

## Web-Driven vs. Machine Translation in the Medical Field

Patrizia GIAMPIERI  
University of Camerino  
University of Malta

**Abstract:** *The web as corpus is considered helpful in translation training, although the literature argues that Internet data might be too chaotic and badly organised. For this reason, web concordancers are often resorted to in order to carry out web searches as they are more linguistically oriented and user-friendly. At the same time, machine translation (MT) has made huge strides forward in the last decades, and scholars have increasingly focused on MT in technical translation training. However, so far little has been investigated on the quality of MT vis-à-vis translations via web concordancers. This paper wishes to bridge this gap and aims at exploring whether and to what extent the web-based translation of a medical article carried out by consulting the Leeds web-concordancer is qualitatively more satisfactory than a translation using MT, or vice-versa. To this aim, a classroom observation study is undertaken with 9 Italian final-year bachelor's students. The students' Leeds-driven translations are compared with the automatic translations performed by two MT tools (DeepL and MateCaT). The paper findings highlight the overall correctness of the machine-translated sentence structures as well as of the related grammar. However, MT presents lexical and collocational issues mostly due to infrequent word usages. By contrast, the web-driven students' translations present a wider variety of lexical and collocational choices, despite some grammatical shortcomings. The paper suggests further research to investigate whether more training in translations based on web concordancers can help students feel at ease with online searches and reduce inaccuracies.*

**Keywords:** *the web as corpus; machine translation; web-based translation; medical translation; translation in the medical field*

### Introduction

#### *The web as corpus*

The web as corpus is argued to be useful in language learning (Fletcher; Giampieri *The web as corpus*) and in translation training (Gatto *From body to web*, Giampieri, Forshaw Labruzzo 23-38). It is resourceful in corpus linguistics, because it allows users to notice language variation, dialects and phraseology (Lindquist, Levin). The web is considered a vast *repertoire* of authentic language (Gatto *The 'body' and the 'web'* 38). It is, in fact, rich in “genuine communicative events” (ibid.), which are constantly updated and accessible to anyone (Fletcher). However, surfing the Internet for language learning or for translation purposes may have some drawbacks. The

information it contains is not always reliable (Gatto *From body to web* 47), it changes continuously and is not organised or displayed in a way that fosters language analyses (Gatto *Web as corpus*). For these reasons, the web is defined as a gargantuan collection of texts (Gatto *From body to web*), which lacks the main features and purposes of a corpus. Hence, there is no possibility to carry out accurate linguistic investigations (Zanettin 56). In addition, users must be taught how to navigate it (Gatto *The 'body' and the 'web'*) in order to eschew unauthentic or irrelevant materials.

To cater for such shortcomings, scholars have developed software solutions or platforms which “interrogate” the web and produce reliable results by using specific algorithms. In this way, they dispose Internet data and information in a way that is suitable for linguistic processing. Examples of such software tools are the WebCorp (Kehoe, Renouf) and the Leeds (Wilson et al.) platforms. The WebCorp and Leeds interfaces are also referred to as “web concordancers”, as they generate concordances from web queries and provide linguistically relevant results (Gatto *From body to web* 80; Giampieri *The web as corpus* 92). For example, the Leeds is equipped with POS (Part-Of-Speech) tagging, which helps users search for words on the basis of their grammatical function (Gatto *Web as corpus* 17). The WebCorp allows the restriction of site domains. Hence, it is possible to narrow down Internet searches to specific websites, or domains. Both interfaces generate collocations and lemmas.

It is evident that users must be acquainted with the query search syntax of web concordancers. In this regard, scholars have dedicated several research papers to teaching and explaining how to use them effectively. For example, Cirillo presents a classroom observation study where students consult the Leeds platform in order to find collocations in the tourism sector. Giampieri (*The web as corpus*) shows how to use the WebCorp and the Leeds interfaces for both language learning and teaching purposes, as well as for technical translations. Giampieri and Forshaw Labruzzo discuss how to navigate the Leeds to verify translation options in the business, economic, financial and medical sectors. Therefore, it can be argued that, as long as users know how to carry out Internet searches mindfully, the web as corpus can be considered a resourceful language tool in translation training.

### ***Machine translation***

Many scholars claim that machine translation (MT) can be beneficial in translation training (Montalbán Martínez, Lee, Rivera-Trigueros). Thanks to the huge strides forward made in the last decades, the literature has increasingly advocated its reliability and usefulness (Montalbán Martínez, Takakusagi et al.). For example, Montalbán Martínez (91) posits that although MT does not yet resemble human translation, it makes the

translation process easier. Also, the author posits that the MateCaT MT platform is particularly useful in the translation training classroom because it is equipped with translation memories. Other scholars highlight the advantages of using MT in various settings, including the medical field (Takakusagi et al.). In fact, in the medical sector the relationship between physicians and patients are greatly enhanced by automatic translation tools (Randhawa et al.). For example, Takakusagi et al. show the advantages of the DeepL MT platform when addressing the English-Japanese language pair.

Many scholars advocate the usefulness of MT, especially in the translation training classroom and/or in post-editing trial lessons (Jia et al., Loock). Jia et al., for example, argue that MT post-editing activities make the students' translation process faster, although it reduces their cognitive effort.

Despite the acclaimed benefits, MT has also some pitfalls and drawbacks, such as the necessity of human intervention (i.e., post-editing) for quality purposes, as noted by Chan. In particular, MT post-editing can be perceived as challenging due to word ambiguity and/or the users' lack of experience in post-editing. In this regard, Loock claims that MT may produce language samples that are not in line with authentic linguistic usage. Based on this assumption, however, he describes how “machine-translationese” can be exploited successfully in translation training to raise the students' awareness of the gap between machine-translated texts and naturally sounding texts. In this way, students are prepared for the post-editing phase.

On the basis of these argumentations, it can be asserted that, to some extent, MT can be beneficial in the translation process, on condition that users are able to (or are taught how to) discern word authentic usages.

### **Research question**

Given the usefulness of machine translation in the translation training classroom (Montalbán Martínez), but the claimed lack of “authenticity” of the language it produces (Loock), and given the attested authenticity of web language as a corpus (Gatto *From body to web*, Gatto *The 'body' and the 'web'*) and its usefulness in translation training and practice (Gatto *From body to web*, Giampieri, Forshaw Labruzzo), this paper is aimed at exploring the quality and reliability of students' web-based translations *vis-à-vis* an automatic translation of the same source text.

In pursuit of this objective, the research questions that this paper aims to pose are the following: is the web-based translation of a medical article qualitatively more (or less) satisfactory than a machine translation? To what extent can MT be relied on, if compared to a web-based translation? Can web concordancers be used as post-editing tools in MT?

### **Methodology**

In order to address the questions posed above, this paper presents and discusses the translations (from Italian into English) of an extract of a medical article (100 words approximately) performed by 9 students enrolled in the last year of a bachelor's program in Translation Studies. In particular, it compares the students' web-based translations with two MT outputs (DeepL and MateCaT's) of the same source text. The article in question addresses initial treatments against the coronavirus at the beginning of the pandemic.

The translation observation study was carried out at the end of the second semester of the students' last year. At the time the study was conducted, the students (whose mother tongue was Italian) had been trained in consulting the Leeds platform, searching for single and multiple words, listing collocations, and looking for lemmas. They had also become acquainted with the Leeds POS tagging syntax.

As mentioned, this paper compares the students' web-based translations with two machine translations of the same source text. To this aim, the DeepL and the MateCaT MT platforms are used. This paper addresses two MT interfaces in order to tentatively yield multiple results and better assess the translation options proposed by both students and machines. DeepL, for example, allows users to select the variety of English desired (e.g., American vs British English), whereas MateCaT offers the possibility to choose the field of reference. As for the former, the American English variety was selected, whereas in the latter, the medical field was chosen.

As far as the web-driven translations are concerned, as already mentioned, the students consulted the Leeds (Collection of Internet Corpora) platform. For the purposes of the classroom observation study, the Leeds interface was preferred over the WebCorp platform thanks to its POS tagging system. The search for collocations and/or colligations (i.e., for lexical and/or syntactical categories, respectively) is also easier and more direct in the Leeds interface. Although the Leeds search syntax may appear more complex than the one of the WebCorp, it is equipped with a variety of search tools. For example, users can apply the Boolean operators in their searches (e.g., the “|” symbol represents the OR operator), or they may specify a gap between words (e.g., two dots “..” represents a gap between two or more words, where any alphanumerical or non-alphanumerical character can be retrieved) (see also Giampieri, Forshaw Labruzzo 30-37).

Therefore, this paper is aimed at exploring the quality and reliability of the two MT platforms compared to web-based translations. The quality and reliability of the automatic and web-based translations are assessed at both grammatical and lexical level, as well as on the basis of the naturalness of the language produced.

## **Analysis**

Table 1 below reports the source text. Field-specific and technical language is underlined.

***Il grande test dell’OMS contro il coronavirus***  
*La pandemia da coronavirus sta portando alla realizzazione di uno dei più grandi test clinici a livello globale mai tentati. L’iniziativa si chiama SOLIDARITY ed è stata ideata dall’Organizzazione Mondiale della Sanità, con l’obiettivo di sperimentare i quattro trattamenti che al momento sembrano essere più promettenti contro la COVID-19, la malattia causata dal coronavirus. I test coinvolgono farmaci già esistenti e usati, e potranno essere svolti in migliaia di ospedali in giro per il mondo, senza particolari complicazioni per il personale medico alle prese con la più grande crisi sanitaria di questo secolo.*

Table 1. The source text. Sourced and abridged from: [www.ilpost.it](http://www.ilpost.it)

As can be seen, the source text contains instances of technical and field-specific language, such as *realizzazione di un dei più grandi test clinici* (back-translation: “realization of one of the biggest clinical tests”); *sperimentare i quattro trattamenti* (back-translation: “experiment the four treatments”); *promettenti contro la Covid-19* (back-translation: “promising against Covid-19); *i test coinvolgono farmaci già esistenti* (back-translation: “the tests involve already existing drugs”); *(test) svolti* (back-translation: “tests carried out”); *complicazioni per il personale medico* (back-translation: “complications for the medical staff”), and *alle prese con la più grande crisi sanitaria* (back-translation: “coping/grappling with the greatest health crisis”). It is interesting to explore how the students and the MT tools addressed the technical words and the collocations of the source text.

To investigate and discuss the quality and reliability of the machine and web-based translations, the medical article is divided into sections. Each section is presented in tables showing the source text (first column), the various translation options written by the students (second column, where each translation proposal is numbered) and the related machine translations (third column). Relevant issues are underlined and commented after each table. Table 2 shows the first section of the medical article, i.e., the title.

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>Il grande test dell’OMS contro il</i>	1.The <u>large</u> WHO’s trial <u>against</u> coronavirus 2.The WHO <u>great</u> trial <u>against</u> the	1.WHO's <u>big</u> test <u>against</u> coronavirus

<p><i>coronavirus</i></p>	<p>coronavirus            3.The <u>great</u> WHO test <u>on</u> the coronavirus            4.The <u>great</u> test of WHO <u>against</u> the coronavirus            5.The <u>big</u> test of WHO <u>against</u> coronavirus            6.The <u>large</u> trial by WHO <u>against</u> coronavirus            7.The <u>large</u> trial of the WHO <u>against</u> coronavirus            8.The <u>great</u> WHO trial <u>for</u> coronavirus            9.The <u>large</u> trial of the WHO <u>against</u> the coronavirus</p>	<p>2.Who's <u>Great</u>            Coronavirus Test</p>
---------------------------	--	---

Table 2. First section of the medical article: the title [author's contribution]

As can be seen in Table 2 above, the students' translations vary quite substantially. The first challenging phrase is *grande test*, which was translated as “large trial”, “great trial/test” and “big test” by the students, and “big test” or “great test” by MT. By writing *large/great/big trial%/test%* in the Leeds search field, it is possible to verify whether the modifiers “large”, “great”, or “big” collocate with the lemmas “trial” or “test”. The results mainly show “big test” and “large trial”, although also “great trial” and “great test” are listed. Therefore, the students' translations can be considered correct.

As regards the word *contro*, the majority of the students proposed “against” (the same as DeepL). However, some also suggested “on” or “for”. If the phrase *test%/trial% on/against/for* is written in the Leeds search field, it appears that the word “trial” is intended as “(criminal) proceedings”. Therefore, “trial” must be excluded from the search, as there is no possibility to specify that this word should be considered as a synonym of “test”. The results of *test% on/against/for* mainly show “test for” and “test on”, with some instances of “test against”. This is confirmed by the query *test% [pos="IN/RP/RB"] .. .. virus*, which helps look for the lemma “test” followed by any preposition or adverb with a gap of words, and then the term “virus”. Therefore, as above, the students' translations were almost accurate.

Among the students' translations, however, some grammar inaccuracies come to the fore, such as the omission of the definite article in the phrase “the great/big test of WHO” (instead of “the great/bit test of the WHO”). As argued by many scholars, the definite article is a recurrent issue for many non-native speakers (Seidlhofer 220, Giampieri *Academic English*

10).

By contrast, the machine-translated texts do not present any grammatical shortcoming. Nevertheless, the students' translations were more varied, despite some grammatical issue. Table 3 below reports the next article section.

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<p><i>La pandemia da coronavirus sta portando alla realizzazione di uno dei più grandi test clinici a livello globale</i></p>	<p>1.The pandemic, due to coronavirus disease, is leading on global level to the <u>creation</u> of one of the largest clinical <u>trials</u></p> <p>2.The coronavirus pandemic is leading to the <u>development</u> of one of the greatest clinical <u>trials</u></p> <p>3.The coronavirus pandemic has led to the <u>carrying out</u> of one of the biggest clinical <u>trials</u></p> <p>4.The pandemic of coronavirus is leading to the <u>realization</u> of one of the greatest clinical <u>trials</u> at global level</p> <p>5. The coronavirus pandemic is leading to the <u>realization</u> of one of the biggest clinical <u>trials</u> on a global level</p> <p>6.The coronavirus pandemic is leading to the <u>achievement</u> of one of the largest global clinical <u>trials</u></p> <p>7.The coronavirus pandemic is leading to one of the biggest global clinical <u>trials</u></p> <p>8.The coronavirus pandemic is leading to the <u>development</u> of one of the greatest clinical <u>trial</u></p> <p>9.The COVID-19 pandemic is leading to the <u>development</u> of one of the largest clinical <u>trials</u> on a global level</p>	<p>1.The coronavirus pandemic is leading to the <u>implementation</u> of one of the largest global clinical <u>trials</u></p> <p>2.The coronavirus pandemic is leading to one of the largest global clinical trials</p>
<p><i>mai tentati</i></p>	<p>1-5.never tested before</p> <p>2.ever <u>experimented</u> at a global level.</p> <p>3.ever <u>attempted</u> globally.</p> <p>4.which has never been <u>tried</u> before</p>	<p>1-2.ever <u>attempted</u>.</p>

	6-7. ever <u>conducted</u> . 8. ever <u>attempted</u> on a global scale. 9. ever <u>attempted</u> .
--	---

Table 3. Second section of the medical article [author's contribution]

As can be noticed, the original noun phrase *realizzazione di uno dei più grandi test* was rendered as “implementation of one of the largest (...) trials” by DeepL, and “one of the largest global clinical trials” by MateCaT. The students' Leeds-based translations presented varied options, such as “the creation of one of the largest clinical trials”; “the development of one the greatest clinical trials”; “the carrying out of one of the biggest clinical trials” and “the achievement of one of the largest global clinical trials”. If the Leeds platform is consulted and the following query is written in the search field: *creation/implementation/development/realization of .. trial%/test%*, the results obtained show phrases such as “development of tests”; “creation of tests” and “implementation of tests”. Also, if the search string is changed into *carrying out of .. trial%/test%*, three hits are obtained with “carrying out of tests”. Therefore, although to a minor extent, the phrase “carrying out of tests” can be considered acceptable. The phrase *achievement of .. trial%/test%*, written by one student (student 6), produces no results. Perhaps the student applied wrong search syntax, or considered “achievement” as a correct collocate of “test” or “trial” without investigating its usages in context. In light of these results, it can be argued that many students provided accurate translation solutions.

As concerns the expression *mai tentati*, MTs suggest “ever attempted”, whereas the students' translation proposals varied from “ever experimented” to “ever attempted”, and “ever conducted” or “tried”. If the search string *test%/trial% .. ever attempted/conducted/experimented/tried* is queried, the results mostly show phrases with “trials ever conducted”. In order to verify whether there are other acceptable alternatives, the search string *test%/trial% ever [pos="V.\*"]* is queried. In this way, it is possible to explore any verb following the lemmas “test” and “trial”, together with the adverb “ever”. The phrases “trial ever performed”, “trials ever conducted” and “trials ever done” come to the fore. These results do not prove that the verbs “attempted”, “experimented” and “tried” are wrong, but, perhaps, that they are less common. In light of these findings, some of the students' translations can be considered satisfactory. The reason why not all the students proposed the most frequent translation candidates can be due to a lack of training or of confidence in the consultation of the Leeds platform. Alternatively, it could be argued that some students lacked confidence or knowledge in the specific language pair (see also similar comments in the



research paper by Giampieri *Can corpus consultation compensate* 14).

As in the previous section, some of the students' web-based translations presented grammatical issues concerning the use of the definite and indefinite article (e.g., “due to coronavirus disease” and “on global level”, instead of “due to the coronavirus disease” and “on a global level”, both written by student 1). Other shortcomings revolved around word order (e.g., “leading on global level to the creation of”, instead of “leading to the creation of .. on a global level”, student 1). There was also a wrong use of the adverbs “ever/never” (see students 1, 4 and 5: “never attempted”), and a wrong plural noun (“one of the greatest clinical trial”, instead of “one of the greatest clinical trials”, student 8). The latter, in particular, is argued to be a frequent issue amongst non-native speakers of English (Giampieri *Academic English* 48-49).

Given the analyses above, it can be claimed that also the second section of the article was translated in varied ways by the students, although the target texts sometimes contained grammatical or word order issues, which MT does not present.

The next section of the article is broken down into smaller parts and analysed in different tables (Tables 4-7).

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>L'iniziativa si chiama SOLIDARITY</i>	1.The initiative, <u>named</u> SOLIDARIETY, 2.This initiative, whose name is SOLIDARITY, 3-4-6-8-9.The initiative is <u>called</u> SOLIDARITY, 5.The World Health Organization has launched an initiative <u>called</u> SOLIDARITY. 7.The initiative, <u>called</u> SOLIDARITY,	1-2.The initiative is <u>called</u> SOLIDARITY

Table 4. Analysis of the third section of the article (first part) [author's contribution]

As regards Table 4 above, in the first line it is possible to notice the verb *si chiama*, referring to an *iniziativa*. The two MT tools indicate “the initiative is called”. The students' Leeds-based translations ranged from “is called” to “whose name is”, and “is named”. By querying *initiative% .. called/named*, the results obtained mostly show “initiative called”. The search string *initiative% ..whose name*, by contrast, produces no hits. Therefore, MT and some of the students' translations were accurate.

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>ed è stata ideata dall'Organizzazione Mondiale della Sanità</i>	1-2.has been <u>launched</u> by the World Health Organization 3.and it was <u>launched</u> by the World Health Organization 4-7.(and it) has been <u>conceived</u> by the World Health Organization 6.and was <u>conceived</u> by the World Health Organization 8.and has been <u>created</u> by the World Health Organisation 9.and it was <u>created</u> by the World Health Organization	1-2.and was <u>conceived</u> by the World Health Organization

Table 5. Analysis of the third section of the article (second part) [author's contribution]

In Table 5, the verb *ideata* refers to *iniziativa* and is rendered as “conceived” in the two machine-translated texts. The students proposed “launched”, “created” and “conceived”. By writing *initiative% .. be% .. conceived/launched/created*, or *initiative% ... .. conceived/launched/created* in the Leeds search field, the verbs “launched” and “created” are found. Hence, there is no evidence of “conceived”. In this case, students probably considered “conceived” as an acceptable equivalent without verifying its usages in contexts and/or its collocational pattern. By searching for *initiative% .. conceived*, in fact, no hits are obtained in the Leeds platform. Nonetheless, “conceived” is the word rendered by MT. By searching for “initiative” followed or preceded by any verb in the British News Corpus (available in the Leeds platform under Collection of English Corpora), for example, the following verbs come to the fore: “announce”, “design”, or “launch”. Thus, there are no instances of “conceive”, and it could be argued that automatic translation produces less frequent patterns of language.

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>con l'obiettivo di sperimentare i quattro trattamenti</i>	1. <u>aiming to try out</u> the four treatments 2. and it <u>aims</u> at <u>testing</u> the four treatments	1. with the <u>goal</u> of <u>testing</u> the four treatments

	<p>3.with the <u>objective</u> of testing the four treatments</p> <p>4-6-8-9.with the <u>aim</u> of testing the four treatments</p> <p>5.Its aim is to <u>experiment with</u> the four treatments</p> <p>7.It <u>aims</u> to test the four treatments</p>	<p>2.with the <u>aim</u> of <u>experimenting with</u> the four treatments</p>
--	---	---

Table 6. Analysis of the third section of the article (third part) [author's contribution]

In Table 6 above, the source text mentions the noun *obiectivo* and the verb *sperimentare* which collocates with *trattamenti*. The MT platforms offered “goal” and “aim” as translation options of *obiectivo*. They also suggested “testing” and “experimenting with” as collocations of “treatments” to render *sperimentare i (...) trattamenti*. With regard to the students' translations, some indicated “aim” and “objective” as equivalents of *obiectivo*. Alternatively, they turned the noun into a verb and used “aim at/to” (students 1, 2 and 7). Concerning *sperimentare i (...) trattamenti*, they mainly suggested “test”, together with “try out” and “experiment with” as collocates for “treatments”.

In order to verify the most frequent and naturally sounding translation(s) of *obiectivo* collocating with *sperimentare* in the Leeds, the search string *aim/goal/objective .. to/of .. test%/experiment%/try%* can be written in the search field. The results show a clear preponderance of “goal” collocating with “try”. However, the verb “try” has a broader meaning as it does not only mean “experiment” or “test”. Sample phrases are “our goal is to try to make it blend”, and “the goal when trying to understand”. Therefore, the lemma “try” should be excluded from the search and the new query would be *aim/goal/objective .. to/of .. test%/experiment%*. The results of this second search are almost equally distributed among “aim”, “goal” and “objective” collocating with “experiment” (not “experimenting”) and “test”. Finally, the search string *aim/goal/objective .. to/of .. test%* is written. In this way, it is possible to notice that “aim”, “goal” and “objective” are, again, almost equally distributed with “test”. Sample phrases are “the aim of the trial was to test”; “we aim to test”; “the goal of the test”; “our goal was to test”, and “the (...) main objective is to design and test”. Finally, to explore the lemma “try out” (as written by student 1), the following query is written: *aim/goal/objective .. to/of .. try% out*. Nonetheless, the results are inconsistent, as they show phrases with different meanings, such as “the goal to have students to try to figure out”. Therefore, it could be claimed that some students did not accurately verify word usages in context and/or collocations.

This could be due to their inexperience and/or to little exposure to the Leeds search syntax.

As concerns *sperimentare i (...) trattamenti*, if the string *test% .. treatment%* is queried, there are many concordances with the lemma “test” collocating with “treatment(s)”. The same occurs to the phrase *try% .. treatment%*, as many hits are obtained with “try treatments”, although no results are shown with the lemma “try out”. Finally, if *experiment% with .. treatment%* is written in the Leeds search field, a few hits with “experiment with” and “treatment” are obtained.

As can be seen in Table 6, there are no grammatical issues in the students' translations. Therefore, it is possible to speculate that targeted web-driven searches helped students fine-tune their target phrases from a grammatical perspective, at least in this case. Table 7 below explores the last part of the third section of the medical article.

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>che al momento sembrano essere più promettenti contro la COVID-19 la malattia causata dal coronavirus</i>	<p>1.They currently <u>seem</u> to be <u>more</u> promising against the disease caused by coronavirus, COVID-19</p> <p>2-4.that currently <u>seem</u> to be the <u>most</u> promising against COVID-19 the disease caused by the coronavirus</p> <p>3.that right now <u>appear</u> to be the <u>most</u> promising in the fight against COVID-19 the disease caused by the coronavirus</p> <p>5.which right now <u>are</u> <u>very</u> promising against COVID-19 the disease caused by coronavirus</p> <p>6-9.that currently <u>seem</u> to be <u>more</u> promising against COVID-19 the disease caused by coronavirus</p> <p>7.that, at present, <u>seem</u> to be the <u>most</u> promising against the COVID-19 disease caused by the coronavirus</p> <p>8.that, at the present moment, <u>seem</u> to be the <u>most</u> promising for COVID-19 the disease caused by coronavirus</p>	<p>1.that currently <u>appear</u> to <u>hold the most promise</u> against COVID-19 the disease caused by the coronavirus.</p> <p>2.that currently <u>seem</u> to be <u>the most promising</u> against COVID-19 the disease caused by the coronavirus.</p>

Table 7. Analysis of the third section of the article (fourth and last part of the section) [author's contribution]

In Table 7, the phrase *sembrano essere più promettenti* is rendered as “appear to hold the most promise” and “seem to be the most promising” by the two MT tools. The students mostly suggested “appear to be the most promising”, “seem to be the most promising”, and “are very promising”. By querying the Leeds platform with the following string: *seem%/appear%/be% .. most/more promising*, there is the preponderance of the lemma “be” followed by either “more” or “most”, then “promising”. If the lemma “be” is excluded from the search and only *seem%/appear% .. most/more promising* are written, the verb “appear” prevails and is followed by either “more” or “most” before “promising”. If the phrase *hold the most promise* is searched for (as suggested by DeepL), no hits are obtained.

The next section of the article is broken down into smaller parts and analysed in different tables (Tables 8-11).

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>I test coinvolgono farmaci già esistenti e usati,</i>	<ol style="list-style-type: none"> <li>1. These trials, <u>including</u> existing drugs used</li> <li>2. These trials, that <u>involve</u> already existing drugs used</li> <li>3. The tests <u>involve</u> drugs that already exist and have been used in the past,</li> <li>4. These tests <u>involve</u> pre-existing drugs which have already been used,</li> <li>5. These clinical trials <u>involve</u> existent drugs that are already used,</li> <li>6. The trials <u>include</u> existing drugs that have already been used</li> <li>7. The tests <u>employ</u> pre-existing drugs used,</li> <li>8. These trials have <u>used</u> existing drugs that have been previously employed</li> <li>9. The tests <u>involve</u> drugs that already exist and that are used,</li> </ol>	<ol style="list-style-type: none"> <li>1. The tests <u>involve</u> drugs that already exist and are used</li> <li>2. The tests <u>involve</u> existing and used drugs</li> </ol>

Table 8. Fourth section of the medical article (first part) [author's contribution]

As can be seen from Table 8 above, the finite *coinvolgono*, referring to *test*, is rendered as “involve” by the two MT interfaces. The students produced many translation options, such as “include”, “involve”, “employ” and “use”. By

writing *test%/trial% employ%/involve%/include%* in the Leeds search field, the results only show the lemma “trial” collocating with “include”. If “trial” is excluded from the search, the lemmas “test” and “include” come to the fore. Then, if also “include” is ruled out (search strings: *test%/trial% employ%/involve%* and *test% employ%/involve%* ), the results are narrowed down to “trial” and “involve”, or “test” and “involve”. Therefore, the most frequent word pairs are “trial” and “include”, “test” and “include”, “trial” and “involve”, and “test” and “involve”. It could be noticed that the collocations indicated by MT (i.e., “test” and “involve”) are not the most frequent.

As already noticed above, there are also some grammatical issues in the students' translations. In this case, the major shortcoming revolved around an incorrect use of the present perfect (“these trials have used existing drugs that have been previously employed”, student 8). As reported in the literature, the present perfect is another recurrent issue for non-native speakers of English (Giampieri *Academic English* 34-35).

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<p><i>e potranno essere svolti in migliaia di ospedali in giro per il mondo,</i></p>	<p>1.may be <u>medicated</u> in thousands of hospitals worldwide,                  2.may be <u>conducted</u> in thousands of hospitals all over the world,                  3.It will be possible to <u>conduct</u> the tests in thousands of hospitals all around the world                  4.it will be able to <u>conduct</u> these tests in thousands of hospitals throughout the world,                  5.They are <u>conducted</u> in thousands of hospitals throughout the world                  6-7.(and) they/They will be <u>carried out</u> in thousands of hospitals around the world,                  8.The trials will be <u>conducted</u> in thousands of hospitals all around the world                  9.and that can be <u>performed</u> in thousands of hospitals around the world</p>	<p>1.and will be <u>carried out</u> in thousands of hospitals around the world                  2.and can be <u>carried out</u> in thousands of hospitals around the world</p>

Table 9. Fourth section of the medical article (second part) [author's contribution]

Table 9 above shows the past participle *svolti* collocating with *test*. Both MT tools generate “carried out” as a translation of *svolti*. The students also mentioned “conducted”, “performed” and “medicated”. If the query *test%/trial% .. conducted/performed/carried* is written in the Leeds search field, the following pairs are obtained: “trial conducted”, “trial performed”, and, to a minor extent, “trial carried out”. If the lemma “trial” is excluded from the search (i.e., *test% .. conducted/performed/carried*), there seems to be a more equal distribution among “conducted”, “performed” and “carried out” collocating with “test”. Also in this case, MT does not seem to account for the most frequent word pairs, albeit producing grammatically correct phrases.

As regards the students' shortcomings, together with the wrong collocation “medicated” (student 1), there was also an inconsistent phrase, i.e., “it will be able to conduct these tests”, which shows an incorrect use of the dummy subject (student 4). As posited, phrase subjects are generally perceived as problematic by second language learners (Giampieri *Academic English* 47-48).

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>senza particolari complicazioni per il personale medico</i>	1.without <u>significant complications</u> for <u>medical personnel</u> , 2.without any <u>particular complications</u> for <u>medical personnel</u> 3.without being of any <u>hindrance</u> to the <u>medical staff</u> 4-9.without (any) <u>particular complications</u> for the <u>medical staff</u> 5.without <u>severe complications</u> for the <u>medical staff</u> 6.without <u>particular complications</u> for <u>medical staff</u> 7.with no <u>significant complications</u> for the <u>medical staff</u> 8.without any <u>particular problems</u> for the <u>medical staff</u>	1.without any <u>particular complications</u> for <u>medical staff</u> 2.without <u>particular complications</u> for <u>medical personnel</u>

Table 10. Fourth section of the medical article (third part) [author's contribution]

Table 10 above presents two interesting word pairs; i.e., *particolari complicazioni* and *personale medico*. The former is rendered as “particular

complications” by both MT tools, whereas the latter is translated as “medical staff” and “medical personnel”. Concerning the students' rendering of *particolari complicazioni*, almost all proposed either “particular complications” or “significant complications”. Some mentioned “any hindrance”, “severe complications” and “particular problems”. As concerns *personale medico*, the majority of the students preferred “medical staff” over “medical personnel”.

By searching for *particular/significant/severe complication%*, the Leeds produces mostly “severe complications” and “significant complications”. There are no results with the word “particular” as suggested by MT. By reading some concordances with the modifier “severe”, however, it is possible to infer that it refers to health issues, as in the phrases “severe complications or death”, or “severe complications such as cirrhosis”. Therefore, it may be argued that “severe” is not the right collocates in the given context. Finally, if *medical staff/personnel* is queried, both “medical staff” and “medical personnel” are frequent. As in the previous cases, MT does not seem to provide the most frequent words. Table 11 below explores the last part of the fourth section of the medical article.

Source Text	Web-based students' translations	MT (1.DeepL; 2.MateCaT)
<i>alle prese con la più grande crisi sanitaria di questo secolo.</i>	1. <u>facing</u> the <u>gravest</u> sanitary crisis of this century. 2-5-6-9. (that is) <u>struggling</u> with the <u>biggest</u> health crisis of this century. 3. that right now is busy <u>dealing with</u> the <u>biggest</u> sanitary crisis of the century 4. who is <u>facing</u> the <u>biggest</u> health crisis of the 21st century. 7. who is <u>dealing with</u> the <u>major</u> health crisis of this century. 8. who is already <u>struggling</u> with the <u>biggest</u> health crisis of this century.	1. <u>struggling</u> with the <u>biggest</u> health crisis of this century 2. <u>dealing with</u> the <u>greatest</u> health crisis of this century

Table 11. Fourth section of the medical article (fourth and last part of the section) [author's contribution]

Table 11 addresses an interesting idiomatic expression; i.e., *alle prese con*. The MT interfaces produce “struggling” and “dealing with”; the students suggested “facing”, “struggling” and “dealing with”. By querying *struggle%/deal%/face% ... crisis%*, the results only show the lemma “face”. If it is excluded from the search, the lemma “deal with” is found. If



“struggle” is searched for together with the lemma “crisis” (i.e., *struggle% .. .. crisis%*), inconsistent phrases are obtained, such as “the struggling financial crisis”, or “struggling before this crisis took place”. Therefore, this is another case where MT does not address recurrent collocations.

Finally, the phrase *la più grande crisi sanitaria* is rendered as “the biggest health crisis” and “the greatest health crisis” by MT. The students wrote “biggest health crisis”; “major health crisis”; “gravest sanitary crisis”, and “biggest sanitary crisis”. If the string *biggest/major/greatest crisis%* is queried, the modifier “major” prevails over “greatest” and “biggest”. Although only one student indicated “major” as the best collocate for “crisis”, this alone should be considered successful from a training perspective. The Leeds platform, in fact, proved to provide the right translation options to trained (or intuitive) students. Regarding the modifier “health” or “sanitary”, the search string *sanitary/health crisis%* confirms the preponderance of “health”, although “sanitary” is also present.

## Discussion

From the analyses carried out above, it is noticeable that some of the students consulted the Leeds platform quite satisfactorily and found a variety of recurrent and *ad hoc* translation candidates. They suggested many translation options which they mostly corroborated by consulting the Leeds. Some examples are the phrases “big test”, “large trial”, “great trial” and “great test” to translate *il grande test*. In this respect, MT only produces “big test” and “great test”.

Other times, the students rendered words more effectively (or “naturally sounding”) than MT, as in the case of *iniziativa concepita*, which they mostly translated as “initiative launched” or “initiative created”, instead of the less frequent “initiative conceived”, as proposed by MT (and by some students) (see Table 5). In this respect, the phrases retrieved by the students can be considered as more naturally sounding, or authentic, as the web as corpus produced evidence of naturally occurring language.

Sometimes, the options or the collocations suggested by MT were not particularly frequent, as in the case of “test” collocating with “involve”, instead of the more recurrent “test/trial” and “include” (Table 8). Other cases in point were the pairs “particular complication” (Table 10) and “biggest/greatest crisis” (Table 11), which were not as frequent as “significant complications” (in the first case), or “major crises” (in the second case), as written by some students.

Not all the students performed web-targeted searches in the same satisfactorily manner. Some collocational shortcomings were noticed in the students' web-based translations, such as “the *achievement* of one of the (...) clinical trials” (Table 3), and “tests (...) *medicated*” (Table 9). There were

also grammatical issues in the students' translations, revolving around the use of definite/indefinite article (Tables 2 and 3); singular/plural nouns (Table 3); the present perfect (Table 8), and the dummy subject (Table 9). Such shortcomings have long been highlighted in the literature, and they typically hallmark second language acquisition (Seidlehofer; Giampieri *Academic English*). They are part of the learning process, where learning a language, as posited by Nunan (233), is more like “growing a garden than building a wall”, because students do not learn one thing at a time, but “numerous things simultaneously (and imperfectly)” (Nunan 233). Also, it may be argued that while performing the translation task, some students did not feel confident with the Leeds tool, with the translation process itself, or with the specific language pair.

As pointed out above, in certain cases only a few students proposed accurate translation candidates (see, for example, Table 11: “major health crisis” written by one student). This, however, can be considered quite successful, as it means that some students were able to consult the Leeds platform effectively to either find translation options or corroborate their intuition.

In light of the above, it can be argued that, to some extent, web-based translations helped produce realistic target language, despite grammatical issues. By contrast, MT output seemed accurate as far as grammar and word order are concerned, but it was not completely satisfactory from a lexical and collocational point of view.

## **Conclusions**

The paper findings highlight that the machine-translated texts sometimes performed less satisfactorily than the students' web-driven translations, albeit grammatically correct and with no evident word order issues. The word pairs that the MT tools produced, in fact, were not always the most recurrent ones, according to Leeds web evidence. In some cases, the MT translation options were not present in the Leeds results, thus questioning the quality and reliability of MT itself. As regards the students' translations, by contrast, they mostly wrote naturally sounding words and/or recurrent word pairs. However, some grammatical issues emerged, highlighting the need for more training in several fields, such as in the consultation of the Leeds platform, in the specific language pair, and/or in translation practice. To a minor extent, some word pairs were also incorrect. It could be speculated that some infrequent collocations were due to a lack of experience in using the new tool, as well as in translation practice, or in the language pair proposed.

In answering the first research question (“Is the web-based translation of a medical article qualitatively more (or less) satisfactory than a machine translation?”), it is evident that web concordancers are satisfactory translation

tools, as they produce instances of naturally occurring language. To some extent, hence, they are more effective than MT, as they provide a variety of language samples and several collocational options. It goes without saying that users must be acquainted with the tool search syntax.

The second question was “To what extent can MT be relied on, if compared to a web-based translation?”. The analyses of this paper proved that non-post-edited MT output should never be relied on, as MT target texts may present infrequent language patterns or unusual collocations, despite being grammatically correct.

The answer to last question (“Can web concordancers be used as post-editing tools in MT?”) is obviously affirmative, as the Leeds proved to be a reliable tool which shows several patterns of authentic language in use.

The limit of this paper lies in the web-concordancing platform chosen. Considering more web concordancers might have produced broader results. For example, exploring the potentialities of the WebCorp platform in translation training could have been helpful, especially in view of its domain restriction options. The Leeds, in fact, does not permit any domain restriction; hence, it could be argued that the language results retrieved are too broad and do not address the medical field. Nonetheless, the article dealt with could be considered as a popularised web article, whose subject-matter and language are neither too technical nor too field-specific.

Future research could embrace other sectors or fields and/or other web concordancers, such as the WebCorp. Other scholars could investigate whether users with more experience in translation practice, or students with more training in the web tool analysed, might produce different or more insightful results. Finally, papers addressing future advancements in MT algorithms could investigate the lexical and collocational issues discussed in this paper.

## Works Cited

- Chan, Sin-wai. *The Human Factor in Machine Translation*. New York: Routledge, 2018.
- Cirillo, Letizia. “From the Stinking Bishop to the Abbucciato Aretino (and back) Using corpora in the translation classroom.” *inTRAlinea* Special Issue: Translation And Interpreting for Language Learners (TAIL) (2018) <http://www.intralinea.org/archive/article/2297>
- Fletcher, William H. “Corpus analysis of the world-wide-web”. *Encyclopedia of Applied Linguistics*. Ed. Carol A. Chapelle. Oxford: Wiley-Blackwell, 2012. 1339-1347.
- Gatto, Maristella. *From body to web: An introduction to web as corpus*.

- Rome: Laterza, 2008.
- Gatto, Maristella. “The 'body' and the 'web': The web as corpus ten years on.” *ICAME* 35 (2011): 35–58.
- Gatto, Maristella. *Web as corpus: Theory and practice*, 2014.
- Giampieri, Patrizia. 2017. *Academic English*. Turin: UTET, De Agostini
- Giampieri, Patrizia. 2019. The web as corpus in ESL classes: A case study. *International Journal of Language Studies*, 13 (2): 91-108.
- Giampieri, Patrizia. 2021. Can corpus consultation compensate for the lack of knowledge in legal translation training? *Comparative Legilinguistics* 46: 5-35.
- Giampieri, Patrizia and Claudia Mary Forshaw Labruzzo. 2021. *Technical Translations: a corpus approach for Italian and English speakers*. Turin: Celid.
- Jia, Yanfang, Carl, Michael and Wang, Xiangling. “How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study.” *The Journal of Specialised Translation* 31 (2019): 60-86.
- Kehoe, Andrew and Renouf, Antoinette. “WebCorp: Applying the Web to linguistics and linguistics to the Web.” *Proceedings of the WWW 2002 Conference*. Hawaii: Honolulu, 7 – 11 May 2002.
- Lee, Sangmin-Michelle. “The effectiveness of machine translation in foreign language education: a systematic review and meta-analysis.” *Computer Assisted Language Learning* (2021). DOI: 10.1080/09588221.2021.1901745
- Lindquist, Hans and Magnus, Levin. *Corpus Linguistics and the Description of English (2nd edition)*. Edinburgh: Edinburgh University Press, 2018.
- Loock, Rudy. “No more rage against the machine: how the corpus-based identification of machine-translationese can lead to student empowerment.” *The Journal of Specialised Translation* 34 (2020): 150-170.
- Montalbán Martínez, Nicolás. “Using Cat (Computer Assisted Translation) tools in scientific and tourist texts - El uso de herramientas TAO (traducción asistida por ordenador) en textos científicos y turísticos.” *Verbeia* V(4) (2019): 74-95.
- Nunan, David. *Learner-centered English Language Education: The Selected Works of David Nunan*. New York / London: Routledge, 2012.
- Randhawa, Gurdeeshpal, Mariella Ferreyra, Rukhsana Ahmed, Omar Ezzat and Kevin Pottie. “Using machine translation in clinical practice.” *Canadian family physician Medecin de famille canadien* 59(4) (2013): 382-3.
- Rivera-Trigueros, Irene. “Machine translation systems and quality assessment: a systematic review.” *Lang Resources & Evaluation*

- (2021) <https://doi.org/10.1007/s10579-021-09537-5>
- Seidlhofer, Barbara. “Research perspectives on teaching English as a lingua franca.” *Annual Review of Applied Linguistics* 24 (2004): 209–239.
- Takakusagi, Yosuke, Oike Takahiro, Shirai Katsuyuki, Sato Hiro, Kano Kio, Shima Satoshi, Tsuchida Keisuke, Mizoguchi Nobutaka, Serizawa Itsuko, Yoshida Daisaku, Kamada Tadashi, and Katoh Hiriyuki. “Validation of the Reliability of Machine Translation for a Medical Article From Japanese to English Using DeepL Translator.” *Cureus* 13(9) (2021).  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8494522/>
- Wilson, James, Hartley, Anthony, Sharoff, Serge and Stephenson, Paul. 2010. “Advanced corpus solutions for humanities researchers.” *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*. Ed. Ryo Otaguro, Kiyoshi Ishikawa, Hiroshi Umemoto, Kei Yoshimoto and Yasunari Harada, 2010. 769-778. <http://anthology.aclweb.org/Y/Y10/Y10-1089.pdf>
- Zanettin, Federico. *Translation-driven corpora*. Abingdon / New York: Routledge, 2014.

### **Online resources**

DeepL: [www.deepl.com](http://www.deepl.com)

Leeds – collection of Internet corpora: <http://corpus.leeds.ac.uk/internet.html>

MateCaT: [www.matecat.com](http://www.matecat.com)