproved, since it is not possible to differentiate between necessary, superficial and random modifications (see, for example, the first raw in Table 11: 'starts' is unreasonably transformed in 'begins' in both RawRTT and RevRTT).

3.6 Comparison with Grammarly

Ultimately, to validate my results, I compared the scores obtained in the experiments with the scores that Grammarly, a commercial automatic writing assistant, produces (see Subsection 2.4.4 for more information on this tool). The comparison, presented in Figure 11, shows an agreement between the results of the RTT-based method, tested in the present work, and the score provided by Grammarly (the yellow bars in Figure 11). This result confirms the validity of the approach proposed in this thesis. Furthermore, it suggests that Grammarly can be used in conjunction with the present method for validating possible RTT/BLEU score thresholds, as will be discussed in the next chapter.

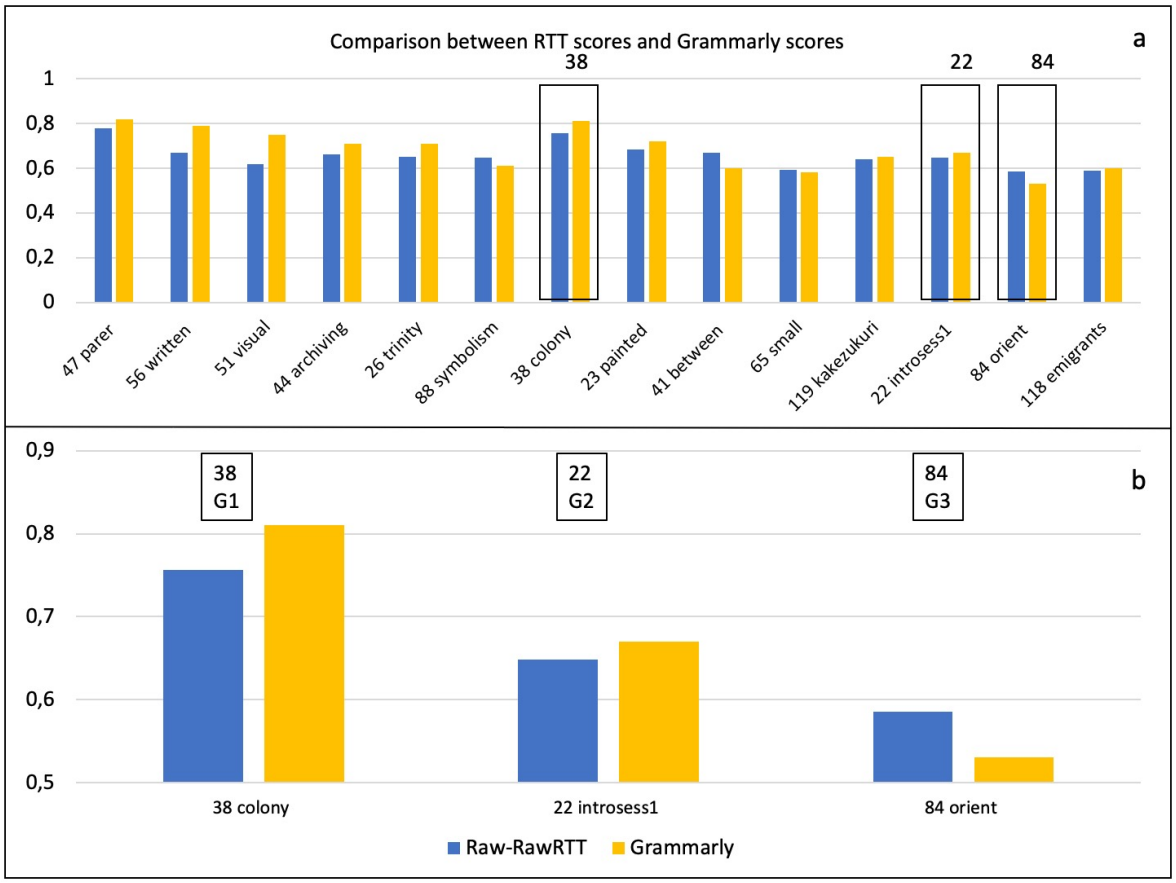

Figure 11: a) Comparison of the Raw-RawRTT score (blue bars) with the score provided by Grammarly (yellow bars). b) Details of the results of the experiments on three sample texts with IRev_S 1,2, and 3, compared with the Grammarly score (yellow bars).

3.7 Conclusion

It can be said that round-trip translation has proved to be able to provide an adequate similarity reference for the evaluation of revision difficulty. RTT can be easily produced automatically by the machine translator systems currently available online. Modern machine translators are generally based on artificial intelligence algorithms, which exploit artificial neural network systems, and are trained with corpora derived from the web (e.g., English Wikipedia) and partially from data provided by vast communities of users through the cloud. Thus, users indirectly contribute to the improvement of the training of automatic translators, and this could mean that the MT systems increasingly assimilate the characteristics of the most used varieties of English, among

which ELF can surely be placed. Therefore, the generation of a similarity reference through round-trip translation performed with these MT systems guarantees a relatively neutral reference with respect to the limits inherent to the concept of native English, as I have thoroughly discussed in the second section of Chapter 2. On the other hand, the experiments have shown that the similarity between the texts and their round-trip translation, measured with a simple method such as BLEU, correlates with the difficulty of revision.

However, Experiment 4 shows some limitations to the possibility of applying the method of round-trip translation as a starting point for revision. The hypothesis that the round-trip translation can actually improve the quality of the manuscripts remains unconfirmed. These as well as other problematic aspects will be further explored in Chapter 4.

The application of round-trip translation to evaluation of revision difficulty is, hence, still at an experimental stage (see Section 4.2 for reflections on its limitations). Nevertheless, the experiments described in this chapter have shown that the method can be used to differentiate texts which need substantial revision from those that need a limited amount of it. This is the reason why, if further developed and implemented in a publisher's workflow, this method could provide a solution – albeit still partial = to problems concerning revision time and costs estimation.

Chapter 4

Limitations, Future Development and Possible Applications

4.1 Introduction

In this last chapter, I will briefly describe the limitations of the application of round-trip translation here presented as a method for identifying the difficulty of revision.

Then, the possibility of obtaining absolute BLEU score values from the results of the experiments in Chapter 3 will be introduced. These values will be used as thresholds for estimating revision difficulty of files belonging to an unrevised dataset will be introduced. For this purpose, a new dataset, called New_Medit, has been compiled, featuring texts from a journal written in English as a lingua franca and published by BUP without revision, due to budget issues[51].

Lastly, a possible application of round-trip translation for other purposes will be briefly presented.

[51] https://newmedit.iamb.it/

4.2 Limitations and Further Developments

In the last part of the discussion regarding the experiments described in Chapter 3 (Subsection 3.5), I have noticed that the outputs deriving from the process of round-trip translation are very standardized texts. In this sense, it is worth remembering that translation is essentially a process of transformation (Vanmassenhove et al., 2021). In our case, the transformation of the manuscript can have both positive and negative side effects. If a text necessitates major changes and revision, the double operation of translating it and back-translating it into the source language can provide a more correct version of the text. Indeed, while text similarity scores can help us prove that the text need significant revision (as I have demonstrated in Experiment 3), these metrics tell us very little about the quality of the new text – the RawRTT. As Experiment 4 has shown, the comparison between RawRTT and Rev texts result in low scores (see Subsection 3.4.6. This fact highlights that RTT may also produce unnecessary changes to the manuscript. Linking this observation to serious reflections about text standardization and even ethical issues on the influences of language technologies on cognition (Vanmassenhove et al., 2021), I would conclude that round-trip translation should not be used as a fully automated technique for language quality improvement, and its use as a starting point for revision – despite possibly being time-saving – might undermine both the author and the revisor's linguistic creativity.

Another limitation of this study is that it is difficult to obtain absolute values for the evaluation of revision difficulty from the experiments that I have conducted. The method of round-trip translation appears potentially valid for the resolution of the problem here addressed. However, this method needs further testing before it can be robustly applied for revision difficulty evaluation of an unknown dataset. Despite this, I am going to explore the possibility of identifying some thresholds that allows for a broad categorization of unrevised texts according to their revision difficulty in the next section.

One last factor that should be further explored relates to the impossibility of discerning

to what degree the eventual use of machine translation from the manuscripts' authors could have influenced the results of the experiments. This reflection is linked to the concept of linguistic interference, which has been treated in Subsection 2.2.2.

4.3 Towards the Identification of Thresholds and Absolute Scores: Working on Unrevised Data

The experiments discussed in Chapter 3 allowed me to investigate the relationships between texts written in English as a lingua franca by non-native authors and their round-trip translations used as similarity references. To better highlight the variation of the similarity between a raw text and its RTT as a function of the revision difficulty, I designed a further experiment, in which I added artificial noise to a subset of sample texts through the random cancellation of some words (see Subsection 3.4.5). Moreover, thanks to the experiments conducted on the CIHA dataset, I was able to observe the behavior of the BLEU score between the raw texts and their RTTs as a function of the variation of the human evaluation of revision difficulty, which I have represented through the IRev_H index (see Subsection 3.3.2). Furthermore, I have proposed an automatic revision difficulty index (IRev_A). This index, properly normalized, amplifies the distance between texts with low revision needs and texts which necessitate major revision.

The behaviors of the calculated scores and indices were consistent with the starting hypothesis, however, the score values range in a narrow interval that apparently does not seem to allow a clear separation between texts with high difficulty of revision and those with low or medium difficulty of revision.

To verify this aspect, i.e., the possibility of applying the scale obtained from the score between the raw texts of the CIHA dataset and their RTTs for the classification of a test set consisting of new texts written in English as a lingua franca, I arbitrarily chose two thresholds based on the score of the CIHA texts with their RTTs (Figure 12).

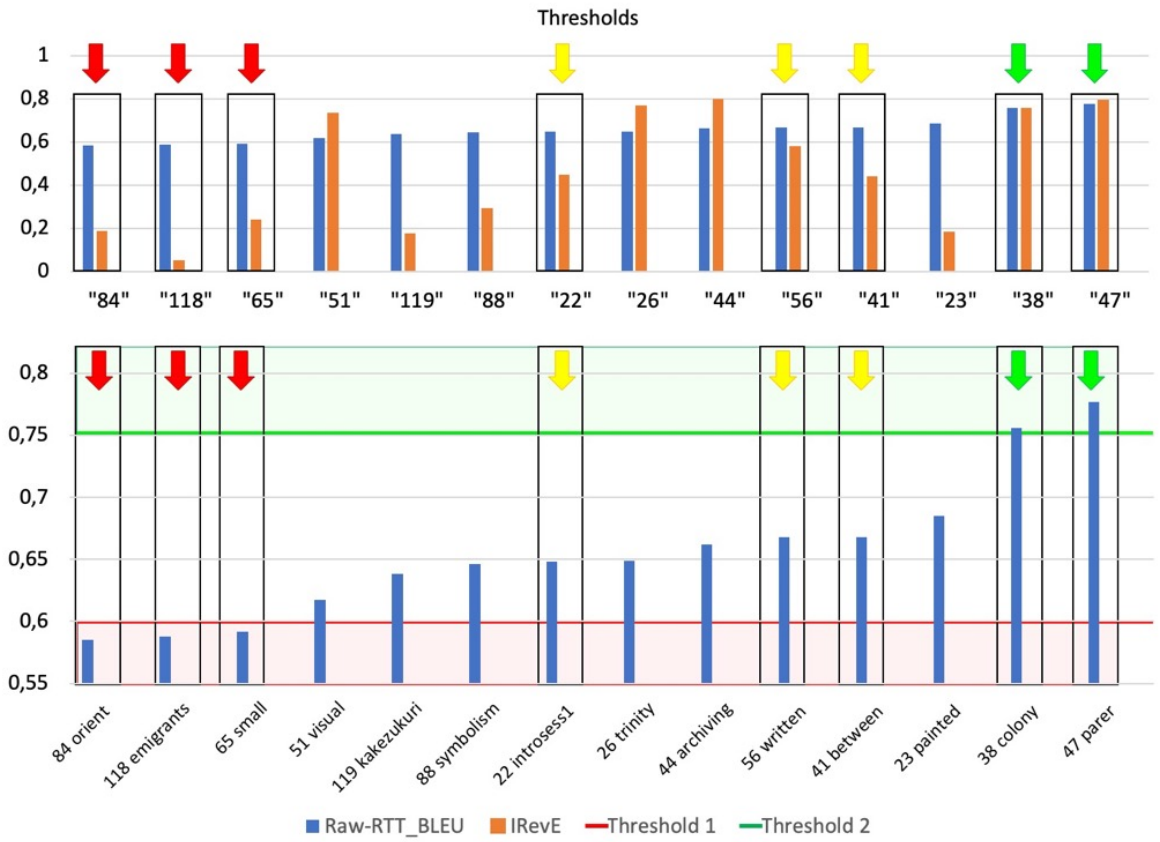

Figure 12: Identification of thresholds based on the results of Experiment 3.

Figure 12 also shows IRevE (index of revision ease), which is Irev_H inverted and normalized on a scale from 0 to 1, to be comparable with the BLEU scores – these latter values had to be kept unaltered for the identification of BLEU thresholds. As can be seen, I have distinguished between a field of texts with high revision difficulties characterized by a score between the text itself and its RTT$\leq$ 0.6 (red line in Figure 12) and a field of texts with low revision difficulty characterized by score $\geq$ 0.75 (green line in Figure 12). The score interval between 0.6 and 0.75 groups texts with intermediate revision difficulties or with unpredictable revision difficulties.

On the basis of these thresholds, I identified two manuscripts of the CIHA dataset that fall into the field of texts with low difficulty of revision (47 parer and 38 colony) and three manuscripts with high difficulty of revision (65 small, 84 orient and 118 emigrants). The remaining manuscripts of this dataset fall within the intermediate range; therefore, they would be associated with a medium revision difficulty or an

indeterminacy of this parameter. In fact, among them there are also manuscripts 22, 41 and 56 originally identified by the reviewer as belonging to group 2, that is the group of manuscripts associated with medium difficulty of revision.

As test set, I have chosen three articles published on the journal New Medit (Mediterranean Journal of Economics, Agriculture and Environment), which were written in English as a lingua franca by non-native speakers, and which had not been subject to any kind of revision or proofreading prior to their publication, due to budget limitations. I compiled the dataset extracting texts from three articles (Markou et al., 2020; Bernabéu et al., 2020; Adinolfi et al., 2020). I followed the standard procedure for preparing the texts described in Chapter 3 and I generated the round-trip translations of each text using Modern MT with the language pair En-Es (English / Spanish), in accordance with what had been identified as the best practices for the application of this method in the Experiments described in Chapter 3. Table 12 shows some samples segments and their round-trip translation.

Table 12: Sample segments extracted from New_Medit Dataset

Text ID	**Raw**	**RawRTT**
new_medit_1	This study drew on this gap and attempted to identify the UTPs that exist in the Cypriot food supply chain, assess their impact on the involved stakeholders and provide guidelines that will assist the transposition of EU relevant Directive to the national law.	This study was based on this gap and attempted to identify UTPs that exist in the Cypriot food supply chain, assess their impact on the involved stakeholders and provide guidelines that will help the transposition of the relevant European Directive into national law.
new_medit_2	Pragmatically speaking, we can say that ethnocentrism is the name given to the sense of belonging felt by any individual. More importantly, it explains the reason why a group accepts certain choices rather than others, which, in the case that concerns us here, refers to accepting certain buying behaviours rather than others (Witkowsky, 1998).	Pragmatically, we can say that ethnocentrism is the name given to the sense of belonging felt by any individual. More importantly, it explains why a group accepts certain choices over others, which in this case refers to the acceptance of certain purchasing behaviours over others (Witkowsky, 1998).
new_medit_3	This choice is in order to both take into account the nature of the decisions investigated and appropriately manage variables that can be complementary.	This choice aims both to take into account the nature of the decisions studied and to manage appropriately the variables that may be complementary.

For what concerns the contents of Table 12, it should be noticed that, although the RTTs seem to have improved the texts, a qualitative evaluation of the actual quality of the RTT outputs has not been conducted.

Then, I calculated the BLEU score for the three manuscripts with their respective round-trip translations and I used Grammarly to evaluate the method efficiency, with the results reported in Table 13.

Table 13: BLEU scores of the New_Medit texts included in the test set (Raw-RawRTT_BLEU) and the respective Grammarly scores

ID text	**Raw-RawRTT_BLEU**	**Grammarly**
new_medit_1	0.645	0.720
new_medit_2	0.717	0.710
new_medit_3	0.528	0.510

The score of the text new_medit_3 (0.528) is lower than the red threshold (0.60), meaning that it falls into the field of texts with high difficulty of revision. On the other hand, the texts new_medit_2 and new_medit_1 have obtained BLEU scores of 0.717 and 0.645, respectively, thus falling into the intermediate field, which groups texts with intermediate or unpredictable revision difficulties (Figure 13).

Although these thresholds need further testing, the results are overall consistent with the scores provided by Grammarly, which in turn had already given similar results to the ones obtained through our method in the previous experiments (see Section 3.5). This suggests that the possibility of developing thresholds for revision difficulty evaluation through the method of round-trip translation is concrete.

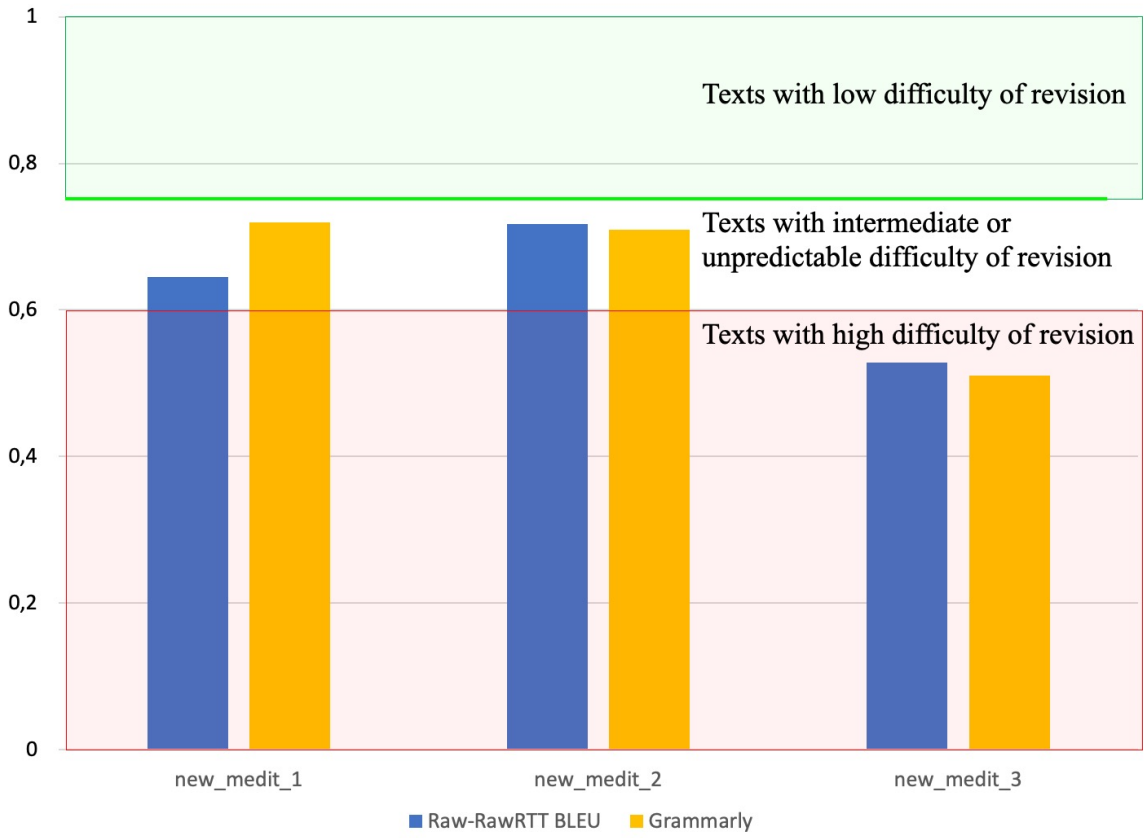

Figure 13: BLEU and Grammarly scores of New Medit Test Set.

4.4 Round-Trip Translation for Second Language Acquisition

At this point, I would like to present some possible applications of round-trip translation beyond the aims of the present study – namely, in the field of second language acquisition (SLA).

Talking about language learning in the context of a study on revision and academic writing is not unusual. Indeed, even throughout the present work, many references have been made to the relationship between ELF academic writing and ELF learning and their possible comparison.

In such an interdisciplinary framework, several studies which relate teaching and learn-

ing practices to editing of academic writing have flourished. For instance, Mišak et al. (2005) describe the experience of a scientific journal in the context of undergraduate medical education, in which students learned academic writing through peer-reviewing; another example is the study by Pospelova (2016) about the influence of self-editing on the improvement of writing micro-skills. Theories on students' agency, on the importance of self-evaluation and self-correction for the development of language acquisition (Ciliberti, 2019) underlie studies as the ones just mentioned.

Furthermore, a similar consideration can be formulated with regard to the relationship between second language acquisition and language technologies. Nowadays, corpus-informed approaches are definitely surpassing prescriptivism in English textbooks and, in general, in English teaching, which has become increasingly inclusive of the different varieties of English – with a particular attention to English as a lingua franca (Osborne, 2000; Johansson, 2009; Römer, 2011; McEnery & Hardie, 2012); many studies have shown the possibility of implementing corpus linguistics in language teaching contexts (see, for example, Mair, 2002; Bernardini, 2000; Bernardini, 2002). Among these, Yoon (2008) has noticed the improvements in learners' academic papers after they had repeatedly been using corpora as a support for their writing activity, thus interconnecting corpus linguistics, second language acquisition and English for academic purposes.

In general, the education world has lately lived a process of digital transformation (Guerra, 2010; Rivoltella & Rossi, 2019), which has been greatly accelerated by the Covid-19 pandemic.

Besides corpora, also research in computational linguistics has lately been committed to solving problems related to teaching, such as assessment and feedback (see Subsections 2.4.2, 2.4.4). Computer-assisted language learning (CALL) is the branch of linguistics which 'seeks to employ computers in order to improve language learning' (Mitkov, 2003: 677). Among the NLP technologies which traditionally contributes to CALL, Mitkov mentions concordancing, speech recognition and morphological and syntactic processing. Here I would like to suggest that the method of round-trip translation could be employed in technological tools for self-learning.

The inspiration, in this sense, came from a demo of a 'Neural Text Improving Tool for English' recently published by Claudio Fantinuoli[52]. In private conversations I had with some high-school Chinese and English teachers based in Naples, the topic of automatic machine translation came up – whether students should be allowed to use it, whether it should be used during lessons etc.

What the technological development has taught us so far is that avoiding or forbidding the use of technology in educational environments can only cause unawareness and misuse of such systems. For this reason, I believe that notions on automatic machine translation, pre- and post-editing, should be included in language teaching, so that students can truly be empowered by the use of such tools. Round-trip translation, in particular, could be employed for two main reasons. First, the method could help students to deductively acquire information and knowledge about the power and limitations (e.g., language standardization) as well as the power and MT potentials. Secondly, this method could work as an automatic feedback provider. In a self-learning environment this could replace the simple use of machine translation; instead of using their native language as input, students might test their hypotheses on the use of certain words and structures. On the other hand, in the classroom environment comparison between input and RTT output could work as a method of eliciting self-correction and generalization of linguistic structures.

4.5 Conclusion

This chapter gives an overview of the possibilities linked to the implementation of round-trip translation for different purposes. The path is still long, and some questions remain unsolved, nevertheless this method could represent an innovative point of view on various topics in linguistics research.

[52] https://www.claudiofantinuoli.org/apps/COR/index.html

Chapter 5

Conclusion

This dissertation has focused on the revision of texts written in English as a lingua franca for academic purposes. In particular, I have faced the problem of assessing and predicting the difficulty of revision that an expert revisor may encounter prior to the beginning of the revision process itself. Indeed, this information is essential for planning the revision workflow and estimating its costs and duration. As with many aspects concerning linguistic issues, also in this case an assistance for the solution of the problem has been found in the modern computer technologies for automatic text processing. In order to engage with this topic as correctly and comprehensively as I could, this work has embraced an interdisciplinary approach. I had to exploit skills and knowledge deriving from different disciplines, such as sociolinguistics, computational linguistics, automatic text processing and artificial intelligence.

Overall, the results of the experiments have satisfied the main hypothesis of my thesis, confirming that RTT and BLEU can be used to evaluate the difficulty of revision. What follows is a brief reconstruction of the main steps that have led to this conclusion.

First of all, it was necessary to investigate the current role of English as a lingua franca. The idea of English as a means to perpetrate cultural hegemony has been countered by a view of ELF as a contact language through which people of different cultures can interact. This aspect is crucial for a contextual framing of revision activities and of the revisor's role within the publishing industry's landscape. When revisors perform their

delicate work correctly, they manage to create an improved version of the manuscripts under revision, while remaining close to the words and intentions of the authors. This distance between a text and its revision can be quantitatively measured by applying text processing methods. Therefore, the analysis of the theoretical aspects linked to linguistics, the work of revision and natural language processing oriented the objectives of the experimental work on which the second part of this dissertation has focused.

Chapter 3 was, thus, devoted to the design and implementation of computer experiments aimed at *a priori* estimation of revision difficulty. These experiments were based on the round-trip translation technique. The experimental method exploits the possibility of automatically creating a reference similarity for a source text written by non-native speakers, generating its round-trip translation (RTT), and to relate the distance between the source text and the RTT with the *a priori* revision difficulty. The development of this method required a preliminary evaluation of the available machine translators and text similarity estimation techniques. The experimentation of the method required the compilation of datasets composed of texts in English as lingua franca and their revisions.

Thanks to the first round of experiments the method as well as some starting hypotheses were refined. In this initial phase, a dataset was compiled with texts extracted from a volume on topic of the economic crisis caused by the COVID-19 pandemic. This volume was chosen because it had already been the focus of a study on revision conducted by another trainee at BUP (Wang, 2021). Using techniques from information retrieval and text processing, the files were cleaned and encoded in UTF-8; manuscripts, revisions and their RTTs produced using different machine translation systems were aligned. Both BLEU and BERTScore were tested as measures of text similarity between different pairs of files (manuscript and revision, manuscript and corresponding RTT, RTT of the manuscript and revision). While the results were generally coherent with the initial hypotheses, they also showed that some criteria needed to be better defined. For example, the choice of MT system partially affected results, therefore a more cautious decision was to be made on this matter. On the other hand, the scores calculated by the two metrics consistently followed the same trend, there-

fore the second set of experiments were with one score only, namely BLEU. Moreover, it was understood that some pre-categorization of the files included in the dataset is essential to verify the outcomes of the experiments with round-trip translations.

At this point, the second set of experiments was conducted, compiling a dataset similar to the first one. It included texts related to the XXXV CIHA Congress, all of which were revised by the same revisor. The CIHA dataset, accompanied by more in-depth information on the revision, was used to create human revision difficulty indexes, necessary for the evaluation of the performance of the method. For this reason, the second dataset was used for experiments aimed at testing the method on specific aspects.

The RTT method to estimate the difficulty of revision can be structured in stages. The first task is dedicated to the generation of automatic similarity references through round-trip translation while the second task involves the measurement of text similarity between the source text and the RTT. The score which derives from this comparison represents an indicator of the *a priori* difficulty of revision: the higher this value, the lower the revision difficulty.

The experiments carried out using the second dataset have shown that the method could successfully predict the difficulty of revision for most of the analyzed texts. Among the results of the experiments, it was found that the BLEU scores between source texts and their revisions, which were carried out by an expert revisor at BUP, are always high and systematically higher than the Raw-RawRTT scores. This indicates that the revisor was indeed able to generate an improved version of the raw text, without departing from it. At the same time, the method proves to be effective in relation to the main application proposed in this dissertation, that is, to evaluate revision difficulty.

I compared the results of the experiments both with the human revision, through the indices I created, and with the scores calculated by a cloud-based writing assistance software, namely, Grammarly. The comparison showed a good performance of the method.

This allowed me to identify two thresholds of the BLEU score between the text and its

round-trip translation that broadly define three levels: the first level relates to texts with high difficulty of revision, characterized by a score less than 0.60; the second level is where texts a medium or unpredictable difficulty of revision are placed, characterized by a score between 0.60 and 0.75; lastly, the third level groups texts with low revision difficulty, which obtained a score greater than 0.75.

I tested the scale thus obtained with a test set composed of three texts extracted from New Medit, a journal which is published in an unrevised version by BUP, due to budget constraints. I obtained congruent results. This suggests that the RTT method for the *a priori* estimation of the revision difficulty, after further tests and refinements, can be a valid and simple tool to support revision practices related to texts written in English as a lingua franca. The experiments also showed that specific commercial software currently have a score comparable to that obtained with our method. Therefore, depending on the style and individual preferences of the revisors and publishing companies, the RTT method could be preferable to ad hoc commercial ones.

Overall, this study does also tell us something more about revision practices and why it is so hard to estimate revision difficulty: because revising is ultimately the art of negotiation. As an opening of this dissertation, I have quoted the very first sentences from 'Dire quasi la stessa cosa', the famous work where Umberto Eco (2003) has exposed his philosophical views on translation. Translation and revision are practices that share many common features: first and foremost, they both involve engagement with a text written by somebody else. With this parallelism in mind, the technological method here presented can be seen as a useful aid to approach manuscripts in a more objective and functional way.

Bibliography

Abney, S. (2011). Data-intensive experimental linguistics, *Linguistics Issues in Language Technology*, 6(2), 1–27.

Adinolfi, F., Capitanio, F., De Rosa, M. and Vecchio, Y. (2020). Gender differences in farm entrepreneurship: comparing farming performance of women and men in Italy, *New Medit*, 19(1), 69–82.

Ahmadnia, B. and Dorr, B.J. (2019). Augmenting neural machine translation through round-trip training approach, *Open Computer Science*, 9(1), 268–278.

Anthonio, T. and Roth, M. (2020). What can we learn from noun substitutions in revision histories?, in: *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1359–1370, International Committee on Computational Linguistics, Barcelona, Spain (Online).

Ashcroft, B., Griffiths, G. and Tiffin, H. (2002). *The Empire Writes Back: Theory and Practice in Post-Colonial Literatures*, Taylor & Francis.

Aslin, R.N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure, *WIREs Cognitive Science*, 8(1-2), e1373.

Audrain, L., Mussinelli, C. and Pellegrino, G. (2021). European Accessibility Act: report on the state of the art of publishing standards.

Avner, E.A., Ordan, N. and Wintner, S. (2016). Identifying translationese at the word and sub-word level, *Digital Scholarship in the Humanities*, 31(1), 30–54.

Baccolini, R. (2005). *Le prospettive di genere. Discipline, soglie e confini*, Bologna University Press, Bologna, Italy.

Baird, R., Baker, W. and Kitazawa, M. (2014). The complexity of ELF, *Journal of English as a Lingua Franca*, 3(1), 171–196.

Baroni, M. and Bernardini, S. (2006). A new approach to the study of translationese: Machine-learning the difference between original and translated text, *Literary and Linguistic Computing*, 21(3), 259–274.

Bathia, V.K. (1993). *Analysing Genre: Language use in professional settings*, Routledge, London and New York.

Belcher, D. (1997). An argument for nonadversarial argumentation: On the relevance of the feminist critique of academic discourse to L2 writing pedagogy, *Journal of Second Language Writing*, 6(1), 1–21.

Bell, J.S. (2012). *Revision and self-editing for publication: techniques for transforming your first draft into a novel that sells*, Cincinnati: Writer's Digest, 2nd edition.

Bellettini, G. and Goldstein, A. (editors) (2020). *The italian economy after Covid-19. Short-term costs and long-term adjustments*, Bologna University Press, Bologna, Italy.

Bennett, K. (2009). English academic style manuals: A survey, *Journal of English for Academic Purposes*, 8(1), 43–54.

Bennett, K. (2013). English as a lingua franca in academia, *The Interpreter and Translator Trainer*, 7(2), 169–193.

Bermann, S. and Porter, C. (editors) (2014). *A Companion to Translation Studies*, Wiley Blackwell, Chichester, UK.

Bernabéu, R., Oliveira, F., Rabadán, A. and Díaz, M. (2020). Influence of ethnocentrism on consumer preference patterns: the case of olive oil in Portugal, *New Medit*, 19(1), 55–68.

Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners, in: B. McEnery (editor), *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the third international conference on Teaching and Language Corpora*, Peter Lang, Frankfurt am Main.

Bernardini, S. (2002). Exploring new directions for discovery learning in teaching and learning by doing corpus analysis, in: M. Kettemann (editor), *Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000*, Rodopi B.V., Amsterdam, New York.

Billingham, J. (2002). *Editing and Revising Text*, Get Ahead in, Oxford University Press, Oxford,.

Boden, M.A. (2018). *Artificial Intelligence: A very short introduction*, Oxford University Press.

Borghi, A., Scorolli, C., Caligiore, D., Baldassarre, G. and Tummolini, L. (2013). The embodied mind extended: using words as social tools, *Frontiers in Psychology*, 4.

Buranyi, S. (2017). Is the staggeringly profitable business of scientific publishing bad for science?, *The Guardian*.

Callison-Burch, C., Osborne, M. and Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research, in: D. McCarthy and S. Wintner (editors), *11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 249–256, Association for Computational Linguistics, Trento, Italy.

Canagarajah, A.S. (1999). *Resisting Linguistic Imperialism in English Teaching*, Oxford Applied Linguistics, Oxford University Press, Oxford, 1st edition.

Canagarajah, A.S. (2002). *A Geopolitics of Academic Writing*, University of Pittsburg Press, Pittsburg, USA.

Canagarajah, A.S. (2005). *Reclaiming the Local in Language Policy and Practice*, Routledge, 1st edition, ISBN 9780805845938.

Canagarajah, A.S. (2012). Teacher development in a global profession: An autoethnography, *TESOL Quarterly*, 46(2), 258–279.

Chakhachiro, R. (2005). Revision for quality, *Perspectives*, 13(3), 255–238.

Charles, M., Pecorari, D. and Hunston, S. (editors) (2009). *Academic writing: At the interface of corpus and discourse*, Continuum Intl. Pub. Group, New York.

Chew, P.G.L. (2010). *Emergent lingua francas and world orders, the politics and place of English as a world language*, Routledge, London.

Chomsky, N. (1965). Persistent topics in linguistic theory, *Diogenes*, 13(51), 13–20.

Church, K. (1988). A stochastic parts program and noun phrase parser for unretricted texts, in: *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas.

Ciliberti, A. (2019). *Glottodidattica - per una cultura dell'insegnamento linguistico*, Carocci Editore, Roma.

Cogo, A. and Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation*, Continuum, London.

Collada, I. and Alvarez, P.C. (2018). Quality improvement in English as a second language teacher training in Cuban higher education, in: *English Public Policies in Latin America: Looking for Innovation and Systemic Improvement in Quality English Language Teaching*, British Council Mexico and CIDE.

Cotos, E. (2009). Designing an intelligent discourse evaluation tool: Theoretical, empirical, and technological considerations, in: C.A. Chapelle, G.H. Jun and I. Katz (editors), *Developing and evaluating language learning materials*, Proceedings of the 6th Annual TSLL Conference, pages 103–127, Ames, IA: Iowa State University.

Cotos, E. (2010). *Automated Writing Evaluation for Non-Native Speaker English academic writing: the case of IADE and its formative feedback*, Proquest llc, ph.d. dissertation, Iowa State University, USA.

Couture, B. (1985). A systemic network for analyzing writing quality, in: J. Benson and W. Greaves (editors), *Systemic perspectives on discourse*, volume 2, pages 67–87, Ablex, Norwood, NJ.

Crystal, D. (2003). *Englis as a Global Language*, Cambridge University Press, Cambridge, UK, 2nd edition.

Cushing Weigle, S. and Malone, M.E. (2016). Assessment of english for academic purpose, in: K. Hyland and P. Shaw (editors), *The Routledge Handbook of English for Academic Purposes*, chapter 45, pages 608–620, Routledge, London.

de Beaugrande, R.A. and Dressler, W.U. (1981). *Introduzione alla linguistica testuale*, Il Mulino, Bologna, traduzione in italiano a cura di Muscas, S. (1984).

Delisle, J. (1988). *Translation: an interpretative approach*, University of Ottawa Press, Ottawa, Canada.

DeRose, S. (1988). Grammatical category disambiguation by statistical optimization, *Computational Linguistics*, 14(1), 31–39.

Derrida, J. and Ronell, A. (1980). The law of genre, *Critical Inquiry*, 7(1), 55–81.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding, preprint arXiv:1810.04805.

Duh, K. and Guzmán (editors) (2021). *Proceedings of Machine Translation Summit XVIII: Research track*, Association for Machine Translation in the Americas, Virtual.

Dumouchel, P. and Damiano, L. (2019). *Vivere con i robot*, Raffaello Cortina Editore, Milano.

Eckman, M. (2021). *Learning Deep Learning Tensorflow: Theory and Practice of Neural Networks, Computer Vision, Natural Language Processing, and Transformers Using TensorFlow*, Addison-Wesley Professional.

Eco, U. (2003). *Dire quasi la stessa cosa: Esperienze di traduzioni*, Bompiani, Milano.

Eco, U. (2013). Encyclomedia, online.

Ehrensberger-Dow, M. (2006). Revising and editing English in Switzerland: Some considerations for native speakers, *Hieronymous*, 4, 20–25.

Englander, K. (2006). Revision of scientific manuscripts by nonnative-English-speaking scientists in response to journal editors' criticism of the language, *Journal of Applied Linguistics*, 3(2), 129–161.

Erkan, G. and Radev, D.R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization, *Journal of artificial intelligence research*, 22, 457–479.

European Accessibility Act (2019). Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services, Official Journal of the European Union.

Faietti, M. and Wolf, G. (editors) (2021). *Motion: Transformation*, Bologna University Press, Bologna, Italy.

Feng, C. (2002). Post-Colonialism in China, or, using English will have our Chinese identity lost?, *Comparative Literature: East & West*, 4(1), 88–104.

Ferraresi, A. (2019). How specialized (or popularized)? Terminological density as a clue to text specialization in the domain of food safety, *Lingue e Linguaggi*, 29, 17–39.

Ferrari, A. (2014). *Linguistica del testo*, Carocci Editore, Roma.

Forstner, M., Lee-Jahnke, H. and Schmitt, P.A. (editors) (2008). *CIUTI-Forum 2008. Enhancing Translation Quality: Ways, Means, Methods*, Peter Lang, European Academic Publisher, Bern, Switzerland.

Foti, M. (2016). Machine Translation service. Service Offering Description - CEF eTranslation, European Commission, DIGIT - Connecting Europe Facility, version 1.2.

Foucault, M. (1972). *The Archaeology of Knowledge*, Pantheon Book, New York.

Foucault, M., Senellart, M. and Burchell, G. (2008). *The birth of biopolitics*, Foucault, M.: Lectures at the Collège de France, 1978-1979, Basingstoke, England.

Frow, J. (2005). *Genre*, Te New Critical Idiom, Routledge, London and New York, 1st edition.

Fyfe, A., Coate, K., Curry, S., Lawson, S., Moxham, N. and Røstvik, C.M. (2017). Untangling academic publishing: a history of the relationship between commercial interests, academic prestige and the circulation of research, Zenodo n.546100.

Galloway, N. and Rose, H. (2015). *Introducing Global Englishes*, Taylor & Francis.

Galvao, G.C.T. (2009). *Linguistic interference in translated academic texts: A case study of Portuguese interference in abstracts translated into English*, Växjö University. School of Humanities, Bachelor's Course, Sweden.

Gass, S.M. and Selinker, L. (1992). *Language Transfer in Language Learning*, John Benjamins, Philadelphia, PA.

Global 50 (2021). The World Ranking of the Publishing Industry, Wischenbart, R. and Fleischhacker, M. A. (ed.), Wischenbart Content and Consulting, editions 2012-2021.

Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publisher, California.

Gosden, H. (1995). Success in research article writing and revision: A social-constructionist perspective, *English for Specific Purpose*, 14(1), 37–57.

Gotou, T., Nagata, R., Mita, M. and Hanawa, K. (2020). Taking the correction difficulty into account in grammatical error correction evaluation, in: *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2085–2095, International Committee on Computational Linguistics, Barcelona, Spain (Online).

Groves, M. and Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes, *English for specific Purposes*, 37, 112–121.

Guerra, L. (editor) (2010). *Tecnologie dell'educazione e innovazione didattica*, Educazione e tecnologie, Junior.

Haar, C. (2016). Definitions and distinctions, in: A. Horning and A. Becker (editors), *Revision: History, Theory, and Practice*, pages 10–24, Parlor Press, The WAC Clearinghouse.

Hall, C.J. (2013). Cognitive contributions to plurilithic views of English and other languages, *Applied Linguistics*, 34, 211–231.

Halliday, M.A.K. (1961). Categories of the theory of grammar, *Word*, 17(2), 241–292.

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

Hamp-Lyons, L. and Lockwood, J. (2015). Delta writing: can academic writing be scored automatically?, in: *Proceedings of 2013 BALEAP conference*, volume 1, pages 199–206.

Haraway, D. (1988). Situated knowledges: The science question in Feminism and the privilege of partial perspective, *Feminist Studies*, 14(3), 575–599.

Hartse, J.H. and Kubota, R. (2014). Pluralizing english? variation in high-stakes academic texts and challenges of copyediting, *Journal of Second Language Writing*, 24, 71–82.

Haywood, T. (1998). Global networks and the myth of equality: trickle down or trickle away?, in: B.D. Loader (editor), *Cyberspace Divide: Equality, Agency and Policy in the Information Society*, chapter 2, pages 19–34, Routledge, London.

Heimgärtner, R. (2014). Intercultural user interface design, in: K. Blashki and P. Isaias (editors), *Emerging Research and Trends in Interactivity and the Human-Computer Interface*, chapter 1, pages 1–33, IGI Global, Hershey, PA.

Held, D., McGrew, A., Goldblatt, D. and Perraton, J. (1999). *Contents and Introduction in Global Transformations: Politics, Economics and Culture*, Stanford University Press, Stanford.

Hood, S. (2016). Systemic functional linguistics and EAP, in: K. Hyland and P. Shaw (editors), *The Routledge Handbook of English for Academic Purposes*, chapter 15, Routledge, London.

Huang, C.H., Yin, J. and Hou, F. (2011). A text similarity measurement combining word semantic information with TF-IDF method, *Jisuanji Xuebao (Chinese Journal of Computers)*, 34(5), 856–864.

Hülmbauer, C., Böhringer, H. and Seidlhofer, B. (2008). Introducing English as a lingua franca (ELF): Precursor and partner in intercultural communication, *Synergies Europe*, 3, 25–36.

Hyland, K. (2002). Authority and invisibility: authorial identity in academic writing, *Journal of Pragmatics*, 34(8), 1091–1112.

Hyland, K. (2006). *English for Academic Purposes. An Advanced Resource Book*, Routledge applied linguistics, Routledge, London.

Hyland, K. (2011). Learning to write: Issues in theory, research and pedagogy, in: R.M. Manchón (editor), *Learning-to-write and writing-to-learn in an additional language*, pages 17–36, John Benjamins, Amsterdam.

Hyland, K. (2015). *Academic publishing: Issues and challenges in the construction of knowledge*, Oxford Applied Linguistics, Oxford University Press, UK.

Hyland, K. and Shaw, P. (editors) (2016). *The Routledge Handbook of English for Academic Purposes*, Routledge, London, 1st edition.

Hymes, D. (editor) (1974). *Foundations in Sociolinguistics: An ethnographic approach*, Routledge, London, 1st edition.

Intento (2021). The state of machine translation, Report, https://try.inten.to/machine-translation-report-2021/.

Ito, T., Kuribayashi, T., Kobayashi, H., Brassard, A., Hagiwara, M., Suzuki, J. and Inui, K. (2019). Diamonds in the rough: generating fluent sentences from early-stage drafts for academic writing assistance, preprint arXiv:1910.09180.

Jafaritazehjani, S., Lecorvé, G., Lolive, D. and Kelleher, J. (2020). Style versus content: A distinction without a (learnable) difference?, in: *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2169–2180, International Committee on Computational Linguistics, Barcelona, Spain (Online).

Jahankhani, H., Akhgar, B., Cochrane, P. and Dastbaz, M. (2020). *Policing in the Era of AI and Smart Societies*, Springer International Publishing, Germany.

Jarvis, S. and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*, Routledge, New York and London, 1st edition.

Jenkins, J. (2013). The spread of English as a Lingua Franca, in: *English as a Lingua Franca in the International University*, chapter 2, Routledge, 1st edition.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a lingua franca, *Englishes in Practice*, 2(3), 49–85.

Jenkins, J. (2018). Introduction, in: J. Jenkins, W. Baker and M. Dewey (editors), *The Routledge Handbook on English as a Lingua Franca*, Routledge.

Jenkins, J., Baker, W. and Dewey, M. (editors) (2018). *The Routledge Handbook on English as a Lingua Franca*, Routledge.

Johansson, S. (2009). Some thoughts on corpora and second language acquisition, in: K. Aijmer (editor), *Corpora and Language Teaching*, volume 33 of *Studies in Corpus Linguistics*, pages 33–44, John Benjamin Publishing Company.

Joshi, Y. (2011). Native and non-native speakers of English as copy-editors of research papers, *European Science Editing*, 37(2), 38–39.

Jurafski, D. and Martin, J.H. (2020). Regular expressions, text normalization, edit distance, in: *Speech and Language Processing*, chapter 2, pages 2–28, Draft of December 30, 2020.

Kachru, B.B. (editor) (1992a). *The Other Tongue: English Across Cultures*, University of Illinois Press, Urbana and Chicago, 2nd edition.

Kachru, B.B. (1992b). The second diaspora of English, in: S.C.T. Machan T. W. (editor), *English in its social contexts. Essays in historical linguistics*, volume 230, page 252, Oxford University Press, New York.

Kalchbrenner, N. and Blunsom, P. (2013). Recurrent continuous translation models, in: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1700–1709, Association for Computational Linguistics, Seattle, Washinghton, USA.

Karpenko-Seccombe, T. (2020). *Academic writing with corpora. A resource book for data-driven learning*, Routledge, London, 1st edition.

Knowles, G. (1997). *A Cultural History of the English Language*, Routledge, 1st edition.

Ko, Y., Park, J. and Seo, J. (2004). Improving text categorization using the importance of sentences, *Inf. Process. Manag.*, 40, 65–79.

Koponen, M., Mossop, B., Robert, I.S. and Scocchera, G. (editors) (2020). *Translation Revision and Post-editing: Industry Practices and Cognitive Processes*, Routledge, London.

Krenn, B. and Samuelsson, C. (1997). *The linguist's Guide to Statistics Don't Panic*, Los autores.

Kruse, O. (2013). Perspectives on academic writing in european higher education: genres, practices, and competences, *Revista de Docencia Universitaria*, 11(1), 37–58.

Larsen-Freeman, D. (2018). Complexity and ELF, in: J. Jenkins, W. Baker and M. Dewey (editors), *The Routledge Handbook on English as a Lingua Franca*, chapter 4, Routledge.

Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives)*, Cambridge University Press, Cambridge.

Léon, J. (2005). Claimed and unclaimed sources of corpus linguistics, *Henry Sweet Society for the History of Linguistic Ideas Bulletin*, 44(1), 36–50.

Léon, J. (2021). *Automating Linguistics*, History of Computing, Springer Nature, Switzerland, 1st edition.

Lepori, M. and McCoy, R.T. (2020). Picking BERT's brain: Probing for linguistic dependencies in contextualized embeddings using representational similarity analysis, in: *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3637–3651, International Committee on Computational Linguistics, Barcelona, Spain (Online).

Li, J. (2000). *Revision strategies in English-second-language academic writing*, Master's thesis, Queen's University, National Library of Canada, Ottawa.

Lichtarge, J., Alberti, C., Kumar, S., Shazeer, N., Parmar, N. and Tong, S. (2019). Corpora generation for grammatical error correction, in: J. Burstein, C. Doran and T. Solorio (editors), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3291–3301, Association for Computational Linguistics, Minneapolis, Minnesota.

Lillis, T. and Curry, M.J. (2010). *Academic writing in a global context. The politics and practices of publishing in English*, Routledge, London, New York.

Liuzza, M.T., Cimatti, F. and Borghi, A.M. (2010). *Lingue, corpo, pensiero: le ricerche contemporanee*, Carocci Editore, Roma.

Loader, B.D. (editor) (1998). *Cyberspace Divide: Equality, Agency and Policy in the Information Society*, Routledge, London, 1st edition.

Lorés Sanz, R. (2020). Science on the web: The exploration of European research websites of energy-related projects as digital genres for the promotion of values, *Discourse, Context & Media*, 35, 100389.

Lowe, C. and Zemliansky, P. (editors) (2010). *Writing Spaces: Readings on Writing, Volume 1*, WritingSpaces.org, Parlor Press, WAC Clearinghouse, Colorado.

Madnani, N. and Cahill, A. (2018). Automated scoring: beyond natural language processing, in: E.M. Bender, L. Derczynski and P. Isabelle (editors), *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1099–1109, Association for Computational Linguistics, Santa Fe, New Mexico, USA.

Mair, C. (2002). Empowering non-native speakers: the hidden surplus value of corpora in continental English departments, in: M. Kettemann and G. Marko (editors), *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000*, volume 42 of *Language and Computers*, Rodopi B.V., Amsterdam, New York.

Maringe, F. and Foskett, N. (2010). *Globalization and internationalization in higher education. Theoretical, strategic and management perspectives*, Continuum International Publishing Group, London.

Markou, M., Stylianou, A., Giannakopoulou, M. and Adamides, G. (2020). Identifying business-to-business unfair trading practices in the food supply chain: the case of Cyprus, *New Medit*, 19(1), 19–34.

Mauranen, A., Hynninen, N. and Ranta, E. (2010). English as an academic lingua franca: The ELFA project, *English for Specific Purposes*, 29(3), 183–190.

Mazzoleni, M. (1996). Un modello di interpretazione testual-proposizionale: la "semantica a scene e cornici", in: *Quaderni di semantica*, volume 33 XVII/1, pages 149–161, Edizioni dell'Orso, Alessadria, Italy.

Mazzoleni, M. (2002). Classificazioni "tipologiche" e classificazioni "generiche" in prospettiva traduttiva, in: M.G. Scelfo (editor), *Le questioni del tradurre: comunicazione, comprensione, adeguatezza traduttiva e ruolo del genere testuale, Atti del Convegno Internazionale (Forlì, 29 novembre - 1 dicembre 2001)*, pages 150–159, Edizioni Associate Editrice Internazionale, Roma.

Mazzoleni, M. and Menin, R. (2011). Traduzione e conoscenze (non) condivise, *Mediazioni, Università di Bologna*, 11, 1–24.

McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*, Cambridge textbooks in linguistics, Cambridge University Press, Cambridge.

McGuigan, G.S. and Russel, R.D. (2008). The business of academic publishing: A strategic analysis of the academic journal publishing industry and its impact on the future of scholarly publishing, *E-JASL 1999-2008 (volumes 1-10)*, page 105.

McKinley, J. and Rose, H. (2018). Conceptualizations of language errors, standards, norms and nativeness in English for research publication purposes: An analysis of journal submission guidelines, *Journal of Second Language Writing*, 42, 1–11.

McKinley, J. and Rose, H. (2019). Standards of English in academic writing: The authors respond, *Journal of Second Language Writing*, 44, 114–116.

McNamara, D.S. and Graesser, A.C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing, in: P. McCarthy and C. Boonthum-Denecke (editors), *Applied Natural Language Processing: Identification, Investigation and Resolution*, chapter 11, pages 188–205, IGI Global.

Melitz, J. (2018). English as a lingua franca: Facts, benefits and costs, *The World Economy*, 41(7), 1750–1774.

Mežek, Š. (2011). Academic writing in a global context: The politics and practices of publishing in English, Theresa Lillis, Mary Jane Curry (Book Review), *Ibérica*, 22, 179–198.

Mikhailov, M. and Cooper, R. (2016). *Corpus Linguistics for Translation and Contrastive Studies: A Guide for Research*, Routledge Corpus Linguistics Guides, Routledge, London.

Mikolov, T. (2010). Recurrent neural network based language model, in: *Interspeech*, volume 2, page 3.

Milroy, J. and Milroy, L. (2012). *Authority in Language: Investigating Standad English*, Routledge.

Mišak, A., Marušić, M. and Marušić, A. (2005). Manuscript editing as a way of teaching academic writing: Experience from a small scientific journal, *Journal of Second Language Writing*, 14(2), 122–131.

Mitkov, R. (2003). *The Oxford handbook of computational linguistics*, Oxford University Press, Oxford, iSBN 0-19-823882-7.

Moon, J., Cho, H. and Park, E.L. (2020). Revisiting round-trip translation for quality estimation, preprint arXiv:2004.13937.

Mordido, G. and Meinel, C. (2020). Mark-Evaluate: assessing language generation using population estimation methods, in: *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1963–1977, International Committee on Computational Linguistics, Barcelona, Spain (Online).

Mossop, B. (2001). *Revising and editing for translators*, Translation Practices Explained, St. Jerome Publishing, Manchester.

Muangkammuen, P., Xu, S., Fukumoto, F., Runapongsa Saikaew, K. and Li, J. (2020). A neural local coherence analysis model for clarity text scoring, in: M. Ptaszynski and B. Ziolko (editors), *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2138–2143, International Committee on Computational Linguistics (ICCL), Barcelona, Spain (Online).

Murphy, A.C. (2012). *Editing specialised texts in English A corpus-assisted analysis II edition*, Traduzione Testi e Strumenti, LED.

Murphy, R. (2008). *English Grammar in Use*, Cambridge University Press, Cambridge.

Nesi, H. (2016). Corpus studies in EAP, in: K. Hyland and P. Shaw (editors), *The Routledge Handbook of English for Academic Purposes*, chapter 16, Routledge, London.

Odlin, T. (1989). *Language Transfer: Cross-linguistic influence in language learning*, Cambridge Applied Linguistics, Cambridge University Press, Cambridge.

Osborne, J. (2000). What can students learn from a corpus?: building bridges between data and explanation, in: B. McEnery (editor), *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the third international conference on Teaching and Language Corpora*, Peter Lang, Frankfurt am Main.

Paetzold, G. and Specia, L. (2016). Collecting and exploring everyday language for predicting psycholinguistic properties of words, in: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1669–1679, The COLING 2016 Organizing Committee, Osaka, Japan.

Pal, S., Naskar, S.K. and van Genabith, J. (2016). Multi-engine and multi-alignment based automatic post-editing and its impact on translation productivity, in: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2559–2570, The COLING 2016 Organizing Committee, Osaka, Japan.

Paltridge, B. and Starfield, S. (editors) (2013). *The Handbook of English for Specific Purposes*, Wiley-Blackwell, Boston.

Paltridge, B., Starfield, S. and Tardy, C.M. (2016). *Ethnographic Perspectives on Academic Writing*, Oxford Applied Linguistics, Oxford University Press, Oxford.

Paolucci, C. (editor) (2007). *Studi di semiotica Interpretativa*, Bompiani, Milano.

Paolucci, C. (editor) (2010). *Strutturalismo e interpretazione*, Bompiani, Milano.

Paolucci, C. (2020). *Persona, soggettività nel linguaggio e semiotica dell'enunciazione*, Bompiani, Milano, ISBN 9788830103337.

Papineni, K., Roukos, S., Ward, T. and Zhu, W.J. (2002). Bleu: a method for automatic evaluation of machine translation, in: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, Association for Computational Linguistics.

Paradowski, M.B. (2013). Barbara Seidlhofer: Understanding English as a lingua franca: A complete introduction to the theoretical nature and practical implications of English used as a lingua franca (review article), *The Interpreter and translator trainer*, 7(2), 312–320.

Pesciarelli, F., Scorolli, C. and Cacciari, C. (2019). Neural correlates of the implicit processing of grammatical and stereotypical gender violations: A masked and unmasked priming study, *Biological Psychology*, 146(107714).

Phillipson, R. (2006a). English, a cuckoo in the European higher education nest of languages?, *European Journal of English Studies*, 10(1), 13–32.

Phillipson, R. (2006b). Figuring out the Englishisation of Europe, in: C. Leung and J. Jenkins (editors), *Reconfiguring Europe: The Contribution of Applied Linguistics*, pages 65–86, Equinox, London.

Pinfield, S. (2015). Making open access work: The "state-of-the-art" in providing open access to scholarly literature, *Online Information Review*, 39(5), 604–636.

Pojanapunya, P. and Todd, R.W. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis, *Corpus Linguistics and Linguistic Theory*, 14(1), 133–167.

Popovic, M. (2021). On nature and causes of observed MT errors, in: *Proceedings of Machine Translation Summit XVIII: Research Track*, pages 163–175, Association for Machine Translation in the Americas, Virtual.

Pospelova, T. (2016). The influence on micro skills development in academic writing in English as a second language, *Journal of Language and Education*, 2(1), 30–38.

Pradhan, N., Gyanchandani, M. and Wadhvani, R. (2015). A review on text similarity technique used in IR and its application, *International Journal of Computer Applications*, 120(9).

Prinz, J.J. (2012). *Beyond human nature: how culture and experience shape the human mind*, Norton, W. W.

Prodromou, L. (2009). English as a lingua franca: a corpus based analysis, *ELT Journal*, 63(1), 78–81.

Rambelli, P. (2021). Course material, università di bologna, unpublished.

Reeder, F., Miller, K., Doyon, J. and White, J. (2001). The naming of things and the confusion of tongues: an MT metric, in: *Workshop on MT Evaluation*, Santiago de Compostela, Spain.

Ricento, T. (editor) (2005). *An Introduction to Language Policy: Theory and Method*, Language and Social Change, Wiley-Blackwell.

Rivera-Trigueros, I. (2021). Machine translation systems and quality assessment: a systematic review, *Lang Resources & Evaluation*.

Rivoltella, P.C. and Rossi, P.G. (2019). *Tecnologie per l'educazione*, Pearson, Torino.

Römer, U. (2009). English in academia: Does nativeness matter?, *Anglistik: International Journal of English Studies*, 20(2), 89–100.

Römer, U. (2011). Corpus research applications in second language teaching, *Annual Review of Applied Linguistic*, 31, 205–225.

Rubino, R., Degaetano-Ortlieb, S., Teich, E. and van Genabith, J. (2016). Modeling diachronic change in scientific writing with information density, in: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 750–761, The COLING 2016 Organizing Committee, Osaka, Japan.

Saberi, D., Lee, J. and Webster, J. (2020). Automatic assistance for academic word usage, in: M. Ptaszynski and B. Ziolko (editors), *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2163–2168, International Committee on Computational Linguistics (ICCL), Barcelona, Spain (Online).

Salton, G. (1971). *The SMART retrieval system-experiments in automatic document processing*, Prentice-Hall, Inc.

Saraceni, M. (2009). Relocating English: towards a new paradigm for English in the world, *Language and Intercultural Communication*, 9(3), 175–186.

Sarnoff, I. (1970). Social attitudes and the resolution of motivational conflict, in: M. Jahoda and N. Warren (editors), *Attitudes*, pages 279–284, Penguin, Harmondsworth.

Schäffner, C. and Adab, B. (2001). The idea of the hybrid text in translation revisited, *Across Languages and Cultures*, 2(2), 227–302.

Schneider, E.W. (2012). *English Around the World: An Introduction*, Cambridge University Press.

Scocchera, G. (2015). Computer-based collaborative revision as a virtual lab of editorial/literary translation genetics, *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 14, 168–199.

Seidlhofer, B. (2009). Accommodation and the idiom principle in English as a Lingua Franca, *Intercultural Pragmatics*, 6(2), 195–215.

Seidlhofer, B. (2011). *Understanding English as a lingua franca*, Oxford University Press, Oxford.

Seidlhofer, B. (2017). Standard English and the dynamics of ELF variation, in: J. Jenkins, W. Baker and M. Dewey (editors), *The Routledge Handbook of English as a Lingua Franca*, chapter 7, Routledge.

Shah, D. (2020). Machine translation with the seq2seq model: Different approaches, Towards data science (online).

Shannon, C.E. (1948). A mathematical theory of communication, *The Bell system technical journal*, 27(3), 379–423.

Shaw, P. (2016). Genre analysis, in: K. Hyland and P. Shaw (editors), *The Routledge Handbook of English for Academic Purposes*, chapter 19, Routledge, London.

Simon, S. (1996). *Gender in Translation*, Routledge, London and New York.

Sinclair, J. and Sinclair, L. (1991). *Corpus, Concordance, Collocation*, Computational linguistics, Oxford University Press, Oxford, UK.

Smith, J.W.T. (1999). The deconstructed journal : a new model for academic publishing, *Learned Publishing*, 12(2), 79–91.

Somasundaran, S., Riordan, B., Gyawali, B. and Yoon, S.Y. (2016). Evaluating argumentative and narrative essays using graphs, in: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1568–1578, The COLING 2016 Organizing Committee, Osaka, Japan.

Sperber, M., Neubig, G., Niehues, J., Stüker, S. and Waibel, A. (2016). Lightly supervised quality estimation, in: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3103–3113, Osaka, Japan.

Stapleton, P. (2019). Standards of English in academic writing: A response to McKinley and Rose, *Journal of Second Language Writing*, 44, 110–113.

Stevenson, M., R., S. and de Glopper, C. (2006). Revising in two languages: A multidimensional comparison of online writing revisions in L1 and FL, *Journal of Second Language Writing*, 15(3), 201–233.

Sugar, C. (editor) (2004). *Home-Work: Postcolonialism, Pedagogy, and Canadian Literature*, Reappraisals: Canadian Writers, University of Ottawa Press.

Sullivan, H. (2013). *The Work of Revision*, Harward University Press.

Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*, Cambridge Applied Linguistics, Cambridge University Press, Cambridge, UK, 1st edition.

Takumi, G., Nagata, R., Mita, M. and Hanawa, K. (2020). Taking the correction difficulty into account in grammatical error correction evaluation, in: M. Ptaszynski and B. Ziolko (editors), *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2085–2095, International Committee on Computational Linguistics (ICCL), Barcelona, Spain (Online).

The Wellcome Trust (2003). Economic analysis of scientific research publishing. A report commissioned by the Wellcome Trust, SQW Limited, Enterprise House, Vision Park, Histon, Cambridgeshire, UK.

Theobald, M., Siddharth, J. and Paepcke, A. (2008). Spotsigs: robust and efficient near duplicate detection in large web collections, in: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 563–570.

Thompson, P. and Diani, G. (editors) (2015). *English for academic purposes: Approaches and implications*, Cambridge Scholars Publishing, Lady Stephenson Library, Newcastle upon Tyne, UK.

Todirascu, A., François, T., Bernhard, D., Gala, N. and Ligozat, A.L. (2016). Are cohesive features relevant for text readability evaluation?, in: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 987–997, The COLING 2016 Organizing Committee, Osaka, Japan.

Tribble, C. (2006). Written in, written out: Who sets the standards for academic writing?, in: A. Usó-Juan E. andMartínez-Flor and C. Tribble (editors), *Current Trends in the Development and Teaching of the four Language Skills*, pages 447–472, De Gruyter Mouton, New York.

Troia, G.A., Olinghouse, N.G., Mo, Y., Hawkins, L., Kopke, R.A., Chen, A., Wilson, J. and Stewart, K.A. (2015). Academic standards for writing: To what degree do standards signpost evidence-based instructional practices and interventions?, *The Elementary School Journal*, 116(2), 291–321.

Turing, A. (1950). Computing machinery and intelligence, *Mind*, 59(236), 433–460.

van der Lee, C., Verduijn, B., Krahmer, E. and Wubben, S. (2018). Evaluating the text quality, human likeness and tailoring component of PASS: A Dutch data-to-text system for soccer, in: *Proceedings of the 27th International Conference on Computational Linguistics*, pages 962–972, Association for Computational Linguistics, Santa Fe, New Mexico, USA.

Vanmassenhove, E., Shterionov, D. and Gwilliam, M. (2021). Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation, in: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2203–2213, Association for Computational Linguistics, Online.

Vasconcelos, S.M.R., Sorenson, M.M. and Leta, J. (2007). Scientist-friendly policies for non-native English-speaking authors: timely and welcome, *Braz. J. Med. Biol. Res.*, 40(6), 743–747.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett (editors), *Advances in neural information processing systems*, volume 30, pages 5998–6008, Curran Associates, Inc.

Ventola, E. (1991). *Functional and Systemic Linguistics*, volume 55 of *Trends in Linguistics. Studies and Monographs*, De Gruyter Mouton, Berlin, New York.

Volansky, V., Ordan, N. and Wintner, S. (2013). On the features of translationese, *Digital Scholarship in the Humanities*, 30(1), 98–118.

von Floton, L. (2017). On the challenges of transnational feminist translation studies, *Traduction, Terminologie, Rédaction (TTR)*, 30(1-2), 173–194.

Wagman, M. (1991). *Cognitive Science and Concepts of Mind: Toward a General Theory of Human and Artificial Intelligence*, New York Praeger.

Wang, X. (2021). Academic English editing in the era of COVID-19: A corpus-assisted case study, Università di Bologna, Laurea Magistrale, Corso di Studio in Specialized Translation [LM-DM270], Forlì, Documento ad accesso riservato.

Widdowson, H.G. (2017). Historical perspectives on ELF, in: J. Jenkins, W. Baker and M. Dewey (editors), *The Routledge Handbook of English as a Lingua Franca*, chapter 8, Routledge.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*, The Massachussets Institute of Technology, Cambridge, Massachussets, USA, 1st edition.

Willey, I. and Tanimoto, K. (2013). "convenience editors" as legitimate participants in the practice of scientific editing: An interview study, *Journal of English for Academic Purposes*, 12(1), 23–32.

Xue, S., Abdel-Hamid, O., Jiang, H., Dai, L. and Liu, Q. (2014). Fast adaptation of deep neural network based on discriminant codes for speech recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(12), 1713–1725.

Yang, Y. (2001). Chinese interference in English writing: Cultural and linguistic differences, Harward Graduate School of Education, Quincy, MA.

Yannakoudakis, H. (2013). Automated assessment of English-learner writing, Technical Report UCAM-CL-TR-842, University of Cambridge, Computer Laboratory, United Kingdom.

Yoon, H. (2008). More than a linguistic reference: the influence of corpus technology on L2 academic writing, *Language Learning & Technology*, 12(2), 31–48.

Zabrodskaja, A. (2012). Crosslinguistic influence in language and cognition. Scott Jarvis and Aneta Pavlenko (2008) New York and London: Routledge. Pp 287. ISBN 0805838856, *Sociolinguistic Studies*, 6(3), 603–607.

Zauberga, I. (2001). Discourse interference in translation, *Across Languages and Cultures*, 2(2), 265–276.

Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q. and Artzi, Y. (2020). Bertscore: Evaluating text generation with BERT, preprint arXiv:1904.09675.

Zhu, Y., Kiros, R. andZemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A. and Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books, in: *Proceedings of the IEEE international conference on computer vision*, pages 19–27.

Zhu, Y., Yue, S., Yu, C. and Shi, Y. (2017). Cept: Collaborative editing tool for non-native authors, in: *CSCW '17: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 273–285.

Appendix A

The English Monolingual Parallel Corpus of Academic Writings and Revisions - readme.txt

DESCRIPTION

This corpus represents the use of English as a Lingua Franca in academic writing for the Italian publishing company Bononia University Press. It is a monolingual parallel corpus because it includes both the raw and the final versions of each text. It includes also a small sub-corpus named 'Files with no revision', that contains texts of which the final version was not available.

Raw and final texts are identified with the same number as filename, but for raw texts the number is followed by '_', whereas for final versions it is followed by '.'.

The corpus is available in three versions. Version 1 includes the texts in .txt format and is composed of the following sub-corpora, each of which has two folders, one with the raw texts and one with the final versions: ArchAlp_N6, containing texts from the international journal ARCHALP n. 6 on the subject of architecture in the Alps region; CIHA, with texts on the subject of history of art from the proceedings of the 2019 CIHA conference; EU-DISASTER RESPONSE LAW, which is a the PhD thesis in European law; The Italian Economy after COVID-19, featuring texts from a volume about economics and sociology written by different authors. Version 2 features texts organised in the same way as version 1, but with useful information encoded as metadata. Version 3 is built as a traditional parallel corpus in the .xls format.

This corpus is 'ad hoc', in the sense that it was compiled for a research purpose – developing a system that automatically evaluates how well an academic text is written, and how much revision it needs in terms of time and specialisation.

The corpus is also meant to be a useful resource for the revisors at BUP, because it constitutes a 'revision memory' of their previous work.

DOMAIN AND GENRE

The corpus contains academic texts with different degrees of specialisation in different domains. The nature of this corpus itself implies the assumption that academic writing has certain charateristics that remains across domains and genres. The corpus includes different text typologies: chapters taken from the same volume, essays, a PhD dissertation, abstracts, information documents and figure captions.

SELECTION CRITERIA

- Texts written for academic reasons by writers with a high degree of expertise in the subject (i.e. academic texts).
- Texts published or that will be published at Bononia University Press.
- Texts edited and revised by revisors at BUP.
- Texts written in English by non-native English speakers.

PREFERENCE CRITERIA

- Italian as the author's native language for Italian texts.
- Genre of texts was not a selection criteria, but a good balance between the types of texts included in the corpus was strongly taken into consideration.

CONSTRUCTION PROCEDURE

Texts were copy-pasted from .docx files.

Texts were cleaned and any irrelevant parts (such as bibliographies and tables containing numeric information) were eliminated. Some descriptions of figures were kept because they were long and contained meaningful stretches of discourse, in this cases a file with the same ID number followed by 'a' was created, containing the figure captions to broaden the possibilities of consultation.

Texts were converted using the any2utf8 tool.

Each file containing a raw text was named with a number shared with the respective final version, so to identify each piece of writing with one identification number.

Texts were aligned manually, using regular expressions.

For example, '\.[ ]' was replaced with '\.\n'; '\n \n' was replaced with '\n'; occurences of '[0-9]+\.\n[0-9]+' (i.e. numeric values) were found and '\n' was delated; 'ch\.\n' was replaced with 'ch\.'; 'A\.\nD\.\n' was replaced with 'A\.D\.'; etc.

TOTAL NUMBER OF DOCUMENTS = 313
TOTAL NUMBER OF TOKENS = 802,070
NUMBER OF TOKENS OF THE RAW TEXTS = 399,417
NUMBER OF TOKENS OF THE FINAL VERSIONS = 399,471

Appendix B

Dataset IE

Table 14: IE Dataset – Complete samples

Text ID	Raw	Raw RTT	Revision	Revision RTT
0 introduction	In March, two weeks after February 21, when Adriano Trevisan was the first victim of the pathogen in Vo' Euganeo, Italy held the daunting title of the country with the most deaths for 23 days.	In March, two weeks after February 21, when Adriano Trevisan in Vo' Euganeo was the first victim of the pathogen, Italy held the frightening title of the country with the most deaths for 23 days.	In March, two weeks after February 21, when Adriano Trevisan was the first victim of the pathogen in Vo' Euganeo, Italy held the daunting title of the country with the most deaths for 23 days.	In March, two weeks after February 21, when Adriano Trevisan in Vo' Euganeo was the first victim of the pathogen, Italy held the frightening title of the country with the most deaths for 23 days.
1 pandemic	The confinement measures in Italy and elsewhere successfully flattened the pandemic curve, allowing the health system to absorb the shock but, as in other countries, they will contribute to the largest contraction of output in modern history.	Containment measures in Italy and elsewhere have successfully flattened the pandemic curve and allowed the health care system to absorb the shock, but as in other countries, they will contribute to the largest decline in output in modern history.	The confinement measures in Italy and elsewhere successfully flattened the pandemic curve, allowing the health system to absorb the shock, but, as in other countries, they will contribute to the largest contraction of output in modern history.	Containment measures in Italy and elsewhere have successfully flattened the pandemic curve and allowed the health care system to absorb the shock, but as in other countries, they will contribute to the largest decline in output in modern history.

Continued on next page

Table 14 – *Continued from previous page*

Text ID	Raw	Raw RTT	Revision	Revision RTT
1a figures	The decline in output is assumed to be: between 50 and 100% in the most severely hit sectors; 30% in the other sectors in the left panel, while 15% in the right panel.	The decline in output is assumed as follows: between 50% and 100% in the most affected sectors, 30% in the other sectors in the left panel, and 15% in the right panel.	The decline in output is assumed to be between 50 and 100% in the most severely hit sectors, 30% in the other sectors in the left panel, and 15% in the right panel.	Output declines are assumed to be between 50% and 100% in the most affected sectors, 30% in the other sectors in the left panel, and 15% in the right panel.
2 demography	This chapter aims at collecting the current knowledge, inevitably based on limited data, and indicating what can be expected in the case of Italy.	The aim of this chapter is to compile the current state of knowledge, which is inevitably based on limited data, and to show what can be expected in the case of Italy.	This chapter aims at collecting the current knowledge, inevitably based on limited data, and indicating what can be expected in the case of Italy.	The aim of this chapter is to compile the current state of knowledge, which is inevitably based on limited data, and to show what can be expected in the case of Italy.
10 banking	In recent years, Italian banks have also greatly improved their asset quality as a result of the significant regulatory tightening on impaired loans promoted from 2014 and implemented according to the 2017 ECB Guidelines on the management of Non-Performing Exposures (NPE) in 2017, the Addendum of 2018 and the recent Calendar Provisioning.	In recent years, Italian banks have also significantly improved their asset quality, due to the significant tightening of impaired loan rules promoted since 2014 and implemented in line with the ECB's 2017 Guidelines for the Management of Non-Performing Loans (NPE), the 2018 Addendum and the recent calendar provision.	In recent years, Italian banks have also greatly improved their asset quality as a result of the significant regulatory tightening on impaired loans promoted since 2014 and implemented according to the 2017 ECB Guidelines on the management of Non-Performing Exposures (NPEs) in 2017, the Addendum of 2018 and the recent Calendar Provisioning.	In recent years, Italian banks have also significantly improved their asset quality, due to the significant tightening of the prudential rules on non-performing loans, promoted since 2014 and implemented in accordance with the 2017 ECB Guidelines for the management of non-performing loans (NPEs), the 2018 Addendum and the recent calendar provision.

Continued on next page

Table 14 – *Continued from previous page*

Text ID	Raw	Raw RTT	Revision	Revision RTT
13 trade	Inadequate policy coordination, insufficient preventive stockpiling of critical equipment, and lack of cooperation in implementing responses were almost the rule, at the global level but also within the European Union.	Poor policy coordination, inadequate preventive stockpiling of critical equipment, and a lack of cooperation in implementing measures were almost the norm, both at the global level and within the European Union.	Inadequate policy coordination, insufficient stocks of critical equipment, and lack of cooperation in implementing measures were almost the norm at the global level, but also within the European Union.	Inadequate policy coordination, insufficient stocks of critical equipment, and lack of cooperation in response have been almost the norm at the global level, but also within the European Union.
17 why	Since the seminal work of Schumpeter, entrepreneurs and their distinctive characteristics have attracted theoretical and empirical attention in economics, psychology, sociology, management, and finance.	Since the seminal work of Schumpeter, entrepreneurs and their distinctive characteristics have attracted theoretical and empirical attention from economics, psychology, sociology, management, and finance.	Since the seminal work of Schumpeter, entrepreneurs and their distinctive characteristics have attracted theoretical and empirical attention in economics, psychology, sociology, management, and finance.	Since the seminal work of Schumpeter, entrepreneurs and their distinctive characteristics have attracted theoretical and empirical attention from economics, psychology, sociology, management, and finance.
117 preface	Public intervention has been on restoring strong demand, adopting short-term, defensive measures, in the face of the huge drop in production and employment, but also longer-term measures, such as the State acquiring shares in companies affected by the crisis on a temporary basis to support the transition.	The public sector has focused on restoring strong demand and has taken short-term defensive measures in the face of massive declines in output and employment, as well as longer-term measures such as the temporary acquisition by the state of stakes in crisis-hit companies to support the transition.	Public intervention has been on restoring strong demand, adopting short-term, defensive measures, in the face of the huge drop in production and employment, but also longer-term measures, such as the State acquiring shares in companies affected by the crisis on a temporary basis to support the transition.	The public sector has focused on restoring strong demand and has taken short-term defensive measures in the face of massive declines in output and employment, as well as longer-term measures such as the temporary acquisition by the state of stakes in crisis-hit companies to support the transition.