

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

## **WORD ORDER VARIATION IN ADVANCED STUDENTS' TRANSLATION TASKS: A LEARNER CORPUS-BASED ANALYSIS**

Rudy Loock

University of Lille and CNRS Research unit  
'Savoirs, Textes, Langage'

### **Abstract**

In this article, we analyze a learner corpus of English-to-French translation tasks produced by advanced students enrolled in their final year of a professional translation master's program. Specifically, we investigate their use of so-called non-standard, non-SVO word order structures: clefting, pseudo-clefting, dislocation, extraposition, and inversion. We aim to confirm students' tendency to overuse SVO word order in their translations in comparison with original French and to provide a finer-grained analysis of their (non-)use of non-standard structures. Complementary analyses on a corpus of machine-translated texts and a corpus of professional translations provide further comparisons. Thanks to an approach where corpus material is used to assist students in the development of their translation skills, all these results are meant to have pedagogical value, by highlighting the specificities of student

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

translations to help them write more authentic texts that take into account language use related to word order. The comparison with machine-translated texts is intended to help students develop post-editing skills.

Keywords: translation training, translation quality, word order variation, specialized translation, post-editing

## **1. INTRODUCTION**

In this study, we aimed to analyze a learner corpus of advanced students' translation tasks to investigate their use of non-standard (also non-canonical) word order. Specifically, we analyzed a corpus of English-to-French translations of specialized documents by students enrolled in a professional translation master's program to uncover their use of word order differing from the standard, canonical Subject-Verb-Object (SVO) word order. Our focus was on their use in translated French of cleft and pseudo-cleft structures, right and left dislocation, extraposition, as well as subject-verb inversion.

Our approach is pedagogical, as our study aimed to confirm the observation of a tendency to underuse such structures in our students' translations, by sticking to the standard SVO word order under the influence of English source texts. Following Loock's (2020a) pilot study focusing on multiple translations of a few

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

short texts, we conducted an analysis on individual, longer translation tasks (ca. 2,600 words on average) produced by students in their final year of training (MA2 students). Also, as our pedagogical aim is to make students aware of their underuse of non-SVO sentences, comparisons were drawn with different corpora: (i) a corpus of French original texts, (ii) a corpus of English-French machine-translated texts, as we aim to teach our students how to define their added value over machine translation, and also (iii) English-French professional translations.

The article is organized as follows. In the second part, we explain our starting point and aims, and we define the linguistic features investigated in this study. The third part presents the learner corpus, the methodology used to analyze it, as well as the results. A fourth and final part provides a discussion and supplementary analyses: comparisons with machine-translated texts and professional translations.

## **2. DEFINITIONS, STARTING POINT, AND AIMS**

### **2.1. Non-standard word order in French**

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

The linguistic feature under investigation here is the use of non-standard word order in French-translated texts. In particular, we focus on so-called 'non-canonical' word order constructions deviating from the standard Subject-Verb-Object (SVO) word order (1): clefting (2a), pseudo-clefting (2b), left/right dislocation (2c), pseudo-clefting+dislocation (2d), extraposition (2e), and subject-verb inversion (2f).

(1) La traduction automatique est un défi pour les étudiants.

'Machine translation is a challenge for students.'

(2) a. C'est la traduction automatique qui est un défi pour les étudiants.

'It is machine translation that is a challenge for students.'

b. Ce qui est un défi pour les étudiants est la traduction automatique.

'What is a challenge for students is machine translation.'

c. La traduction automatique, elle est un défi pour les étudiants./Elle est un défi pour les étudiants, la traduction automatique.

'Machine translation, it is a challenge for students./'It is a challenge for students, machine translation.'

d. Ce qui est un défi pour les étudiants, c'est la traduction automatique.

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

‘What is a challenge for students, it is machine translation.’

e. Il est évident que la traduction automatique est un défi pour les étudiants.

‘It is obvious that machine translation is a challenge for students.’

f. Comme l’ont montré plusieurs études, la traduction automatique est un défi pour les étudiants.

‘As it have shown several studies, machine translation is a challenge for students.’  
(As has been shown by several studies, machine translation is a challenge for students.)

Examples (2a)-(2f) deviate from the so-called standard, SVO word order provided in (1). In clefting, the sentence is divided into two parts, with a focused element being introduced by *ce/c’* + the verb *être*, followed by a structure similar to a relative clause. In pseudo-clefting, the subject or the complement of the SVO structure is a *ce qu-* nominal relative clause while in left/right dislocation, an element of the sentence is preposed/postposed and a referential pronoun is kept in the element’s original position. Pseudo-clefting and dislocation are often combined in French (2d). Extraposition is a case where an element, generally a clause, is postposed and replaced with a substitute form (impersonal or ‘dummy’

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

*il*). Finally, subject-verb inversion covers optional cases where the verb appears before the subject (note that cases of constrained inversion in interrogative clauses or reporting verbs to introduce quotations were not included to focus on stylistic subject-verb inversions).

All the non-canonical constructions mentioned in (2a)-(2f) exist in English and French. However, the discourse constraints for them to appear felicitously and their frequencies are not similar (e.g. Birner & Ward, 1998 for English; Lambrecht, 1994, Carter-Thomas, 2002, and Loock, 2020a for cross-linguistic comparisons). Within a functional approach to language, word order variation is never free but governed by pragmatic constraints generally related to the information status of the different components in the sentence, such as the newness/oldness or thematic/rhematic status of the information. In particular, the French language shows a higher frequency for such structures than English, where the SVO, canonical word order is more frequently used. Our corpus analysis stems from the regular observation of English to French translation tasks produced by our students, enrolled in a specialized translation program at the master's level. Noticing a tendency to resort to a canonical word order on a very regular basis, that is Subject-Verb-Object (SVO), we wanted

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

to confirm that this was indeed the case, adopting a corpus-based approach to a learner corpus. After a pilot study on the translation of short texts (Loock, 2020a) confirmed our intuition, the present study is meant to go further by analyzing a larger corpus of longer translation tasks, with each student translating a different text. A learner corpus was specifically collected and investigated to retrieve all occurrences of the non-standard word order sentences listed above. Extra analyses with specific corpora dealt with texts translated thanks to machine translation and by professional translators.

## **2.2. Transfer, translation quality, and linguistic homogenization**

In this article, we focus on English-French translations to observe our students' use of non-standard word order in their translated texts. If students underuse structures listed in (2) in their translations, this might be due to some linguistic transfer under the influence of the word order used in the English source sentences.

In such conditions, a comparison with original French data can confirm or infirm the underuse of non-SVO word order, following the corpus-based translation studies (CBTS) approach of

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

comparing original texts with translated texts to uncover the specificities of the translated language. Many CBTS studies, following Baker's (1996) seminal paper, have shown that original and translated language differ in many respects, which has led Baker and other researchers to posit the existence of the so-called "translation universals", or less controversially "translation features" like exploitation, simplification, or normalization. Source language interference has also been uncovered with such an approach.

Our corpus study is specific herein in that the learner corpus contains French texts translated from one source language only, English. It thus sounds reasonable to say that some linguistic transfer or interference is a reasonable explanation that might account for our results: under the influence of English, students underuse clefting or inversion for instance. While transfer is a general, neutral term, interference is generally considered to be negative transfer, up to gross deviations or translation errors (Mauranen, 2004, pp. 71-72), which in turn raises the thorny question of translation quality. In our case, we are dealing here with a case of "pair-wise interference" (Mauranen, 2004, p. 69) specific to the English-French language pair. What we argue is that to achieve high-quality translations



Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

meeting the demands of today's translation market for authentic, as fluent as possible texts, that read as if written directly in the language (to the point where translators should be invisible), language use should be taken into account. The use of non-standard word order should be found in French-translated texts, ideally in similar proportions to what is found in original French. We, therefore, promote linguistic homogenization between translated and original French, and the results of our corpus study are meant to measure, and possibly later on reduce, the gap between our students' use of the language and original French.

### **3. METHODOLOGY**

#### **3.1. Corpus material**

All translated texts in the English-French translation learner corpus were produced by advanced students enrolled in the second and final year of the specialized translation master's program at the Université de Lille, France (MA2 students), namely the *Traduction spécialisée multilingue* (TSM) program. The translation tasks were performed within a specific class, "Translation Project", in which students are supervised by professional

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

translators<sup>1</sup>. Each student works on their translation project, with a different text to translate from English into French with all texts belonging to the popular science genre or specialized press. Topics include 1970s cinema, animal rights, online streaming platforms, violence in the US, machine translation, HIV treatment, polar bears, women in US politics, and sustainability in the fashion industry. Students are allowed to use whatever translation aids (including machine translation) they want, as the aim of the lecture is to let them work autonomously. Contrary to the pilot study by Loock (2020a), which focused on MA1 students, all students attended a French-English comparative grammar class (described in Loock, 2019) aimed at uncovering differences in language use between the two languages, with a focus on a series of grammatical features, including word order. This class aims to teach them that grammatically correct sentences are not sufficient and that natural-sounding texts should consider usage. The learner corpus comprises 50 English texts translated into French between 2016 and 2020, by 47 students, all of them native speakers of

---

<sup>1</sup> We would like to thank Guillaume Deneufbourg, a professional translator teaching this specific English-French translation class, for providing us with both the original texts and the students' translations.

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

French (as some texts are shorter, some students translated two texts instead of one), for a total of 108,439 words for English source texts and 130,367 words for French target texts. This means an average of 2,169 words per text for English originals and 2,607 for French-translated texts. Table 1 provides a summary of the data. All texts were anonymized.

Table 1. The learner corpus

	English originals	French translations
Number of words	108,439	130,367
Minimum	347	451
Maximum	4,328	4,953
Average	2,169	2,607
Median	1,586	1,885

### 3.2. Analysis of the learner corpus

Working on non-standard word order means that automatic retrieval of occurrences can prove a tedious, time-consuming task. While French clefting examples can be retrieved with the aid of automatic queries including *ce* or *c'*, followed by the verb *être* ('to be'), combined with manual weeding out to remove noisy hits, subject-verb inversions require a thorough reading of the corpus. This explains the relatively small size of our learner corpus, which was carefully read to retrieve in the translated texts all occurrences of non-standard

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

word order constructions listed in the previous section. This also explains why we did not use statistical tools to analyze our data. Each example was labelled according to the type of construction and the number of the text in which it was found, which allowed us to observe general as well as individual results. A distinction was also made between the non-standard word order introduced in the target sentence, not triggered by non-standard word order in the source sentence (3) and the non-standard word order already used in the source sentence (4):

(3) a. As shareholders are aware, during 2017 Barclays disclosed a whistleblowing incident involving allegations made in connection with the hiring of a senior management team member.

Comme le savent nos actionnaires, suite à une dénonciation anonyme de la part de l'un de nos collaborateurs, des allégations en lien avec le recrutement d'un membre de l'équipe de direction a déclenché un incident en 2017. (introduction of inversion, 'as are aware our shareholders')

b. When Germany invaded Poland on September 1, 1939, the "Jewish question" became urgent.

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

C'est quand l'Allemagne envahit la Pologne le 1er septembre 1939 que la « question juive » devient urgente. (introduction of clefting, 'it is when Germany invaded Poland [...] that...')

(4) a. Around the table were 15 men representing government agencies necessary to implement so bold and sweeping a policy.

Autour de la table se trouvaient 15 représentants d'agences gouvernementales nécessaires à la mise en œuvre d'une politique aussi audacieuse que radicale. (maintained inversion)

b. The basic principle for the Swedish Right of Public Access is that it is the existing landscape with its characteristic features and traits and associated enjoyment that is “accessible”, as long as the tolerance limits for what one can and cannot do, principally “disturb not, destroy not”, are not transgressed.

Le principe de base pour le droit de libre accès à la nature suédois est que c'est le paysage existant, avec ses caractéristiques propres et son appréciation qui est « accessible », tant que les seuils de tolérance concernant ce que l'on peut ou ne pas faire, principalement « ne pas déranger, ne pas détruire » ne sont pas franchis. (maintained clefting)

### 3.3. Quantitative results

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

Table 2 below provides frequencies, both raw (Raw F) and normalized per thousand words (*Norm F*), for each of the 50 translated texts and the learner corpus as a whole. For better readability, all instances where no occurrences were found have been left blank.

The first result provided by this table is that the use of non-standard word order shows some variation: some texts do not or hardly deviate from the standard, SVO word order (e.g., texts 3, 4, 6, 22, 32, 43, and 49), while others show a significant use of different non-standard word orders (e.g., texts 1, 11, 12, 16, 29, 35, and 45). This could be due to the length of texts: the fact that some texts show little or even no word order variation, may then be due to chance. However, we do see some longer texts with no non-SVO constructions (22, 49). A second result is that some structures are more frequent than others: extraposition is the most frequent (139 occurrences), while clefting and inversion were used by students 63 and 76 times respectively; however, pseudo-clefting and dislocation are rarer (14 and 12 occurrences respectively).

Look Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

Table 2. Frequencies (raw and per thousand words) for each construction in the learner corpus

Text	Number of words EN	Number of words FR	Extrapolation		Clefting		Inversion		Pseudo-clefting		Pseudo-clefting + Dislocation		Dislocation	
			Raw F	Norm F	Raw F	Norm F	Raw F	Norm F	Raw F	Norm F	Raw F	Norm F	Raw F	Norm F
1	3 186	3 867	21	5.43										
2	1 579	1 828	4	2.19			1	0.55					1	0.55
3	3 601	4 254												
4	1 599	2 030												
5	1 373	1 646	2	1.22	3	1.82	3	1.82					1	0.61
6	1 106	1 266												
7	929	1 109			1	0.90	2	1.80					1	0.90
8	3 775	4 694	3	0.64	1	0.21	3	0.64						
9	1 592	1 941	1	0.52	5	2.58								
10	2 127	2 584	1	0.39	2	0.77	2	0.77						
11	3 407	3 836	1	0.26	5	1.30	4	1.04						
12	1 154	1 466	3	2.05	4	2.73								
13	3 909	5 057	2	0.40	1	0.20	1	0.20						
14	740	766	1	1.31			1	1.31						
15	3 999	5 247	1	0.19			4	0.76						
16	3 349	4 060	9	2.22	7	1.72	2	0.49						
17	1 024	1 169	2	1.71										
18	4 035	4 930			1	0.20	6	1.22			5	1.01	1	0.20
19	975	1 210	2	1.65										
20	874	1 020	3	2.94	1	0.98	2	1.96						
21	1 280	1 544	4	2.59							1	0.65	1	0.65
22	1 880	2 180												
23	687	841												
24	1 020	1 330					2	1.50					1	0.75
25	3 380	3 971	6	1.51	2	0.50								
26	721	828			1	1.21								
27	3 410	4 019	1	0.25	3	0.75	3	0.75	1	0.25				
28	860	950	2	2.11	1	1.05								
29	3 637	4 924	6	1.22	1	0.20	6	1.22			1	0.20		
30	3 805	4 259	3	0.70	4	0.94	2	0.47						
31	3 423	4 495	12	2.67	3	0.67	4	0.89						
32	347	451												
33	3 938	4 535			1	0.22	1	0.22						
34	3 696	4 487	5	1.11	1	0.22	8	1.78			2	0.45		
35	2 776	3 120	5	1.60	3	0.96	2	0.64			2	0.64		
36	1 456	1 615	3	1.86										
37	3 701	4 340	10	2.30	2	0.46								
38	969	1 133	3	2.65	3	2.65								
39	959	1 114												
40	4 045	5 085	3	0.59	2	0.39	2	0.39						
41	765	901	2	2.22	2	2.22							1	1.11
42	4 328	4 953	2	0.40	1	0.20	5	1.01			1	0.20		
43	812	993												
44	1 226	1 388			1	0.72								
45	3 606	4 458	8	1.79			6	1.35					4	0.90
46	575	599									1	1.67	1	1.67
47	1 318	1 527	2	1.31										
48	1 104	1 257	6	4.77			2	1.59						
49	3 410	3 926												
50	972	1 164			1	0.86	2	1.72						
<b>Total</b>	<b>108 439</b>	<b>130 367</b>	<b>139</b>	<b>1.07</b>	<b>63</b>	<b>0.48</b>	<b>76</b>	<b>0.58</b>	<b>1</b>	<b>0.01</b>	<b>13</b>	<b>0.10</b>	<b>12</b>	<b>0.09</b>

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

What is relevant for our study is a comparison of these results with original French. We provide below the results based on the analysis of a comparable sample of 113,778 words of original French press articles extracted from the *TSM press corpus* (Loock, 2019)<sup>2</sup> and provided in Loock (2020a), but adapted here with normalized frequencies per thousand words. Table 3 and Figure 1 provide these results.

Table 3. Frequencies (per thousand words) in the learner corpus and a corpus of original French

	Extra position	Clefting	Inversion	Pseudo-clefting	Dislocation
Original French	0.58	0.66	0.71	0.12	0.24
Learner corpus EN-FR translations	1.07	0.48	0.58	0.11	0.09

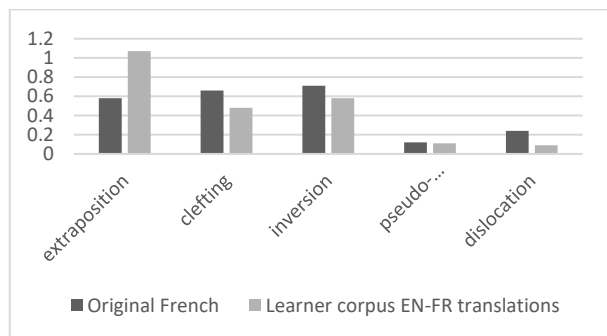
---

<sup>2</sup> The TSM (“Traduction Spécialisée Multilingue”) press corpus is (in 2023) a 4.4-million-word comparable corpus which contains original press texts in American English, British English, and (French) French collected from quality newspapers (e.g. *The Guardian*, *The New York Times*, *Le Monde*, *Libération*) for different domains (business, environment, technology, culture, politics, etc.).



Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

Figure 1. Frequencies (per thousand words) in the learner corpus and a corpus of original French



What this intra-language comparison showed is that the students tended to underuse non-standard word order, except extraposition that was overused (the normalized frequency is almost twice as high). These results are in line with what was found with the analysis of shorter translations in Loock (2020a) on MA1 translations, although slightly different: extrapositions were less frequent this time, and clefting was used more frequently (0.26 occurrences per thousand words in shorter translations), as were dislocation and pseudo-clefting which were not present in shorter translations. However, cases of verb-subject inversion were rarer. The length of texts thus seems to have had some consequences on the results, which were closer to the norms found in original French. Attending the comparative

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

grammar class may also (hopefully) have played a role.

### 3.4. Qualitative analysis

The observation of the learner corpus beyond the general quantitative results provided above was meant to provide us with deeper insight into our students' translation strategies. It provided information on what triggered the use of non-standard word order when none was used in the original sentence and also enabled the identification of cases where non-standard word order would have been felicitous and could have been expected.

When considering clefting, we found it interesting to note that out of 63 occurrences, only 19 (30%) were a direct transfer from the original sentences in which cleft structures were already present. The students, therefore, used clefting to translate SVO sentences, for instance, to introduce a temporal element or to assess a situation, contexts in which English clefting would be typically infelicitous (5):

(5) a. [T]heir popularity within the Anglo-speaking world for at least three centuries before this has just now come to light.

[C]e n'est que récemment qu'on prit conscience de l'ampleur du succès qu'elles

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

rencontrèrent au sein du monde anglo-saxon, au cours des trois derniers siècles précédant celui-ci.

‘It is only recently that...’

b. Absolute and relative pitch are best learned in very early youth.

C’est dans la prime jeunesse que l’oreille absolue et l’oreille relative s’apprennent le mieux.

‘It is in very early youth that...’

c. About 80 to 100 feature films, documentaries and animation co-productions are made with support from the fund every year.

Chaque année, ce sont environ 80 à 100 films, documentaires et coproductions d’animation qui sont réalisés grâce à ce soutien financier.

‘ (...) it is about 80 to 100 feature films (...) that...’

d. We added nearly 500 new Azure capabilities in the past year alone, focused on both existing workloads and new workloads such as IoT and Edge AI.

L’an dernier, ce sont près de 500 nouvelles fonctionnalités consacrées aux nouvelles et anciennes charges de travail qui ont été ajoutées à Azure, telles que IoT et Edge AI.

‘ (...) it is nearly 500 new capabilities...’

There was an even sharper contrast for cases of inversion: 72 out of 76 examples were actually

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

introduced by the students (94%). This is a clear illustration first of all of the scarcity of subject-verb inversion in original English as opposed to original French, but also of the students' ability to introduce inversions in their translations. However, a qualitative analysis showed that among the 72 cases of inversion that were introduced, exactly half of them corresponded to one syntactic configuration, namely comparative clauses introduced by *comme*, which trigger an optional but very frequent case of inversion in French (6). In our corpus, these were very often translations of clauses introduced by:

(6) a. as shareholders are aware > comme le savent nos actionnaires

'as are aware our shareholders'

b. as shown in Figure 1 > comme le montre la Figure 1

'as shown Figure 1'

c. as the Credit Mobilier scandal revealed > comme l'a révélé le scandale du Credit Mobilier

'as revealed the Credit Mobilier scandal'

d. consistent with regulatory commitments > comme l'exige la réglementation en vigueur

'as demand the regulatory commitments'

What we can conclude from this is that there might be a lack of variety concerning the types

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

of context in which inversions were introduced by the students, although we did find some occurrences where the subject-verb inversion was a very authentic stylistic improvement (7):

(7) a. That's why many colleagues decide to keep their public social media feeds completely free of anything controversial and off-topic, which is fair enough and probably a wise choice.

C'est pourquoi nombreux sont les collègues qui refusent d'alimenter leurs profils publics sur les réseaux sociaux avec des publications controversées et hors-sujet, ce qui paraît être une décision sage et justifiée.

'That is why many are the colleagues who...'

b. This is a problem attributable to both the nature of the information they wish to convey and also to what Noam Chomsky (Achbar & Wintonick, 1992) calls "concision".

Sont en cause la nature des informations qu'ils souhaitent transmettre et ce que Noam Chomsky (Achbar & Wintonick, 1992) appelle la « concision ».

'Are in cause the nature of the information...'

As far as dislocations are concerned, it is striking to note that all of those found in the translated texts (n=25) were introduced by the students themselves, half of the time (n=13)

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

when translating a pseudo-cleft structure (8), an optional but more natural choice in French:

(8) a. These are all valid criticisms, but what really grinds my gears is that Pollan offers a false etymology for the Greek term *mageiros*.

Il s'agit de critiques recevables mais ce qui me fait vraiment grincer des dents, c'est la fausse étymologie qu'offre Pollan pour le terme grec *mageiros*.

'(...) what makes me grind my teeth, it is the false etymology...'

b. What ancient Greek medicine emphasized—and what modern dietary advice, with its “one size fits all” recommendations, gets wrong—is the idea that everybody, and every body, was different.

Ce que la médecine grecque de l'Antiquité soulignait - et là où les conseils diététiques modernes font fausse route - c'est l'idée que chaque individu et chaque corps sont différents : il n'y a pas de modèle universel.

'What ancient Greem medicine emphasized (...), it is the idea that...'

c. [W]hat mattered instead was whether or not a person ate too much or too little of any given type of food for their own constitution.

Ce qui comptait, c'était plutôt de savoir si un individu mangeait trop ou pas assez d'un type d'aliment précis par rapport à sa propre constitution.

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

‘What mattered, it was rather whether...’

d. What we really need is fish oil.

Ce qu’il nous faut vraiment, c’est de l’huile de poisson.

‘What we need really, it is fish oil.’

Other cases of dislocation included examples in (9), with more spontaneous and oral-like sentences:

(9) a. Perhaps most importantly of all, we enter 2018 in a strong capital position.

Mais le plus important, c’est sans doute que nous débutons 2018 avec un capital solide.

‘But the most important, it is without doubt that...’

b. Going the extra mile to benefit society and Barclays is what our Shared Growth Ambition is all about.

Aller plus loin pour le bien de la société et de Barclays, c’est l’objectif même de notre Plan de croissance partagée.

‘Going further for the good of society and Barclays, it is the goal itself...’

c. Ads are expensive though.

Le problème avec la publicité, c’est ce que cela coûte cher.

‘The problem with advertising, it is that it is expensive.’

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

A word also needs to be said about extrapositions, which showed a much higher frequency in the student translations than in original French (1.07 vs. 0.58 occurrences per thousand words). These were not the result of transfer, since the frequency of extraposition in original English is 0.59 occurrences per thousand words (Loock, 2020a, p. 87), very similar to the frequency in original French; also, out of the 139 cases of extraposition, 72 were introduced by the students, meaning no extraposition was to be found in the original sentences. This resulted in an overuse of extraposition in the French translated texts. As observed in the pilot study, many cases of extraposition were triggered by the translation of a modal auxiliary (10), a probable consequence of a translation strategy often recommended by translation textbooks or even some English grammar books written in French (see examples in Vinay & Darbelnet, 1995, p. 142; Wecksteen-Quinio, Mariaule and Lefebvre-Scodeller, 2015, p. 118-119). This is a very good translation strategy to limit the use of the verbs *pouvoir*, *devoir*, or  *falloir* in French, which are not as frequent as the English modal auxiliaries. However, if the strategy is used on a recurrent basis, it leads to an overuse of extraposition within impersonal structures.



Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

(10) a. You **can't** change the language of your primary profile once you've set it up.

Il est impossible de changer la langue de votre profil principal une fois que vous l'avez configuré.

'It is impossible to change the language...'

b. (S)ome SH bears **may** have expanded their hunting forays, leading to competition for food with WH bears.

Il est donc probable que certains individus de la côte sud aient élargis leur terrain de chasse.

'It is thus probable that some SH bears...'

c. An apparent tendency towards late spring warming **can** be derived by examining the period from 1981 to 1999, illustrated by the dashed trend curve in Fig. 1b.

Sur la période 1981 - 1999, il est possible d'établir une tendance visible à l'augmentation des températures à la fin du printemps, représentée par la courbe de tendance à tirets dans la Fig. 1b.

'(...) it is possible to establish a tendency...'

Finally, we noticed some sentences where using a non-standard word order would have improved the translation (11):

(11) a. What ARVAs offer (whether it be naked girls, art installations, or undercover footage) are not "spectacles," in the Debordian sense but

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

rather are “disruptive image events” (Derville, 2005: 531).

Ce que les MVPA offrent (qu’il s’agisse de femmes nues, d’installations artistiques ou d’images d’infiltration) ne sont pas des « spectacles » au sens débordien du terme, mais plutôt des représentations visuelles dérangeantes.

b. That’s the same approach the European Union takes when it’s regulating chemicals.

C'est la même approche que l'Union européenne adopte lorsqu'elle réglemente les produits chimiques.

In (11a), the quite long sentence could have been improved by introducing first an inversion (*ce qu’offrent les MPVA*, ‘what offer ARVAs’), but above all a left dislocation (*ce que...*, *ce ne sont pas des spectacles...*, ‘what..., it is not spectacles...’).

In (11b), inverting the subject and the verb would have made the sentence more fluent (*C’est la même approche qu’adopte l’Union européenne...*, ‘It is the same approach that adopts the EU’). We did not notice any particular pattern, however, which could explain why students did not use a non-canonical construction.

#### 4. SUPPLEMENTARY ANALYSES

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

#### **4.1. Comparison with machine-translated texts**

Many studies (e.g. Lapshinova-Koltunski, 2015; Macketanz et al., 2017; Vanmassenhove, Shterionov & Way, 2019; Loock, 2018, 2020b; de Clercq et al., 2021 for recent examples on neural machine translation systems) have compared the linguistic features of machine-translated texts with original texts to uncover “machine-translationese”, that is to say the specificities of texts obtained thanks to a machine translation (MT) tool. What such corpus-based analyses have uncovered is that machine-translated texts show, for example, a lesser lexical richness and an overuse or underuse of some specific language features (e.g., overuse of derived adverbs or existential constructions in English texts machine-translated into French). From a pedagogical point of view, such results make students aware of the limits of machine translation and help them define their added value as human translators; they also provide information on what can be checked during the post-editing (PE) process (Loock, 2020, pp. 159-160).

For such a comparison, a sample of English original texts was extracted from the *TSM press corpus* and translated with DeepL ([www.deepl.com](http://www.deepl.com)), a generic neural machine

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

translation tool freely available online, with a reputation for providing natural-sounding output. A series of 50 English press articles, for a total of 32,605 words, was machine-translated into French, providing a 38,030-word corpus of translated French (see Loock, 2020a for more information). Each non-standard construction (see 2a-f above) was collected, and frequencies were normalized per thousand words for comparison with both original French and student translations. Results are provided in Table 4 and Figure 2. What the analysis of this small-size corpus showed is that cases of non-canonical word order were even more under-represented in machine-translated texts, in particular for clefting and inversion. Also, as noted in Loock (2020a), all occurrences of clefting were triggered by the use of clefting in the original sentences, while all but one example of inversion was not triggered by the use of inversions in the source texts. Occurrences of dislocation and pseudo-clefting were found, with an overuse of pseudo-clefting in comparison with original texts, a rather intriguing result. There is, therefore, a gap between the student translations and the machine translations, the former being generally closer to the expected norms in original French. The differences observed in

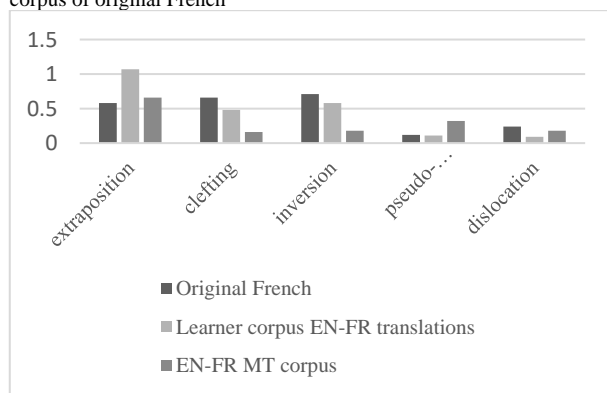
Look Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

our learner corpus can help students become aware of their added value over the machine: for example, their use of clefting or inversion is closer to what is to be found in original French, meaning that their translations may be closer to actual usage, and therefore more natural-sounding.

Table 4. Frequencies (per 1,000 words) in the learner corpus and a corpus of original French

	Extraposi- tion	clefting	inversion	pseudo- clefting	Disloca- tion
Original French	0.58	0.66	0.71	0.12	0.24
Learner corpus EN-FR translations	1.07	0.48	0.58	0.11	0.09
EN-FR MT corpus	0.66	0.16	0.18	0.32	0.18

Figure 2. Frequencies (per thousand words) in the learner corpus and a corpus of original French



Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

Also, comparing such machine-translated data with data from texts directly written in French allowed us to measure the existing gap between machine-translated texts and original language, thus providing valuable information for the post-editing process. Once again, if translator invisibility is the goal to be achieved, linguistic homogenization between original and translated language should be optimal.

#### 4.2. Comparison with professional translations

Comparing our learner corpus with professional translations can also have some important pedagogical value: the comparison of translations produced by students and by professional translators can measure the gap that still needs to be filled, the premise being that professionals typically have more experience and produce higher quality translations. Such comparisons are not that frequent due to the non-availability of data produced by professional translators and the lack of information on their profiles and methods.

We focus here on the analysis of a corpus of English-French professional translations for the use of clefting. These were collected from two different sources: the French website of National Geographic (<https://www.nationalgeographic.fr/>), which

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

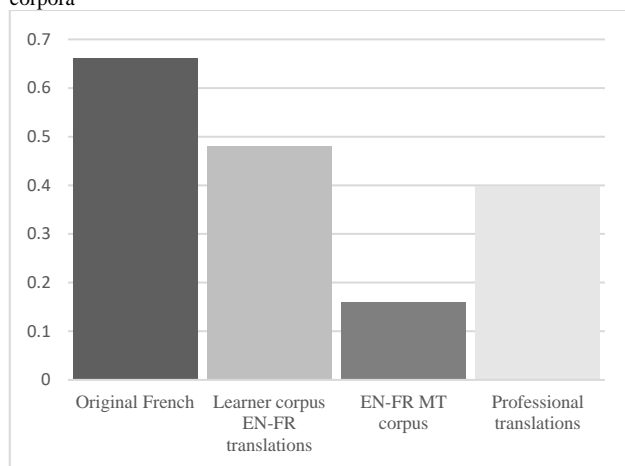
provides translations into French of texts published on the American website, and the French version of the online magazine Slate (<http://www.slate.fr/>), where some of the articles are translations from <https://slate.com/> and identified as such. Although we have no information on how these texts were translated, particularly on the translation tools and aids used (including the possible use of machine translation), we know that the texts are produced by professional translators. For this supplementary analysis, 38 different texts were collected on the two websites in late 2019 and early 2020, for a total of 52,758 words.

The analysis that we decided to conduct focused on the use of clefting, for which we noticed an underuse among our students and also in machine-translated texts. Our initial hypothesis was that professional translators' use of clefting would be closer to that found in original French than student translators' use. Surprisingly, this was not the case: as shown in Figure 3, with 21 occurrences of clefting in our 52,758-word corpus, the normalized frequency of clefting was 0.39 occurrences per thousand words, higher than what was found in our machine-translated texts (0.16) but lower than the frequency observed in original French (0.66) and slightly lower than the one observed in student translations (0.48). This is an

Look Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

intriguing result that requires further investigation and might be explained by the very small size of our corpus.

Figure 3. Frequencies (per thousand words) for clefting in the four corpora



## 5. CONCLUSION

In this paper, we analyzed a 130,367-word learner corpus of student translation tasks, with a focus on their use of specific non-standard word order constructions. The results provided by the analysis are quite encouraging and in contrast with what was found in our pilot study. Although the students did not use such structures with the same frequencies as what is expected in original French, we noted frequent



Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

use of such structures, and not only due to their presence in the original English sentences. The students in particular did introduce clefting, inversion, as well as dislocation with pseudo-clefting. These results contrast with what machine translation provides, highlighting the need for post-editing and human translators' added value; also, we noted that the students' use of clefting was more frequent than what was found in a small-size corpus of professional translations. The fact that the learner corpus contains translation tasks produced by students near the end of their training, and also the existence of a comparative grammar class aimed at making them aware of the importance of considering language use, might explain these better results. There is, nevertheless, still room for improvement: extrapositions were still overused by students, who excessively translated sentences containing modal auxiliaries with impersonal structures triggering extrapositions. Also, the qualitative analysis of our learner corpus uncovered cases where the use of a non-canonical structure would have improved the fluency of the target text. Similarly, we noted that the use of non-standard word order constructions was not to be found across all translation tasks, some of them showing no occurrences.

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

Our study has some methodological limitations. In particular, because of the linguistic features investigated which require a thorough reading of the data, the different corpora (students' translation tasks, original French, machine-translated texts, professional translations) are pretty small and not necessarily representative. Further studies should therefore confirm our results. Also, some results remain unaccounted for, like the overuse of pseudo-clefting in machine-translated texts or the higher frequency of clefting in students' translations compared with professional translations. This is left open for future research.

## REFERENCES

- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175-186). John Benjamins. <https://doi.org/10.1075/btl.18>
- Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English*. John Benjamins. <https://doi.org/10.1075/slcs.40>

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

Carter-Thomas, S. (2002). Theme and information structure in French and English: A contrastive study of journalistic clefts. Paper presented at the 14th Euro-International Functional Systemic Workshop, Lisbon University.

de Clercq, O., de Sutter, G., Loock, R., Cappelle, B. & Plevoets, K. (2021). Uncovering machine translationese Using corpus analysis techniques to distinguish between original and machine-translated French. *Translation Quarterly*, 101(1), 2-45.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge University Press.

<https://doi.org/10.1017/CBO9780511620607>

Lapshinova-Koltunski, E. (2015). Variation in translation: Evidence from corpora. In C. Fantinuoli, & F. Zanettin (Eds.), *New directions in corpus-based translation studies* (pp. 93-114). Language Science Press.

<https://doi.org/10.17169/langsci.b76.64>

Loock, R. (2018). Traduction automatique et usage linguistique: une analyse de traductions anglais-français réunies en corpus (Machine translation and language

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

use: A corpus analysis of English-French translation). *Meta: le journal de traducteurs* 63(3), 785-805.

<https://doi.org/10.7202/1060173ar>

Loock, R. (2019). Parce que “grammaticalement correct” ne suffit pas: Le respect de l’usage grammatical en langue cible (Because “grammatically correct” is not enough: Respecting grammatical use in the target language). In M. Berré, B. Costa, A. Kefer, C. Letawe, H. Reuter & G. Vanderbauwhede (Eds.), *La formation grammaticale du traducteur: Enjeux didactiques et traductologiques (Grammatical training for translators: A didactics and translation studies perspective)* (pp. 179-194). Presses Universitaires du Septentrion.

Loock, R. (2020a). It's a non-canonical word order that you should use! A corpus approach to avoiding standardized word order in translated French. In S. Granger & M.A. Lefer (Eds.), *Translating and comparing languages: Corpus-based insights, Selected proceedings of the fifth Using Corpora in Contrastive and Translation Studies Conference* (pp. 81-97). Presses Universitaires de Louvain.

Loock, R. (2020b). No more rage against the machine: How the corpus-based

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

identification of machine-translationese can lead to student empowerment. *The Journal of Specialised Translation* 34, 150-170.

Macketanz, V., Avramidis, E., Burchardt, A., Helcl, J. & Srivastava, A. (2017). Machine translation: Phrase-based, rule-based and neural approaches with linguistic evaluation. *Cybernetics and Information Technologies* 17(2), 28-43.

<https://doi.org/10.1515/cait-2017-0014>

Mauranen, A. (2004). Corpora, universals and interference. In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 65-82). John Benjamins. <https://doi.org/10.1075/btl.48>

Vanmassenhove, E., Shterionov D. & Way, A. (2019) Lost in translation: Loss and decay of linguistic richness in machine translation. In M. Forcada, A. Way, B. Haddow, & R. Sennrich, (Eds.), *Proceedings of the Machine Translation Summit XVII* (pp. 222–232), Dublin, Ireland, 19-23 August 2019.

Vinay, J.-P. & Darbelnet, J. (1958/1995). *Comparative stylistics of French and English: A methodology for translation* (J. Sager & M.-J. Hamel, Trans. & Eds.). John Benjamins.

<https://doi.org/10.1075/btl.11>

Loock Rudy (2023). Word order variation in advanced students' translation tasks: A learner corpus-based analysis. *Current Trends in Translation Teaching and Learning E*, 10, 338 – 375. <https://doi.org/10.51287/cttl202311>

Wecksteen-Quinio, C., Mariaule, M., & Lefebvre-Scodeller, C. (2015). *La traduction anglais-français (English-French translation)*. De Boeck Supérieur.