



Translating Ulyssei@s: Work in progress between general and specialised translation in light of NMT

Françoise Bacquelaine

University of Porto

Porto, Portugal

franba@letras.up.pt / shirleybac@gmail.com

<https://orcid.org/0000-0001-8055-1678> 

Elena Zagar Galvão

University of Porto

Porto, Portugal

egalvao@letras.up.pt

<https://orcid.org/0000-0003-1361-1727> 

Abstract: It is common knowledge that the publication of translation projects carried out by students is a very motivating factor for training. Since the second semester of 2015-2016, undergraduate translation students of the Faculty of Arts and Humanities of the University of Porto have participated in the development of the French and English versions of the digital encyclopaedia Ulyssei@s, created by the Institute for Comparative Literature Margarida Losa (ILCML) in the early 2010s. Besides motivating students, this translation project published on the Internet allows us to develop translation skills, foster teamwork, consolidate the three-phase methodology proposed by Gouadec (2002, 2007), and experiment with several essential tasks in language services: collaborative glossary-making, translation, revision of human translation and post-editing neural machine translation (NMT). Some of the challenges posed by the translation of Ulyssei@s are common to all types of translation, while others are specific to the translation of entries of this *sui generis* encyclopaedia, such as a variety of textual genres (biography, literary criticism, prose, poetry, theatre, essay) in hybrid texts and the need to find the original or the official translation of the works cited. This paper focuses on the data collected in the spring semester of 2022 and is based on the results of a survey submitted to students at the end of the project. It aims to share the experience of developing an inevitably unfinished project and adapting training and learning to the latest advances in Human Language Technology, since NMT has become a valuable tool for human translators, and Ulyssei@s is a digital encyclopaedia regularly enriched by ILCML collaborators with new entries.

Keywords: translator training; digital encyclopaedia; multi-generic text; CAT tools; neural machine translation (NMT).



1. Introduction

A project-based approach in translator training is most effective when the project leads to the publication of the translated product (Kiraly, 2012; Kiraly & Massey, 2019). This article focuses on a translation project carried out at the Faculty of Arts and Humanities of the University of Porto (FLUP, *Faculdade de Letras da Universidade do Porto*) since 2016. It combines general and specialised translation as well as human and machine translation (MT). Our main goal is to share our teaching and learning experiences. We, therefore, make no scientific claims but rather hope to show how we have responded to the challenge of integrating MT, in particular neural machine translation (NMT), into our classes, while at the same time using a constructivist, project-based collaborative approach (Kiraly 2012; Kiraly & Massey, 2019) that mobilises a whole range of skills.

First, we present *Ulyssei@s*, a digital encyclopaedia initially written in European Portuguese. Then, we focus on the project experience in two translation classes taught in the academic year 2021-2022. Based on the university data and students' responses to a survey, we present this four-month experience in terms of human resources (class make-up, degree programmes, language combinations), objectives, and methodology. The results of the survey about motivation, challenges and CAT tools in the broadest sense, i.e., including NMT, are presented in the third part. Some examples of the challenges faced by students and NMT are discussed in the last section.

2. *Ulyssei@s*, a digital encyclopaedia

Most translation projects arise from the need to make content available in other languages for other audiences. In this case, the need is to internationalise *Ulyssei@s*, a digital encyclopaedia of “displaced” writers of the 20th and 21st centuries launched in the early 2010s to explore the different cultural and literary aspects of “displacement”. It is managed and regularly updated by a group of researchers at the Margarida Losa’s Institute for Comparative Literature (ILCML, hereafter “the client”). It is a successful endeavour, as testified by the fact that it is featured in the list of readings recommended by the Portuguese Ministry of Education (*Plano Nacional de Leitura*, 2021-2022).

All the entries are available in European Portuguese, and some of them are also available in English and French, thanks to collaborative projects carried out over the years in Portuguese-English and Portuguese-French general translation classes. This paper focuses on the project developed in two of these classes, General Translation Portuguese into English and General Translation Portuguese into French, both in the spring semester of the academic year 2021-2022.

3. *Ulyssei@s* translation project: spring 2022

The following sections discuss the main characteristics of the students involved in the project, the general and the new objectives in light of NMT, and the methodology adopted according to class size.

Two brief considerations are in order before looking at class make-up according to the students’ dominant or first language. One has to do with the notions of *general* and *specialised translation*, and the other with directionality. What is general translation? How general is it? How



specialised is specialised translation? As Kelly (2014) puts it, the core curriculum of translation programmes at European universities usually places general translation at undergraduate level and specialised translation at master’s level. There is no clear definition of what *general* is supposed to mean and “the underlying assumption is that general translation is easier than specialised translation, a dubious assumption many professional translators would question, if by ‘general texts’ or ‘general translation’ we mean the translation of literary or media texts, for example” (Kelly, 2014, p. 68). It is generally agreed that even highly specialised texts may be easier to translate than highly culture-bound, expressive texts, often referred to as “general”. According to Kelly (2014, p. 68), “there is some consensus regarding the kinds of text best used in the initial stages of translator training (highly conventionalised or standardised, short, and meaningful to the learner, for example), and ‘general’ is definitely not the best description that can be found for them”. With this project, students soon realise that the distinction between general and specialised translation is not clear-cut and that the so-called general translation can indeed be very challenging. They also realise that texts are often hybrid, i.e., the same text may fulfil various functions and span different genres, registers, and even styles.

As for directionality, in these classes, students translate only into English and French. For the majority of Portuguese-speaking students, it is the first time they have translated out of their first language. In 2021-2022, most English-speaking students attending the class under the Erasmus mobility programme (hereafter “mobility students”) had never taken a translation class, while all French-speaking mobility students had already translated, but never post-edited.

3.1 Human resources: class make-up, fields of degree programmes, language combinations

The project was conducted in two classes: *Tradução Geral Português-Ingês* (TGPI) and *Tradução Geral Português-Francês* (TGPF). Figure 1 presents the total number of participants and the percentage of mobility students and FLUP students for each class.

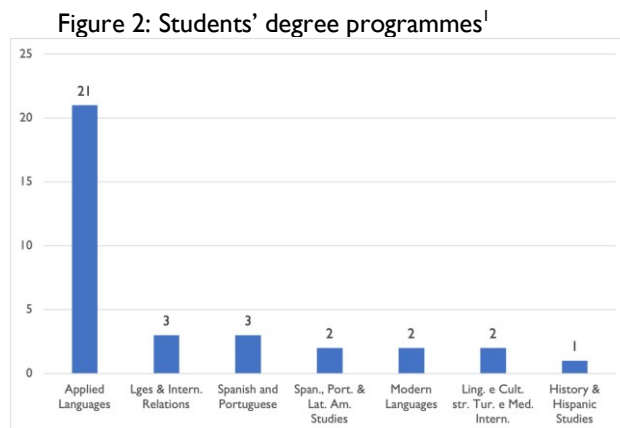
Figure 1: Class make-up (2021-2022)



Source: Authors (2024)

TGPI had 26 regularly attending students: roughly 60% Portuguese and 40% mobility students. TGPF had eight students involved in the project, approximately 60% mobility and 40% Portuguese. All students were enrolled in the third and final year of their undergraduate degrees. Both classes are compulsory in the undergraduate programme in Applied Languages with a major in Translation.

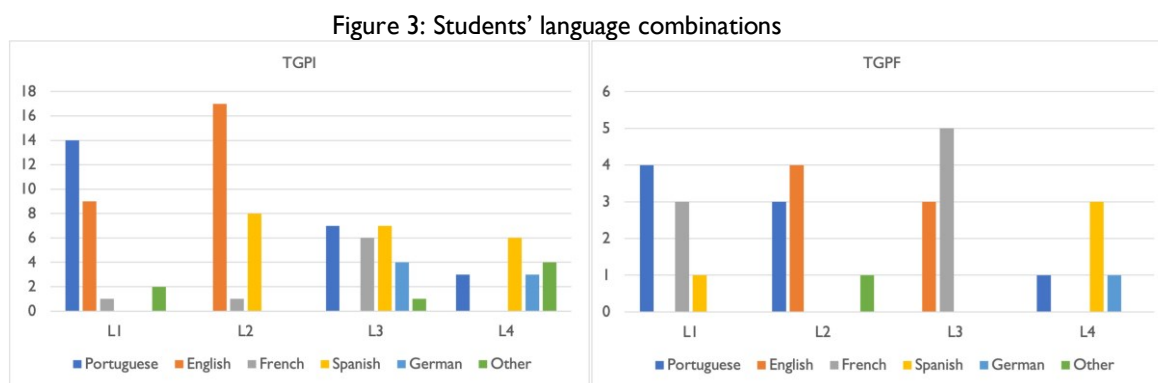
As a result, most of the students taking these two classes were enrolled in Applied Languages, 17 at FLUP and four at French universities. However, there were also students from other undergraduate degree programmes who were taking these classes as electives, as shown in Figure 2.



Source: Authors (2024)

According to the official description, FLUP's undergraduate degree in Applied Languages (Translation major) provides essential training in translation-related areas to prepare students for more specialised studies in translation at master degree level. These classes are offered in the sixth and last semester of the programme, so students are supposed to have attained C1.2 level in English and B2.1 in French, as described in the Common European Framework of Reference for Languages (Council of Europe, 2020). Their writing skills in these languages, however, may vary from excellent to fairly poor.

Figure 3 shows how students described their language combinations in the survey.



Source: Authors (2024)

Portuguese was the most common Language I (LI) in both classes, followed by English and French, as was to be expected. Indeed, mobility students from Britain and France tend to choose these classes, which are taught in English and French, to improve their Portuguese language skills rather than to learn how to translate. Other first languages were Russian and Italian in TGPI and

¹ Full names of the degree programmes: Languages and International Relations; Spanish, Portuguese and Latin American Studies; *Lingue e culture straniere per il Turismo e la Mediazione Internazionale* (Foreign languages and culture for tourism and international mediation).

Spanish in TGPF. English clearly dominated as L2, while Portuguese, Spanish, French and German were L2, L3 or L4.

What we have outlined up to now is the backdrop against which learning activities had to be designed and the project carried out. The main difference is the size of the two classes, which explains why in TGPI this was a group project, and in TGPF it was an individual project.

3.2 General and specific objectives

This project aims to give all the students an equal opportunity to develop the range of skills set out in the objectives of the TGPI and TGPF syllabi, which in turn reflect those currently required of a language service provider. We thus aim to motivate students with the prospect of publication, to apply the three-phase methodology described by Gouadec (2002, 2007), to ensure the quality of the translated product, to carry out tasks that are considered essential in the language service market, as well as to develop team spirit, critical thinking, interpersonal communication skills, problem-solving strategies and organisational skills.

Considering the rise of NMT, we also aim to raise awareness of its advantages and disadvantages and of how useful NMT may be depending on source text features and translation brief. The idea is for students to use NMT in their translation process. NMT has reached very good quality standards in terms of fluency, but its accuracy and adequacy are not always reliable. It is therefore crucial to incorporate its use in our teaching/learning experiences. Students should be able to use any resources available, and, given the high quality of some NMT engines, it is important that they learn how to make the best out of them. This entails being able to weigh the relevance of several factors, such as text genre and type, author's style, register, domain-specific terminology, as well as the purpose of the translation (how and by whom it will be used).

3.3 Methodology

As mentioned above, we followed Gouadec's three-phase methodology: pre-translation, translation and post-translation (Gouadec, 2002, 2007). According to this author, "pre-translation includes anything that takes place up to the moment the translator actually receives the material for translation" (Gouadec, 2007, p. 13), and "post-translation covers all activities that usually follow delivery of the translated material to the work provider or client" (Gouadec, 2007, p. 14). As to the translation phase, it "is divided into three stages:" pre-transfer, transfer and post-transfer (Gouadec, 2007, p. 13-14):

Pre-transfer includes all operations leading up to the actual 'translating', including preparation of the material, documentary searches, alignment, memory consolidation, terminology mining, deciding on options, etc.

Transfer is the well-known core activity of shifting to another language-culture combination. *Post-transfer* covers any operation carried out after translating to meet the quality requirements and criteria prior to delivery of the translated material. It mostly pertains to quality control and upgrading. It also includes formatting and various preparations for delivery (Gouadec, 2007, p. 14).



The two teachers involved, the authors of this article, and the client, ILCML, conducted the post-translation phase and guided the students through the pre-translation and translation phases.

In the pre-translation phase, the project was presented to the class. The TGPI class was divided into groups including native Portuguese and English speakers. In the TGPF class, each student translated an entry and a native French-speaking student revised it. Students had time to explore and become familiar with the website in Portuguese, French and English, before choosing one or more untranslated encyclopaedia entries to work on. A brief explanation of the structure of the entries is now in order. They all include the same sections in the same order: Surname, Name (e.g., Bessa-Luís, Agustina) and picture of the artist with a biography and description of their main works related to the theme of 'displacement', Travels, Quotations, Selected primary bibliography, Discography and Filmography (if applicable), Selected critical bibliography. The 'Quotation' section may include excerpts from works of fiction, non-fiction, essays, poetry, and even entire poems. According to the survey we conducted, the main factor influencing the students' choices of entry/ies in both classes was their interest in the artist. In TGPI, the second factor was the number of words to be translated, the third was the writing style of the entry's author, the fourth was the place of birth or the language of the artist, and the fifth was having ready access to the original language texts included in the 'Quotation' section. In TGPF, the second factor was the artist's place of birth and the third factor was one of the following: access to the quotations in the original language, the writing style of the entry's author or the artist's language. Once the groups (in TGPI) or each individual student (in TGPF) had chosen their entries, they could organise the next steps of the project, according to the translation brief (methodology, tools, deadlines, terminology to be used, special treatment of quotations, etc.).

The translation phase, including pre-transfer, transfer and part of the post-transfer, was carried out by the students. The pre-transfer stage included a thorough text analysis to help students identify the specific issues they would have to address in the transfer stage. The students then conducted documentary research and created lexical, terminological and phraseological databases. The transfer stage corresponds to the translation proper, whose quality is highly dependent on the quality of the pre-transfer. In the post-transfer stage, the target texts were revised by the native English- and French-speaking students and the teachers carried out an additional revision and final proofreading.

After delivering the project by the agreed deadline, students were asked to complete a survey on Google Forms, which provided us with the quantitative data presented above and the qualitative data presented below.

4. Motivation, challenges and CAT tools

The survey included closed and open-ended questions and focused on three main areas: the factors that had motivated the students throughout the project, the most complex challenges they had faced and the tools they had used. As regards motivating factors, they could choose one of three options: 1) their taste for translation challenges; 2) the prospect of seeing their translations published online and being able to include this information in their CVs; 3) the weighting of the project in the overall class assessment (50% in TGPI and 40% in TGPF). Alternatively, they could provide their own personal answer in a text box, but none of them did. To identify the most complex challenges,



the students could select up to three options out of six² or answer an open-ended question. As for CAT tools and other resources used during the project, the students were asked to select from a list of thirteen options³ and to assess the usefulness of Smartcat (an open-source, free-of-charge CAT tool), MT engines, dictionaries and encyclopaedias. Here we shall focus on one resource type: NMT.

4.1 Students' feedback about motivation and hardest challenges

According to the survey, the students' most significant motivating factor was their taste for translation challenges (50%), followed by the prospect of seeing their translations published online and being able to include this information in their CVs (29.5%), and the weighting of the project in the overall class assessment (mentioned by 20.5% of the students).

Over 70% of the students stated that one of the most complex challenges was finding suitable phraseological equivalents, such as collocations and other multiword units that needed to be translated *en bloc*. Understanding and translating long sentences in the source text proved a major challenge for 65% of the students. Less than half (47%) regarded finding terminological equivalents as a significant challenge, while 38% found it difficult to find lexical equivalents. Using the proper registers was perceived as difficult by only 15% of the students. Teamwork was not viewed as a challenge. These results confirm Kelly's (2014) comments on the challenges posed by the so-called 'general translation': like many other texts, Ulyssei@s' entries are hybrid and pose phraseological, syntactical, terminological, lexical and discursive challenges. As a learning experience, therefore, it seems to function well as a transition between general and specialised translation, between undergraduate and graduate levels.

4.2 CAT tools and other resources used in the project

During the pre-transfer phase, students used mainly comparable texts in English and French, as well as online monolingual and bilingual dictionaries. In the transfer phase, TGPI students used Smartcat with NMT (Google Translate), and TGPF students used several free online NMT engines (DeepL, Reverso and Google Translate).

Students' opinions on the usefulness of NMT vary. Only 13% of the students, however, did not find it useful, while nearly 49% found it useful and 38% very useful. This means that almost 87% of the students recognised its usefulness. Overall, their comments show that they are becoming aware of the strengths and weaknesses of NMT. Saving time is often mentioned as an advantage, while lack of accuracy or misinterpretation of the source text, and inappropriate register are considered major weaknesses in NMT.

² Options available: understanding and translating long sentences, finding adequate lexical equivalents, finding adequate terminological equivalents, finding adequate phraseological equivalents, working as a team, using a proper register, NMT, other option.

³ Options available: Smartcat; other CAT tool; DeepL; Google Translate; Reverso; other NMT engine; Translation memories; Portuguese dictionary; English (or French) dictionary; Portuguese-English (or Portuguese-French) dictionary; comparable texts in English (or French); encyclopaedia(s); other option.



5. NMT facing source text challenges

Despite extraordinary advances, NMT still faces source text challenges (Koehn, 2020). In this study, some are common to the two language pairs considered here, while others are more language specific. As to common challenges, we can name the obvious mistakes in the source text (punctuation, typos, etc.), lexical ambiguities, and named entities (book titles, toponymy, personalities). Named entities, in particular, raise several questions: should they be translated or not? Should they be written in capital letters or in italics? What about potential cultural gaps? Other common challenges are collocations and other phraseological issues, as well as syntactic challenges, such as syntactic ambiguity and long sentences (in Portuguese), where long-distance relationships between sentence elements usually lead to faulty cohesion networks.

To illustrate these challenges, we selected a few examples from the entry about the Portuguese writer Agustina Bessa-Luís⁴, the only one translated in both classes. In TGPI, a group formed by two Portuguese-speaking students and an English-speaking student translated the entry from Portuguese into English and then revised it. In TGPF, a Portuguese-speaking student translated the entry into French and a French-speaking student revised the translation. We use Lefer *et al.*'s (2022) terminology to classify the students' post-edits: *value-adding edit*, *successful edit*, *unnecessary edit*, *incomplete edit*, *error-introducing edit*, *unsuccessful edit*, *missing edit*.

Example 1 illustrates the post-editing of a long Portuguese sentence containing cohesion network challenges. The source text is in the first column, the translation (Google Translate) in the second, and the post-edited text in the third. Successful edits are highlighted in green. Unsuccessful edits and error-introducing edits are highlighted in yellow.

Table 1: Post-edition of PT-EN NMT, successful, unsuccessful and error-introducing edits

Sónia Serrano Pujalrás Agustina Bessa-Luís Ulyssei@s	Smartcat with NMT (Google Translate)	Post-edited by students
Embaixada a Calígula surge do convite, feito em 1959, para participar num encontro de escritores europeus, sendo ela a única representante de Portugal, no âmbito do Congresso para a Liberdade da Cultura, em Aix-en-Provence que virá a ser conhecido como o "Encontro de Lourmarin", nome do castelo onde decorreram as sessões (Pujalrás, 2021).	Embassy to Caligula comes from the invitation, made in 1959, to participate in a meeting of European writers, being her the only representative of Portugal, in the scope of the Congress for the Freedom of Culture, in Aix-en-Provence, which would become known as the "Lourmarin Meeting", name of the castle where the sessions took place.	<i>Embaixada a Calígula</i> is born from the invitation made in 1959 for Agustina to participate in a meeting of European writers, in which she was the sole representative of Portugal. This meeting happened in the context of the Congress for the Freedom of Culture, in Aix-en-Provence, that would later come to be known as "The Lourmarin Meeting", as a nod to the name of the castle where the sessions took place.

Source: Authors (2024)

In this first example, students made five successful edits, two error-introducing edits, and one unsuccessful edit. The named entity *Embaixada a Calígula* was italicised, following English writing conventions; the long-winded Portuguese sentence was split into two shorter sentences; the

⁴ <https://ulyssseias.ilcml.com/pt/termo/bessa-luis-agustina/>

cohesion network was re-established; and naturalness was partly improved through the use of the idiomatic expression *as a nod to*. Though recognising that the verb *comes from* was not the best solution for *surge*, the students were unable to find a suitable translation and actually introduced another inadequate solution (*is born*). They also failed (as did Google Translate) to recognise the need to change the verb tense from the Portuguese *Indicativo Presente*, (often used as a historical present, as in this text) to the English Simple Past. By replacing *which* by *that* to introduce a non-defining relative clause, the students actually introduced a grammar mistake. In addition, by inserting the first name of the author (*Agustina*) and not her full name or her surname in the first sentence, they did not comply with the target language writing conventions. All this clearly shows how challenging post-editing of so-called general texts can be.

The second example consists of the post-editing of NMT from Portuguese into French of the same sentence. The table includes a fourth column, which presents the reformulation of the excerpt.

Table 2: Incomplete, successful, unsuccessful, and missing edits of PT-FR NMT requiring reformulation

Pujalr�s, S. "Agustina Bessa-Lu�s" Ulysei@s	DeepL (free version)	Post-edited NMT	Reformulation
Embaixada a Cal�gula surge do convite, feito em 1959, para participar num encontro de escritores europeus, sendo ela a �nica representante de Portugal, no �mbito do Congresso para a Liberdade da Cultura, em Aix-en-Provence que vir� a ser conhecido como o "Encontro de Lourmarin", nome do castelo onde decorreram as sess�es (Pujalr�s, 2021).	L'ambassade � Caligula vient de l'invitation, faite en 1959, � participer � une r�union d'�crivains europ�ens, �tant la seule repr�sentante du Portugal, au sein du Congr�s pour la libert� de la culture, � Aix-en-Provence, qui deviendrait connu comme la "Rencontre de Lourmarin", nom du ch�teau o� les sessions ont eu lieu.	L' <i>Embaixada a Caligula</i> � d�coul� de l'invitation, faite en 1959, � participer � une r�union d'�crivains europ�ens, �tant la seule repr�sentante du Portugal, <i>dans le cadre du</i> Congr�s pour la Libert� de la Culture, � Aix-en-Provence, qui deviendrait connue comme « rencontre de Lourmarin », nom du ch�teau o� les sessions ont eu lieu.	<i>Embaixada a Caligula</i> r�sulte d'une rencontre d'�crivains europ�ens qui s'est tenue en 1959 � Aix-en-Provence dans le cadre du Congr�s pour la libert� de la culture. Agustina Bessa-Lu�s est la seule repr�sentante du Portugal � �tre invit�e � cette rencontre connue sous le nom de "Rencontre de Lourmarin" en r�f�rence au ch�teau o� elle a eu lieu.

Source: Authors (2024)

The translator and the revisor made three superficial successful edits and an unsuccessful one (*a d coul *), but they do not seem to have fully realised how challenging this long Portuguese sentence is. The first edit is incomplete: they italicized the title in PT, but maintained the definite article (*L'*) added to the book title by DeepL in contrast to French usage. Apart from the cohesion challenges, the sentence in Portuguese includes several named entities – a book title and an event that took place in France. Thus, the translator needs to use the original designation of the event (*Congr s pour la libert  de la culture*) and the name by which it came to be known later (*Rencontre de Lourmarin*), both requiring changes in the use of capitalisation. The sentence also contains a gerund (*sendo*) and some phraseological challenges, such as the collocation *fazer um convite* and the preformed construction unit (Schmale, 2013) *vir a ser conhecido como*. In this case, NMT was not useful, and a reformulation was necessary.



In the next example, the cohesion network of *outros textos [...] que [...] a autora escreveu a propósito do Brasil* was not identified by the NMT engines, probably because the relative pronoun *que* is distant from its antecedent *outros textos*, and from the rest of the relative clause (i.e., *a autora escreveu a propósito do Brasil*). To make matters worse, a comma is missing after the relative pronoun *que* in the source text. These challenges give rise to a mistranslation in the NMT output in both English (*the main work, the author wrote about Brazil*) and French (*l'œuvre principale, que l'auteur a écrite sur le Brésil*). Though the English post-edited version contains minor inaccuracies, TGPI students understood the need for reformulating the target text by moving and simplifying the segments *that are not part of the main work* and *the author wrote about Brasil*. TGPF students, on their part, did not notice the mistranslation. They trusted NMT and left it unchanged. Reformulation in French by the teacher is shown in Example 3, under the reformulation by the TGPI group.

Table 3: Long-distance relationships inducing poor quality NMT – successful edits in TGPI and missing edits in TGPF

Pujalrás, S. "Agustina Bessa-Luís" Ulysei@s	NMT (Google Translate and DeepL)	Reformulation
Na edição consultada estão incluídos <i>outros textos</i> (artigos e discursos), publicados em revistas e jornais <i>que</i> não fazendo parte da obra principal, <i>a autora escreveu a propósito do Brasil</i> (Pujalrás, 2021).	In the edition consulted are included other texts (articles and speeches), published in magazines and newspapers that are not part of the main work, the author wrote about Brazil (Google Translate).	In the edition we have consulted there are other texts about Brazil that aren't part of Agustina's main work (articles and speeches), which were published in magazines and newspapers (TGPI students).
	L'édition consultée comprend d'autres textes (articles et discours), publiés dans des magazines et des journaux qui ne font pas partie de l'œuvre principale, que l'auteur a écrite sur le Brésil (DeepL).	L'édition consultée comprend d'autres textes sur le Brésil (articles et discours) publiés dans des magazines et des journaux. Ces textes ne font pas partie de son œuvre principale (TGPF teacher).

Source: Authors (2024)

At lexical level, false friends and lexical ambiguities can be challenging for NMT engines as well as human translators and post-editors. In example 4, there are two occurrences of *cenário* in the source text. In Portuguese, *cenário* is a term in the domain of theatre/cinema, where it means 'scenery' or 'setting', but it is often used as a word in general language, as in example 4.

Table 4: False friend and undetected lack of consistency – value-adding, missing, unnecessary, and error-introducing edits

Pujalrás, S. "Agustina Bessa-Luís" Ulysei@s	Smartcat with NMT (Google Translate) PT-EN	Post-edited PT-EN	DeepL NMT PT-FR	Post-edited PT-FR NMT



Nasceu em Vila Meã, Amarante, em 1922, e essas paragens, entre o Douro e o Minho, passando pelo Porto, ofereceram-lhe o privilegiado cenário de muitos dos seus romances, [...]. (Pujalrás, 2021).	He was born in Vila Meã, Amarante, in 1922, and those places, between the Douro and the Minho, passing through Oporto, offered him the privileged scenery for many of his novels [...].	Agustina was born in Vila Meã, in the district of Amarante, in 1922. The Douro and Minho regions, and the city of Oporto offered her the privileged setting of many of her novels, [...].	Elle est née à Vila Meã, Amarante, et ces lieux, entre le Douro et le Minho, en passant par Porto, lui ont fourni le cadre privilégié de nombre de ses romans, [...].	Elle est née dans la ville portugaise de Vila Meã, dans la commune d'Amarante en 1922, et ces escales, entre le Douro et la région du Minho, en passant par le district de Porto, lui ont fourni le cadre privilégié de nombre de ses romans, [...].
[...] o Brasil para o português será sempre uma pátria partilhada, onde chega por diversas vezes a vislumbrar cenários que lhe evocam o Douro, [...].	[...] for the Portuguese, Brazil will always be a shared homeland, where he sometimes glimpses scenarios that remind him of the Douro [...].	[...], for the Portuguese, Brazil will always be a shared homeland, where she manages to catch a glimpse of scenarios that remind her of the Douro region, [...].	[...] le Brésil sera toujours pour les Portugais une patrie partagée, où elle entrevoit parfois des scénarios qui lui rappellent le Douro, [...].	[...] le Brésil sera toujours pour les Portugais une patrie commune, où elle entrevoit parfois des scénarios qui lui rappellent le Douro, [...].

Source: Authors (2024)

The first occurrence of *cenário* was translated adequately into French (*cadre*) by the NMT engine, but the second occurrence was rendered by the term *scenario*, which in French means ‘script’ and is, therefore, a false friend. This example illustrates the common lexical and terminological inconsistencies of free NMT.

Neither TGPI nor TGPF students noticed the lack of lexical consistency and they did not post-edit the false friend (*scenarios* and *scénarios*) in the second occurrence. TGPI students did, however, replace *scenery* with *setting* in the first occurrence, thus improving the target text. TGPF students introduced a value-adding edit (*dans la commune d’*) and four error-introducing edits in the first segment of example 4, as well as an unnecessary edit (*commune*) in the second segment. Moreover, the Portuguese noun *paragens* in the first segment is polysemic. In this particular context, it means ‘regions’ or ‘places’, so the students’ post-editing solution *escales* (‘stopover’) is even worse than the hypernym *lieux* (‘places’) suggested by DeepL. Adding *dans la ville portugaise de, la région du* and *le district de* gives rise to a mistranslation: in this case, *Vila Meã* (about 5,000 inhabitants) is not a city (*ville*), Douro and Minho are rivers, not regions, and Porto is the city, not the district. A possible translation for the French-speaking target audience could be *cette région comprise entre les fleuves du Douro et du Minho en passant par Porto*. Note that *essas paragens, entre* has been omitted in English. Instead, the students chose to split the original sentence into two separate sentences. They simplified and clarified the original by adding *in the district of*, eliminating the phrase *passando por*, and moving the concept of *paragens* (regions) to the second sentence: *The Douro and Minho regions*.

Another translation challenge in the text about Agustina Bessa-Luís is the lexical ambiguity of the Portuguese noun *estação*, which occurs five times in the entry, as illustrated in Example 5⁵.

Table 5: Polysemy and puns

Pujalrás, S. "Agustina Bessa Luís" Ulyssei@s	Smartcat + NMT (Google Translate)	Post-edited	DeepL (free version)	Post-edited NMT
<i>As Estações da Minha Vida</i>	<i>As Estações da Minha Vida</i>	<i>As Estações da Minha Vida</i> [lit. The Stations of My Life]	<i>As Estações da Minha Vida</i>	<i>As Estações da Vida</i> [littéralement : Les étapes de ma vie]
<i>As Estações da Vida</i>	<i>As Estações da Vida</i>	<i>As Estações da Vida</i> [lit. The Stations of Life]	<i>As Estações da Vida</i>	<i>As Estações da Vida</i>
O comboio era quase um modo de vida, marcava o horário dos dias, o ritmo das estações, [...] (Pujalrás, 2021).	The train was almost a way of life, it marked the schedule of the days, the rhythm of the seasons, [...].	The train was almost a way of life. It marked the time of day, the rhythm of the stations, [...].	Le train était presque un mode de vie, il marquait l'horaire des journées, le rythme des gares, [...].	Le train était presque un mode de vie, il marquait l'horaire des journées, le rythme des gares, [...].

Source: Authors (2024)

The book title *As Estações da Vida* is a pun, where the writer plays with the polysemic nature of the word *estações*, which can mean “railway stations” (the subject of the book) as well as the four seasons of the year and, figuratively, the “seasons” or “stages” of life. It is to be noted that the first occurrence of the title in the source text contains an extra word (*Minha*). All students transferred the wrong Portuguese title into the target text and translated it literally, within brackets, as *The Stations of My Life* and *Les étapes de ma vie*. Faced with the impossibility of keeping the ambiguity of the Portuguese title, TGPI and TGPF students tackled this dilemma differently, opting for the literal meaning in English and the figurative meaning in French. TGPI students also translated the second occurrence of the title, but TGPF students did not.

Apart from the book title, two occurrences of *estação* are unambiguous and were adequately translated by NMT engines. The first refers explicitly to Portuguese railway stations (*estações ferroviárias portuguesas*). The second compares railway stations to airports (*a estação, como hoje os grandes aeroportos*). The other occurrence, however, may be regarded as ambiguous: *O comboio era quase um modo de vida, marcava o horário dos dias, o ritmo das estações, ...* (The train was almost a way of life. It marked the time of day, the rhythm of the stations, [...]). In the phrase *ritmo das estações*, the literal meaning of *season* (*saison* in French) is more natural, but the subject *comboio* (train) points to a possible ambiguity. TGPI students replaced Google Translate most probable output *seasons* by *stations*, which probably seemed more adequate in this context, while DeepL selected *gares* and TGPF students left this segment unchanged, though the choice may be questionable.

Example 6 concerns syntactic ambiguity in a citation from Bessa-Luís’s travel book *Embaixada a Caligula*: “*uma sensação de vida isolada e profunda*” (Bessa-Luís, 2009, p. 11). In Portuguese, both

⁵ We excluded from the table two unambiguous occurrences and two occurrences of an untranslated book title (see discussion below).

nouns (*sensação* and *vida*) are feminine, hence the possible syntactic ambiguity for MT and humans: what is isolated and deep? Is it life or is it the sense of life? Is there a difference? The NMT engines rendered this phrase as *a sense of isolated and deep life* and *un sentiment de vie isolé et profond*. Thus, it could be argued that both life and a sense of life are isolated and deep in the English translation, whereas in the French translation the sense of life is isolated and deep, as the adjectives *isolé* and *profond* agree in gender and number with the masculine noun *sentiment* and not with the feminine noun *vie*. In this case, the students may not have detected the ambiguity and, as a result, did not question the NMT outputs in English (Google Translate) or French (DeepL).

Terminological consistency is particularly important when it comes to the terms used for the headings, which are found in all the entries in the encyclopaedia. Example 7 shows how students successfully post-edited literal translations of terms by NMT engines.

Table 6: Terminological consistency across texts

Pujalrás, S. "Agustina Bessa Luís" Ulyssei@s	Smartcat + NMT (Google Translate)	Post-edited	DeepL (free version)	Post-edited NMT
Passagens	Tickets	Travels	Billets	Lieux de passage
Bibliografia Ativa Selecionada	Selected active bibliography	Selected primary bibliography	Bibliographie active sélectionnée	Bibliographie primaire (sélection)
Bibliografia Crítica Selecionada	Selected critical bibliography	Selected secondary bibliography	Bibliographie critique sélectionnée	Bibliographie secondaire (sélection)

Source: Authors (2024)

The section *Passagens* lists the names of all the countries visited by the artist. The meaning of this polysemic word in this context is ‘places one passes by’ and obviously not *tickets* in English or *billets* in French, as might be the case in the purchase of a plane ticket, for example. During the pre-translation phase, when negotiating the translation brief, students had been instructed to stick to the terminology already in use for headings in Ulyssei@s since 2016. They complied with these instructions and successfully post-edit NMT outputs.

6. Conclusion

Thanks to the hybridity of Ulyssei@s’ entries as to genres (biography, literary criticism, prose, poetry, theatre, essay), domains (literature, painting, geography, etc.), and registers, the students were able to appreciate how difficult and indeed artificial it is to compartmentalise general and specialised translation. They started to develop a sense of their target audience’s needs such as the literal translation of the original title in Example 5. Moreover, they explored and assessed up-to-date CAT tools and NMT engines, as well as online dictionaries and comparable texts. Their general feedback was positive and they enjoyed participating in the project.

As expected, since more data are available in language pairs composed of English and another language, and the majority of research in MT focuses on English and another language, NMT fares



better in the pair PT-EN than PT-FR. However, regardless of the language pair, it proves to be a useful tool provided that you are aware of its limitations (O'Brien & Ehrensberger-Dow, 2020; Wilks 2009). An NMT engine is regularly subject to human intervention to (try to) fine-tune the balance between fluency and accuracy, but NMT works with inevitably finite data while language possibilities are infinite (Melby & Warner, 1995).

Neural networks have undoubtedly increased fluency in translation, but the examples discussed above show that lexical and syntactic ambiguity, named entities, cohesion networks within long sentences and across texts, terminological consistency, collocations and other multiword units, domain coverage and accuracy remain a challenge, especially in the case of generic NMT available free of charge. Whilst students should know that customised NMT systems for specific clients or domains are able to deliver very high-quality output, they should also become aware of the scope and limits of NMT so as to be better equipped to identify potential lexical, phraseological, syntactic, and semantic mistakes. We, as teachers, must promote this awareness, a crucial element in MT literacy (Bowker & Ciro, 2019; Ehrensberger-Dow *et al.*, 2023; O'Brien & Ehrensberger-Dow, 2020). Now that ChatGPT and other similar AI-driven tools promise to usher students into an era of superfast easy solutions to translation problems, it is ever more important to focus on text analysis skills, translation quality assessment, post-editing, and, above all, critical thinking. Students will thus be better prepared to decide when and how to post-edit NMT output and to justify their decisions.

Acknowledgements

Our most sincere thanks to all the students who participated in this project.

References

- Bessa-Luís, A. (2009). *Embaixada a Caligula*. Guimarães Editores.
- Bowker, L., & Ciro, B. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. <https://doi.org/10.1108/9781787567214>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: learning, teaching, assessment – Companion volume*. Council of Europe Publishing. www.coe.int/lang-cefr
- Ehrensberger-Dow, M., Delorme Benites, A., & Lehr, C. (2023). A new role for translators and trainers: MT literacy consultants. *The Interpreter and Translator Trainer*, 17(3), 393–411. <https://doi.org/10.1080/1750399X.2023.2237328>
- Gouadec, D. (2002). *Profession : traducteur*. Paris: La Maison du Dictionnaire.
- Gouadec, D. (2007). *Translation as a profession*. John Benjamins.
- Instituto de Literatura Comparada Margarida Losa (n.d.). *Ulyssei@s*. *Enciclopédia digital*. Available at: <https://ulyssseias.ilcml.com/en/> Accessed in: Jun. 28, 2023.
- Kelly, D. (2014). *A Handbook for Translator Trainers*. Routledge. <https://doi.org/10.4324/9781315760292>
- Kiraly, D. (2012). Growing a Project-Based Translation Pedagogy: a fractal perspective. *Meta*, 57(1), 82–95. <https://doi.org/10.7202/1012742ar>
- Kiraly, D., & Massey, G. (2019). *Towards Authentic Experiential Learning in Translator Education* (2nd ed.). Cambridge Scholars Publishing.
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.



- Lefer, M. A., Piette, J., & Bodart, R. (2022). *Machine Translation Post-Editing Annotation System (MTPEAS) manual*. UCLouvain.
- Melby, A. K., & Warner, T. (1995). *The Possibility of Language: A discussion of the nature of language, with implications for human and machine translation*. Benjamins Translation Library.
- O'Brien, S., & Ehrensberger-Dow, M. (2020). MT Literacy – A cognitive view. *Translation, Cognition & Behavior*, 3(2), 145–164. <https://doi.org/10.1075/tcb.00038.obr>
- Pujalr s, S. S. (2021). Agustina Bessa-Lu s. In Instituto de Literatura Comparad Margarida Losa. *Ulyssei@s. Enciclop dia digital*. <https://ulyssseias.ilcml.com/pt/termo/bessa-luis-agustina/>
- Schmale, G. (2013). Qu'est-ce qui est pr fabriqu  dans la langue ? : R flexions au sujet d'une d finition  largie de la pr formation langagi re. *Langages*, 189, 27–45. <https://doi.org/10.3917/lang.189.0027>
- Wilks, Y. (2009). *Machine Translation: Its Scope and Limits*. Springer.

Notes

Authorship contribution

Conceptualization: F. Bacquelaire, E. Galv o

Data curation: F. Bacquelaire, E. Galv o

Formal Analysis: F. Bacquelaire, E. Galv o

Investigation: F. Bacquelaire, E. Galv o

Writing – review & editing: F. Bacquelaire, E. Galv o

Research dataset

Survey, entry of Ulyssei@s in European Portuguese, Smartcat (Google Translate) bilingual PT-EN file, post-edited and revised text in EN, PT-FR NMT by DeepL, post-edited and revised text in FR

Funding

This study was supported by national funds by FCT – Funda o para a Ci ncia – e a Tecnologia, I.P., within the scope of project UIDB/00022/2020.

Image copyright

Not applicable.

Approval by ethics committee

Not applicable.

Conflicts of interest

Not applicable.

Data availability statement

The data from this research, which are not included in this work, may be made available by the author(s) upon request.

License

The authors grant *Cadernos de Tradu o* exclusive rights for first publication, while simultaneously licensing the work under the Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) 4.0 International License. This license enables third parties to remix, adapt, and create from the published work, while giving proper credit to the authors and acknowledging the initial publication in this journal. Authors are permitted to enter into additional agreements separately for the non-exclusive distribution of the published version of the work in this journal. This may include publishing it in an institutional repository, on a personal website, on academic social networks, publishing a translation, or republishing the work as a book chapter, all with due recognition of authorship and first publication in this journal.

Publisher

Cadernos de Tradu o is a publication of the Graduate Program in Translation Studies at the Federal University of Santa



Cadernos de Tradu o, 44(Special Issue 1), 2024. e95214
Graduate Program in Translation Studies
Federal University of Santa Catarina. ISSN 2175-7968
DOI <https://doi.org/10.5007/2175-7968.2024.e95214>

Catarina. The journal *Cadernos de Tradução* is hosted by the [Portal de Periódicos UFSC](#). The ideas expressed in this paper are the responsibility of its authors and do not necessarily represent the views of the editors or the university.

Special issue editors

Andréia Guerini – Fernando Ferreira Alves – Orlando Grossegese

Section editor

Willian Moura

Technical editing

Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

Article history

Received: 29-06-2023

Approved: 17-02-2024

Revised: 12-03-2024

Published: 04-2024

