

A corpus-based study of opinions of advocates general of the court of justice of the European Union

Mattioli, Virginia; McAuliffe, Karen

DOI:
[10.1515/ijld-2021-2047](https://doi.org/10.1515/ijld-2021-2047)

License:
None: All rights reserved

Document Version
Peer reviewed version

Citation for published version (Harvard):
Mattioli, V & McAuliffe, K 2021, 'A corpus-based study of opinions of advocates general of the court of justice of the European Union: changes in language and style', *International Journal of Legal Discourse*, vol. 6, no. 1, pp. 87-111. <https://doi.org/10.1515/ijld-2021-2047>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

A Corpus-Based Study of Opinions of Advocates General of the Court of Justice of the European Union: Changes in Language and Style

Virginia Mattioli, Pontificia Universidad Católica de Valparaíso

Karen McAuliffe, University of Birmingham*

1. Introduction

This paper presents a Corpus Linguistics study of lexical features in Opinions of Advocates General (AGs) of the Court of Justice of the European Union (CJEU).¹ The study aims, primarily, to explore the language of some AGs' Opinions, before and after the introduction of changes in the CJEU's linguistic regime relating to the language(s) in which Opinions should normally be drafted. The study is interdisciplinary insofar as it incorporates concepts and methodologies from three main fields: the conceptual framework is drawn from questions arising in the context of law and language studies on the workings of the CJEU; the concepts supporting the methodological choices are based in translation studies (Universals of Translation); and the methodology itself is grounded in corpus-based applied linguistics/corpus linguistics.

The results of the analysis carried out in this study demonstrate that certain changes in the linguistic and stylistic nature of AGs' Opinions can be observed post-2004. On the one hand, those changes corroborate the study's primary hypothesis that AG Opinions drafted after 2004 in non-mother tongue languages are stylistically simpler and less 'fluent' than those

* The authors would like to thank the anonymous reviewer(s) of this volume for their detailed and constructive feedback. The research for this paper was generously funded by the European Research Council under the FP7 Starting Grant funding scheme (project number 313353 – www.llecj.karenmcauliffe.com).

¹ This study forms part of a larger project investigating the impact of multilingualism and translation on the case law of the CJEU, *The Law and Language at the European Court of Justice (LLECJ) Project*. See www.llecj.karenmcauliffe.com for further information.

drafted (in AGs' mother tongues) before 2004. On the other hand, and perhaps more interestingly, however, the results also indicate that AG Opinions drafted after 2004 in mother tongue languages are similarly becoming simpler and less 'fluent'. These results are inherently interesting in terms of Corpus Linguistics research. However, in order to have a value outside of that field, they are best considered as a basis for more nuanced research questions, which can be investigated through interdisciplinary methods taking account of the factors of production of AG Opinions.

2. Conceptual framework

The conceptual framework set out here is developed from two theoretical bases: scholarship on language and multilingualism at the CJEU, which shines a light on the otherwise invisible processes and institutional culture within that court that have a significant impact on the jurisprudence that it produces; and linguistic theories on the universals of translation (UT), those specific features which characterize translations vis-à-vis original/source language texts. In the study set out in this paper, UT are used as a point of departure to identify the linguistic characteristics of AGs' Opinions produced in the multilingual context of the CJEU.

2.1 Language and Multilingualism at the CJEU

The CJEU is a multilingual court, producing case law in up to 24 different languages, which is applicable throughout the 27 Member States of the European Union (EU). In the hybrid environment of the CJEU, issues of language, multilingualism and translation are at the forefront of the development of its jurisprudence. The CJEU comprises many departments, and various actors are involved in the production of its jurisprudence. This paper focuses only on Advocates General, and specifically on the Opinions they produce.

There are currently 11 AGs at the CJEU. The role of the AG is to deliver a reasoned Opinion on a case, prior to deliberations on and delivery of the final judgment. Opinions are not

delivered in every case before the CJEU. Article 20 of the Statute of the CJEU allows that Court to determine a case without an AG's Opinion where no new points of law are raised.² However, the role of the AG is an important one insofar as the development of the CJEU's case law is shaped and guided by Opinions of AGs. Whereas CJEU judgments tend to be expressed in formulaic 'Lego-like' language (McAuliffe, 2011), AG Opinions are more prone to being discursive, speculative, almost academic-style documents. One of the main reasons for such a discursive style is that, historically, AG Opinions were drafted in the language of the respective AG (Borgsmidt, 1988), meaning that AGs were not constrained in any way by language or a formulaic style of drafting. However, since 2004, the CJEU has requested AGs to draft their Opinions in one of the five 'pivot languages' of that Court (French, English, German, Spanish and Italian – and, since 2018, Polish). These 'pivot' languages were introduced at the CJEU as a result of the 2004 EU enlargement in order to reduce the number of possible translation combinations and thus to reduce the translation workload. This change in convention relating to the drafting language(s) of AGs' Opinions means that some AGs now draft in a non-mother tongue language.³ Consideration of the linguistic aspect of the role of AGs raises the question of whether that change in convention regarding drafting language(s) may affect the style of some AG Opinions.

2.2 Universals of translation

Theories of universals of translation first emerged in the 1990s, proposed by Mona Baker in her 1993 analysis of translated text corpora. Baker claimed certain features could be considered universals of translations owing to their intrinsic relation to translated texts (vis-à-vis original texts), independent from both source and target languages (Baker, 1993: 242-247). That theory has been further developed and, today, the main universals of translation can be summarized as follows:

² Treaty on the Functioning of the European Union Protocol (No 3) On the Statute of the Court of Justice of the European Union, OJ C 202, 7/6/2016, p. 210-229.

³ It must be noted that the change in convention regarding drafting language affects only some AGs. Many (both permanent and non-permanent) AGs continue to draft Opinions in their mother tongue.

- I. Explicitation: translated texts appear to be more explicit than original/source texts on a number of different linguistic levels. From a syntactic point of view, for example, in the case of sentences introduced by the reporting verbs *tell* and *say*, translated texts demonstrate a greater use of the *that*-connective instead of the zero-connective preferred in original texts (Olohan & Baker, 2000).
- II. Simplification: the language of a translated text is always simpler than the one used in texts originally written in the relevant target language, at the levels of syntax, lexical features and style (Zanettin, 2013: 3; Xiao & Yue, 2009:250). According to Laviosa (1998), the simplicity of a text can be assessed through its lexical variety and density, together with its 'readability'. With the same aim, Puurtinen (2003:395) considers texts' 'speakability', defined as "the ease of reading aloud".
- III. Normalization/conventionalization: translations demonstrate a "tendency to conform to target language rather than source language patterns and norms, producing more conventional rather than unusual target strings" (Zanettin, 2013: 4). Again, this feature can be related to each language level: lexically (as to creativity and formality), syntactically (concerning distribution of typical features) and semantically (according to the range of lexis referred to a certain concept) (Zanettin, 2013:4).
- IV. Avoidance of repetitions: in translated texts repetitions tend to be avoided and substituted with synonyms. This feature that Baker (1998: 289) considers as an aspect of stylistic simplification, is, according to Toury (1991: 188) "one of the most persistent, unbending norms in translation in all languages studied so far".
- V. Interference/transfer: translated texts present a tendency to transfer phenomena typical of the source texts to the target ones. This feature, already well-known in translation and language acquisition studies, was first described as a translation universal by Toury (1995:275). Interference/transfer can occur on either a positive or negative level, and depends on the translator's mental process involved in translation (Baker, 1998: 291).
- VI. Translation of unique items: elements that are not typical in a source language culture are not frequently found in translations, even where those elements are typical in the target language culture (Zanettin, 2013:4). The resulting absence of such unique items

in translated texts has led to some authors referring to this particular universal of translation as ‘underrepresentation’ (Xiao & Yue, 2009:251; Zanettin, 2013:4).

The introduction of the theory of universals of translation was controversial and divided translation scholars into two separate camps: those who maintain the validity of the theory (e.g. Baker, 1993; Toury, 2004; Chesterman, 2004; Xiao & Yue, 2009) and those who contest the existence of features common to all translated texts (e.g. Tymockzco, 1998). Critics of the theory highlight the impossibility of the existence of universal prototypes and descriptions generally (Tymockzco, 1998). However, such criticism is countered by reference to its statistical or representative character: just because a feature is considered a universal of translation does not mean that it must be identified in *all* existing translations, but merely in a predominant percentage of cases (Toury, 2004).

The texts considered in this study are not translated texts in the sense that they have undergone a formal translation process. However, as non-mother tongue/second language texts they inevitably demonstrate translational properties since “[second language] writing is never detachable from translation” (Lee, 2018: 201) and features of translation, as a sort of third code occupying an in-between state between two languages, are not unique to translations only, but can be found in texts produced under conditions of bi- or multilingual language activation (Kruger, 2018: 9). It is therefore reasonable to expect to observe markers for universals of translation in such texts, and fewer such markers in mother tongue texts.

3. The present study

The overarching aim of the study considered here is to explore whether the changes to the conventions relating to AGs’ drafting languages, introduced in 2004, have had an impact on the linguistic and stylistic nature of the relevant AGs’ Opinions. Although language is rarely explicitly addressed in scholarship on AGs’ Opinions, most EU legal scholars tend to accept that language and linguistic tactics are important; moreover, that the discursive, speculative nature of AGs’ Opinions is due in no small part to the fact that historically those documents were drafted in the language of the relevant AG, and can therefore be expressed in their own

individual and eloquent style (Borgsmid, 1988).⁴ If the style of AGs' Opinions change, that may well have consequences for the impact of those Opinions, on CJEU case law and EU legal scholarship more generally. While it is impossible to satisfactorily explore that overarching aim using one specific research method,⁵ Corpus Linguistics allows us to interrogate the texts of Opinions pre- and post-2004 and search for markers that may highlight stylistic changes. This paper sets out the Corpus Linguistics study that was carried out in order to identify such markers.

The hypothesis on which this study is based is: AG Opinions drafted after 2004 in non-mother tongue languages are simpler and less fluent than those drafted (in AGs' mother tongues) before 2004. And in order to assess that hypothesis, the following research questions were devised:

- i. Which linguistic features represent 'fluency' and stylistic simplicity/complexity in a text?
- ii. How can such features be identified in the relevant Opinions?
- iii. Do Opinions drafted after 2004 present more features related to stylistic simplicity and less representing fluency than those drafted before 2004?

On the basis of those research questions, the study itself could then be set out:

- i. Compiling the relevant corpora of Opinions.
- ii. Determining the linguistic features related to stylistic simplicity and fluency
- iii. Searching for those features in each corpus
- iv. Comparing the results obtained from the relevant corpora in terms of stylistic complexity and fluency.

The first two steps are set out in detail in the methodology section below. The final two are then discussed in terms of analysis of the data obtained through such methodology.

⁴ It must be noted that the change in convention regarding drafting language affects only some AGs. Many (both permanent and non-permanent) AGs continue to draft Opinions in their mother tongue.

⁵ Such a discussion is beyond the scope of the present paper. Cf McAuliffe, Muntean and Mattioli, 2021.

4. Methodology

4.1 Corpus compilation

The corpora designed for this study included two sets of texts: all Opinions drafted in English (EN_OPINIONS) and French (FR_OPINIONS), from 1993-2003 and 2005-2015 i.e. ten years either side of 2004, when the new language convention was introduced at the CJEU.⁶ All relevant Opinions were retrieved and downloaded from the EUR-Lex website.⁷ The texts were then grouped according to the language of drafting, the mother tongue of the relevant AG, and the year of production.⁸

Consequently, each corpus (EN_OPINIONS and FR_OPINIONS) is divided into three comparable subcorpora that distinguish the Opinions from two different points of view: diachronic (before/after 2004) and linguistic (drafted in a native/non-native language). The three subcorpora resulting from such a distinction include:

- i. Opinions drafted in the mother tongue of the relevant AGs between 1993 and 2003
- ii. Opinions drafted in a non-mother tongue language between 2005 and 2015
- iii. Opinions drafted in the mother tongue of the relevant AGs between 2005 and 2015

During the compilation process an unexpected issue arose, which required the sampling procedure to be revisited. Far fewer Opinions were drafted, in the relevant languages, in the ten years before 2004 than in the 10 years after 2004. Consequently, in order to maintain the balance of the corpora, the pre-2004 sample was compiled from Opinions drafted between 1988 and 2003.

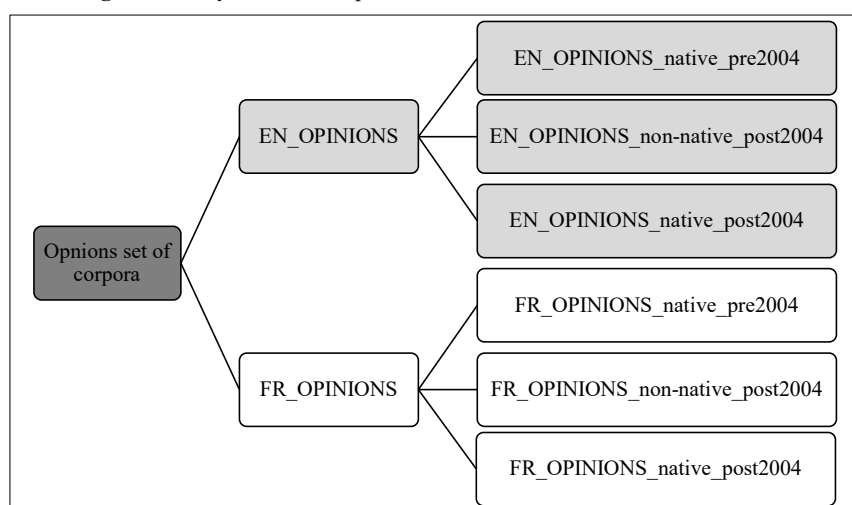
⁶ Opinions drafted in 2004 were excluded to ensure the selection of only those Opinions from the period before the introduction of the language convention (1993-2003) and once that convention had been embedded in the workings of the CJEU (2005-2015).

⁷ <https://eur-lex.europa.eu/homepage.html?locale=es>

⁸ The topics on which Opinions were written were not taken into account in the sampling process for two reasons: first, since we were analysing language use and language change in general, it was appropriate to ensure that *all* Opinions from the relevant time periods were included; and second, since the subject matter of AGs' Opinions is always dependent on the legal issues in the relevant case before the CJEU.

Once compiled, the Opinions were converted into .html or .txt. The archive of corpora resulting from the compilation process is represented in Figure 1.

Figure 1: analyzed set of corpora



The archive of corpora represented in Figure 1 is formed by 2 078 Opinions, including 82 092 types (the number of different words included in the corpus) and 18 135 987 tokens (the total number of words included in the corpus) divided among the different corpora and subcorpora as follows:

OPINIONS_EN: 984 Opinions; 42 797 types; 7 967 766 tokens

OPINIONS_EN_MT_pre2004: 341 Opinions; 21 943 types; 1 914 195 tokens

OPINIONS_EN_nonMT_post2004: 367 Opinions; 27 157 types; 3 319 145 tokens

OPINIONS_EN_MT_post2004: 276 Opinions; 23 519 types; 2 734 426 tokens

OPINIONS_FR: 1 094 Opinions; 55 667 types; 10 168 221 tokens

OPINIONS_FR_MT_pre2004: 410 Opinions; 28 588 types; 2 591 349 tokens

OPINIONS_FR_nonMT_post2004: 345 Opinions; 34 699 types; 3 775 458 tokens

OPINIONS_FR_MT_post2004: 339 Opinions; 33 952 types; 3 801 414 tokens

4.2 Determination of linguistic features of stylistic simplicity and fluency.

Two factors support the application of the theory of UT to the texts in our corpora. First, as discussed in Section 3 above, non-mother tongue texts have important similarities with translated texts, as both are “characterized by diverse communicative constraints” as a result of linguistic mediation between two codes (Kruger, 2018). Second, the UT of simplification is highly relevant to the initial hypothesis set out in Section 3 above. Once the linguistic features which represent the characteristic of simplification were identified, a corpus linguistic methodology was designed in order to identify those features in the relevant corpora.

The features usually related to the characteristic of simplification are: lexical variety, lexical density, and the presence of hypotactic structures (as a syntactic representation of the relationship within the sentences) (Baker, 1993; Laviosa, 1998, and Baroni & Bernardini, 2003). Fluency, on the other hand, appears to be related to sentence length (Tai, 2015 and Xiao & Yue, 2009: 250). According to Translation Studies scholars, the greater the presence of such determined features in the texts, the higher the degree of stylistic complexity and fluency. Thus, identifying the four main features of lexical variety, lexical density, length of sentences, and presence of hypotactic structures would tend to corroborate the initial hypothesis. In addition, each subcorpus was examined further in order to explore the existence of other features related to simplicity or to translated texts. Namely: repetitions⁹, non-finite clauses, and suspending periods.¹⁰ Finally, the lexical specialization of each subcorpus was assessed by determining the idiosyncrasy of the specialized vocabulary used. It must be noted, however, that those additional examinations were only a first attempt to explore the possibility of similarity between translated texts and Opinions drafted in a second language, as for features related to simplicity. Further studies are needed to corroborate the initial outcomes.

5. Analysis

Analysis of each subcorpus comprised twelve steps:

⁹ According to Baker (1998), repetitions should be less frequently observed in translated texts as a consequence of their stylistic simplification.

¹⁰ According to Vanderauwera (1985), the minor use of such features is a common process employed by translators to simplify syntactic structures.

- i. Creation of a frequency list.
- ii. Calculation of the lexical variety of the corpus departing from the number of tokens and types, using the Standardized Types/Tokens Ratio formula, provided by WordSmith Tools (Scott, 2017)¹¹.
- iii. Annotation of the average sentence length of each subcorpus.
- iv. Distinction between the lexical and the functional words among the items with a frequency equal or higher than 10 occurrences, through a manual selection.
- v. Calculation of the lexical density through the relationship between lexical/functional words.
- vi. Observation and registration of the number of occurrences of the most frequent items.
- vii. Manual selection of the specialized legal items among the lexical words with a frequency equal or higher than 500 occurrences.
- viii. Search for the most frequent conjunctions in English and French (according to the exhaustive lists compiled by the Linguistics and English Language of the University of Edinburgh¹² and by the “Études Littéraires” website¹³, respectively). In both languages, the conjunction *that* (and its French correspondence *que*) was not considered in the analysis because it has several functions.
- ix. Search in the concordance list for each subordinate conjunction included in the list followed by a non-finite verb. This step differs for the English and the French subcorpora. In the English sets of Opinions, the non-finite verbs were identified by looking for each conjunction followed by the non-finite verbs ending in *-ed* and *-ing* or the ones introduced by *to*. In French, the non-finite verbs are infinitive, gerund and participle, hence the regular forms of such tenses could be automatically searched in the concordance list.
- x. Looking for the suspension points in the concordance list, in order to know the number of suspending periods.

¹¹ The formula and the variables used are explained at http://www.lexically.net/downloads/version5/HTML/?type_token_ratio_proc.htm

¹² <http://www1.msje.edu/apps/pub.asp?Q=1702>

¹³ <https://www.etudes-litteraires.com/grammaire-conjonctions.php>

- xi. Looking for the most frequent expressions introducing explicitation, to assess the degree of explicitation of the subcorpora in analysis. Again, this analysis is realized differently for the two sets of texts, according to language. In the case of English, the search includes *that means, example, for example, such as* and *like as*, whereas for French *c'est-a-dire, à savoir, comme* (only when it is not used at the beginning of the clause as a conjunction) *exemple* and *par exemple*.
- xii. Creation of a keyword list comparing non-mother tongue Opinions vis-à-vis mother tongue Opinions in each corpus (for example, OPINIONS_EN_MT_pre2004 is compared to OPINIONS_EN_nonMT_post2004). The functional words, lexical words and specialized legal words were then manually distinguished.

Each of these steps was applied to each subcorpus and the results compared across different subcorpora of each language (English/French)¹⁴

7. Results

The results obtained following the methodology described above are set out as follows (i) primary results (those required to corroborate or refute the initial hypothesis); (ii) diachronic results, permitting observation of changes in the style of Opinions during the analyzed periods; and (iii) qualitative results, which introduce new questions about the nature and characteristics of Opinions.

7.1 Primary results: corroboration of the initial hypothesis

¹⁴ To compare the frequency of the items identified across subcorpora of different sizes, a Log Likelihood (LL) calculation was used: the higher the LL value, the more significant the difference between two frequency scores. This study accepted the minimum LL value of 6.63, representing a significant difference of $p < 0.01$. The statistical significance was calculated using the LL calculator developed by the UCREL research group of the University of Lancaster, available at <http://ucrel.lancs.ac.uk/llwizard.html>. It must be noted that, in order to allow comparison between corpora, the LL value can be calculated only considering the on the basis of the tokens in a corpus (the total number of words used in the corpus) so that we can based only on this value to compare the corpora. As a consequence Thus, in the tables shown in the following pages below, the results across different groups of texts can be compared only on the basis of LL values (relating to differences in tokens across referred to the difference of tokens across corpora and subcorpora) can be used to compare the results across the different groups of texts.; whereas the quantity of types are absolute numbers that can be considered only within the corpus or subcorpus to which they belong.

Commented [A1]: I added the explanation of the calculation of LL value in the footnote (we need it for explications in the next pages and footnotes)... is it clear?

Commented [A2R1]: I've just tweaked the wording for English – can you check to make sure it still makes sense?!

The primary results were obtained by analysing the four main features related to text complexity and fluency (steps (i-v) and (viii) above). The results corroborate the initial hypothesis, demonstrating that Opinions drafted in a non-mother tongue language present a simpler style and less fluency than those drafted in an AG's mother tongue. In fact, when compared with mother tongue Opinions, those produced in a second language demonstrate three of the four features related to simplicity: lower lexical variety, shorter sentences and fewer hypotactic and complex structures. Tables 1, 2 and 3 show the quantitative data corresponding to these results.

Table 1: Lexical variety of each analyzed subcorpus, expressed in Standardized Types/Tokens Ratio

| OPINIONS_EN | | | OPINIONS_FR | | |
|-------------|----------------|-------------|-------------|----------------|-------------|
| MT_pre2004 | nonMT_post2004 | MT_post2004 | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| 31.87 | 29.23 | 30.04 | 32.96 | 31.66 | 32.11 |

Table 2: Mean sentence length of each analyzed subcorpus

| OPINIONS_EN | | | OPINIONS_FR | | |
|-------------|----------------|-------------|-------------|----------------|-------------|
| MT_pre2004 | nonMT_post2004 | MT_post2004 | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| 33.63 words | 52.01 words | 55.07 words | 33.92 words | 57.32 words | 64.13 words |

As Table 2 demonstrates, there is an evident chronological decrease of the mean sentence length before and after 2004. However, analysis of texts drafted after 2004 demonstrates longer sentences in mother tongue Opinions than in non-mother tongue Opinions.

Table 3: Conjunctions identified in each analyzed subcorpus

| Type of conjunction | OPINIONS_EN | | | OPINIONS_FR | | |
|---------------------|-------------|----------------|-------------|-------------|----------------|-------------|
| | MT_pre2004 | nonMT_post2004 | MT_post2004 | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| | 004 | 2004 | 004 | 004 | 2004 | 004 |

| | | | | | | |
|-----------------------------------|---|------------------------------|--|--|------------------------------|--|
| Total conjunctions | Types: 31 Tokens: 125 245 LL: + 362.41 | Types: 31 Tokens: 202 770 | Types: 31 Tokens: 174 458 LL: + 176.18 | Types: 26 Tokens: 76 551 LL: + 131.35 | Types: 27 Tokens: 105 617 | Types: 25 Tokens: 110 153 LL: + 53.75 |
| Coordinate conjunctions | Types: 6 Tokens: 68 478 LL: - 3.10 | Types: 6 Tokens: 119 745 | Types: 6 Tokens: 106 623 LL: + 2034992.6 2 | Types: 7 Tokens: 58 600 LL: - 3.62 | Types: 7 Tokens: 86 252 | Types: 7 Tokens: 88 724 LL: +19.25 |
| Subordinate conjunctions | Types: 25 Tokens: 56 767 LL: + 965.49 | Types: 25 Tokens: 83 025 | Types: 25 Tokens: 67 835 LL: - 2.58 | Types: 19 Tokens: 17 951 LL: + 835.03 | Types: 20 Tokens: 19 365 | Types: 18 Tokens: 21 429 LL: + 90.81 |
| Relative subordinate conjunctions | Types: 5 Tokens: 17 720 LL: + 531.44 | Types: 7 Tokens: 24 435 | Types: 7 Tokens: 22 550 LL: + 150.78 | Types: 16 Tokens: 5 485 LL: + 482.34 | Types: 17 Tokens: 5 222 | Types: 15 Tokens: 6 771 LL: + 190.15 |

Table 3 demonstrates that mother tongue Opinions present a greater quantity of conjunctions than non-mother tongue Opinions (as demonstrated by the positive LL values. The only negative values are lower than 6.63, the critical value accepted in this study, and the differences are therefore not significant)¹⁵. Since each conjunction introduces a sentence, a lower number of conjunctions indicates a lower use of coordinate and subordinate clauses, i.e. less complex structures, and a greater use of simpler independent unrelated sentences. In particular, non-mother tongue Opinions demonstrate a smaller quantity of relative conjunctions and consequently, of relative clauses.

The only feature that does not corroborate the initial hypothesis is the lexical density, as set out in Table 4.

¹⁵ In LL calculation, the non-MT_post2004 subcorpus of each language was used as reference to calculate the over or underuse of each type of conjunction in the subcorpora of mother tongue Opinions. Consequently, in Table 3 and the following tables LL values appear only in the columns representing the mother tongue Opinions indicating the difference between MT_pre2004/nonMT_post2004 and between MT_post2004/nonMT_post2004, respectively. Even though the statistical significance is always a positive number, a positive or a negative sign preceding the LL value in each column indicates the overuse or the underuse of the analyzed item in the relevant subcorpus in relation to nonMT_post2004.

Table 4: Lexical density of each analyzed subcorpus.

| OPINIONS_EN | | | OPINIONS_FR | | |
|-------------|----------------|-------------|-------------|----------------|-------------|
| MT_pre2004 | nonMT_post2004 | MT_post2004 | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| 4 | 04 | 4 | 4 | 04 | 4 |
| 0.48 | 0.51 | 0.50 | 0.49 | 0.49 | 0.49 |

These results demonstrate that lexical density does not follow any fixed pattern but is similar across the subcorpora. The only variation in lexical density is a small variation that can be observed only in the English language corpus (greater in non-mother tongue Opinions than in mother tongue Opinions). These data, taken together with the divergences between the two analyzed languages, indicates that the lexical density of Opinions is not related to the mother tongue of the relevant AG.

The greater level of simplicity of non-mother tongue Opinions (demonstrated by three of the four examined linguistic features) is further supported by results obtained by the additional analysis carried out (see Section 4.2 above), in particular in relation to the English language corpus. Other features related to syntactic simplicity, such as a low quantity of non-finite clauses, were identified in non-mother tongue Opinions (see Tables 5 and 6 below).

Table 5: non-finite clauses identified in English language subcorpora

| Type of non-finite clause | EN_OPINIONS | | |
|-----------------------------------|----------------------|----------------|-----------------------|
| | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| Total non-finite clauses | 7 133 LL: + 0.80 | 12 204 | 11 393 LL: + 91.91 |
| -ING non-finite clause | 4 304 LL: + 1.42 | 7 294 | 7 236 LL: - 308.90 |
| -ED non-finite clauses | 1 450 LL: - 25.49 | 2 952 | 1 997 LL: - 46.77 |
| TO + base form non-finite clauses | 1 379 LL: +31.89 | 1 958 | 2 160 LL: +87.66 |
| Subordinate non-finite clauses | 3 666 LL: - 4.51 | 6 640 | 5 446 LL: - 0.06 |

| | | | |
|---------------------------------|----------------------|-------|----------------------|
| Coordinative non-finite clauses | 3 467 LL: + 12.71 | 5 564 | 5 947 LL: +194.87 |
|---------------------------------|----------------------|-------|----------------------|

As shown in the Table 5, English language non-mother tongue Opinions demonstrate a lower quantity of non-finite clauses. Specifically, this difference appears to be due to a greater presence of the structures *to + base form* in Opinions drafted in a second language than in mother tongue Opinions, as demonstrated by the data related to the three kinds of non-finite clauses analyzed. This difference is particularly evident in coordinate clauses.

Table 6: non-finite clauses identified in each French analyzed subcorpus

| Type of non-finite clause | FR_OPINIONS | | |
|-------------------------------|------------------------|----------------|-------------------------|
| | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| Total non-finite clauses | 16 435 LL: - 826.62 | 31 466 | 18 718 LL: - 3362.00 |
| Gerund non-finite clauses | 872 LL: - 326.70 | 2 502 | 3 678 LL: + 217.17 |
| Infinitive non-finite clauses | 14 146 LL: -607.26 | 26 557 | 13 264 LL: - 4615.32 |
| Participle non-finite clauses | 1 417 LL: - 21.28 | 2 407 | 1 776 LL: - 99.92 |

In the French language corpus, the total number of non-finite clauses is greater in non-mother tongue Opinions than in mother tongue Opinions. Among those non-finite clauses, only the quantity of those that indicate gerund verbs is lower in the former than in the latter. Moreover, the statistical significance of the difference in quantity of gerundial clauses between mother tongue and non-mother tongue Opinions increases with time. This demonstrates that fewer gerund non-finite clauses are found in non-mother tongue Opinions than in mother tongue Opinions after 2004.

In relation to the English language corpus, the data demonstrate that non-mother tongue Opinions present a higher quantity of tokens (statistically significant) when comparing lexical words with a frequency higher than 500 in each subcorpus. These results are presented in Table 7.

Table 7: most frequent lexical words of each analyzed English subcorpus

| EN_OPINIONS | | | |
|--|--|---------------------------------|---|
| | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| Total lexical words | 6 091 types 872 152 tokens LL: +353.76 | 7 702 types 1 474 300 tokens | 6 824 types 1 212 559 tokens LL: - 1.89 |
| Lexical words with a frequency higher than 500 | 349 types 455 178 tokens LL: - 8851.73 | 588 types 927 474 tokens | 498 types 718 350 tokens LL: - 1545.93 |
| STTR | 31.87 | 29.23 | 30.04 |

These data imply a high quantity of repetitions. Calculation of lexical variety corroborates that finding. In fact, the Standardized Types/Tokens Ratio is higher in the subcorpus representing mother tongue Opinions than in the subcorpus representing non-mother tongue Opinions. According to Laviosa (1998), the tendency towards repetitions is a further characteristic of the simpler syntax found in translated texts. Thus, these data, not only support the main hypothesis of the present study, but also suggest the possibility of an effective similarity, as for further features, between translations and texts produced in a second language.

7.2 Diachronic results

From a diachronic perspective the comparisons of the LL values obtained contrasting the data resulting from the analysis of the three corpora demonstrate that mother tongue Opinions are gradually becoming more similar to non-mother tongue Opinions in terms of style and fluency. Specifically, this gradual stylistic assimilation is indicated by an increasing number of frequent lexical words (a frequency higher than 500 occurrences), and a decreasing number of relative clauses, of lexical variety and of lexical words. After 2004, the amount of lexical words becomes respectively smaller than before that date and equal to that in non-mother tongue Opinions. Moreover, the English language subcorpora also show a particularly evident gradual decrease of the number of lexical words and of suspending periods in mother tongue Opinions. (see Table 8 below).

Table 8: diachronic change of some of the analyzed features¹⁶

| Feature | OPINIONS_EN | | | OPINIONS_FR | | |
|--|---------------------------------|---|---|-----------------------------------|--|--|
| | MT_pre2004 | nonMT_post2004 | MT_post2004 | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| Lexical words with a frequency higher than 500 | Types: 349 Tokens: 455 178 | Types: 588 Tokens: 927 474 LL: + 8851.73 | Types: 498 Tokens: 718 350 LL: + 2786.06 | Types: 471 Tokens: 667 555 | Types: 639 Tokens: 993 LL: +179.87 | Types: 635 Tokens: 1 003 186 LL: +233.52 |
| Lexical variety | 31.87 | 29.23 | 30.04 | 32.96 | 31.66 | 32.11 |
| Relative clauses | Types: 5 Tokens: 17 720 | Types: 7 Tokens: 24 435 LL: - 531.44 | Types: 7 Tokens: 22 550 LL: - 131.88 | Types: 16 Tokens: 485 | Types: 17 Tokens: 5 222 LL: - 482.34 | Types: 15 Tokens: 6 771 LL: - 88.70 |
| Lexical words | Types: 6 091 Tokens: 872 152 | Types: 7 702 Tokens: 1 474 300 LL: - 353.76 | Types: 6 824 Tokens: 1 212 559 LL: - 372.09 | Types: 8 425 Tokens: 1 197 342 | Types: 9 671 Tokens: 1 623 641 LL: - 3536.62 | Types: 9 674 Tokens: 1 621 942 LL: - 4354.13 |
| Suspending periods | 414 | 172 LL: - 280.03 | 148 LL: - 243.72 | 891 | 662 LL: - 174.74 | 1 335 LL: + 0.24 |

Commented [A3]: Again, just changed the wording for clarity/English. Can you check I haven't messed up the meaning in the footnote!

These results were obtained by comparing the statistical difference resulting from the comparison of the data obtained from the three subcorpora representing each language, excluding lexical variety.

If we calculate the statistical difference of each subcorpora of mother tongue Opinions in respect to the subcorpora of non-mother tongue Opinions, the decrease in the LL values from

¹⁶ [Since the compared corpora differ in size, comparisons were realized using LL values to consider the results proportionally according to the size of each corpus \(see note 14 above\).](#) In order to highlight the diachronic change of the displayed features, in Table 8 the displayed LL were calculated for each subcorpus of Opinions drafted after 2004 relative to the subcorpus MT_pre2004. [Even though/Although the statistical significance is always a positive number, a positive or a negative sign preceding the LL value in each column indicates the overuse or the underuse of the analyzed item in the relevant subcorpus in relation to MT_pre2004. As a consequence, a LL value preceded by a positive sign indicates an increase of the considered feature in the relevant subcorpus in respect to the subcorpus of Opinions drafted before 2004; on the contrary whereas, a LL value preceded by a negative sign indicates a decrease of the considered feature in the relevant subcorpus in respect to the subcorpus of Opinions drafted before 2004.](#)

MT_pre2004 to MT_post2004 shows that non-mother tongue Opinions present a greater difference when compared with mother tongue Opinions before 2004 than when compared to mother tongue Opinions after 2004. Thus, mother tongue Opinions appear to show a gradual assimilation to non-mother tongue Opinions (see Table 9).

Table 9: statistical difference (LL) of the diachronic change displayed in Table 8 of Opinions MT_pre2004 and MT_post2004 in respect to Opinions nonMT_post2004

| Feature | OPINIONS_EN | | | OPINIONS_FR | | |
|--|--|-----------------------------------|--|---|-----------------------------------|---|
| | MT_pre2004 | nonMT_post2004 | MT_post2004 | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| Lexical words with a frequency higher than 500 | Types: 349 Tokens: 455 178 LL: -8851.73 | Types: 588 Tokens: 927 474 | Types: 498 Tokens: 718 350 LL: -1545.93 | Types: 471 Tokens: 667 555 LL: +179.87 | Types: 639 Tokens: 993 441 | Types: 635 Tokens: 1 003 186 LL: +8.47 |
| Lexical variety | 31.87 | 29.23 | 30.04 | 32.96 | 31.66 | 32.11 |
| Relative clauses | Types: 5 Tokens: 17 720 LL: +531.44 | Types: 7 Tokens: 24 435 | Types: 7 Tokens: 22 550 LL: +150.78 | Types: 16 Tokens: 485 LL: +482.34 | Types: 17 Tokens: 5 222 | Types: 15 Tokens: 6 771 LL: +190.15 |
| Lexical words | Types: 6 091 Tokens: 872 152 LL: +353.76 | Types: 7 702 Tokens: 1 474 300 | Types: 6 824 Tokens: 1 212 559 LL: -1.89 | Types: 8 425 Tokens: 1 197 342 LL: +3536.62 | Types: 9 671 Tokens: 1 623 641 | Types: 9 674 Tokens: 1 621 942 LL: -41.70 |
| Suspending periods | 414 LL: +280.03 | 172 | 148 LL: +0.15 | 891 LL: +174.74 | 662 | 1 335 LL: -226.72 |

Since the primary results demonstrate that non-mother tongue Opinions present a simpler style and less fluency than mother tongue Opinions, it can be claimed that mother tongue Opinions are undergoing a progressive change towards less 'fluency', and a simpler syntax and style.

The concrete words used in the texts of mother tongue and non-mother tongue Opinions also appear to be becoming gradually more similar. Indeed if we consider the first 20 most frequent lexical words in each subcorpus, the number of items common to mother tongue and

non-mother tongue Opinions is lower before 2004 than after 2004. This difference is particularly evident in the French corpus. See Table 10.

Table 10: first 20 most frequent words of each analyzed subcorpus

| OPINIONS_EN | | | OPINIONS_FR | | |
|-------------|----------------|-------------|-------------|----------------|-------------|
| MT_pre2004 | nonMT_post2004 | MT_post2004 | MT_pre2004 | nonMT_post2004 | MT_post2004 |
| article | Article | article | pas | article | article |
| case | Court | court | article | pas | pas |
| Commission | case | directive | commission | point | cette |
| court | Directive | member | cette | directive | commission |
| member | commission | case | arrêt | droit | droit |
| regulation | Law | law | cour | paragraphe | directive |
| community | Member | commission | point | cette | point |
| directive | Paragraph | regulation | règlement | commission | paragraphe |
| state | Regulation | state | droit | cour | arrêt |
| states | State | states | paragraphe | arrêt | état |
| law | See | paragraph | directive | règlement | cour |
| paragraph | States | see | fait | ainsi | règlement |
| national | National | national | décision | voir | décision |
| other | Question | general | même | union | ainsi |
| question | General | other | plus | membres | union |
| council | Judgment | council | effet | décision | membres |
| treaty | Also | whether | question | tribunal | membre |
| judgment | Other | decision | traité | fait | tribunal |
| decision | Decision | there | selon | même | voir |
| view | Council | question | cas | membre | fait |

7.3 Qualitative results

Finally, analysis of idiosyncrasy and the degree of specialization of the lexicon used in the three subcorpora (methodological steps xii and vii, respectively) has produced some qualitative results. Analysis of lexical words with a frequency higher than 500 obtained from each subcorpus (methodological step vii) demonstrates that non-mother tongue Opinions

present a greater quantity of high frequency specialized legal terms than mother tongue Opinions, as shown in Table 11.

Table 11: specialized legal lexical word with a frequency higher than 500 of each analyzed subcorpus

| OPINIONS_EN | | | OPINIONS_FR | | |
|--------------------|--------------------|-------------------|-------------------|--------------------|--------------------|
| MT_pre20 04 | nonMT_post2 004 | MT_post200 4 | MT_pre200 4 | nonMT_post2 004 | MT_post20 04 |
| Types: 78 | Types:109 | Types: 91 | Types: 79 | Types: 111 | Types: 105 |
| Tokens: 159 125 | Tokens:282 15 5 | Tokens:214 606 | Tokens:184 962 | Tokens: 291 016 | Tokens: 286 254 |
| LL: - 50.93 | | LL: - 780.18 | LL: - 672.15 | | LL: - 78.72 |

Contrary to that, the keyword lists obtained from the comparisons between subcorpora (methodological step xii) show that mother tongue Opinions present more idiosyncratic items related to context-specific social, political or cultural aspects than non-mother tongue Opinions (see Table 12 below).

Table 12: keywords related to context-specific social, political or cultural aspects identified in each analyzed subcorpus¹⁷

| OPINIONS_EN | | | OPINIONS_FR | | |
|-----------------|--------------------|-----------------|------------------|--------------------|-----------------|
| MT_pre20 04 | nonMT_post20 04 | MT_post20 04 | MT_pre20 04 | nonMT_post20 04 | MT_post20 04 |
| 17 types | 13 types | 17 types | 17 types | 11 types | 18 types |
| 9 758 tokens | 7 371 tokens | 8 092 tokens | 12 020 tokens | 6 714 tokens | 7 862 tokens |

Legal specialized terminology includes all the words semantically related to the legal field (such as *article*, *law*, *council* or *commission*) rather than general words, whereas items related to context-specific social, political or cultural aspects are topics of more general interest (for example, the terms belonging to the semantic category ‘primary sector production’, such as *farmer*, *agricultural*, *animal*, and *land*). The qualitative results set out in Table 11 and 12 demonstrate a difference in writing style, which may suggest that the authors of non-mother

¹⁷ No statistical difference is displayed for these data because to generate a keyword list a statistical method is used to compare the two relevant corpora (in this case, Log Likelihood).

tongue Opinions pay more attention to the stylistic aspect of their writing insofar as they use a greater quantity of specialized legal words than are found in mother tongue Opinions.

7. Conclusions

The results set out above lead to two kinds of conclusions, methodological and conceptual, linked respectively to the methodology (Corpus Linguistics), and to the conceptual framework of the study (Legal and Translation Studies), each suggesting new research questions and proposals for further research.

From a methodological point of view, the results allow for an array of conclusions to be drawn. Since the methodological design, based on an interdisciplinary model, has proved successful in corroborating the primary hypothesis of the study, it could be used/adapted to compare the stylistic features of different text types. Furthermore, the methodological design suggests an alternative use of the Log Likelihood calculation employed to assess the diachronic change of a phenomenon through its comparison with a fixed referential element (as, in this case, the set of Opinions drafted by non-native AGs). It is, however, important to note that the additional analysis undertaken in the present study was exploratory only. In accordance with that aim, that analysis was designed to allow an opening observation which would include a wide range of possible results. In order for the methodology to be fully transferable to further studies, certain steps could be taken to strengthen its application. For example: morphological tagging could be carried out, in order to consider all occurrences of subordinate and non-finite clauses, independently from the presence of an introductory conjunction; Further and more specific analysis investigating the semantic relationships between the words included in the corpora could be carried out, focusing on those items related to context-specific social, political or cultural aspects as well as on other lexical elements; The analysis could be extended to identify specialized legal words to the entire set of corpora, in order to reach a deeper knowledge of the degree of specialization of language used in AGs' Opinions (in the present study this was applied only to the first 500 occurrences in each subcorpus).

From the conceptual point of view, the study described above first expands the scope of Translation Studies theory by applying Universals of Translation in the fields of Legal/Law

and Language studies and Applied Linguistics in order to analyse texts from a novel perspective. Second, the analysis demonstrates that certain changes in the linguistic and stylistic nature of AGs' Opinions can be observed over the timeframe of the study. Those changes correlate with the introduction of changes to the conventions relating to AGs' drafting languages, insofar as they demonstrate that, overall, AG Opinions drafted after 2004 in non-mother tongue languages are simpler and less 'fluent' than those drafted (in AGs' mother tongues) before 2004. However, the results also demonstrate changes in the language of post-2004 mother tongue Opinions, which indicate that AG Opinions drafted after 2004 in mother tongue languages are becoming similarly simpler and less 'fluent'. In the sense of Corpus Linguistics analysis, these results are interesting in and of themselves. However, it would be inappropriate to conclude that such results can, of themselves, answer the overarching question of whether linguistic changes introduced at the CJEU in 2004 have had an impact on the linguistic and stylistic nature of the relevant AGs' Opinions. We cannot comment, on the basis of Corpus Linguistics analysis alone, as to any causal link between the introduction of linguistic changes at the CJEU and the observed changes in the language of Opinions. There are a multitude of factors in the production of AG Opinions that Corpus Linguistics analysis simply cannot take into account. For example, the analysis described here compared the language of Opinions based on the mother tongue of the relevant AG. However, recent empirical studies have shown that the process of producing an AG Opinion is far more collaborative than it would appear at face value (McAuliffe, Muntean and Mattioli, 2021). AGs are assisted, in the writing of their Opinions, by their teams of référendaires (legal clerks), who may or may not have the same mother tongue as 'their' AG. Furthermore, lawyer-linguists provide linguistic assistance to those AGs who choose to draft in a language other than their mother tongue, and it is very difficult to establish the extent of linguistic assistance that may have been provided in each case (McAuliffe, Muntean and Mattioli, 2021). Finally, language and law scholarship on the CJEU reveals a myriad of constraints and cultural compromises involved in the production of texts within that Court (McAuliffe, 2016), none of which are visible in any measurable sense in the texts themselves (McAuliffe, Muntean and Mattioli, 2021). Any conclusions as to the impact of the introduction of changes to linguistic conventions at the CJEU must take account of the context in which AG Opinions are drafted and the factors of their production. For the analysis

described here to have a real value beyond the field of Corpus Linguistics itself, we need to move beyond observing lexical items in the texts. The results set out here can thus form the basis of research questions which could be investigated through interdisciplinary methods (e.g. interviews, legal analysis). For example: What are the factors of production of AGs' Opinions?; Which of those factors may contribute to the observed changes in linguistic style? Answering such research questions, and triangulating the results of such interdisciplinary analysis, would then allow conclusions to be drawn as to whether linguistic changes introduced at the CJEU in 2004 have had an impact on the linguistic and stylistic nature of the relevant AGs' Opinions, and more generally what that might mean for the administration of justice in the EU context. Exploring such research questions in an holistic manner may allow conclusions to be drawn in relation to the impact of legal linguistics on EU case law, and thus on the development of EU law more generally.

References

- Baker, M. (1998). *Routledge encyclopedia of translation studies*. Routledge.
- Baker, M. (1993) "Corpus Linguistics and Translation Studies – Implications and Applications". In Baker, M., G. Francis & E. Tognini-Bonelli, E. (Ed.) (1993) *Text and Technology: In Honour of John Sinclair*, Amsterdam and Philadelphia: John Benjamins.
- Baroni, M. & Bernardini, S. (2003). "A preliminary analysis of collocational differences in monolingual comparable corpora." In D. Archer, P. Rayson, A. Wilson and T. McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 Conference, UCREL*, Technical Paper number 16, Special issue, Lancaster: UCREL, Lancaster University, 82–91.
- Borgsmidt, K. (1988). "The advocate general at the European Court of Justice: A comparative study" *European Law Review*, 13, 106.
- Chesterman, A. (2004). «Beyond the particular» en Mauranen, A. y P. Kujamäki (ed.) *Translation universals: Do they exist ?* John Benjamins Publishing, 33-49.
- Kruger, H. (2018). Expanding the third code: Corpus-based studies of constrained communication and language mediation. *CECL Papers*, 9.

- Laviosa, S. (1998). «Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose», *META* (43-4), 557-570.
<https://www.erudit.org/fr/revues/meta/1998-v43-n4-meta169/>
- Lee, Y. C. (2018). "The hallmarks of L2 writing viewed through the prism of translation universals". *Linguistic Research*, 35, 171-205.
- McAuliffe, K. (2011), 'Hybrid Texts and Uniform Law? The Multilingual Case Law of the Court of Justice of the European Union', *International Journal for the Semiotics of Law* 24(1): 97-115.
- McAuliffe, K. (2016). "Hidden Translators: The Invisibility of Translators and the Influence of Lawyer-Linguists on the Case Law of the Court of Justice of the European Union". *Language and Law= Linguagem e Direito*, 3(1).
- McAuliffe, K., L. Muntean, & Mattioli, V. (2021). "Through the Lens of Language: Using Interviews and Corpus Linguistics Analysis to Examine Opinions of Advocates Generals". In: *Researching the European Court of Justice: Methodological Shifts and Law's Embeddedness*, Cambridge University Press.
- Olohan, M. & M. Baker (2000). "Reporting that in Translated English: Evidence for Subconscious Processes of Explication?". *Across Languages and Cultures* 1(2), 141-158.
- Puurinen, T. (2003) , "Genre-specific Features of Translationese? Linguistic differences between Translated and Non-translated Finnish Children's Literature". *Literary and Linguistic Computing* 18(4), 389-406, <https://doi.org/10.1093/lc/18.4.389>
- Scott, M. (2017) *WordSmith Tools version 7*, Stroud: Lexical Analysis Software.
 <<http://lexically.net/wordsmith/>>
- Tai, H. Y. (2015). "Writing development in syntactic complexity, accuracy and fluency in a content and language integrated learning class." *International Journal of Language and Linguistics*, 2(3), 149-156.
- Toury, G. (1991). What are descriptive studies into translation likely to yield apart from isolated descriptions?. *Translation Studies: The State of the Art*. Amsterdam: Rodopi, 179-192.
- Toury, G. (1995) *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: John Benjamin Press
- Toury, G. (2004). "Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals?". In Mauranen, A. y P. Kujamäki (eds.)

Translation universals: Do they Exist? Amsterdam y Philadelphia: John Benjamins, 15-32.

Tymoczko, M. (1998). "Computerized corpora and the future of translation studies". *Meta* 43(4), 652-660

Vanderauwera, R. (1985) *Dutch Novels Translated into English: The Transformation of a 'Minority' Literature*, Amsterdam: Rodopi

Xiao, R. & M. Yue (2009). "Using Corpora in Translation Studies: The State of the Art". In Baker, P. (ed.) (2009) *Contemporary Corpus Linguistics*. London: Continuum, 237-262.

Zanettin, F. (2013). "Corpus Methods for Descriptive Translation Studies", V International Conference on Corpus Linguistics (CILC2013) *Procedia - Social and Behavioral Sciences*. Elsevier Ltd.